

UNIVERSIDADE DE SÃO PAULO  
ESCOLA DE ENGENHARIA DE SÃO CARLOS

KALYL GOMES CALIXTO

Relações causais e persistências na dinâmica da disponibilidade hídrica  
para a formulação de modelos hidrológicos conceituais

São Carlos

2020



KALYL GOMES CALIXTO

Relações causais e persistências na dinâmica da disponibilidade hídrica  
para a formulação de modelos hidrológicos conceituais

Dissertação apresentada à Escola de  
Engenharia de São Carlos da Universidade de  
São Paulo, como requisito para a obtenção do  
Título de Mestre em Ciências: Engenharia  
Hidráulica e Saneamento

Orientador: Prof. Tit. Edson Wendland

VERSÃO CORRIGIDA

São Carlos

2020

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTA TRABALHO,  
POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS  
DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da  
EESC/USP com os dados inseridos pelo(a) autor(a).

C153r Calixto, Kalyl Gomes  
Relações causais e persistências na dinâmica da  
disponibilidade hídrica para a formulação de modelos  
hidrológicos conceituais / Kalyl Gomes Calixto;  
orientador Edson Wendland. São Carlos, 2020.

Dissertação (Mestrado) - Programa de  
Pós-Graduação em Engenharia Hidráulica e Saneamento e  
Área de Concentração em Hidráulica e Saneamento --  
Escola de Engenharia de São Carlos da Universidade de  
São Paulo, 2020.

1. Disponibilidade hídrica. 2. Informação  
mútua. 3. Modelo hidrológico conceitual. 4.  
Sensoriamento remoto. 5. Sistema Aquífero Guarani. 6.  
Teste de causalidade de Granger. I. Título.



## FOLHA DE JULGAMENTO

Candidato: Engenheiro **KALYL GOMES CALIXTO**.

Título da dissertação: "Relações causais e persistências na dinâmica da disponibilidade hídrica pra a formulação de modelos hidrológicos conceituais".

Data da defesa: 13/03/2020.

### Comissão Julgadora:

Prof. Titular **Edson Cezar Wendland**  
**(Orientador)**  
(Escola de Engenharia de São Carlos/EESC)

Prof. Associado **José Rodolfo Scarati Martins**  
(Escola Politécnica/EP-USP)

Prof. Dr. **Walter Collischonn**  
(Universidade Federal do Rio Grande do Sul/UFGRS)

### Resultado:

Aprovado

Aprovado

Aprovado

Coordenador do Programa de Pós-Graduação em Engenharia Hidráulica e Saneamento:

Prof. Dr. **Eduardo Mario Mendiondo**

Presidente da Comissão de Pós-Graduação:

Prof. Titular **Murilo Araujo Romero**



## AGRADECIMENTOS

Por ter tido saúde e condições suficientes para o desenvolvimento deste trabalho.

À minha mãe e à Letícia pela compreensão, carinho e apoio diário nesses últimos anos.

Ao Professor Edson pela dedicação ao grupo de pesquisa, pelas oportunidades concedidas, e pelas ideias, sugestões e contribuições ao trabalho e à minha formação.

Aos colegas do Laboratório de Hidráulica Computacional e do Programa de Hidráulica e Saneamento pelo compartilhamento de conhecimento e ideias e pela boa convivência em São Carlos.

Aos funcionários da Escola de Engenharia de São Carlos e do Departamento de Hidráulica e Saneamento pelo suporte em questões administrativas e logísticas, em especial, ao técnico Bergamo, pelas contribuições nos trabalhos de campo.

À Agência Nacional de Águas (ANA), ao Centro de Recursos Hídricos e Estudos Ambientais (CRHEA-USP) e à NASA- Land Processes Distributed Active Archive Center (LP DAAC) pela disponibilização de dados utilizados neste trabalho.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (processo nº 131086/2018-9; vigência: 03/2018 a 11/2018) e à Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP (processo nº 2018/19222-7; vigência: 12/2018 a 02/2020) pelas bolsas de mestrado outorgadas.

À FAPESP pelo Auxílio à Pesquisa Regular (processo nº 2015/03806-1).

E a todos (colegas, amigos, professores e familiares) que me acompanharam ou fizeram parte dessa etapa acadêmica.



## RESUMO

CALIXTO, K. G. **Relações causais e persistências na dinâmica da disponibilidade hídrica para a formulação de modelos hidrológicos conceituais**. 2020. 182p. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2020.

O desenvolvimento e a aplicação de modelos hidrológicos com forte poder preditivo são uma das estratégias para orientar a formulação de políticas relacionadas ao controle e gerenciamento dos recursos hídricos e do uso do solo em direção à sustentabilidade. Um mundo em mudança, com incertezas climáticas e crescente população, requer estimativas confiáveis de disponibilidade de água sob uma variedade de cenários. Nessas estimativas, o conhecimento de dependências entre fluxos e armazenamentos de água pode ser necessário para preservar as consistências físicas e causais entre os processos. Embora essas dependências a curto e longo prazo sejam frequentemente identificadas, poucos estudos adotam essas características do sistema hidrológico para avaliação de modelos. O objetivo deste estudo é construir um modelo hidrológico conceitual concentrado para quantificar a disponibilidade hídrica de uma bacia instrumentada localizada em área de afloramento do Sistema Aquífero Guarani. A precipitação e a evapotranspiração de referência (FAO-56) foram estimadas a partir dos dados monitorados em superfície e de métodos de interpolação espacial. O escoamento no curso de água foi estimado a partir de modelos de transformação de nível de água em vazão com parâmetros variáveis no tempo. O teste de causalidade de Granger e a informação mútua com defasagem foram avaliados e utilizados para caracterizar as dependências entre variáveis do sistema hidrológico no domínio temporal. O modelo conceitual foi desenvolvido a partir de resultados anteriores, em uma estrutura que permitiu a parametrização flexível dos fluxos dominantes. Índices de vegetação obtidos por sensoriamento remoto (*enhanced vegetation index*) foram incluídos no modelo para estimar a evapotranspiração real. O método de integração de Runge-Kutta com passo de tempo adaptativo foi utilizado para resolver o sistema de equações diferenciais do modelo. A calibração foi conduzida em uma abordagem multi-objetivo com o algoritmo NSGA-II. Os objetivos da calibração foram a reprodução da variabilidade de vazões e de armazenamento subterrâneo, avaliado a partir de níveis freáticos de poços de monitoramento. Os resultados obtidos foram satisfatórios e compatíveis com o de trabalhos anteriores, que utilizaram modelos mais complexos ou consolidados. Além disso, o modelo reproduziu adequadamente as dependências entre variáveis hidrológicas no domínio temporal.

**Palavras-chave:** Disponibilidade hídrica; Informação mútua; Modelo hidrológico conceitual; Sensoriamento remoto; Sistema Aquífero Guarani; Teste de causalidade de Granger.



## ABSTRACT

CALIXTO, K. G. **Causality and persistence in the water availability dynamics for the development of hydrological conceptual models**. 2020. 182p. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2020.

The development and application of hydrological models with strong predictive power are one of the strategies to guide policy-making processes related to the control and management of water resources and land use towards sustainability. A changing world, with climate uncertainties and increasing population, requires reliable water availability estimates under a variety of scenarios. In these estimates, the knowledge of the dependencies between water fluxes and storages may be needed to keep the physical and causal consistencies between processes. Although short and long-term dependencies are often identified, a few studies make proper use of these characteristics for model evaluation. The aim of this study is to develop a lumped conceptual hydrological model to represent the water availability of a monitored catchment located in an outcrop zone of the Guarani Aquifer System. As input data, precipitation and reference evapotranspiration (FAO-56) were estimated from ground-based data and spatial interpolation (IDW). The streamflow was estimated from water level data and dynamic rating curves to consider changes in the flow conditions. Parametric Granger causality test and time-delayed mutual information were evaluated and applied to characterize dependencies between hydrological variables in the time domain. The hydrological conceptual model was developed taking previous conceptions as the bottom line and was parameterized accordingly to the dominant flows in the study area. Vegetation indices obtained by remote sensing were used as input data to estimate actual evapotranspiration. Runge-Kutta integration method with an adaptive time step was used to solve the system of ordinary differential equations of the model, which took the continuous state-space representation. The model calibration was conducted in a multi-objective approach (NSGA-II algorithm was used). The objectives were to reproduce the streamflow dynamics at the outlets and the storage in the saturated zone, evaluated from the observed groundwater levels. The results obtained were satisfactory and comparable with previous simulation studies, which used more complex or consolidated models. In addition, the model adequately reproduced the dependencies between hydrological variables in the time domain.

**Keywords:** Water availability; Mutual information; Conceptual hydrological model; Remote sensing; Guarani Aquifer System; Granger causality test.





## LISTA DE ILUSTRAÇÕES

Figura 1. Relação entre entropia e informação mútua. ....	42
Figura 2. Localização da bacia do Ribeirão da Onça, Brotas – SP.....	45
Figura 3. Instrumentação em operação (2019) na bacia hidrográfica do Ribeirão da Onça. ....	47
Figura 4. Pontos de monitoramento de dados meteorológicos na bacia em 2019. ....	47
Figura 5. Estações pluviométricas identificadas próximas à área de estudo.....	48
Figura 6. Dados brutos de precipitação mensal de algumas estações próximas à área de estudo. Meses em que uma ou mais estações apresentaram falhas foram deixados vazios. ....	49
Figura 7. Extensões das séries de dados meteorológicos utilizadas neste estudo. ....	50
Figura 8. Fluxograma para estimativa dos valores de precipitação média. ....	51
Figura 9. Mapa de uso do solo da área de estudo no ano de 2017 (classificação de Alcântara (2018) e pontos amostrados para definição de séries temporais do índice de vegetação EVI correspondentes a diferentes coberturas vegetais.....	55
Figura 10. Principais usos do solo e coberturas vegetais na bacia do Ribeirão da Onça (2018/2019). ....	56
Figura 11. Exemplo de distribuição espacial de EVI na área de estudo. ....	57
Figura 12. Seção de medição de vazão e estação de bombeamento de água. ....	59
Figura 13. Perturbações nas relações entre nível d'água e vazão provocadas pelo barramento a jusante da seção de medição de vazão. Gráficos indicam períodos em que relações bem definidas entre nível e vazão existem (entre novembro e dezembro de 2017) e períodos perturbados por efeito de barramento do curso d'água (agosto de 2017 e abril de 2018). ....	59
Figura 14. Variabilidade geométrica do leito do curso d'água onde a medição de vazão é conduzida. ....	60
Figura 15. Calha Parshall 2. ....	61
Figura 16. Resumo do modelo de transformação de nível d'água em vazão proposto para a seção de monitoramento Santa Maria. ....	62
Figura 17. Modelo conceitual referente aos processos da zona saturada (P = precipitação; Rd = recarga direta;. 76	
Figura 18. Representação de processos dominantes no modelo proposto. ....	78
Figura 19. Representação esquemática da estrutura do modelo proposto. ....	78
Figura 20. Distribuição espacial da precipitação diária de 12/01/2016 a 13/01/2016 usando o método IDW e os dados de quatro estações pluviométricas locais. Códigos indicados fazem referência à Figura 5 (CRHEA: 10001; NMA: 10002; CAA: 2247185; FRV: 2247197). ....	89
Figura 21. Séries de precipitação média sobre as duas bacias estudadas (CP2 e SM).....	92
Figura 22. Comparação entre os métodos de cálculo horário e diário para determinação da evapotranspiração de referência. Dados da estação Monte Alegre tomados como exemplo. ....	93
Figura 23. Evapotranspiração de referência diária de algumas estações utilizadas neste estudo. ....	94
Figura 24. Evapotranspiração de referência acumulada mensal nas estações e média da superfície resultante do processo de interpolação na área SM (NMA: Monte Alegre; Santo Ignácio STO; SM: Santa Maria). ....	95
Figura 25. Série de evapotranspiração de referência média diária na área de estudo. ....	96
Figura 26. Séries temporais de EVI, entre o ano de 2004 e 2019, em pontos com coberturas representativas da área de estudo. As respectivas coordenadas de aquisição (SIRGAS2000) estão indicadas. ....	97

Figura 27. Relações estimadas entre EVI e coeficiente de cultura em diferentes condições. Limite inferior obtido com $c = 0,320$ , e limite superior obtido com $c = 0,125$ .	98
Figura 28. Séries de valores médios de EVI (a cada 16 dias) na área de estudo.	99
Figura 29. Séries de coeficiente $KVI = ETa/ETo$ (a cada 16 dias) para a área SM. Gráficos apresentam resultados de diferentes valores do parâmetro 'c'.	99
Figura 30. Totais anuais em mm na área SM: precipitação (média = 1390 mm), evapotranspiração de referência (média = 1180 mm) e de evapotranspiração estimada (média = 975 mm).	100
Figura 31. Métodos de ajuste por otimização global e por mínimos quadrados ordinários (OLS).	101
Figura 32. Nível de água ajustado entre 2016 e 2018, período em que o método de estimativa de vazão foi avaliado.	102
Figura 33. Resultados do método proposto de transformação de nível de água em vazão e dados observados pelo Sontek IQ-Plus (ADCP). Sombreamento indica o desvio do valor estimado pelo modelo por mais ou menos a raiz quadrada da variância.	103
Figura 34. Resultados do método proposto de transformação de nível de água em vazão para o período completo de dados de nível de água disponível.	104
Figura 35. Distribuição de frequência dos números de pontos adotados nos ajustes por cada método.	105
Figura 36. Separação do escoamento de base pelo método proposto por Duncan (2019).	106
Figura 37. Estimativa de vazão total e escoamento de base na calha Parshall 2 (CP2).	108
Figura 38. Balanço hídrico ( $\Delta S = P - ET - Q - R_p$ ) correspondente à área SM para diferentes condições de evapotranspiração real baseadas na alteração do parâmetro $c$ da equação (20). PMA: precipitação média anual. $R_p$ : percolação profunda. Nível de água indicado é a média dos níveis freáticos observados.	109
Figura 39. Funções wavelet-mãe ( $\psi$ ) do tipo Haar e Symlets 5.	111
Figura 40. Decomposição da série de precipitação média diária na seção SM usando 'Haar' como wavelet-mãe e até o nível cinco. Eixo das abscissas corresponde ao número de dias após 05-12-2008. 'A-j' e 'D-j' indicam as aproximações (baixa frequência) e detalhes (alta frequência) no nível de decomposição 'j'...	111
Figura 41. Decomposição da série de evapotranspiração real média na área da bacia usando 'Symlets 5' como wavelet-mãe e até o nível cinco. Eixo das abscissas corresponde ao número de dias após 05-12-2008. 'A-j' e 'D-j' indicam as aproximações (baixa frequência) e detalhes (alta frequência) no nível de decomposição 'j'.	112
Figura 42. Resultados da simulação referentes ao problema sintético S1-1. NMI: informação mútua com defasagem; LR e valor p: razão de verossimilhança e probabilidade de significância relacionadas ao teste de Granger. $\Delta t = 1$ dia.	113
Figura 43. Resultados da simulação referentes ao problema sintético S1-2. $\Delta t = 1$ dia.	114
Figura 44. Resultados da simulação referentes ao problema sintético S1-3. $\Delta t = 1$ dia.	115
Figura 45. Resultados da simulação referentes ao problema sintético S1-4. $\Delta t = 1$ dia.	116
Figura 46. Resultados da simulação referentes ao problema sintético S1-5 (agregação temporal).	117
Figura 47. Resultados da simulação referentes ao problema sintético S1-6 (análise de multirresolução). $\Delta t = 1$ dia.	118
Figura 48. Resultados das relações entre precipitação e escoamento superficial e escoamento de base. Dados do período entre 12/2008 e 02/2019. $\Delta t = 1$ dia.	120

Figura 49. Resultados das relações entre precipitação e níveis freáticos (poços P4, P9, P13 e P19). .....	121
Figura 50. Resultados das relações entre precipitação e níveis freáticos (poços P14, P15, P16 e P19). .....	122
Figura 51. Resultados de relações entre precipitação, níveis freáticos, índice de vegetação e anomalias de grande escala. <i>GWL</i> : média aritmética de níveis freáticos dos poços 4, 5, 13, 14, 15, 16, 19. <i>EVI</i> : Enhanced Vegetation Index médio da área SM; <i>ONI</i> : Oceanic Niño Index. $\Delta t = 1$ dia. ....	123
Figura 52. Efeitos simultâneos entre média dos níveis freáticos e vazão de base. $\Delta t = 1$ dia. ....	124
Figura 53. Resultados das relações entre evapotranspiração e níveis freáticos (poços P4, P9, P13 e P19). ....	125
Figura 54. Resultados das relações entre evapotranspiração e níveis freáticos (poços P5, P14, P15 e P16). ....	126
Figura 55. Avaliação do esquema de integração numérica RK45 com passo de tempo adaptativo e com dados de entrada admitidos constantes no intervalo de discretização. ....	127
Figura 56. Resultados referentes à área SM correspondente à população da última geração do algoritmo NSGA-II. Linhas vermelhas e amarelas diferenciam os períodos utilizados para calibração (01/2009 a 09/2014) e validação (09/2014 a 02/2019), respectivamente. Barras verticais invertidas indicam precipitação média diária. Cada linha (total de 100) indica um resultado da simulação para um conjunto de parâmetros diferente. ....	128
Figura 57. Escoamento de base simulado e estimado pelo filtro digital referente à área SM. As cores vermelha e amarela indicam os períodos de calibração e validação, respectivamente. Cada linha (total de 100) indica um resultado da simulação para um conjunto de parâmetros diferente. ....	130
Figura 58. Reprodução da curva de permanência de vazões na seção SM exemplificada para um vetor de parâmetros calibrados. ....	130
Figura 59. Avaliação de características simuladas na área SM relacionadas à disponibilidade de água no solo e à evapotranspiração. Curvas indicam os resultados referentes a 100 soluções (resultados estão sobrepostos). ....	131
Figura 60. Resultados referentes à área CP2 correspondentes à população (100 indivíduos) da última geração do algoritmo de otimização (NSGA-II). Linhas vermelhas e amarelas diferenciam os períodos utilizados para calibração (05/2017 a 09/2018) e validação (09/2018 a 02/2019). Cada linha (total de 100) indica um resultado da simulação para um conjunto de parâmetros diferente. ....	132
Figura 61. Escoamento de base simulado e estimado pelo filtro digital referente à área CP2. As cores vermelha e amarela indicam os períodos de calibração e validação, respectivamente. Cada linha (total de 100), vermelha ou amarela, indica um resultado da simulação para um conjunto de parâmetros diferente. ....	133
Figura 62. Avaliação de características simuladas na área CP2 relacionadas à disponibilidade de água no solo e à evapotranspiração. Cada linha (total de 100) indica um resultado da simulação para um conjunto de parâmetros diferente. ....	134
Figura 63. Valores das funções objetivo (fase de calibração) referentes às soluções não dominadas na geração $NG = 10.000$ do NSGA-II. ....	135
Figura 64. Avaliação do modelo na área SM para relações entre evapotranspiração de referência, índice de vegetação e armazenamento subterrâneo. Gráficos indicam as 100 soluções da última geração da otimização. Escala temporal $[\Delta t]$ : diária. ....	139

Figura 65. Avaliação do modelo na área SMF para relações entre precipitação, vazão, vazão de base e armazenamento subterrâneo. Gráficos indicam as 100 soluções da última geração da otimização. Escala temporal: diária. .... 141

## LISTA DE TABELAS

Tabela 1. Caracterização dos dados de monitoramento fluviométrico utilizados.....	58
Tabela 2. Fontes dos dados de monitoramento de nível freático.....	69
Tabela 3. Problemas sintéticos para avaliar efeitos de agregação temporal e suavização de séries temporais.....	72
Tabela 4. Simulações com o problema sintético [1]. A máxima defasagem adotada foi de 300 dias (d). Para a DWT, ‘sym5’ foi utilizada para y e ‘haar’ foi utilizada para P.....	73
Tabela 5. Simulações com o problema sintético [2]. A máxima defasagem adotada foi de 300 dias (d).....	73
Tabela 6. Análises de causalidade e dependência entre variáveis observadas. Variáveis referentes à área de controle SM e aos poços de monitoramento 4, 5, 9, 13, 14, 15, 16 e 19.....	74
Tabela 7. Parâmetros do modelo.....	86
Tabela 8. Relações de regressão da precipitação média mensal obtidas a partir dos resultados de interpolação espacial no período 01/1988 a 04/2018 na área de drenagem SM.....	90
Tabela 9. Relações de regressão da precipitação média diária obtidas a partir dos resultados de interpolação espacial no período 08/2011 a 04/2018 na área de drenagem SM.....	91
Tabela 10. Relações de regressão da precipitação média diária obtidas a partir dos resultados de interpolação espacial no período 08/2011 a 04/2018 na área de drenagem CP2.....	91
Tabela 11. Relações de regressão da precipitação média horária obtidas a partir dos resultados de interpolação espacial no período 08/2011 a 12/2013 na área de drenagem CP2.....	91
Tabela 12. Resultados da evapotranspiração de referência para cada estação no período total de operação. Notação: média amostral (desvio padrão amostral).....	93
Tabela 13. Relações de regressão da evapotranspiração de referência média mensal obtidas a partir dos resultados de interpolação espacial no período 07/2011 a 12/2013.....	95
Tabela 14. Relações de regressão da evapotranspiração de referência média diária obtidas a partir dos resultados de interpolação espacial no período 06/2011 a 12/2013.....	95
Tabela 15. Relações de regressão da evapotranspiração de referência média horária obtidas a partir dos resultados de interpolação espacial no período 06/2011 a 09/2012.....	96
Tabela 16. Regressões lineares entre precipitação média diária P [mm] (> 0) e nível d’água [m a.s.l.].....	102
Tabela 17. Métricas de avaliação do método proposto.....	103
Tabela 18. Distribuição sazonal do número de pares utilizados para a determinação dos parâmetros.....	106
Tabela 19. Ajustes para o cálculo da carga hidráulica de montante na calha CP2.....	107
Tabela 20. Síntese de métricas de avaliação de desempenho do modelo conceitual. Resultados referentes à população final (100 soluções) de parâmetros calibrados. IQR: intervalo interquartil.....	129
Tabela 21. Resultados da calibração. Valores indicados são as medianas dos 100 vetores de parâmetros calibrados resultantes de uma execução de otimização com 10.000 gerações.....	135



# SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>21</b>
1.1 Contexto e justificativa.....	21
1.2 Objetivos.....	23
1.3 Contribuições pretendidas .....	23
1.4 Estrutura da dissertação .....	24
<b>2 REVISÃO BIBLIOGRÁFICA .....</b>	<b>25</b>
2.1 Hidrologia e processos hidrológicos .....	25
2.1.1 <i>Precipitação pluviométrica (chuva)</i> .....	26
2.1.2 <i>Evapotranspiração de referência</i> .....	27
2.1.3 <i>Evapotranspiração real baseada em índices de vegetação</i> .....	28
2.1.4 <i>Escoamento em cursos de água</i> .....	29
2.2 Gestão de recursos hídricos e o uso de modelos determinísticos .....	30
2.3 Modelos hidrológicos .....	32
2.3.1 <i>Objetivos, abordagens e características desejadas</i> .....	32
2.3.2 <i>Solução numérica</i> .....	33
2.3.3 <i>Calibração e validação</i> .....	34
2.4 Efeitos de memória e inferência causal em sistemas hidrológicos.....	36
2.4.1 <i>Efeitos de longa memória em séries hidrológicas</i> .....	36
2.4.2 <i>Causalidade</i> .....	37
2.4.3 <i>Teste de causalidade de Granger</i> .....	38
2.4.4 <i>Entropia, informação mútua e entropia de transferência</i> .....	40
2.4.5 <i>Análise de multirresolução</i> .....	43
2.5 Síntese da revisão bibliográfica.....	44
<b>3 MATERIAL E MÉTODOS .....</b>	<b>45</b>
3.1 Área de estudo .....	45
3.2 Dados .....	48
3.2.1 <i>Dados de precipitação</i> .....	48
3.2.2 <i>Dados meteorológicos e evapotranspiração de referência</i> .....	49
3.2.3 <i>Espacialização de dados de precipitação e de evapotranspiração</i> .....	50

3.2.4 Índices espectrais de vegetação e estimativa de evapotranspiração real .....	53
3.2.5 Monitoramento fluviométrico .....	57
3.2.5.1 Estação fluviométrica Santa Maria (SM).....	58
3.2.5.2 Estação fluviométrica calha Parshall CP2 .....	60
3.2.6 Transformação de nível d'água em vazão na seção fluviométrica Santa Maria.....	61
3.2.6.1 Descrição geral do método.....	61
3.2.6.2 Correção de nível d'água .....	65
3.2.6.3 Avaliação das incertezas .....	66
3.2.6.4 Avaliação da estratégia .....	66
3.2.7 Separação do escoamento de base.....	67
3.2.8 Níveis freáticos.....	68
3.2.9 Verificação do balanço hídrico no período monitorado .....	69
3.3 Análise de causalidade e conectividade de séries temporais .....	70
3.3.1 Teste de causalidade de Granger.....	70
3.3.2 Informação mútua com defasagem normalizada .....	70
3.3.3 Avaliação das ferramentas e dos efeitos de agregação temporal e de suavização .....	71
3.3.4 Análises pareadas de processos hidrológicos observados na área de estudo.....	74
3.4 Proposição do modelo conceitual hidrológico .....	76
3.4.1 Conhecimentos básicos sobre o comportamento hidrológico da área de estudo.....	76
3.4.2 Modelo perceptivo da bacia.....	76
3.4.3 Considerações sobre o funcionamento da bacia e parametrização dos fluxos .....	79
3.4.3.1 Reservatório de interceptação vegetal [Sc].....	79
3.4.3.2 Reservatório da zona não saturada [Su] .....	80
3.4.3.3 Reservatório da zona saturada [Ss].....	82
3.4.3.4 Composição de fluxos .....	83
3.4.5 Representação matemática e solução numérica.....	84
3.4.5 Calibração, validação e avaliação do modelo .....	85
3.5 Avaliação do modelo pelas métricas de causalidade e dependência .....	88
<b>4 RESULTADOS E DISCUSSÃO .....</b>	<b>89</b>
4.1 Precipitação média .....	89
4.2 Evapotranspiração de referência .....	92
4.3 Índices de vegetação e estimativa da evapotranspiração real .....	97
4.4 Estimativa das séries de vazão total e de vazão de base .....	101



4.4.1 Seção de monitoramento fluviométrico Santa Maria (SM).....	101
4.4.2 Seção de monitoramento fluviométrico calha Parshall 2 (CP2).....	107
4.5 Avaliação da consistência das estimativas dos processos hidrológicos .....	108
4.6 Análise de multirresolução de séries temporais .....	110
4.7 Identificação de conectividades e dependências em problemas sintéticos.....	112
4.8 Identificação de conectividade e de dependências nas séries de processos hidrológicos observados na área de estudo.....	119
4.9 Calibração e validação do modelo conceitual proposto .....	127
4.9.1. Comparação com trabalhos anteriores de modelagem na área de estudo, avaliação geral do modelo proposto e limitações.....	136
4.10 Avaliação do comportamento do modelo .....	138
4.11 Contribuição para prever a dinâmica da disponibilidade hídrica .....	142
<b>5 CONCLUSÕES.....</b>	<b>143</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>147</b>
<b>APÊNDICE A - VARIÁVEIS METEOROLÓGICAS .....</b>	<b>161</b>
Análises de regressão entre variáveis meteorológicas.....	161
Síntese das variáveis meteorológicas.....	165
<b>APÊNDICE B – CÓDIGOS COMPUTACIONAIS .....</b>	<b>167</b>
B1 Determinação de parâmetros variáveis da curva-chave.....	167
B1.1 Parâmetros determinados pelo método de mínimos quadrados ordinários .....	167
B1.2 Parâmetros determinados pelo método de otimização global – algoritmo differential evolution).....	168
B2 Teste de causalidade de Granger e informação mútua.....	170
<b>APÊNDICE C – NÍVEIS FREÁTICOS .....</b>	<b>175</b>
<b>APÊNDICE D – RESULTADOS DO PROBLEMA SINTÉTICO 2.....</b>	<b>179</b>



# 1 INTRODUÇÃO

## 1.1 Contexto e justificativa

Desmatamento de biomas naturais, mudanças no uso do solo, mudanças climáticas e aumento da população e da urbanização são algumas das mudanças globais que têm aumentado as incertezas referentes ao equilíbrio e à sustentabilidade futura do planeta (MARENGO et al., 2018; IPCC, 2019). A disponibilidade hídrica é sensível a essas mudanças e o aumento de conflitos e disputas pelo uso da água tem sido observado (LEVY; SIDEL, 2011). Em conjunto com a questão quantitativa, outros problemas como a degradação da qualidade da água, redução das reservas subterrâneas e aumento da vulnerabilidade a eventos extremos emergem no sentido de ameaçar a segurança hídrica e a integridade de ecossistemas e o desenvolvimento das atividades humanas (WEATHER; GOBER, 2015).

Desastres naturais relacionados a enchentes ou a secas causaram perdas econômicas globais superiores a 1,3 trilhões de dólares americanos entre 1992 e 2012 (UNISDR, 2012). Os cenários futuros, que preveem maiores demandas de água e mudanças nos padrões de precipitação, com aumento da frequência de eventos extremos, indicam que os impactos podem ser intensificados. Estratégias de adaptação e mitigação, relacionadas sobretudo a políticas de controle do uso do solo, à transferência de conhecimento e tecnologia, à educação, ao desenvolvimento de infraestrutura resiliente e à maior eficiência no uso dos recursos naturais estão entre as principais medidas para atenuar os impactos previstos, e para avançar nos objetivos do desenvolvimento sustentável das Nações Unidas (UN GENERAL ASSEMBLY, 2015).

Explicar e modelar os processos observados adequadamente têm um interesse especial na aplicação da Hidrologia, que é o de antecipar (prever) as respostas hidrológicas resultantes de mudanças das condições físicas das áreas de estudo (KIRCHNER, 2006). Entretanto, a ausência de monitoramento hidrológico abrangente de longo termo e o limitado poder preditivo de modelos hidrológicos são ainda algumas das barreiras que dificultam a elaboração de estratégias e políticas de gestão integrada de uso do solo e de uso de recursos hídricos voltadas para a garantia do abastecimento de água, e de forma mais geral, para a manutenção de serviços ecossistêmicos (KŘEČEK; HAIGH, 2018; EVARISTO; MCDONNELL, 2019). Essas duas barreiras podem ser abordadas de maneira conjunta, com o uso, por exemplo, de bacias experimentais para o desenvolvimento de estudos detalhados, que buscam explicações para o

complexo funcionamento dos sistemas hidrológicos e de suas interações com outros sistemas (WHITEHEAD; ROBINSON, 1992; BURT; MCDONNELL, 2015; ARAÚJO, 2016).

Apesar dos avanços computacionais e em tecnologias para observação de processos hidrológicos em superfície, subsuperfície e por sensoriamento remoto, a modelagem hidrológica ainda convive com elevadas incertezas nos dados observados, nas definições dos volumes de controle, nas condições iniciais e de contorno e incertezas paramétricas e estruturais. Não há uma estrutura geral que possibilita a simulação adequada de qualquer sistema hidrológico. Essa ausência motiva o desenvolvimento de estruturas flexíveis capazes de adaptar a modelagem para diferentes objetivos pretendidos e considerando diferentes graus de disponibilidade de dados.

Nesse contexto, surge a questão de como avaliar a adequação da estrutura de um modelo e resolver eventuais inadequações a partir de dados observados. Na perspectiva operacional da engenharia de recursos hídricos, a adequação é fortemente baseada na funcionalidade para um conjunto esperado de condições e na caracterização das incertezas (deficiências do modelo). Por outro lado, em outras comunidades de modelagem, a consistência física com o sistema ou a complexidade podem ser critérios mais relevantes para avaliar a adequação do modelo (GUPTA et al., 2012).

Sob a perspectiva de ser fisicamente consistente com o sistema hidrológico, a avaliação da adequação estrutural é uma etapa que pode colaborar para o aumento do poder preditivo do modelo. A conectividade entre processos observados e modelados, por exemplo, pode ser utilizada para avaliar essa consistência (BENNETT et al., 2019). Uma breve avaliação da equação básica da hidrologia (balanço hídrico) faz perceber a necessidade da existência de relações causa-efeito, conectividade e fluxo de informação entre as componentes do balanço hídrico.

O desenvolvimento deste trabalho é justificado pela demanda por estudos que apoiam o monitoramento contínuo de variáveis hidrológicas em bacias hidrográficas de baixa ordem (responsáveis por grande parte da disponibilidade hídrica superficial de grandes bacias), que objetivam a redução de incertezas preditivas e que exploram o uso de métodos estatísticos de análise de dados observados a favor da modelagem determinística dos processos.

O desenvolvimento de um modelo conceitual local buscou evitar algumas limitações relacionadas ao uso de outros modelos ao mesmo tempo em que buscou usar adequadamente os dados disponíveis na área de estudo e aproveitar o entendimento hidrológico adquirido em estudos anteriores. A aplicação de algum modelo conceitual com estrutura fixa, como o HBV (LINDSTRÖM et al., 1997), SIMHYD (CHIEW; PEEL; WESTERN, 2002), SMAP

(NASCIMENTO; REIS JR.; MARTINS, 2009), ou ainda de alguma estrutura flexível de modelagem, como a SUPERFLEX, proposta por Fenicia, Kavetski e Savenije (2011), poderia apresentar uma ou mais das seguintes limitações: falta de acesso a variáveis de estado; falta de flexibilidade na parametrização de processos ou no estabelecimento de conectividade entre elementos; dificuldade em implementar uma calibração automatizada de parâmetros; e a falta de documentação ou de acesso e/ou entendimento do código computacional, dificultando eventuais adaptações.

## **1.2 Objetivos**

Construir um modelo hidrológico conceitual para prever a dinâmica da disponibilidade hídrica de uma bacia hidrográfica.

As etapas deste trabalho foram:

- a. Quantificar os processos hidrológicos na área de estudo.
- b. Implementar computacionalmente e avaliar, com auxílio de problemas sintéticos, ferramentas de inferência causal e de análise de persistência entre processos.
- c. Caracterizar as relações de causalidade/conectividade e de persistência nas séries hidrológicas estimadas referentes à área de estudo.
- d. Propor um modelo hidrológico conceitual, incluindo: a representação dos processos dominantes; a parametrização das equações de fluxo; a definição do modelo matemático e do método numérico para resolução do sistema de equações de conservação de massa; o processo de calibração; e a avaliação e validação do modelo.
- e. Comparar e discutir as interdependências identificadas nas séries de dados observados com as identificadas nos resultados simulados.

## **1.3 Contribuições pretendidas**

Este trabalho busca produzir informações para orientar a modelagem de processos hidrológicos a partir de uma avaliação exploratória do comportamento do sistema hidrológico visando entender como eventos passados afetam a disponibilidade hídrica presente e futura. Entre as contribuições pretendidas diretamente com a execução desse trabalho estão: i) a

quantificação das conectividades e persistências entre os processos meteorológicos de estímulo e hidrológicos de resposta em escala de bacia hidrográfica; e ii) a proposição de um modelo conceitual que reproduza os processos observados e suas interdependências no domínio temporal.

Quanto à modelagem, uma questão específica foi abordada: “A especificação de defasagens entre processos deve ser explicitada na modelagem matemática de sistemas hidrológicos?”. Essa especificação poderia ser feita, por exemplo, a partir de equações diferenciais funcionais, mais especificamente, com equações diferenciais com defasagem (EDD) (GHIL; ZALIAPIN; THOMPSON, 2008), se contrapondo a esquemas convencionais baseados em sistemas de equações diferenciais ordinárias.

## **1.4 Estrutura da dissertação**

O esquema tradicional foi adotado neste texto. A revisão bibliográfica dos temas centrais do estudo é apresentada no capítulo 2; a área de estudo e os métodos de análise dos dados e modelagem são descritos no capítulo 3; os resultados e discussões são apresentados no capítulo 4; e as conclusões, respostas às contribuições pretendidas e recomendações finais são sintetizadas no capítulo 5.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 Hidrologia e processos hidrológicos

Hidrologia é a Geociência que descreve e prevê a ocorrência, circulação e distribuição de água doce na Terra, e inclui entre seus objetivos a melhor compreensão dos fluxos e interações da água sobre e sob a superfície terrestre e as trocas com a atmosfera (DINGMAN, 2014). Apesar dos progressos continuamente obtidos na capacidade de observar e assimilar grandes quantidades de dados, incertezas aleatórias e epistêmicas ainda dominam essa ciência em função da alta variabilidade natural dos processos envolvidos e da falta de teorias que expliquem adequadamente processos em diferentes escalas espaciais (MCDONNELL et al., 2007).

Quando isolados e analisados na microescala, muitos processos são razoavelmente bem caracterizados e descritos pelas teorias existentes (e.g., equação de Richards para escoamento de água em solos não saturados). Entretanto, na macroescala do mundo real, as heterogeneidades e as conexões e interdependências com fenômenos atmosféricos, com ecossistemas e com as atividades humanas, entre outros, conferem à Hidrologia e a outras geociências um comportamento complexo e de difícil previsão (KUMAR, 2015), levando à necessidade de estudos experimentais e da contínua coleta de dados em diferentes circunstâncias (BURT; MCDONNELL, 2015). Isso revela um caráter essencialmente empírico da Hidrologia, cujo avanço depende da assimilação de múltiplas observações e da proposição de teorias que consigam explicar adequadamente as variabilidades observadas (KIRCHNER, 2006).

O ciclo hidrológico é o componente central da Hidrologia e consiste em uma série de processos (e.g., precipitação, evaporação, transpiração, interceptação, infiltração, escoamento superficial, escoamento subterrâneo) que ocorrem em escalas locais e globais envolvendo a circulação da água entre a atmosfera, oceanos, superfície e subsuperfície terrestre (MAYS, 2010). Nessa circulação, diversas interações, dependências e retroalimentações são estabelecidas com outros ciclos biogeoquímicos (e.g., carbono, nitrogênio), com os ecossistemas e com a energia disponível (KUMAR, 2015).

Sistemas hidrológicos são usualmente avaliados em termos do balanço hídrico (baseado na conservação da massa), aplicável em qualquer escala temporal ou espacial (DINGMAN, 2014). Em geral, as medições ou estimativas das variáveis de entrada, saída e de estado, ou

mesmo as definições da geometria dos volumes de controle adotados para delimitar o sistema hidrológico, possuem grandes incertezas associadas (BEVEN, 2019). Apesar dessas incertezas, muitas aplicações hidrológicas ou relacionadas, desde a mais simples à mais avançada, encontra na equação de balanço hídrico uma forma de restringir e relacionar, de uma forma fisicamente baseada, os processos envolvidos.

### 2.1.1 Precipitação pluviométrica (chuva)

Em grandes escalas, o estudo da dinâmica e interação entre processos atmosféricos e oceânicos assume um caráter mais fisicamente baseado e determinístico, que busca explicar padrões de circulação e as causas de anomalias observadas nas formações de sistemas responsáveis pelo desenvolvimento de chuvas regulares em porções continentais (RODRIGUES et al., 2019).

As irregularidades na formação da zona de convergência do Atlântico Sul (ZCAS), por exemplo, que impactou a disponibilidade hídrica para o abastecimento humano, para a irrigação e para geração de energia hidrelétrica em uma parte do Brasil no verão de 2013/2014 evidenciam a importância do desenvolvimento de estudos nessa área, de modo a esclarecer os riscos e incertezas associadas à variabilidade climática (COELHO et al., 2016; RODRIGUES et al., 2019). Em escala local, a precipitação é admitida como uma variável de natureza estocástica, e os mecanismos responsáveis pela sua geração não são objetos de explicação em grande parte das aplicações em Hidrologia em função da complexidade e aleatoriedade envolvidas.

Em geral, a medição da precipitação ocorre pontualmente, por meio de pluviômetros convencionais, com amostragem temporal diária, ou automáticos, com amostragem temporal sub-horária. Alguns estudos têm demonstrado que mesmo em pequenas escalas espaciais (sub-quilométrica) a precipitação pode apresentar uma grande variabilidade (FAURÈS, 1995; FIENER; AUERSWALD, 2009).

O uso de mais de um ponto de monitoramento pluviométrico simultâneo associado a algum esquema de interpolação espacial é uma das estratégias empregadas para lidar com a questão da variabilidade da chuva. Dirks et al. (1998) e Ly, Charles e Degré (2011), ao compararem alguns dos métodos mais utilizados para interpolar dados de precipitação, concluíram que o método determinístico de ponderação pelo inverso das distâncias (*inverse distance weighting* – IDW), embora simples, permite obter resultados de qualidade similar aos obtidos por métodos mais complexos. Outro método determinístico simples comumente



utilizado é o do vizinho natural (*natural neighbor interpolation* – NNI). No estudo de Alves da Silva et al. (2019), com dados mensais de precipitação do estado de Pernambuco entre 1950 e 2012, o método NNI apresentou um desempenho ligeiramente superior ao método IDW. Em uma análise espacialmente mais extensiva, comparando métodos geoestatísticos, de análise de superfície e determinísticos, com dados meteorológicos registrados entre 1961 e 1990 no continente europeu, Hofstra et al. (2018) não encontraram um método consideravelmente superior aos demais.

### 2.1.2 Evapotranspiração de referência

Evapotranspiração é o processo de transferência de água para a atmosfera, resultante tanto da evaporação de água de superfícies livres quanto da transpiração vegetal, que é o movimento da água do solo para a planta e da planta para atmosfera (NOVÁK, 2012). Em média, cerca de 65 % do total precipitado em superfícies terrestres retornam à atmosfera através da evapotranspiração (MCDONALD, 1961), indicando, assim, a importância dessa componente no ciclo hidrológico.

Por depender do complexo sistema solo-vegetação-atmosfera, sua quantificação é desafiadora (ABTEW; MELESSE, 2012). A seleção do método de estimativa da evapotranspiração depende da disponibilidade e qualidade de dados e da acurácia requerida. O uso combinado de valores de referência (estimados a partir de dados meteorológicos) e de coeficientes de cultura (determinados experimentalmente para diferentes coberturas vegetais) é um dos métodos mais difundidos em termos práticos, principalmente para fins de irrigação (KAMBLE et al., 2013). Entre os métodos de referência, destaca-se o método de Penman-Monteith proposto pela FAO (*Food and Agriculture Organization of the United Nations*) (ALLEN et al., 1998).

$$ET_o = \frac{0,408 \Delta [R_n - G] + \gamma \frac{C}{T + 273} u_2 [e_s - e_a]}{\Delta + \gamma(1 + 0,34u_2)} \quad (1)$$

em que:  $ET_o$  é a evapotranspiração de referência (mm/dia se  $C = 900$  ou mm/h se  $C = 37$ );  $\Delta$  é a declividade da curva de pressão de vapor em função da temperatura ( $kPa \text{ } ^\circ C^{-1}$ );  $[R_n - G]$  é a diferença entre o saldo de radiação total e o fluxo de calor no solo no intervalo de tempo de cálculo ( $MJ/m^2/dia$  ou  $MJ/m^2/h$ );  $T$  é a temperatura média ( $^\circ C$ );  $u_2$  é a velocidade média do

vento a 2 m de altura do solo (m/s);  $[e_s - e_a]$  é o déficit de pressão de vapor do ar (kPa);  $\gamma$  é a constante psicométrica ( $\text{kPa } ^\circ\text{C}^{-1}$ ). A superfície de referência adotada pelo método consiste em uma grama hipotética com altura de 0,12 m, resistência estomática de  $70 \text{ s m}^{-1}$  e albedo de 0,23.

Os dados meteorológicos requeridos para o cálculo de  $ET_o$  são temperatura do ar, umidade relativa do ar, radiação solar e velocidade do vento, que são pontualmente obtidos. A ideia de possuir mais de um ponto de monitoramento desses dados, calcular  $ET_o$  nesses pontos e em seguida interpolar os resultados é uma das abordagens possíveis para estimar a variabilidade da evapotranspiração em uma dada área. Tegos, Malamos e Koutsoyiannis (2015), ao compararem diferentes métodos de interpolação de evapotranspiração, concluíram que o simples método IDW apresenta desempenho compatível com os demais.

### 2.1.3 Evapotranspiração real baseada em índices de vegetação

A utilização de produtos de sensoriamento remoto permite incluir a heterogeneidade e a dinâmica espaço-temporal das condições de uso e cobertura do solo. Em geral, os métodos de sensoriamento mais empregados com esse objetivo fazem o balanço de energia a partir de dados de emissão na região do infravermelho termal e/ou utilizam índices de vegetação em conjunto com dados meteorológicos obtidos em superfície (ANDERSON et al., 1997; NAGLER et al., 2013). A decisão de qual método utilizar pode depender das limitações e restrições de cada localidade e dos objetivos pretendidos.

Entre os benefícios dos métodos que utilizam índices de vegetação está a possibilidade de identificar as diferentes fases de desenvolvimento fisiológico vegetal, incluindo variações sazonais. A assinatura espectral singular da vegetação possibilita a avaliação de processos complexos que ocorrem com retroalimentações (*feedbacks*), como é o caso da evapotranspiração e da disponibilidade hídrica. Nesse sentido, diversos trabalhos têm demonstrado que as relações entre a evapotranspiração real ( $ET_a$ ) e de referência ( $ET_o$ ) podem ser determinadas de forma razoável, em diferentes biomas, a partir de índices de vegetação obtidos por sensoriamento remoto e de medições em superfície para calibrar e validar as relações (modelos) (GLENN et al., 2008; GLENN et al., 2011; KAMBLE et al., 2013; NAGLER et al., 2013). Outra vantagem está associada a não necessidade de mapas auxiliares de cobertura do solo após a validação do modelo para um conjunto estacionário de condições (GUERSCHMAN et al., 2009).

Por outro lado, entre as desvantagens dos métodos baseados em índices de vegetação está a impossibilidade de identificar efeitos de curto prazo relacionados a falta de água no solo e de detectar a parcela de evaporação direta (GLENN et al., 2011; NAGLER et al., 2013).

#### **2.1.4 Escoamento em cursos de água**

A água que passa em uma seção transversal de um manancial superficial é o produto de todas as interações que ocorrem de forma espacialmente distribuída em um determinado volume de controle a montante. A determinação das origens e caminhos que as águas percorrem para formar o escoamento, com auxílio de traçadores isotópicos e geoquímicos, por exemplo, tem sido uma das estratégias de estudos experimentais para avançar a ciência hidrológica (MCGUIRE; MCDONNELL, 2015; CARTWRIGHT; MORGENSTERN, 2018).

A discussão em torno da origem da água que forma o escoamento nos cursos de água vem pelo menos do fim da década de 1970, quando Sklash e Farvolden (1979) demonstraram por meio de medições com isótopos a possível dominância de águas subterrâneas durante eventos de chuva, se contrapondo ao entendimento científico que se tinha (esperava-se que a composição do escoamento fosse dominada pela água da chuva). Atualmente, os mecanismos responsáveis pelo rápido deslocamento de águas armazenadas na bacia (anteriores ao evento de precipitação) para formação do escoamento e a simultânea lenta resposta em termos de variação de solutos dissolvidos no escoamento são mais bem compreendidos (MCDONNELL; BEVEN, 2014; HRACHOWITZ et al., 2016).

O escoamento em cursos de água possui três componentes características principais relacionadas a suas origens espaciais: escoamento de base (contribuição do aquífero); escoamento subsuperficial (contribuição resultante da zona vadosa); e escoamento superficial (capacidade de infiltração excedida, segundo modelo hortoniano). Na escala de bacia hidrográfica, métodos fisicamente baseados ou empíricos ainda enfrentam dificuldades para estimar essas contribuições, em função de heterogeneidades espaciais, alta demanda por dados, e possível inadequação do método para a área de estudo.

O método empírico do *National Resources Conservation Service* (CN-SCS), por exemplo, embora desenvolvido para um limitado número de pequenas bacias rurais nos Estados Unidos para estimar o escoamento superficial, teve seu uso disseminado em diversas regiões do mundo com pouco criticismo ao longo dos últimos 50 anos (HAWKINS, 2014; BURT; MCDONNELL, 2015).

O escoamento é medido em termos de vazão ( $Q$ ), volume de água por unidade de tempo que passa em uma seção de um curso d'água, que é usualmente monitorada de forma indireta a partir de dados do nível d'água ( $h$ ), que são transformados em vazão usando uma função de transformação. Essa função de transformação (curva-chave),  $F(h)$ , é obtida a partir de medições independentes e simultâneas, espaçadas no tempo, de nível d'água e de vazões.

Entre os métodos usuais de medição de vazão está o do molinete hidrométrico (*velocity-area method*), que faz a integração da velocidade média do escoamento ao longo da seção transversal (HERSCHY, 1993). Estruturas hidráulicas simples, como vertedouros e calhas Parshall, com base no princípio de conservação de energia e em estudos empíricos, também permitem estabelecer relações entre a carga hidráulica e a vazão afluyente. Alternativamente, há equipamentos acústicos (perfilador acústico de correntes – ADCP), que, baseados no princípio do efeito Doppler, estimam o perfil de velocidades e a geometria da seção de escoamento.

As incertezas envolvidas no processo de medição de vazão e da transformação de nível d'água em vazão podem ser consideravelmente altas e por isso são objetos de inúmeros estudos, que objetivam tanto sua quantificação quanto estratégias para a sua redução (LE COZ et al., 2012; MCMILLAN; WESTERBERG, 2015).

## **2.2 Gestão de recursos hídricos e o uso de modelos determinísticos**

Uma das dimensões da sustentabilidade da gestão de recursos hídricos é diagnosticada, em diferentes recortes territoriais de planejamento, a partir do balanço contínuo entre disponibilidades e demandas humanas e ambientais. A pequena ou inexistente alteração nas estimativas de disponibilidade hídrica e nos valores de máximas vazões outorgáveis no Brasil, por exemplo, reflete as abordagens estáticas utilizadas na prática da gestão de recursos hídricos, que supõem a estacionariedade dos sistemas hidrológicos (DETZEL; FERNANDES; MINE, 2016).

O uso de métodos estatísticos para cálculo de vazões de referência baseados apenas na frequência de eventos observados, além de poder resultar em inconsistências em função da incompatibilidade das hipóteses assumidas, não permite antecipações a eventos extremos, como, por exemplo, o enfrentado em 2014 no estado de São Paulo – momento em que o persistente baixo volume de precipitações impactou adversamente a disponibilidade de água nos mananciais de uma das regiões mais populosas e economicamente ativas da América Latina. Por não considerar as conexões causais (e.g., dependências com reservas de águas

subterrâneas), os métodos tradicionais estatísticos de estimativa de disponibilidade hídrica não permitem prever alterações em curtos períodos de tempo – o que pode ser cada vez mais necessário em função das evidências de como a variabilidade de eventos e fenômenos climáticos globais e de mudanças no uso do solo em escala regional podem afetar o regime hidrológico em escalas locais (MARENGO; ALVES, 2015; MARENGO et al., 2018; RODRIGUES et al., 2019).

Atualmente, a previsão hidrológica tem uma grande relevância para prover informações a favor da segurança hídrica e da sustentabilidade do uso dos recursos hídricos tendo em vista o cenário futuro de incertezas globais associadas ao crescimento populacional (LUTZ; SAMIR, 2010), à insegurança alimentar (SINGH; BORROK, 2019) e às mudanças climáticas (IPCC, 2019).

Para uma determinada escala espacial, a condição em que os recursos hídricos estão disponíveis e acessíveis em quantidade e qualidade adequadas para atender às demandas humanas e ambientais, de forma sustentável e a longo prazo, é uma das premissas e objetivos para garantir a segurança hídrica. Diferentes indicadores têm sido propostos para avaliar objetivamente essa segurança, incluindo suas dimensões – disponibilidade, acessibilidade, qualidade e gestão. Especificamente, quanto à avaliação da disponibilidade, o escoamento de base e as reservas hídricas são considerados importantes indicadores da vulnerabilidade do sistema hidrológico a eventos de escassez hídrica (GAIN; GIUPPONI; WADA, 2016; PIRES et al., 2017; THOMAS; CAINETA; NANTEZA, 2017; ANA, 2019).

A quantificação dos impactos em mananciais superficiais decorrentes da intensa exploração de águas subterrânea para suprir demandas urbanas e agrícolas, assim como da influência de fenômenos climáticos nas respostas das bacias, ainda são objetos de pesquisa da Hidrologia e representam um dos desafios para promover uma gestão sustentável dos recursos hídricos (SOPHOCLEOUS, 2002; VAN LANEN et al., 2013; LIUZZO et al., 2014; DUVERT et al., 2015; WU et al., 2015). Nesse contexto está inserida a estratégia de modelagem hidrológica, que permite, a partir de um conjunto limitado de observações no tempo e no espaço, prever comportamentos para diferentes condições (e.g., alterações no uso do solo, no clima, nos padrões de precipitação), auxiliando no processo de tomada de decisão e no processo de adaptação para enfrentar possíveis cenários indesejáveis no futuro (BEVEN, 2012).

## 2.3 Modelos hidrológicos

### 2.3.1 Objetivos, abordagens e características desejadas

O objetivo de modelos é em geral extrapolar espacial e temporalmente dados observados de modo a obter alguma informação de interesse (BEVEN, 2007). Por natureza, eles são incompletos e imperfeitos para representar toda a complexidade do mundo real, e assim as incertezas estarão seguramente presentes em qualquer resultado obtido por meio deles (EFSTRATIADIS; KOUTSOYIANNIS, 2010). O avanço no desenvolvimento de novos modelos é motivado pela redução dessas incertezas, pelo aumento do poder preditivo e pelo reconhecimento de que há muito a se fazer para integrar e descrever a dinâmica e as interações dos sistemas hidrológicos, ecológicos e humanos (HRACHOWITZ et al., 2016; BLÖSCHL, 2019).

Dois grupos de modelos determinísticos são tipicamente utilizados na Hidrologia: conceituais concentrados (ou semi-distribuídos) e distribuídos fisicamente baseados. Um modelo conceitual concentrado consiste em uma estrutura simplificada de representação dos processos hidrológicos dominantes, com parâmetros que dependem de dados observados para serem calibrados (WAGENER et al., 2001). Já modelos distribuídos fisicamente baseados, incluem representações físicas dos processos, válidos no microescala e cuja aplicação direta na macroescala é questionável (KIRCHNER, 2006). Por depender ainda de uma grande quantidade de parâmetros, espacialmente distribuídos, modelos fisicamente baseados sofrem com a super-parametrização, o que pode dificultar a sua aplicação.

Algumas características desejadas para o processo de modelagem conceitual são: 1) otimização do uso das informações disponíveis de modo a reduzir as incertezas estruturais (relacionadas a estrutura de representação) e paramétricas (relacionadas à identificação dos parâmetros); 2) equilíbrio entre a complexidade do modelo, suportado pela disponibilidade de dados, com a aplicação desejada; e 3) calibração multiobjetivo de parâmetros baseada em variáveis de fluxo e de estado (SEIBERT, 1999; WAGENER et al., 2001). Na década de 1970, quando Bergström e Forsman (1973) desenvolveram o modelo HBV para bacias localizadas em países nórdicos, algumas dessas ideias pareciam já estar consolidadas na comunidade hidrológica.

Embora seja possível investir recursos para propor modelos localmente mais adequados, a prática indica um uso preferencial por ‘modelos de todos os lugares’ (BEVEN, 2007), que possui suas vantagens específicas. A seleção de modelos, entretanto, parece não ser orientada por questões técnicas, mas, sim, simplesmente por questões regionais. Addor e Melsen (2019)

ao analisar 1529 estudos hidrológicos publicados entre 1991 e 2018, verificaram que em 74% dos estudos o modelo utilizado podia ser previsto baseando-se somente na filiação do primeiro autor. Ainda, os diferentes modelos eram usados para um diversificado conjunto de objetivos, condições climáticas, biomas e escalas espaciais e temporais.

Para superar as potenciais inadequações associadas, a substituição de estruturas fixas de modelagem para estruturas flexíveis, capazes de acomodar a disponibilidade local de dados e a variabilidade geomorfológica e climática, tem sido uma das práticas emergentes para obter modelos mais adequados para as aplicações pretendidas (FENICIA; KAVETSKI; SAVENIJE, 2011; KRAFT et al., 2011; CLARK et al., 2015; ADDOR; MELSEN, 2019).

O modelo SUPERFLEX proposto por Fenicia, Kavetski e Savenije (2011) é um exemplo de como uma abordagem flexível pode ser implementada. Ela permite a construção de hipóteses e o teste de diferentes combinações de componentes, como reservatórios, conexões, equações constitutivas e funções de transferência, equilibrando a complexidade do modelo com a disponibilidade local de dados para reproduzir as respostas hidrológicas (FENICIA; KAVETSKI; SAVENIJE, 2011).

De forma similar, porém com base mais física, Clark et al. (2015) propuseram o SUMMA – *Structure for Unifying Multiple Modeling Alternatives* – que idealiza a combinação de processos físicos dominantes para representar os fluxos de água e energia que afetam a evolução temporal dos estados (termodinâmicos e hidrológicos) de interesse. O conceito de flexibilidade nas parametrizações de equações de fluxo e nas conexões é também empregado.

Outros trabalhos, mais ou menos abstratos, tem disponibilizado meios de testar diferentes estruturas para simular os processos hidrológicos, facilitando a inter-comparação e possibilitando a redução de incertezas estruturais envolvidas na seleção de um modelo hidrológico (KRAFT et al., 2011; KNOBEN et al., 2019). Outras abordagens, que pensam no futuro do processo de modelagem, sugerem ainda a adoção de uma visão sistemática orientada pelo processo de co-evolução do clima, solo, vegetação e topografia, de modo a entender e dar significado a variedade de distintos comportamentos observados (MCDONNELL et al., 2007; HARMAN; TROCH, 2014).

### **2.3.2 Solução numérica**

No contexto de grandes incertezas sobre como modelar e parametrizar os processos hidrológicos, a solução numérica das equações diferenciais governantes acaba recebendo pouca atenção. Clark e Kavetski (2010) demonstraram como decisões inadequadas sobre os métodos numéricos podem afetar e prejudicar a aplicação de modelos hidrológicos conceituais.

Dos resultados de trabalhos relacionados, as principais conclusões foram a contra-indicação de métodos de integração explícitos com passo de tempo fixo, a necessidade de controle de erro numérico e a preferência por métodos com passo de tempo adaptativo (CLARK; KAVETSKI, 2010; FENICIA; KAVETSKI; SAVENIJE, 2011).

O uso de métodos explícitos de Runge-Kutta de ordem 4 e 5 com passo de tempo adaptativo, por exemplo, apresenta um bom desempenho em termos de acurácia numérica e eficiência computacional (FENICIA; KAVETSKI; SAVENIJE, 2011). Nesses métodos, a diferença entre as soluções de ordem 5 e 4 é admitida como o erro da solução de ordem 4, e o passo é então ajustado de modo que o erro máximo tolerado estabelecido (absoluto e relativo) não seja superado. Como simplificação, os dados de entrada são geralmente considerados constantes nos passos de tempo de simulação (SANTOS; THIREL; PERRIN, 2018).

A eficiência computacional (e.g., pequeno tempo de processamento) de métodos numéricos de integração é relevante no contexto de calibração de parâmetros, especialmente quando um elevado número de avaliação de funções é utilizado pelo método de otimização.

### **2.3.3 Calibração e validação**

Em modelos conceituais, os parâmetros não possuem uma correspondência com propriedades físicas, e, portanto, necessitam de dados observados para serem calibrados. A calibração consiste na obtenção de um conjunto de parâmetros que minimiza as diferenças entre os resultados obtidos pelo modelo e os dados observados. Essas diferenças são em geral descritas em termos de funções objetivos (métricas paramétricas ou não paramétricas que estimam o erro do modelo).

O uso de fluxos e estados observados em funções objetivos independentes, em uma abordagem de calibração multiobjetivo, contribui para um melhor condicionamento de parâmetros e para a redução das incertezas estruturais e paramétricas do modelo. Por exemplo, conduzir uma calibração usando apenas dados observados de vazão no exutório da bacia força o modelo a reproduzir esse fluxo especificamente e não garante uma reprodução adequada dos demais fluxos ou de estados (e.g., umidade do solo, nível freático) (SEIBERT, 1999; MOSTAFAIE et al., 2018).

Para modelos conceituais, calibrações de objetivo único conseguem identificar de forma razoável em torno de 6 parâmetros (WAGENER et al., 2001). Assim, a inclusão de funções objetivos independentes também garante maior capacidade de identificação dos parâmetros (EFSTRATIADIS; KOUTSOYIANNIS, 2010).



Em casos de objetivos conflitantes, não existe um único ponto ótimo, mas, sim, um conjunto de pontos denominados ótimos de Pareto. Uma solução é classificada como ótimo de Pareto, ou como não dominada, se não existe outra solução factível que possa melhorar algum objetivo sem causar uma deterioração simultânea em pelo menos uma outra função objetivo. Irregularidades na frente de Pareto podem indicar problemas estruturais na proposição do modelo ou na parametrização das equações de fluxo (EFSTRATIADIS; KOUTSOYIANNIS, 2010).

De forma geral, há várias evidências que tornam as estratégias multiobjetivos mais efetivas que as estratégias baseadas em objetivo único. Em grandes escalas, por exemplo, com a possibilidade de uso de produtos de sensoriamento remoto relacionados ao armazenamento de água, como os resultantes do GRACE (*Gravity Recovery and Climate Experiment*), é possível tentar melhorar as simulações hidrológicas a partir do condicionamento simultâneo de fluxos e estados (MOSTAFAIE et al., 2018).

Quanto aos métodos de otimização global, com restrições, para determinar os parâmetros, o uso de meta-heurísticas – especialmente de algoritmos evolutivos (AEs) – vem se popularizando devido à simplicidade conceitual e à capacidade de lidar com problemas complexos e não lineares (MAIER et al., 2014). O princípio desses algoritmos está baseado nos mecanismos da evolução biológica, envolvendo conceitos similares aos de população, adaptação, introdução de diversidade e seleção natural. Em contraste com métodos orientados por gradientes, AEs não requer que a função objetivo seja diferenciável ou contínua e nem que tenha algum comportamento matemático específico. Entretanto, como desvantagem quando comparado a métodos clássicos de otimização baseados em gradiente, AEs requerem um elevado número de avaliação da função objetivo.

A validação é uma etapa que sucede a calibração e é essencial para orientar o processo de criação de modelos (SEIBERT, 1999) e uma oportunidade para verificar se o modelo fornece as respostas certas pelos motivos corretos. Biondi et al. (2012) indicaram uma série de orientações relacionadas à validação de desempenho e à validação científica de modelos hidrológicos. Para a validação de desempenho, eles recomendam: 1) usar dados independentes dos usados na calibração; 2) usar diferentes métricas de avaliação e visualizar graficamente os resultados; 3) discutir sobre as incertezas dos dados de entrada utilizados; e 4) quando possível, validar o modelo para diferentes escalas espaciais e temporais. Para a validação científica, as recomendações são praticamente coincidentes com as referentes à proposição de modelos: 1) identificar o objetivo do modelo; 2) listar e discutir as hipóteses, considerações e simplificações assumidas; 3) justificar o grau de complexidade adotado; 4) avaliar e discutir hipóteses

alternativas de modelagem; 5) verificar a coerência dos resultados com a base física dos processos simulados; 6) tornar o trabalho reproduzível a partir da disponibilização de dados e códigos computacionais; e 7) identificar os pontos fortes, fracos e as oportunidades de melhoria (BIONDI et al., 2012).

Complementarmente, traçadores isotópicos podem contribuir na validação de modelos, pois permitem obter propriedades relacionadas às origens e caminhos da água e assim possibilitam o teste objetivo das hipóteses e das considerações adotadas sobre a parametrização de fluxos e compartimentação de armazenamentos de água (HRACHOWITZ et al. 2013; KUPPEL et al., 2018).

## **2.4 Efeitos de memória e inferência causal em sistemas hidrológicos**

### **2.4.1 Efeitos de longa memória em séries hidrológicas**

A predição está entre os objetivos de muitos sistemas de modelagem. A presença de propriedades em uma série temporal que tornam possível prever termos futuros a partir do comportamento passado observado é, portanto, relevante. A existência de persistência em variáveis hidrológicas e climatológicas é amplamente reconhecida em diversas escalas temporais. Isso significa, de forma geral, que valores altos tendem a ser sucedidos por valores altos, e valores baixos, por valores baixos.

Zimmermann, Elsenbeer e Moraes (2006) e Tomasella et al. (2008) em estudos experimentais envolvendo alteração de uso do solo na floresta Amazônica, por exemplo, identificaram significantes efeitos de memória em seus resultados. Isso sugeriu que o histórico de uso e ocupação estava influenciando nas respostas hidrológicas. Hurst (1951) foi o precursor na identificação de que a hipótese de independência serial de dados hidrológicos resultava em inconsistências estatísticas, desencadeando uma série de estudos que formalizaram a teoria de processos de memória longa. A discussão central foi baseada na tentativa de explicar a consistente divergência ('fenômeno Hurst') entre resultados empíricos e resultados teóricos previstos.

De forma a modelar matematicamente essa divergência, Mandelbrot e van Ness (1968) sugeriram a presença de uma infinita interdependência entre vazões sucessivas, dando origem à teoria do movimento Browniano fracionário (fBm), capaz de explicar o fenômeno Hurst de forma simples, com apenas um parâmetro. Por não ter sentido físico algum (interdependência infinita), a proposta foi prontamente contestada por hidrólogos, que buscavam explicações

baseadas em mecanismos físicos (ou ao menos fisicamente possíveis) para reproduzir de maneira correta os comportamentos observados (KLEMEŠ, 1974; GRAVES et al., 2017).

Embora não tenha proporcionado avanços à Hidrologia do ponto de vista físico, Manbelbrot e van Ness (1968) contribuíram para o desenvolvimento dos modelos autorregressivos integrados fracionários de médias móveis – ARFIMA (GRANGER; JOYEUX, 1980; HOSKING, 1981), que são atualmente utilizados para modelar processos que exigem memória longa e encontram aplicações práticas e operacionais em algumas áreas.

Alguns autores defendem que um melhor entendimento das dependências de longo termo pode colaborar na previsão de eventos extremos e de suas intensidades (PELLETIER; TUCOTTE, 1997; DEY; MUJUMDAR, 2018), porém faltam estudos que demonstrem como reproduzir tais dependências durante a modelagem de processos físicos, como um sistema dinâmico. A identificação de conectividades e dependências entre processos, possível primeira parte do problema de tentar reproduzir as dependências, pode ser feita por meio de métodos de inferência causal, que objetivam identificar relações de causa e efeito a partir de um conjunto de observações.

#### **2.4.2 Causalidade**

Causalidade é um conceito de difícil definição lógica e matemática relacionado ao fluxo e à conectividade de informação entre processos. A função essencial da análise de causalidade é a de inferir processos futuros a partir de observações passadas, elucidando relações de causa e efeito entre variáveis e eventos (RUSSELL, 1913). O uso de métodos de inferência causal na Hidrologia encontra motivação na complexidade das múltiplas interações e interdependências que existem entre os sistemas hidrológicos, climáticos, ambientais e humanos, nos domínios espaço e tempo (KUMAR, 2015).

Em uma abordagem de causalidade determinística, espera-se que a causa cronologicamente anteceda o efeito e que as condições de necessidade (se X ocorre, então Y deve ocorrer) e suficiência (se Y ocorre, então X deve ter ocorrido) sejam atendidas. Entretanto, grande parte dos processos de interesse relacionados à dinâmica do mundo real não são estritamente determinísticos, sugerindo a reformulação das condições de causalidade para algo no sentido probabilístico (e.g., se X ocorre, então a probabilidade de Y ocorrer aumenta) (HLAVÁČKOVÁ-SCHINDLER et al., 2007).

Não há métodos estatísticos capazes de identificar a causalidade entre processos exclusivamente a partir dos dados, sem nenhuma informação adicional sobre como os dados estão relacionados (PEARL; GLYMOUR; JEWELL, 2016). A expressão ‘correlação (ou

associação) não implica causa’, bem conhecida na estatística, foi consolidada pelos inúmeros casos em que a inadequada premissa contrária (‘associação implica causa’) leva a resultados inesperados.

A adoção de suposições sobre a estrutura causal do conjunto de dados na forma de um modelo estrutural causal, que indica claramente as interações e sentidos de dependência entre os processos, é um possível ponto de partida para lidar com questões de causalidade. Esse modelo consiste em dois conjuntos de variáveis,  $U$  e  $V$ , e de um conjunto de funções  $f$  que atribui a cada variável de  $V$  um valor baseado nos valores das outras variáveis do modelo (PEARL; GLYMOUR; JEWELL, 2016). Com base nessas considerações, estes autores propuseram a seguinte definição de causa: a variável  $X$  é uma causa direta de  $Y$  se  $X$  aparece na função que atribui os valores de  $Y$ . A variável  $X$  é uma causa de  $Y$  se  $X$  é uma causa direta de  $Y$  ou de qualquer causa de  $Y$ . As variáveis em  $U$  são chamadas exógenas (externas ao modelo), e não são objetos de explicação. As variáveis em  $V$  são chamadas endógenas, e são dependentes de pelo menos uma variável do conjunto  $U$ .

Diversos métodos foram desenvolvidos para inferir causalidade (e.g., método de Granger e variações, mapas causais, redes causais, algoritmos baseados na teoria da informação) (RUNGE et al., 2019). A seleção do método mais adequado depende do conhecimento prévio que se tem do sistema, da natureza das variáveis envolvidas e dos objetivos pretendidos.

### 2.4.3 Teste de causalidade de Granger

Uma definição simples e até hoje amplamente utilizada para lidar com a causalidade baseia-se no poder de predição. Granger (1969) propôs que: “o processo  $X$  causa, no sentido de Granger, outro processo,  $Y$ , se valores futuros de  $Y$  podem ser melhor previstos usando valores de  $X$  e  $Y$  ao invés de apenas valores de  $Y$ ”.

Seja  $U$  toda a informação disponível no universo e  $\overline{U - X}$  toda a informação disponível no universo exceto a informação sobre a variável  $X$ . Se  $\sigma^2(Y|U) < \sigma^2(Y|\overline{U - X})$ , então é dito que  $X$  causa  $Y$  no sentido de Granger. A variância ( $\sigma^2$ ) é adotada como critério pela facilidade de ser manipulada e interpretada matematicamente.

Granger (1969) fez uma série de considerações para tornar o método aplicável, entre elas que os dados são gerados segundo um processo autorregressivo linear, gaussiano e estacionário – AR(p). No caso mais simples, bivariado, dois modelos são considerados – o irrestrito (2), com o histórico das variáveis  $X$  e  $Y$ , e o restrito (3), apenas com o histórico da variável  $Y$ .

$$y_{u_t} = c_u + \sum_{j=1}^p [\phi_{1j}y_{t-j} + \phi_{2j}x_{t-j}] + \varepsilon_t, \quad \varepsilon_t \sim N(0, v_u) \quad (2)$$

$$y_{r_t} = c_r + \sum_{j=1}^p \phi_{1j}y_{t-j} + \varepsilon_t, \quad \varepsilon_t \sim N(0, v_r) \quad (3)$$

em que  $c_u$ ,  $c_r$ ,  $\phi_{1j}$  e  $\phi_{2j}$  são parâmetros dos modelos, e  $\varepsilon_t$  é um erro aleatório, não auto-correlacionado, com distribuição normal, média zero e variância  $v$ .

A hipótese nula  $H_0$  do teste é aceita quando  $\phi_{2j} = 0$ , para  $j = 1, 2, \dots, p$ , significando que  $X$  não causa  $Y$  no sentido de Granger. Por outro lado, a hipótese nula é rejeitada quando  $\phi_{2j} \neq 0$  para algum  $j$  entre 1 e  $p$ . Em teste de hipóteses envolvendo modelos aninhados (*nested models*), como é o caso dos modelos restrito e irrestrito, a razão de verossimilhança (LR) é um dos critérios comumente utilizados.

$$LR = -2 \ln \left[ \frac{\mathcal{L}_r(\hat{\theta}_0)}{\mathcal{L}_u(\hat{\theta}_1)} \right] \sim \chi^2 (gl = p_u - p_r) \quad (4)$$

em que LR é razão de verossimilhança,  $\hat{\theta}_i$  indica a estimativa de máxima verossimilhança de  $\theta_i$  (parâmetros do modelo) sob a hipótese  $i$ ,  $\mathcal{L}$  indica a função de verossimilhança, e os subscritos  $u$  e  $r$  fazem referência aos modelos irrestrito e restrito, respectivamente. A LR segue aproximadamente uma distribuição  $\chi^2$  com graus de liberdade igual à diferença entre os números de parâmetros dos modelos irrestrito ( $p_u$ ) e restrito ( $p_r$ ) (WILKS, 1938). A aproximação melhora à medida que o tamanho amostral aumenta.

A falta de um conhecimento sobre o sistema pode resultar na obtenção de resultados inesperados com o uso dessa definição, dando sentido a formulações mais rigorosas (e abstratas), como a proposta por Pearl, Glymour e Jewell (2016). Entretanto, para fins práticos, o teste de Granger é uma das ferramentas estatísticas mais consolidadas para verificar a não existência de relações causa e efeito com defasagens em séries temporais e para determinar a conectividade entre variáveis aleatórias (MARINAZZO; PELLICORO; STRAMAGLIA, 2008; BARNETT; SETH, 2015).

A grande vantagem do método proposto por Granger (1969) consiste na consideração dos efeitos de memória das variáveis de resposta (efeitos), tendo, portanto, aplicação em processos que apresentam comportamentos persistentes. Métodos baseados em correlação linear (seja no domínio temporal ou da frequência) ou em regressão com defasagem entre variáveis de causa e efeito são sujeitos a falhas quando empregados em inferência causal e na detecção de defasagens entre processos, justamente por não levarem em consideração o efeito de memória dos dados (RUNGE; PETOUKHOV; KURTHS, 2014; MCGRAW; BARNES, 2018).

Diversos estudos em sistemas dinâmicos têm aplicado o conceito de causalidade de Granger para extrair informações relacionadas à conectividade causal e à defasagem entre processos: Detto et al. (2012) em redes ecológicas; Papagiannopoulou et al. (2017) para investigar a dinâmica entre clima e vegetação; McGraw e Barnes (2018) em estudos de variabilidade climática; Singh e Borrok (2019) em estudos relacionado a redução de reservas de água subterrânea e produção de alimento em escala global; Huang et al. (2019) em análises entre variáveis climáticas e hidrológicas em regiões hidrográficas brasileiras.

Uma das limitações do teste de causalidade de Granger clássico é a premissa de linearidade. Para superar essa limitação, abordagens distintas têm surgido, e em especial a entropia de transferência, aplicável a sistemas lineares ou não (BARNETT; BARRETT; SETH, 2009).

#### **2.4.4 Entropia, informação mútua e entropia de transferência**

Teoria da informação é uma área do conhecimento que proporcionou grandes avanços na teoria da comunicação, na estatística e na ciência da computação (COVER; THOMAS, 2006). Duas medidas baseadas nessa teoria foram fundamentais para isso: a entropia e a informação mútua.

Entropia está associada a uma medida de incerteza de uma variável aleatória. Matematicamente, a entropia de Shannon  $H(x)$  de uma variável aleatória discreta  $X$  é definida por

$$H(X) = - \sum_{x \in \mathcal{X}} p_X(x) \log_2 p_X(x) \quad (5)$$

em que  $p_X$  é a função massa de probabilidade de  $X$ , e  $\mathcal{X}$  é o espaço amostral.

A entropia é um funcional da distribuição de probabilidade de  $X$  e não depende dos valores tomados pela variável  $X$  (apenas das probabilidades). Ao estender essa definição para um par de variáveis aleatórias, tem-se a definição da entropia conjunta

$$H(X, Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \log_2 p_{XY}(x, y) \quad (6)$$

em que  $p_{XY}$  é a função massa de probabilidade conjunta de  $X$  e  $Y$ .

Para chegar na definição de informação mútua, duas novas definições são necessárias: a de entropia condicional e a de entropia relativa. Por definição, a entropia condicional  $H(Y|X)$  é dada por

$$H(Y|X) = \sum_{x \in \mathcal{X}} p_X(x) H(Y|X = x) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \log_2 p(y|x) \quad (7)$$

e a entropia relativa ou distância de Kullback-Leibler ( $D$ ) entre duas funções de distribuição de probabilidade  $p(x)$  e  $q(x)$  é definida por

$$D(p||q) = \sum_{x \in \mathcal{X}} p(x) \log_2 \frac{p(x)}{q(x)} \quad (8)$$

Conhecido esses conceitos, a informação mútua  $I(X;Y)$  é definida como a entropia relativa entre a distribuição conjunta e o produto das distribuições  $p(x)p(y)$ , isto é,

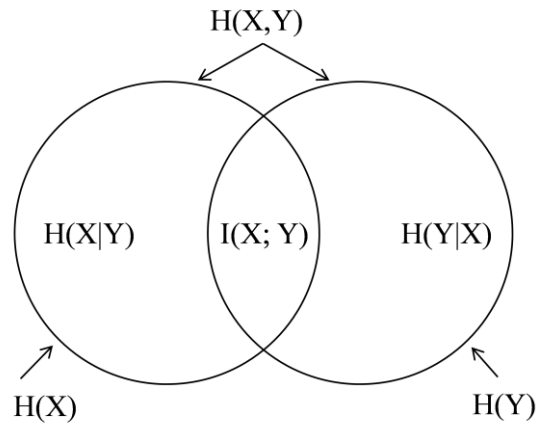
$$I(X; Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \log_2 \frac{p_{XY}(x, y)}{p_X(x)p_Y(y)} \quad (9)$$

Equivalentemente, a informação mútua pode ser definida a partir de entropia condicional como:

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X, Y) \quad (10)$$

A informação mútua  $I(X;Y)$  mede a dependência geral (linear e não linear) entre duas variáveis e pode ser vista como a redução de incerteza da variável  $X$  (ou  $Y$ ) devido ao conhecimento da variável  $Y$  (ou  $X$ ) (FRASER; SWINNEY, 1986). A representação gráfica pelo diagrama de Venn dessas definições, relacionando entropia e informação mútua, é indicada na Figura 1.

Figura 1. Relação entre entropia e informação mútua.



Fonte: Cover e Thomas (2006).

Ao adotar alguma defasagem temporal entre as variáveis em análise, a informação mútua pode ser utilizada para detectar a direção e intensidade de interação entre processos que exibem comportamento linear ou não linear (LI et al., 2018).

$$\text{TDMI}(X; Y, \tau) = - \sum_{x_t} \sum_{y_t} p_{XY}(x_t, y_{t+\tau}) \log_2 \frac{p_{XY}(x_t, y_{t+\tau})}{p_X(x_t)p_Y(y_{t+\tau})} \quad (11)$$

em que TDMI é a informação mútua com defasagem no tempo (não simétrica, i.e., depende de qual variável está sujeita à defasagem), e  $\tau$  é a defasagem temporal entre a variável de causa ( $X$ ) e efeito ( $Y$ ). A identificação da defasagem entre processos de um sistema com dados observados pode ser feita ao construir uma curva relacionando TDMI para diferentes defasagens – o primeiro pico de informação mútua corresponderá à defasagem.

Diferente do teste de Granger e de métodos baseados em correlação, a informação mútua não requer a condição de linearidade do sistema ou qualquer conhecimento prévio sobre o comportamento das variáveis para estudar interações causais. A limitação é a necessidade de



dados representativos e suficientes para a construção das funções de distribuição de probabilidade (LI et al., 2018).

Apesar do potencial da informação mútua, dependências estáticas devido ao histórico comum das variáveis não são ignoradas em sua definição. Para superar isso, Schreiber (2000) propôs uma alternativa: a entropia de transferência (TE). Esta métrica mede a quantidade de informação direcional transferida entre processos. Similarmente à TDMI, a TE não apresenta a propriedade de simetria (transferência de X para Y não têm o mesmo significado de transferência de Y para X). Essa característica confere à TE uma íntima relação com medidas de causalidade (AMBLARD; MICHEL, 2013). Foi demonstrado, por exemplo, que a TE é proporcional à razão de verossimilhança do teste de causalidade de Granger (LR) para variáveis aleatórias gaussianas (BARNETT; BARRETT; SETH, 2009).

Do ponto de vista de aplicação, a estimativa da TE é ainda um problema desafiador e uma ativa área de pesquisa, em função de questões numéricas, alta dimensionalidade e dependência por estimativas acuradas das distribuições de probabilidade (SCHREIBER, 2000; GENÇAÇA, 2018). Em estudos na área de Hidrologia, o uso de métricas baseadas na teoria da informação, e principalmente envolvendo a TE, é ainda pouco explorado. Bennett et al. (2019), em uma das poucas aplicações da TE em estudos hidrológicos publicada, caracterizaram e compararam a dinâmica dos processos observados e modelados. Os autores sugerem que o método proposto, baseado na TE, possa ser utilizado como ferramenta para guiar a seleção e o desenvolvimento de modelos usando a conectividade entre processos como critério de ajuste.

#### 2.4.5 Análise de multirresolução

A transformada *wavelet* (WT) é uma transformação linear que permite uma completa representação, nos domínios do tempo e da frequência, de padrões estacionários ou não de uma série temporal (LABAT, 2010). Na forma de maior interesse prático, a transformada *wavelet* discreta (DWT) permite escrever uma série  $f(t)$  da seguinte forma:

$$f(t) = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(t), \text{ com } \psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k) \text{ e } d_{j,k} = \int_{-\infty}^{\infty} f(t) \psi_{j,k}(t) dt \quad (12)$$

em que:  $\psi$  representa uma família de funções *wavelet*, denominada *wavelet*-mãe, sujeita a algumas condições matemáticas; e  $\psi_{j,k}$  são obtidas de  $\psi$  a partir de dilatações e translações (MORETTIN, 2014). A possibilidade de escrever uma série nesse formato permite obter

aproximações sequenciais, numa abordagem denominada multirresolução, onde operadores filtram termos de alta e baixa frequência sucessivamente (AKANSU; HADDAD, 2001).

A ideia da análise de multirresolução é obter uma sequência crescente de subespaços fechados  $V_j$  ( $V_j \subset V_{j+1}, j \in \mathbb{Z}$ ) constituídos por funções aproximadoras, de modo que uma função  $f$  qualquer de quadrados integráveis ( $f \in L^2(\mathbb{R})$ ) possa ser aproximada pela sua projeção ortogonal sobre cada  $V_j$ . Assim, à medida que  $j$  aumenta (aumento da resolução), mais detalhes são adicionados às aproximações de  $f$  (MORETTIN, 2014). O conjunto de funções aproximadoras é formado pelas funções  $\psi_{j,k}$  ( $j \in \mathbb{Z}, k \in \mathbb{Z}$ ), que formam uma base de  $L^2(\mathbb{R})$ .

Na área de hidrologia, transformadas *wavelet* têm sido utilizadas em problemas de detecção de anomalias, suavização de séries e modelagem de sistemas (KIM et al., 2017). Rathinasamy et al. (2014) demonstram a utilidade da abordagem multirresolução para avaliar, calibrar e melhorar os modelos hidrológicos, quantificando a performance desses modelos em diferentes escalas do tempo.

Entretanto, a aplicação de técnicas de suavização de séries temporais em estudos de causalidade ou fluxo de informação entre processos baseados em esquemas autorregressivos, como o teste de Granger, pode induzir a resultados incorretos devido à introdução de componentes de médias móveis (BARNETT; SETH, 2015). Isso requer um controle dos efeitos de operações de filtragem ou agregação temporal, que são tipicamente utilizadas na prática para reduzir ruídos de dados observados ou para trabalhar em uma escala de tempo com menores incertezas.

## 2.5 Síntese da revisão bibliográfica

Entre os direcionamentos resultantes da revisão da literatura, destacam-se:

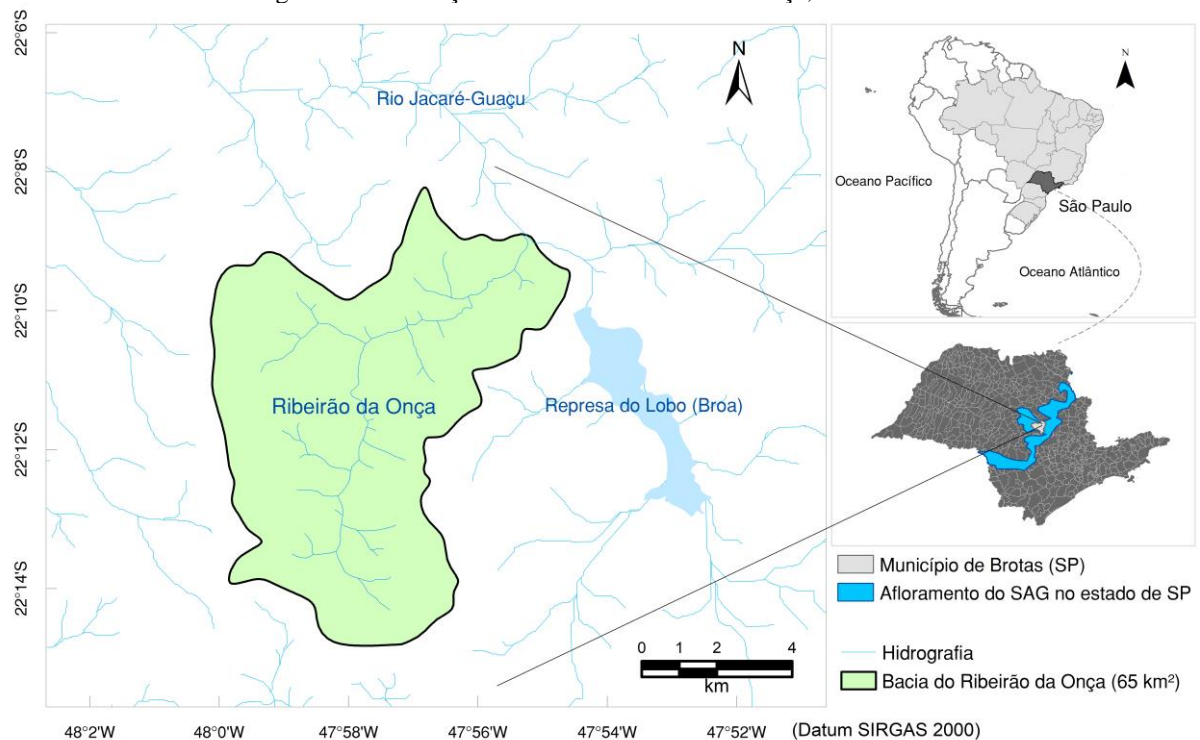
1. A promoção de um uso adequado e abrangente de dados disponíveis, incluindo os obtidos por sensoriamento remoto, na quantificação dos processos hidrológicos e no processo de modelagem para reduzir as incertezas dos dados de entrada, da estrutura e dos parâmetros dos modelos.
2. A utilização do teste de Granger e de métricas estatísticas (informação mútua com defasagem ou a entropia de transferência) como possíveis ferramentas para avaliar a conectividade e o fluxo de informação entre processos.
3. A recomendação de avaliar os efeitos de filtragem e agregação temporal dos dados nas análises de causalidade e conectividade entre processos.

## 3 MATERIAL E MÉTODOS

### 3.1 Área de estudo

A área de estudo é a bacia hidrográfica do Ribeirão da Onça (65 km<sup>2</sup>), afluente da margem esquerda do rio Jacaré-Guaçu, localizada em área de uso predominantemente agrícola na região central do estado de São Paulo (município de Brotas) (Figura 1). Uma característica marcante da bacia é sua representatividade em termos hidrogeológicos por estar localizada em área de afloramento do Sistema Aquífero Guarani (SAG).

Figura 2. Localização da bacia do Ribeirão da Onça, Brotas – SP.



Fonte: Autor.

O SAG é composto por arenitos de origem eólica do período Jurássico (Formação Botucatu) e de origem fluvio-eólica do período Triássico (Formação Piramboia), e em grande parte encontra-se confinado por derramamentos basálticos ocorridos no período Cretáceo (Formação Serra Geral). Ele atende a demanda por água de mais de 90 milhões de pessoas da Argentina, Brasil, Paraguai e Uruguai (KIRCHHEIM et al., 2019), e é uma das mais importantes reservas de água subterrânea da América do Sul. A ocorrência do processo de recarga e a correspondente vulnerabilidade à contaminação antrópica (e.g., por nitrato,

nutrientes, pesticidas) motiva a condução de estudos experimentais em áreas de afloramento do SAG a fim de desenvolver estratégias para proteger e garantir um uso sustentável desse sistema aquífero (OEA, 2009; SINDICO; HIRATA; MANGANELLI, 2018).

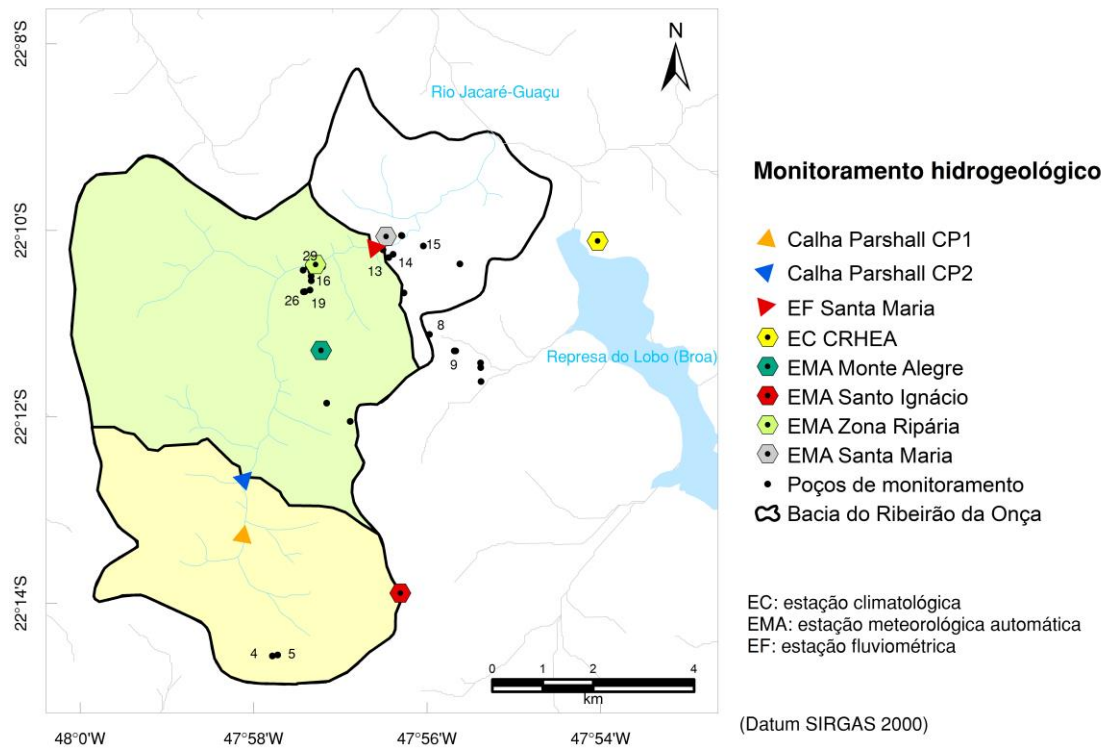
Arenitos resultantes de processos de intemperismo da Formação Botucatu cobrem considerável área da bacia e constituem o material do leito do Ribeirão da Onça e de seus afluentes (CASTRO JÚNIOR; WENDLAND, 2015). Neossolo quartzarênico órtico e latossolo vermelho amarelo alumínico são as classes de solo predominantes na área de estudo (ROSSI, 2017). A bacia apresenta uma declividade média superficial de 0,008 m/m, com um relevo plano a suave ondulado. As elevações estão entre 655 e 825 m acima do nível do mar (a.s.l.).

O semestre chuvoso apresenta cerca de 65 % da precipitação média anual, que é de 1485 mm (CABRERA et al., 2016a). A classificação climática de Köppen-Geiger (ALVARES et al., 2013) indica que a área apresenta clima subtropical úmido com invernos secos (*Cwa*).

Quanto ao uso e ocupação, as atividades agropecuárias e de silvicultura predominam na bacia do Ribeirão da Onça há pelo menos três décadas. Em 1990, os cultivos de eucalipto ocupavam cerca de 30% da área total, pastagens ocupavam 15% e a vegetação nativa (cerrado), 30% (POMPEO, 1990). Baseado na classificação de uso do solo feita por Alcântara (2018), em 2017, os cultivos de eucalipto, citros, cana-de-açúcar e as atividades de pastagem totalizavam aproximadamente 35%, 30%, 10% e 4%, respectivamente, da área total da bacia, com a vegetação nativa remanescente correspondendo a 10% (Figura 9). Solo exposto ou com vegetação rasteira de cerrado perfazem o restante da área da bacia. Periodicamente, cultivos de ciclo rápido, como soja e amendoim, mudam um pouco essas distribuições, porém não de forma significativa. Rotações ou substituições entre cultivos de cana-de-açúcar e citros ou entre áreas de pastagem e plantio de eucalipto são outros tipos de alterações presentes na bacia, porém limitadas a pequenas áreas.

A instrumentação em operação, responsável pelo monitoramento meteorológico e hidrogeológico da bacia está indicada na Figura 3. As vazões na estação fluviométrica Santa Maria e os níveis d'água de alguns poços são monitorados de maneira combinada (manual e automática), com diferentes intervalos de amostragem e instrumentos, desde 2004. Estações meteorológicas automáticas (EMA), em diferentes períodos de operação, têm contribuído desde 2010 com o monitoramento da bacia (Figura 4).

Figura 3. Instrumentação em operação (2019) na bacia hidrográfica do Ribeirão da Onça. Poços não identificados não foram utilizados neste trabalho.



Fonte: Autor.

De forma conjunta, a estação climatológica do CRHEA (1980 - presente), algumas estações pluviométricas da ANA e os trabalhos de monitoramento da bacia do Ribeirão da Onça, feitos com apoio institucional da Escola de Engenharia de São Carlos, foram as principais fontes dos dados utilizados no desenvolvimento desta pesquisa.

Figura 4. Pontos de monitoramento de dados meteorológicos na bacia em 2019.  
(a) EMA Monte Alegre (b) EMA Santo Ignácio (c) EMA zona ripária



Fonte: Autor.

### 3.2 Dados

As fontes de dados utilizados e os métodos associados à quantificação de processos hidrológicos relacionados são apresentadas nesta seção. Estão incluídas as quantificações de: a) precipitação média (sobre as áreas de interesse); b) evapotranspiração de referência média; c) evapotranspiração real média a partir da análise de índices de vegetação obtidos por sensoriamento remoto; d) vazão e e) variações de nível freático em poços de monitoramento.

#### 3.2.1 Dados de precipitação

Parte das estações pluviométricas identificadas na área de estudo, indicadas na Figura 5, foram utilizadas para determinar a precipitação média na bacia. As estações com maior período de operação na região são a do CHREA, Campo Alegre (código ANA 2247185) e Ferroviária (código ANA 2247197), que informam dados diários (7:00 h às 7:00 h) desde a década de 1980 (Figura 6). Dentro da bacia, a estação automática com maior período de operação é a Monte Alegre, que informa dados horários desde agosto de 2011.

Devido à variabilidade espacial identificada, especialmente nas escalas horária e diária, o processo de espacialização indicado na sub-seção seguinte ('Espacialização de dados de precipitação e de evapotranspiração') foi considerado para estimar valores médios nas áreas de drenagem das seções de monitoramento fluviométrico.

Figura 5. Estações pluviométricas identificadas próximas à área de estudo.

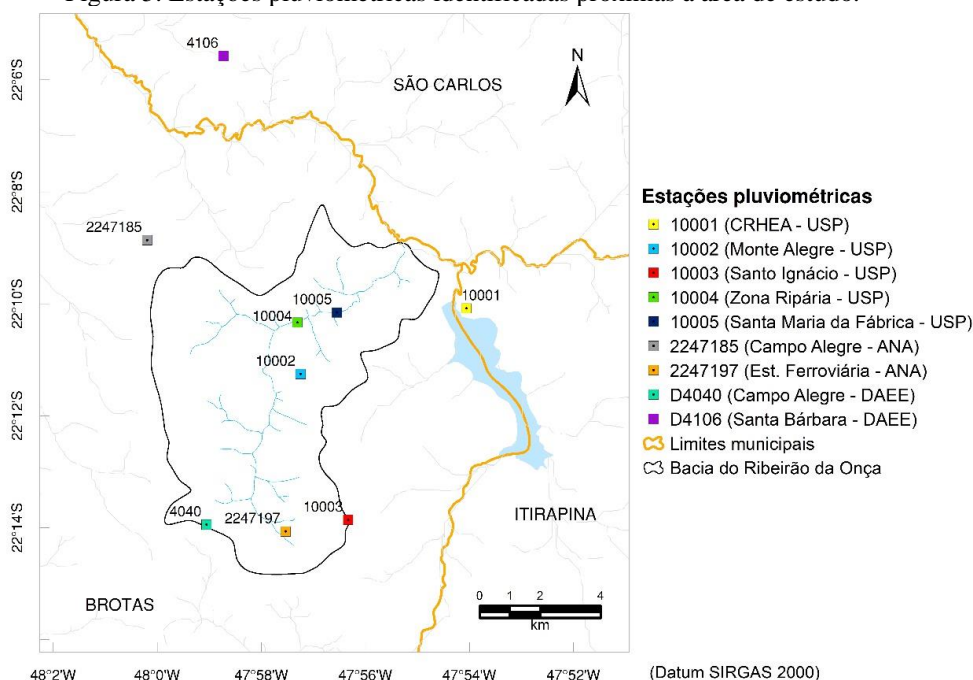
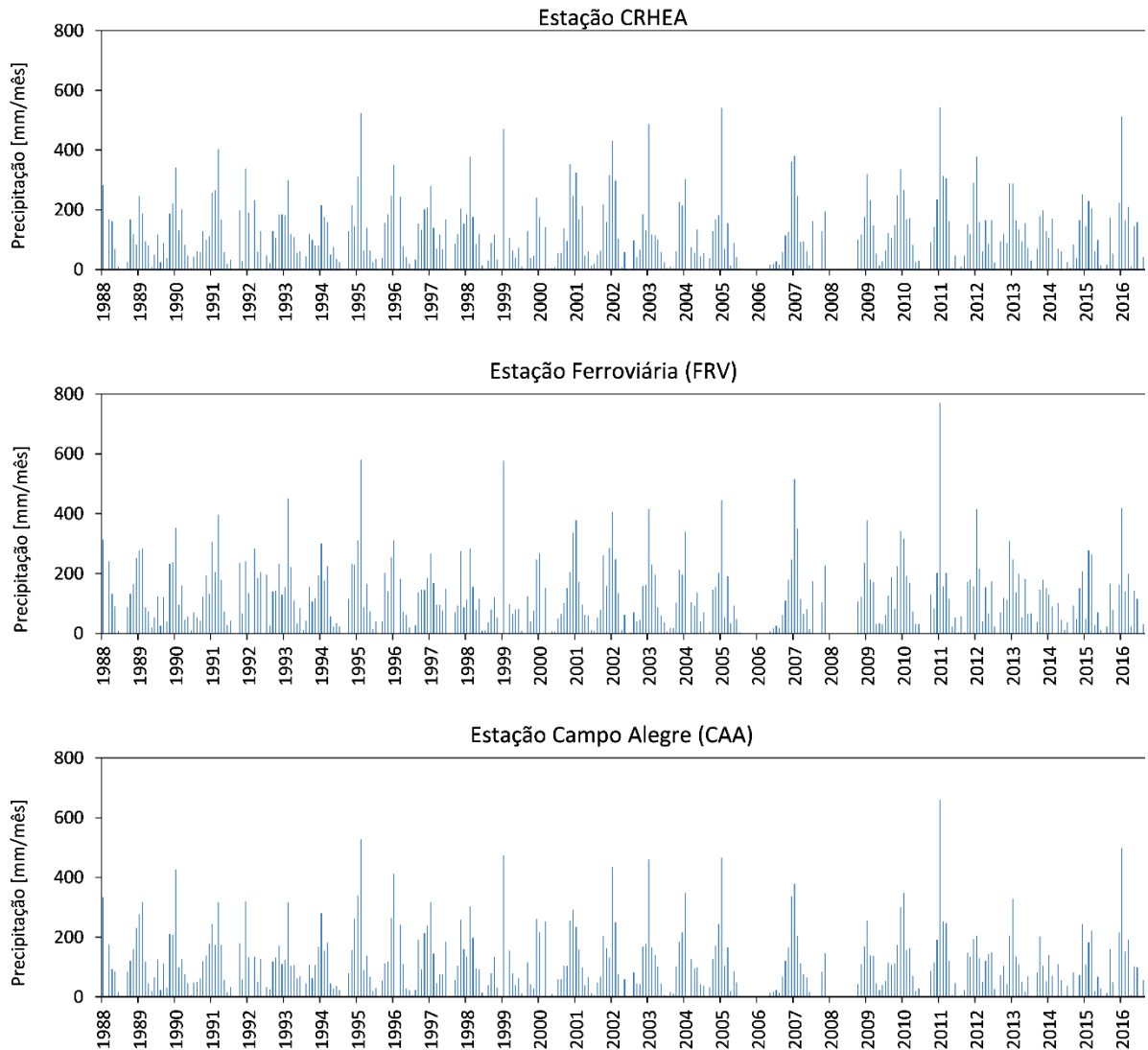


Figura 6. Dados brutos de precipitação mensal de algumas estações próximas à área de estudo. Meses em que uma ou mais estações apresentaram falhas foram deixados vazios.



Fonte: Autor.

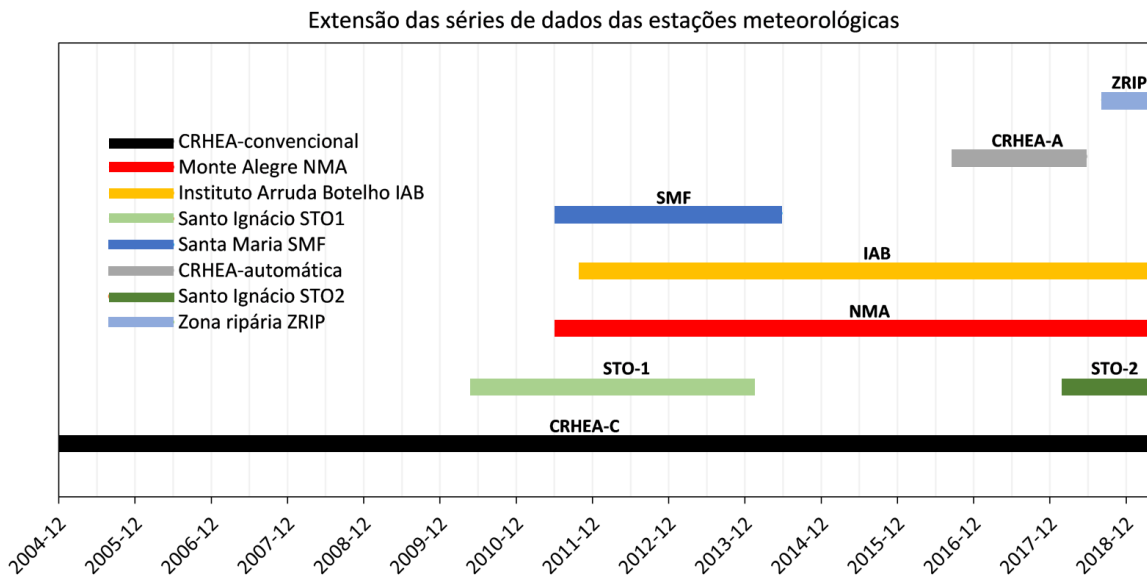
### 3.2.2 Dados meteorológicos e evapotranspiração de referência

Os dados meteorológicos disponíveis na área de estudo passaram inicialmente por uma análise integrada de consistência e de preenchimento. Além das estações meteorológicas inicialmente definidas para conduzir este trabalho (no plano inicial), uma estação automática localizada no Instituto Arruda Botelho - IAB (10 km ao leste da bacia de estudo) foi utilizada como suporte no controle de qualidade dos dados, tendo seu uso justificado pela proximidade e pela disponibilidade de dados no período de interesse. Além disso, para o período em que estão disponíveis apenas dados da estação convencional do CRHEA (Figura 7), a estação meteorológica do INMET-São Carlos (código 83726), localizada 27 km ao norte da área de



estudo, também foi utilizada, porém exclusivamente para fins de preenchimento de falhas (de temperatura, umidade relativa do ar e velocidade do vento).

Figura 7. Extensões das séries de dados meteorológicos utilizadas neste estudo.



Fonte: Autor.

Uma estratégia baseada em regressões lineares entre valores registrados em diferentes estações nos mesmos instantes de tempo (discretização horária para estações automáticas e diária para convencionais) foi adotada, facilitando a identificação de medições fora do padrão de variabilidade. Os resultados desta etapa são apresentados no Apêndice A.

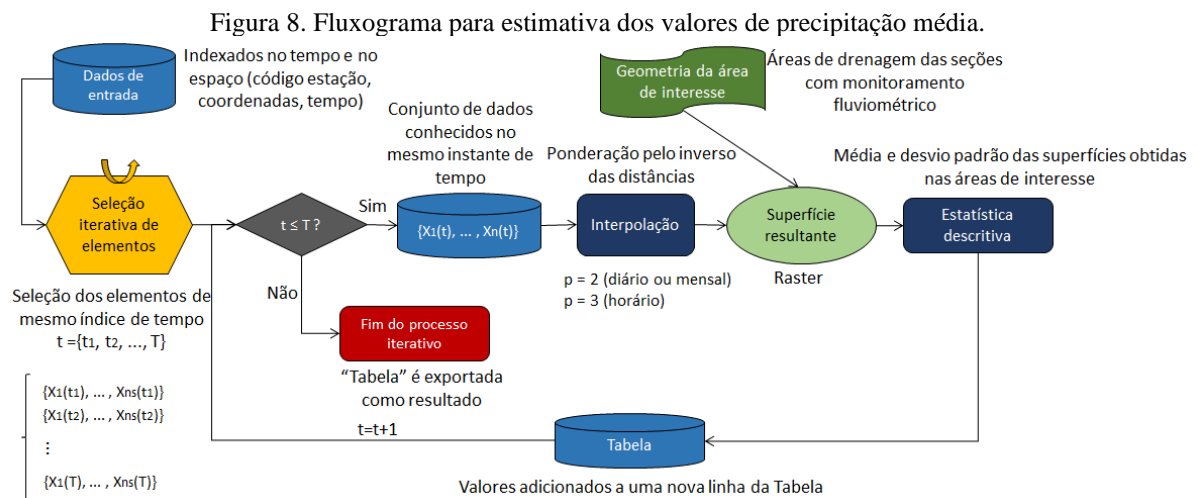
Após o controle de qualidade dos dados meteorológicos, os procedimentos de cálculo padronizados apresentados por Allen et al. (1998) foram seguidos para determinação da evapotranspiração de referência com passo de cálculo horário (estações automáticas) e diário (estação convencional).

### 3.2.3 Espacialização de dados de precipitação e de evapotranspiração

O fluxograma da Figura 8 apresenta o modelo construído em ambiente SIG (ArcGIS®) para automatização do processo de estimativa da precipitação média e de evapotranspiração de referência nas áreas de interesse. No processo inicial, de análise das séries temporais, os recortes espaciais para determinação da precipitação foram as áreas de drenagem correspondentes às seções de monitoramento fluviométrico, indicadas na Figura 3. Para cada discretização de tempo (horária, diária, mensal), um conjunto de estações foi escolhido para o processo de



interpolação espacial, usando o método determinístico de ponderação pelo inverso das distâncias (*inverse distance weighting* – IDW). Seguindo as recomendações de Dirks et al. (1998), a potência  $p$  foi adotada igual a 2 para interpolação de dados mensais e diários, e igual a 3 para interpolar dados horários.



Fonte: Autor.

Para os propósitos deste trabalho, os valores médios de precipitação nas áreas de contribuição são mais relevantes do que sua distribuição propriamente dita (modelos concentrados foram utilizados). Dessa forma, uma vez selecionado um conjunto de estações, as séries de precipitações ou de evapotranspiração de referência médias foram estimadas.

Pelo método IDW, os coeficientes de cada estação para cada posição são calculados como

$$\lambda_{i,X_k} = d_{i,k}^{-p} \left[ \sum_{i=1}^n d_{i,k}^{-p} \right]^{-1}, \text{ com } d_{i,k} = \|X_k - E_i\| \quad (13)$$

em que  $X_k \in \mathbb{R}^2$  é a posição do centro de cada elemento de área  $A_k$  do domínio de discretização da interpolação [células (pixels) de 50 m x 50 m],  $E_i \in \mathbb{R}^2$  é a posição de cada estação meteorológica ou pluviométrica, e  $\| \cdot \|$  é a distância entre a estação e o elemento considerado.

Conhecidos os valores observados em cada estação em diferentes instantes de tempo ( $O_{E_i,t}$ )

$$Z_t(X_k) = \sum_{i=1}^{N_{est}} \lambda_{i,X_k} \cdot O_{E_i,t} \quad (14)$$

em que  $Z_t(X_k)$  é o valor da variável de interesse na posição  $X_k$  obtido pelo método IDW no instante de tempo (discreto)  $t$ , e  $N_{est}$  é o número de estações utilizadas na interpolação.

Para obter valores médios em recortes geográficos de interesse, a seguinte operação foi conduzida

$$\bar{Z}_{B,t} = \frac{1}{A_B} \sum_{A_k \in B} \sum_{i=1}^{N_{est}} \lambda_{i,X_k} O_{E_i,t} A_k, \quad A_B = \sum_{A_k \in B} A_k \quad (15)$$

em que  $\bar{Z}_{B,t}$  é o valor médio da variável de interesse em uma determinada área ( $A_B$ ) no instante  $t$ .

A partir da identificação de períodos em que há o maior número possível de observações simultâneas (mesmo instantes  $t$ ) por diferentes estações, regressões lineares foram estabelecidas da seguinte forma

$$\bar{Z}_{B,t} = a_0 + \sum_{i=1}^{N_{est}} a_i O_{E_i,t} + \varepsilon_t, \quad \varepsilon_t \sim N(\mu = 0, \sigma^2 \cong 0) \quad (16)$$

em que  $a_k$  são coeficientes de regressão linear múltipla e  $\varepsilon_t$  é um erro aleatório com média e variância desprezíveis. Essa situação (de erro nulo) ocorre quando todas as estações utilizadas na interpolação estão contribuindo com alguma informação. Horas, dias ou meses em que não há precipitação ou em que ocorre falha de monitoramento em alguma delas não foram considerados nas análises.

A regressão indicada pela Equação (16) tenta reproduzir o padrão de valores médios na área da bacia a partir das observações existentes e da interpolação espacial conduzida. Em instantes em que um número de estações menor ou igual a  $N_{est}$  registrou dados (situação mais comum), aproximações da seguinte forma foram consideradas

$$\hat{Z}_{B,t} = a_{0j} + \sum_{E_i \in D_j} a_{ij} O_{E_i,t} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_j^2) \quad (17)$$

em que  $\hat{Z}_{B,t}$  é o valor estimado de  $\bar{Z}_{B,t}$ ,  $D_j$  é um conjunto de estações com dados simultaneamente disponíveis no instante  $t$ , cuja análise de regressão resulta nos coeficientes  $a_{k,j}$  e num erro padrão com variância  $\sigma_j^2$ . Como mais de uma combinação de estações pode ser feita em um mesmo instante de tempo, o critério para seleção foi baseado no mínimo  $\sigma_j^2$ .

Alternativamente ao método de regressão aqui seguido, os valores médios poderiam ser obtidos diretamente a partir da interpolação espacial dos dados observados. Por exemplo, em instantes com um único ponto de monitoramento, o valor médio da área é igual ao (único) valor observado. No método proposto, o valor médio resultaria de uma relação linear estabelecida a partir dos dados historicamente observados. As vantagens e desvantagens de cada procedimento, assim como de outros métodos de interpolação (e.g., vizinho natural, métodos geoestatísticos) demandariam um estudo independente e fogem do escopo desta pesquisa.

### 3.2.4 Índices espectrais de vegetação e estimativa de evapotranspiração real

A dinâmica da vegetação é uma variável explicativa relevante na descrição do comportamento hidrológico de bacias, especialmente em pequenas escalas espaciais e temporais ( $A < 500 \text{ km}^2$  e  $1 - 5$  anos) (DONOHUE et al., 2007; WEGEHENKEL, 2009). O monitoramento dessa dinâmica pode ser feito a partir de índices espectrais, obtidos por sensoriamento remoto, que possibilitam detectar mudanças temporais e espaciais em grandes escalas, como é o caso de estudos em bacias hidrográficas. O EVI (*enhanced vegetation index*), por exemplo, é definido da seguinte forma:

$$EVI = 2,5 \frac{\rho_{IVP} - \rho_{ver}}{1 + \rho_{IVP} + 6 \rho_{ver} - 7,5 \rho_{azul}} \quad (18)$$

em que  $\rho_{IVP}$  é o valor de refletância para o infravermelho próximo,  $\rho_{ver}$ , para o vermelho, e  $\rho_{azul}$ , para o azul. Comparado a outros índices espectrais de vegetação existentes (e.g., NDVI), o EVI é menos sensível a alterações de refletância do solo devido a variações de umidade (GLENN et al., 2011) e apresenta bons resultados em trabalhos relacionados à estimativa de evapotranspiração (GUERSCHMAN et al., 2009; NAGLER et al., 2013). Com base nessas características e nos objetivos deste estudo, o EVI foi selecionado para ser dado de entrada no modelo hidrológico e auxiliar na quantificação da evapotranspiração real na bacia.

O produto MOD13Q1 (DIDAN, 2015) gerado a cada 16 dias em resolução espacial de 250 m e disponibilizado no NASA *Land Processes Distributed Active Archive Center* (LP DAAC<sup>1</sup>) foi utilizado. Os dados de refletância são obtidos pelo sensor *Moderate Resolution Imaging Spectroradiometer* (MODIS) a bordo do satélite Terra.

Nagler et al. (2013) desenvolveram um método empírico relacionando a evapotranspiração real ( $ET_a$ ), medida com torres de fluxo (método Eddy-Covariance), com o EVI e com a evapotranspiração de referência ( $ET_o$ ) em diferentes usos do solo no estado do Arizona (EUA). O método segue a equação (19), que possui embasamento na lei de Beer-Lambert para representar a absorção de luz pela vegetação, tendo o EVI como indicador da densidade de absorção.

$$ET_a = K_{c,VI} ET_o = [a(1 - e^{-b \cdot EVI}) - c] ET_o \quad (19)$$

em que  $ET_a$  é a evapotranspiração real estimada pelo método,  $ET_o$  é a evapotranspiração de referência, e a, b e c são parâmetros de ajuste. Nagler et al. (2013) adotaram os valores de 1,65, 2,25 e 0,169 para esses parâmetros respectivamente. Enquanto o parâmetro c produz translações na curva de transformação, os parâmetros a e b determinam a forma de como variações em EVI provoca alterações no coeficiente  $K_{c,VI}$ .

Como na área de estudo não existe um histórico de medições de evapotranspiração real, uma abordagem alternativa foi proposta para estimar a evapotranspiração real a partir do EVI. Essa abordagem pode ser resumida em um processo de três etapas: 1) identificação de pontos com cobertura vegetal conhecida; 2) extração da série temporal de EVI desses pontos (indicados na Figura 9) e 3) comparação/utilização de coeficientes de cultura reportados da literatura.

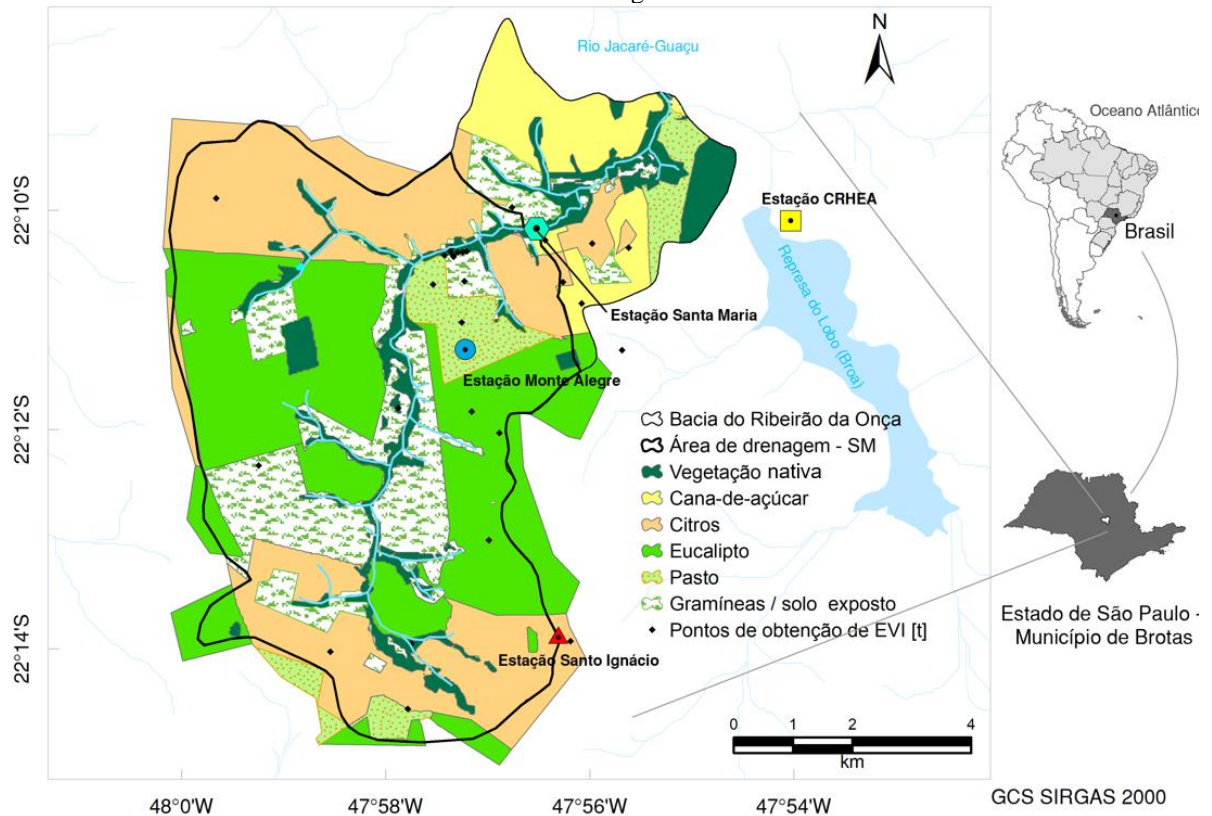
As culturas foram identificadas com base no monitoramento contínuo da área, desenvolvido por meio de visitas quinzenais ao local (e.g., Figura 10) e pela análise de imagens de satélite. A extração da série temporal foi feita em ambiente SIG a partir dos dados obtidos no *NASA's Land Processes Distributed Active Archive Centers* (produto em formato 'geotiff'). As séries obtidas foram comparadas (em termos de valores médios, máximos anuais, mínimos anuais, em ciclos completos e em períodos úmidos) com as relações  $ET_a/ET_o$  reportadas por Allen et al. (1998) – para áreas de citros e cana-de-açúcar, Cabral et al. (2010) – eucalipto, e Anache et al. (2019) – pasto. Os pixels que apresentaram baixa qualidade ou utilidade (cerca

---

<sup>1</sup> NASA LP DAAC: dados obtidos em <https://lpdaac.usgs.gov/>. Acesso: 28/03/2019

de 5 % do total), seguindo os critérios apresentados em Didan (2015), foram removidos da série e substituídos pela média dos valores vizinhos.

Figura 9. Mapa de uso do solo da área de estudo no ano de 2017 (classificação de Alcântara (2018)) e pontos amostrados para definição de séries temporais do índice de vegetação EVI correspondentes a diferentes coberturas vegetais.



Fonte: Autor.

A equação proposta por Nagler et al. (2013) foi ligeiramente adaptada e testada com os pontos definidos pela abordagem anteriormente descrita no seguinte formato (permitindo simples translação):

$$K_{VI} = \max([1,65(1 - \exp(-2,25EVI)) - c]; k_{\min}), \quad (20)$$

em que  $c$  é um parâmetro a ser ajustado para a área de estudo, e  $k_{\min}$  é um valor de modo a evitar que  $K_{VI}$  tome valores negativos (foi adotado  $k_{\min} = 0,05$ ).



Figura 10. Principais usos do solo e coberturas vegetais na bacia do Ribeirão da Onça (2018/2019).



(a) Eucalipto



(b) Citros



(c) Área de pasto e cerrado residual

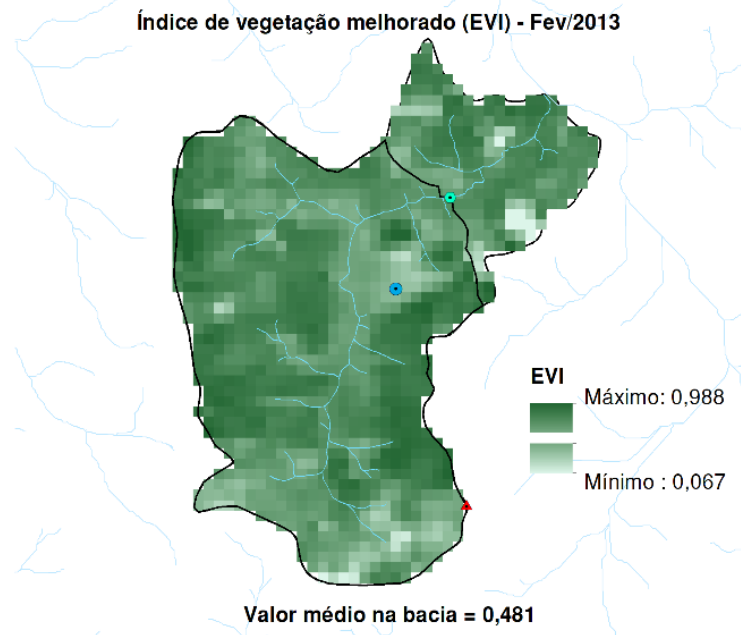


(d) Cana-de-açúcar (à esquerda)

Fonte: Autor.

A distribuição do EVI na área de estudo para o mês de fevereiro (período úmido) de 2013 é apresentado na Figura 11. Processos relacionados à zona ripária podem não ser bem representados em função da resolução espacial (250 m) do produto utilizado (MOD13Q1), que é muito maior que a eventual faixa de interesse.

Figura 11. Exemplo de distribuição espacial de EVI na área de estudo.



Fonte: Autor.

### 3.2.5 Monitoramento fluviométrico

Os dados fluviométricos foram obtidos com os instrumentos indicados na Tabela 1 em duas seções de monitoramento do curso de água principal da bacia (Ribeirão da Onça) (localizações apresentadas na Figura 2). O monitoramento de vazão  $[Q_t]$  ocorre em grande parte do tempo de forma indireta, a partir do monitoramento do nível d'água  $[h_t]$ . As sub-seções seguintes apresentam os métodos empregados na transformação desta variável em estimativas de vazão.

Tabela 1. Caracterização dos dados de monitoramento fluviométrico utilizados.

Seção de monitoramento	Área de drenagem <sup>2</sup> (km <sup>2</sup> )	Instrumento de medição	Período	Frequência
[SM] Est. Fluviométrica Santa Maria (47,941°W; 22,171°S)	52,1	Molinete hidrométrico [Vazão]	12/2008 a 02/2019	Quinzenal
		Régua e medidor de nível [nível de água - NA]		
		Transdutor de pressão (Diver®) [NA]	07/2015 a 05/2018	Sub-horária
		ADCP (SonTek-IQ Plus®) [Vazão]		
[CP2] Calha Parshall 2 W18'' (47,969°W; 22,212°S)	19,2	Régua [NA]	04/2017 a 02/2019	Quinzenal
		Transdutor de pressão (Diver®) (Diver®) [NA]	04/2017 a 02/2019	Sub-horária

Fonte: Autor.

### 3.2.5.1 Estação fluviométrica Santa Maria (SM)

A estação de monitoramento fluviométrico Santa Maria (Figura 12) está localizada imediatamente a montante de um bueiro duplo tubular de concreto de 1,50 m de diâmetro. A jusante do bueiro, há uma estação de captação e bombeamento de água superficial, utilizada para fins de irrigação por proprietários locais, que periodicamente promovem o barramento (represamento) do curso d'água. Este processo provoca consideráveis perturbações nas relações entre nível d'água e vazão (Figura 13). Além disso, há contínuas mudanças na geometria e na resistência do leito arenoso do curso d'água, que ocorrem naturalmente em função de processos fluviais de transporte de sedimentos e do desenvolvimento sazonal de vegetação (Figura 14). A multiplicidade de condições geométricas e de contorno que controlam as relações entre vazão e nível d'água demandou uma estratégia de transformação variável no tempo.

<sup>2</sup> Áreas de drenagem determinadas com base no modelo digital de elevação (MDE) disponibilizado pela Secretaria do Meio Ambiente do Estado de São Paulo em escala 1:50.000 (disponível em <https://www.ambiente.sp.gov.br/cpla/modelo-digital-de-elevacao-mde-do-estado-de-sao-paulo/>, acesso: 16/08/2018).

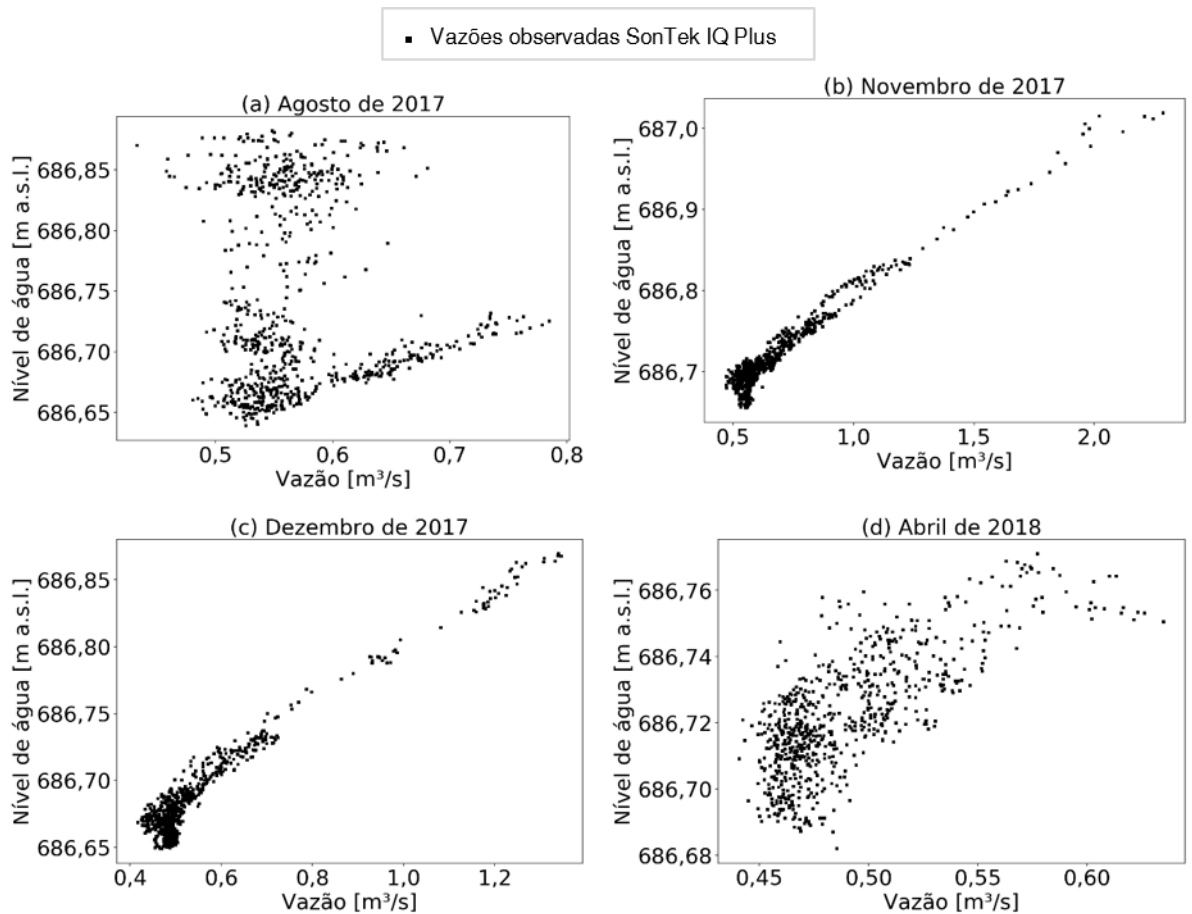


Figura 12. Seção de medição de vazão e estação de bombeamento de água.



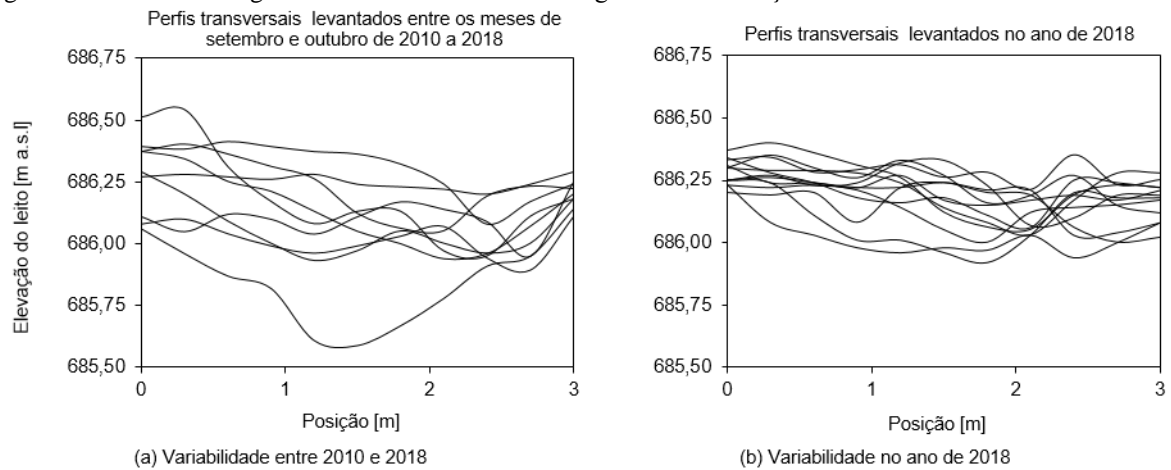
(a) Seção onde nível e vazão são medidos (montante). (b) Local onde o curso d'água é represado (jusante).  
Fonte: Autor.

Figura 13. Perturbações nas relações entre nível d'água e vazão provocadas pelo barramento a jusante da seção de medição de vazão. Gráficos indicam períodos em que relações bem definidas entre nível e vazão existem (entre novembro e dezembro de 2017) e períodos perturbados por efeito de barramento do curso d'água (agosto de 2017 e abril de 2018).



Fonte: Autor.

Figura 14. Variabilidade geométrica do leito do curso d'água onde a medição de vazão é conduzida.



Fonte: Autor.

#### 4.2.5.2 Estação fluviométrica calha Parshall CP2

Na área de montante da bacia, calhas Parshall de dimensão característica  $W = 18''$ , com capacidade de medição de 700 L/s, estão instaladas no curso d'água principal. Dados de nível de água a cada 10 minutos são obtidos a partir de dados registrados por transdutores de pressão (Diver<sup>®</sup>, Van Essen Instruments) instalados em poços de monitoramento hidráulicamente conectados às calhas (Figura 15). O monitoramento foi iniciado em 2017 e dados de qualidade, com poucas falhas, estão presentes desde janeiro de 2018. A estimativa de vazão é feita de forma direta a partir das cargas de montante e da equação empírica de transformação da calha Parshall

$$H_{a,t} = \frac{[p_{w,t} - p_{atm,t}]}{\gamma_w} + c + \varepsilon_t \quad c = \min_c \sum (H_{a,k}(c) - H_{obs,k})^2 \quad (21)$$

$$Q_t = 1,056 \times [H_{a,t}]^{1,538} \quad (22)$$

em que  $Q_t$  [m<sup>3</sup>/s] é a vazão estimada,  $p_{wt}$  [kN/m<sup>2</sup>] é a pressão total registrada pelo transdutor de pressão,  $p_{atm}$  [kN/m<sup>2</sup>] é a pressão atmosférica,  $\gamma_w$  é o peso específico da água [~9,81 kN/m<sup>3</sup>],  $H_{obs,k}$  é uma medição manual de carga hidráulica na seção de controle de montante da calha Parshall (feita a cada 15 dias em média),  $c$  é uma constante determinada por ajuste de mínimos quadrados para relacionar o nível d'água compensado com a carga hidráulica estimada pelos

sensores automáticos ( $H_{a,k}$ ),  $\varepsilon_t$  é um erro aleatório decorrente desse ajuste, e  $t$  é o índice de tempo das variáveis.

Figura 15. Calha Parshall 2.



(a) Montante.

(b) Jusante.

Fonte: Autor.

### 3.2.6 Transformação de nível d'água em vazão na seção fluviométrica Santa Maria

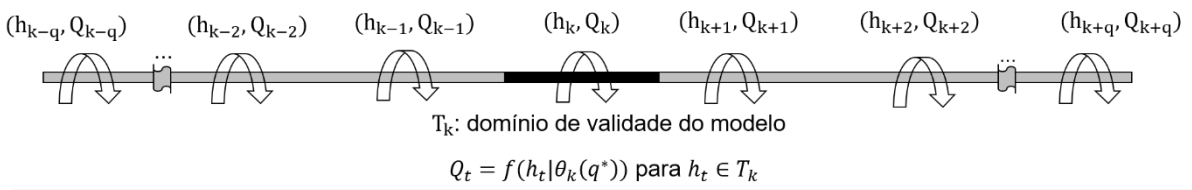
#### 3.2.6.1 Descrição geral do método

Para cada medição de vazão há uma medição correspondente de nível d'água, formando um par ordenado de dados observados no tempo  $(h_t, Q_t)$ . A estratégia de definição dinâmica da curva-chave foi baseada na segmentação da série dessas observações, tomando como referência apenas as medições mais próximas do instante de tempo para o qual há interesse de se estimar a vazão. Um resumo do método é apresentado na Figura 16 e na sequência os detalhes são apresentados.

Seja  $\Omega$  o conjunto de todos os pares ordenados observados de nível d'água e de vazão.

$$\Omega = \bigcup_{k=1}^N p_k, \quad p_k = (h_k, Q_k) = \{\{h_k\}, \{h_k, Q_k\}\} \quad (23)$$

Figura 16. Resumo do modelo de transformação de nível d'água em vazão proposto para a seção de monitoramento Santa Maria.

Espaços amostrais alternativos	Otimização de parâmetros	Função de verossimilhança
$q = 2, 3, \dots, 7$	$\theta_k   \min \varepsilon^T \varepsilon$	$\ell_s(A_k(q^*; \theta_k(q^*))) = \max \ell_j$
1. Variar a quantidade de dados observados utilizados na assimilação	2. Definir os parâmetros do modelo a partir da minimização da soma dos erros quadráticos	3. Selecionar a combinação que resulta na máxima função de verossimilhança
$A_k(q): \{p_{k-q}, \dots, p_{k-1}, p_k, p_{k+1}, \dots, p_{k+q}\}$	$\theta_k(q) = [a_k; b_k; \sigma_k^2   q]$	$\ell_M(A_k(q); \theta_k(q)   q)$
$p_j = (h_j, Q_j)$ : observações pareadas de nível d'água e de vazão [ordenadas no tempo]		
		

Fonte: Autor.

Assumindo uma correspondência do índice 'k' com o tempo, o domínio de validade ( $T_k$ ) de cada curva-chave (modelo de transformação),  $f_k(h_t; \theta_k)$ , a ser determinada, foi limitado ao intervalo centrado na medição  $p_k$  com limites inferior e superior correspondentes aos pontos médios entre as medições imediatamente anterior ( $p_{k-1}$ ) e posterior ( $p_{k+1}$ ) à medição em consideração:

$$\forall k \exists t_k | Q_k = Q(t_k) \text{ e } T_k = \left[ \left[ \frac{1}{2}(t_{k-1} + t_k) \right], \left[ \frac{1}{2}(t_k + t_{k+1}) \right] \right] \quad (24)$$

A equação potência, tradicionalmente recomendada na literatura hidrométrica, foi o modelo padrão de transformação adotado neste estudo:

$$f_k(h_t; \theta_k) = \exp(a_k)(h_t - h_o)^{b_k}, \quad t \in T_k \quad (25)$$

$$Q_t = f_k(h_t; \theta_k) + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_k^2), \quad t \in T_k \quad (26)$$

em que  $f_k$  é a vazão estimada pelo modelo de transformação de nível d'água em vazão,  $\theta_k = [a_k, b_k, \sigma_k^2]$  é o vetor de parâmetros do modelo,  $\sigma_k^2$  é a variância estimada dos resíduos do modelo,  $h_t$  é o nível médio d'água no intervalo considerado e  $h_o$  é um parâmetro da curva-chave, que foi admitido como a média das cotas de fundo dos bueiros tubulares que estão localizados 5 m a jusante da seção onde o nível d'água é medido.

A definição do vetor de parâmetros  $\theta_k$  de cada curva-chave  $f_k(h_t; \theta_k)$  depende da definição de um outro domínio, aqui denominado de domínio amostral ( $A_k$ ), definido da seguinte forma:

$$A_k(q) = \bigcup_{j=k-q}^{k+q} p_j \quad (27)$$

que são os pares observados de vazão e de nível d'água utilizados no método de estimativa do vetor de parâmetros  $\theta_k$ .

A determinação dinâmica da curva-chave foi escrita como um problema de otimização, tendo como variáveis de decisão os parâmetros do vetor  $\theta_k$  e o inteiro  $q$ , relacionado ao número de observações,  $n = 2q + 1$ , contidas no domínio amostral  $A_k$ . A abordagem clássica de determinação de parâmetros baseada no método dos mínimos quadrados foi adotada. A função de verossimilhança foi definida como critério de seleção entre as diferentes combinações possíveis de parâmetros e espaços amostrais. Sob a condição de normalidade, a função de log-verossimilhança ( $\ell$ ) de uma variável aleatória  $z \sim N(\mu, \sigma^2)$  assume a seguinte forma:

$$\ell = \sum_{j=1}^n \ln g(z; \mu, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=1}^n (z_j - \mu)^2 \quad (28)$$

em que  $g(z; \mu, \sigma^2)$  é a função de densidade de probabilidade de  $z$ .

A maximização da função de verossimilhança é usualmente empregada como critério para estimar parâmetros de modelos em condições de espaço amostral total e invariável. Assim, o método proposto é diferente da abordagem tradicional, e tem neste estudo uma oportunidade de ser avaliado empiricamente com dados reais observados em um problema de interesse prático na hidrometria. A hipótese é que no formato proposto o critério seja capaz de: 1) identificar alterações de propriedades que controlam o escoamento (e.g., barramento do curso d'água, geometria, resistência do leito); e 2) decidir pela combinação (de parâmetros e de espaço amostral) mais representativa de cada domínio de validade dos modelos de transformação de nível água em vazão.

Quanto aos métodos de otimização para definição do vetor  $\theta_k$ , duas abordagens foram adotadas: uma de regressão linear simples baseada nos métodos dos mínimos quadrados ordinários (OLS), com resíduos independentes e sem restrição dos parâmetros; e outra baseada

em um método de otimização global (algoritmo evolução diferencial) com restrição dos parâmetros. O método OLS foi escolhido pela simplicidade, generalidade de uso e pequeno tempo exigido no processamento computacional (KIM; LEE; OUARDA, 2014). Já a otimização global com o algoritmo evolução diferencial foi escolhido pela robustez e eficiência do método comprovada na resolução de diversos tipos de problemas (STORN; PRICE, 1997), e por estar implementado no projeto *open source* Scipy (VIRTANEN et al., 2019).

A independência dos resíduos foi admitida com base nos argumentos de McMahon e Peel (2019), sobre as dificuldades de modelar a autocorrelação com passo de tempo variável entre as observações. Além disso, tentativas de especificar o modelo com uma estrutura de autocorrelação dos resíduos resultaram em um comportamento com picos mais acentuados do que a abordagem sem autocorrelação, desmotivando assim essa abordagem.

As otimizações foram executadas de forma independente para diferentes valores de  $q$  (entre 2 e 7). O inteiro  $q$  que resultou na máxima função de verossimilhança e o correspondente vetor de parâmetros  $\theta_k$  foram selecionados para determinação da curva-chave para cada domínio  $T_k$ . O valor mínimo de  $q = 2$  ( $n = 5$  pares de dados observados) foi escolhido para criar alguma redundância na definição dos parâmetros  $\theta_k = [a_k, b_k, \sigma_k^2]$ , e assim evitar que erros de observação tenham dominância sobre o resultado. O valor máximo de  $q = 7$  ( $n = 15$  pares) foi escolhido com base no intervalo entre medições sucessivas na área de estudo – uma a cada duas ou três semanas. Dessa forma, a duração dos intervalos considerados para assimilação de dados observados ficou entre 50 e 250 dias, período razoável para manter um equilíbrio entre representatividade dos dados observados e a quantidade de pontos necessários para definição da curva-chave em condições regulares.

Em ambos os métodos de otimização, a transformação logarítmica das variáveis foi considerada:

$$\ln Q_j = a_k + b_k \ln(h_j - h_o), \quad y_j = \ln Q_j \quad \text{e} \quad x_j = \ln(h_j - h_o) \quad (29)$$

No caso da regressão linear com parâmetros estimados pelo método dos mínimos quadrados ordinários, o objetivo foi minimizar a soma dos quadrados dos resíduos:

$$Y = X\beta_k + \varepsilon, \quad \beta_k^* = \arg \min_{\beta_k} (\varepsilon^T \varepsilon) \quad \text{com} \quad \varepsilon \text{ i. i. d. } N(0, \sigma_k'^2) \quad (30)$$

em que  $\sigma_k'^2$  é a variância da variável transformada ( $\ln Q$ ).

Para a otimização global, o objetivo de minimizar a soma dos erros quadráticos se manteve, porém restrições foram impostas aos parâmetros da curva-chave:  $0,5 \leq a_k \leq 4,5$  e  $0,5 \leq b_k \leq 2,5$ . Esse método foi proposto como uma alternativa para forçar variações paramétricas dentro de uma faixa fisicamente factível, evitando a ocorrência, por exemplo, de expoentes negativos ou excessivamente altos, que fogem do esperado pelo conhecimento hidráulico teórico e empírico. A variância do modelo (na variável transformada) foi estimada da seguinte forma:

$$\sigma_k'^2 = \frac{1}{(2q + 1) - 2} \sum_{j=-q}^q [y_j - (a_k + b_k \ln(h_j - h_o))]^2 \quad (31)$$

No denominador, a subtração por 2 corresponde ao número de parâmetros livres de  $f_k$ . No Apêndice B, alguns trechos do código computacional para a determinação dos parâmetros (métodos dos mínimos quadrados ordinários e método *differential evolution*) são apresentados.

### 3.2.6.2 Correção de nível d'água

O nível d'água é a única variável de entrada no modelo de transformação nível d'água em vazão. Assim, qualquer procedimento de modelagem depende desse dado e de sua qualidade. De forma a assegurar uma correspondência entre os dados registrados automaticamente e os dados registrados manualmente a cada 15 dias, duas abordagens foram adotadas com base nas equações abaixo:

$$h_{rt} = \frac{[p_{w,t} - p_{atm,t}]}{\gamma_w} + 686,122 \quad (32)$$

$$h_t = c_k + d_k \times h_{r,t} \quad (33)$$

em que  $h_{r,t}$  [m a.s.l.] é o nível d'água bruto compensado a partir dos dados de pressão total  $p_{wt}$  [kN/m<sup>2</sup>] e de pressão atmosférica  $p_{atm}$  [kN/m<sup>2</sup>],  $\gamma_w$  é o peso específico da água [ $\sim 9,81$  kN/m<sup>3</sup>],  $h_t$  [m a.s.l.] é o nível d'água ajustado, e  $c_k$  e  $d_k$  são coeficientes determinados de forma similar aos parâmetros da curva-chave, de forma dinâmica. Um número fixo (31 pontos) foi adotado para o cálculo de cada par ( $c_k, d_k$ ). Os métodos de otimização descritos na seção anterior foram

utilizados para o cálculo do correspondente nível d'água ajustado. No método de otimização global, as restrições consideradas foram:  $-0,20 \leq c_k \leq 0,20$  e  $d_k = 1$ . Esta restrição pressupõe que a variação do nível d'água seja igual à estimada pela compensação dos dados do transdutor de pressão.

Períodos com falha de monitoramento (16/09/2010 a 23/11/2010; 04/05/2011 a 15/06/2011; 02/11/2015 a 07/12/2015; 06/04/2016 a 27/06/2016) foram preenchidos a partir de regressões lineares estabelecidas para cada mês com a precipitação média diária.

### 3.2.6.3 Avaliação das incertezas

A incerteza estatística  $\delta_k$  do método para cada instante do tempo foi estimada pela equação de propagação de erros normais:

$$\delta_k \cong \sqrt{\sigma_h^2 \left| \frac{\partial Q}{\partial h} \right|^2 + \sigma_k'^2} \quad \sigma_k = \sigma_k' \times Q_k \quad (34)$$

em que  $\sigma_h$  é a incerteza do nível d'água (obtida a partir do ajuste entre dados manuais e automáticos descrito na sub-seção anterior),  $\sigma_k'$  é a incerteza resultante da variável transformada  $\ln Q_k$  e  $\sigma_k$  é a incerteza da vazão modelada.

### 3.2.6.4 Avaliação da estratégia

Dados de vazão sub-horários medidos pelo ADCP SonTek IQ-Plus, instalado no bueiro do lado esquerdo na área de estudo, foram utilizados para verificar o desempenho do método proposto. Como as cotas de fundo ( $h_o$ ) dos bueiros são ligeiramente diferentes (esquerdo:  $h_{ol} = 686,260$  m a.s.l. e direito:  $h_{or} = 686,286$  m a.s.l.) (MALDONADO, 2009), a seguinte equação foi utilizada para estimar a vazão total:

$$Q_{obs,t} = Q_{a,t} \left[ 1 + \frac{h_t - h_{or}}{h_t - h_{ol}} \right] \quad (35)$$

em que  $Q_{a,t}$  [ $m^3/s$ ] é a vazão medida pelo ADCP no bueiro esquerdo,  $h_t$  [m a.s.l.] é o nível d'água ajustado, e  $Q_{obs,t}$  [ $m^3/s$ ] é a vazão total estimada.



As avaliações do método proposto foram feitas com auxílio de métricas de erro nas escalas de tempo horária e diária.

### 3.2.7 Separação do escoamento de base

O escoamento de base é a contribuição do aquífero ao escoamento superficial total. Como a área de estudo carece historicamente de dados isotópicos ou de outros traçadores para quantificar objetivamente essa componente, um método recursivo baseado apenas nos dados históricos de vazão total foi adotado. Métodos recursivos são objetivos e de fácil implementação, porém possuem a desvantagem de não possuírem embasamento físico (NATHAN; MCMAHON, 1990).

Em um trabalho anterior na área de estudo, Gómez (2016) utilizou o filtro recursivo de Lyne e Hollick (1979), seguindo as recomendações de Nathan e McMahon (1990). Nesse trabalho, um BFI da ordem de 0,93 (entre 0,88 e 0,96 no período de 2003 a 2013) foi estimado. Por outro lado, Batista et al. (2018), em trabalhos conduzidos em locais próximos à área de estudo e em condições geológicas e climáticas similares (áreas de afloramento do SAG, dentro da bacia hidrográfica do rio Jacaré-Pepira), estimaram um índice de escoamento de base (BFI) da ordem de 0,83 por meio de análises de balanço de massa com isótopos estáveis (H-2 e O-18) com amostragens de água superficial mensais ao longo de um ano, em pontos com áreas de drenagem de 8,9 km<sup>2</sup> e 10 km<sup>2</sup>.

Outro resultado relevante é que a estimativa do escoamento de base por meio de filtro inverso apresentou maior correlação com os resultados obtidos via balanço de massa isotópico do que o filtro de Eckhardt (2005), demonstrando que as hipóteses utilizadas na construção destes filtros podem não ser adequadas para a área de estudo (sobretudo, a hipótese de comportamento linear do aquífero).

A equação proposta por Collischonn e Fan (2013) para determinação do parâmetro  $BFI_{max}$  do filtro de Eckhardt (2005), baseada na razão entre os valores  $Q_{90}$  e  $Q_{50}$  da curva de permanência de vazões ( $0,8344 \times Q_{90}/Q_{50} + 0,2146$ ), resulta em 0,83 para os dados de ambas as seções de monitoramento fluviométrico da área de estudo (SM:  $Q_{90} = 0,434 \text{ m}^3/\text{s}$  e  $Q_{50} = 0,587 \text{ m}^3/\text{s}$ ; CP2:  $Q_{90} = 0,115 \text{ m}^3/\text{s}$  e  $Q_{50} = 0,157 \text{ m}^3/\text{s}$ ). Esse resultado é igual ao estimado por Batista et al. (2018) e consideravelmente menor que o estimado por Gómez (2016). Assim, há algumas diferenças marcantes entre os diferentes filtros recursivos e abordagens para estimar o escoamento de base.

Para tentar conciliar alguns desses resultados de trabalhos relacionados, o método alternativo proposto por Duncan (2019), baseado na aplicação de um filtro inverso com posterior simples filtragem, foi selecionado por levar a resultados graficamente satisfatórios e com índices de escoamento de base (BFI) próximos ao determinado por Batista et al. (2018) ou pela equação proposta por Collischonn e Fan (2013).

A primeira etapa do método consiste na aplicação de um filtro inverso da forma:

$$M_t = \min\left(\frac{Q_{t+1} - c}{k_1} + c, Q_t\right) \quad (36)$$

em que  $c$  e  $k_1$  são parâmetros do filtro,  $Q_t$  é a vazão observada, e  $M_t$  é a vazão resultante do filtro. O parâmetro  $k_1$  deve ser ajustado de modo que  $M_t$  reproduza o comportamento esperado da curva mestra de recessão, isto é, envelopando as diversas curvas de recessão.

Em seguida, o filtro recursivo de Lyne e Hollick (1979), de um parâmetro ( $k_2$ ), é utilizado para separar a componente de baixa frequência de  $M_t$  (resposta lenta -  $B_t$ ), que é a usada para estimar o escoamento de base.

$$F_t = k_2 F_{t-1} + (M_t - M_{t-1}) \times \frac{1 + k_2}{2} \quad (37)$$

$$B_t = M_t - F_t \quad (38)$$

em que  $F_t$  é a componente de resposta rápida,  $B_t$  é a componente de resposta lenta e  $M_t$  é a vazão resultante do filtro inverso. Quanto aos parâmetros do filtro, Duncan (2019) sugere que: a constante  $c$  seja fixada em 0 para cursos d'água perenes; resultados satisfatórios podem ser obtidos ao adotar  $k = k_1 = k_2$ ; e que o parâmetro  $k$  seja definido iterativamente por meio de análises gráficas.

### 3.2.8 Níveis freáticos

As fontes dos dados de níveis freáticos utilizados neste trabalho são apresentadas na Tabela 2, e foram selecionadas de modo a dar preferência aos poços com registros automáticos, com maiores períodos de monitoramento e em locais representativos das diferentes condições de uso do solo na bacia hidrográfica do Ribeirão da Onça. As séries temporais são apresentadas graficamente no Apêndice C.

O monitoramento automático de nível d'água é feito com transdutores de pressão Levellogger® Edge 3001 (Solinst) e Diver® DI501 (van Essen Instruments), que registram dados de pressão total a cada 12 h. Os dados de pressão total são compensados com dados de pressão atmosférica. Em seguida, os dados compensados são comparados com dados observados com medidor de nível, a cada 15 dias, para obtenção de constantes de ajuste.

Tabela 2. Fontes dos dados de monitoramento de nível freático

Poço de monitoramento	Cobertura atual	Medição manual (quinzenal)	Medição automática (a cada 12 horas)	
		Início	Início	Interrupções
P04	Pasto	03/2004	08/2018	--
P05	Pasto	04/2004	01/2015	--
P09	Eucalipto	07/2004	12/2013	--
P13	Cana-de-açúcar	03/2004	09/2014	06/2015 a 05/2018
P14	Cana-de-açúcar	07/2004	02/2015	--
P15	Citros	08/2004	09/2014	--
P16	Eucalipto	09/2004	01/2015	--
P19	Eucalipto	09/2004	05/2012	02/2014 a 11/2017

Fonte: Autor.

### 3.2.9 Verificação do balanço hídrico no período monitorado

Como forma de controle global das etapas anteriormente descritas, uma análise de balanço hídrico foi realizada de modo a evitar inconsistências nas estimativas dos processos. O volume de controle adotado foi a região entre a camada superior da vegetação e a base do aquífero freático da bacia:

$$\Delta S = P - ET - Q - R_p \quad (39)$$

em que  $\Delta S$  é a variação aproximada no volume total armazenado (nas zonas vadosa e saturada),  $P$  é a precipitação média,  $ET$  é a evapotranspiração média,  $Q$  é a vazão na seção de monitoramento fluviométrico e  $R_p$  é a percolação (recarga) profunda. Todos os termos são expressos em taxas (e.g., mm/ano).

Tomando dois instantes com condições antecedentes de precipitação similares (índices temporais  $t_a$  e  $t_b$ ), e supondo que a variação de armazenamento na zona vadosa pode ser desprezada nessas condições quando comparada à variação na zona saturada:

$$\Delta S_{ab} \cong S_y \Delta h_{ab} \quad (40)$$

em que  $\Delta S_{ab}$  é a variação aproximada de armazenamento da bacia entre os instantes  $t_a$  e  $t_b$ ,  $\Delta h_{ab}$  é a variação do nível freático médio da bacia no mesmo intervalo, e  $S_y$  é o rendimento específico médio do solo.

Análises gráficas do balanço hídrico foram conduzidas considerando diferentes condições de recarga profunda e diferentes valores do parâmetro de evapotranspiração  $c$  da equação 20 (usando para determinar o coeficiente  $K_{VI}$  a partir do índice de vegetação).

### 3.3 Análise de causalidade e conectividade de séries temporais

#### 3.3.1 Teste de causalidade de Granger

O teste de causalidade de Granger em sua formulação tradicional no domínio do tempo foi utilizado para investigar as relações causais e de conectividade entre as variáveis. O teste está baseado no aumento do poder preditivo de um modelo autorregressivo ao incluir uma variável adicional, candidata a apresentar relação de causa ou conexão, com uma certa defasagem. A significância estatística do teste de Granger foi verificada com o teste da razão de verossimilhança (WILKER, 1938).

A biblioteca de código aberto *Statsmodels 0.9.0* (Seabold e Perktold, 2010) foi utilizada para implementação computacional do método (código apresentado no Apêndice B). O algoritmo implementado consistiu na execução do teste para múltiplas defasagens entre pares de variáveis e avaliação das correspondentes significâncias estatísticas (valor-p) da relação causal no sentido de Granger.

#### 3.3.2 Informação mútua com defasagem normalizada

Complementarmente ao teste de Granger, a informação mútua com defasagem foi utilizada para identificar os graus de dependência entre as variáveis. As análises foram conduzidas em uma forma normalizada:

$$NMI(X, Y; \tau) = \frac{2 I(X, Y; \tau)}{H(X) + H(Y)} \quad (41)$$

em que NMI é a informação mútua entre X e Y com defasagem  $\tau$  normalizada,  $I(X, Y; \tau)$  é a informação mútua entre X e Y com defasagem  $\tau$ , e  $H(X)$  e  $H(Y)$  são as entropias informacionais (Shannon) das variáveis X e Y.

As distribuições de probabilidade foram estimadas a partir dos histogramas (LI et al., 2017; 2018). Apesar das deficiências e vieses introduzidos, o uso de estimadores de probabilidade baseados em histogramas é eficiente computacionalmente e pode ser suficiente para certas aplicações. O número de blocos do histograma foi definido para cada combinação de variável X e Y a partir do critério de Freedman-Diaconis (FREEDMAN; DIACONIS, 1981).

$$w_i = 2 \times \frac{\text{IQR}(i)}{\sqrt[3]{n}}; \quad b_{\text{FD}}(i) = \left\lceil \frac{\Delta_i}{w_i} \right\rceil \quad (43)$$

$$b = \left\lceil \sqrt{b_{\text{FD}}(X) \cdot b_{\text{FD}}(Y)} \right\rceil \quad (44)$$

em que  $w_i$  é a largura do bloco do histograma da variável  $i$ ,  $\text{IQR}(i)$  é o intervalo interquartil (diferença entre o terceiro e o primeiro quartil da variável  $i$ ),  $n$  é o tamanho amostral,  $\Delta_i$  é a diferença entre o maior e o menor valor observado da variável  $i$ ,  $b_{\text{FD}}$  é o número de blocos pela fórmula de Freedman-Diaconis, e, por fim,  $b$  é o número de blocos adotado nos histogramas. Os valores de NMI mudam com a quantidade de blocos utilizados, entretanto, o padrão de distribuição em relação à variável defasagem se mantém. O método foi implementado a partir da adaptação de algoritmos disponibilizados em repositórios abertos (Módulo *NonLinear Time Series Analysis*, <https://github.com/manu-mannattil/nolitsa>) (Apêndice B).

### 3.3.3 Avaliação das ferramentas e dos efeitos de agregação temporal e de suavização

Esta etapa teve como objetivo verificar o adequado funcionamento das ferramentas computacionais implementadas e a influência de diferentes esquemas de agregação temporal e suavização das séries temporais nos resultados. Uma abordagem multirresolução baseada em transformadas *wavelet* discretas foi adotada para o processo de suavização. Este processo foi conduzido com auxílio da ferramenta de código aberto *PyWavelets* (LEE et al., 2019). As séries resultantes de esquemas de agregação temporal foram obtidas a partir da média dos valores observados em períodos múltiplos do intervalo mínimo de amostragem da série original.

Os problemas sintéticos apresentados na Tabela 3 foram considerados. O problema sintético [1] consiste na geração de uma série temporal a partir da combinação (linear se  $c = 1$ ) de tamanho  $n = 2m + 1$  de uma variável independente  $P_t$  com defasagem temporal  $\tau$ . Uma componente aleatória é adicionada a essa parcela determinística, assumindo, arbitrariamente, uma distribuição normal, com parâmetros média zero e variância igual a um múltiplo, fracionário ou inteiro, da variância amostral da variável independente  $P_t$ . Esse problema foi proposto para verificar a sensibilidade do teste de Granger e da NMI na identificação de defasagens entre as variáveis envolvidas e a sensibilidade aos parâmetros:  $m$  (tamanho da janela de combinação de termos sequenciais);  $c$  (expoente da variável precipitação, em eventual não linearidade); e  $\alpha^2 \sigma_p^2$  (variância do erro aleatório) em diferentes esquemas de agregação temporal e em presença ou ausência de filtragem por DWT dos dados.

Tabela 3. Problemas sintéticos para avaliar efeitos de agregação temporal e suavização de séries temporais.

	Problemas propostos	Parâmetros
[1]	$y_t = \max \left( 0, \quad 0,2 + 1 \times 10^{-4} \sum_{j=-m}^m [P_{t-\tau^*+j}]^c + \varepsilon_t \right), \quad \varepsilon_t \sim N(0, \alpha^2 \sigma_p^2)$	$c, m, \alpha \tau^*$
[2]	$y_t = 0,8 \times \sum_{j=-\infty}^{\infty} P_{t-j} u_j + \varepsilon_t, \quad \varepsilon_t \sim N(0, \alpha^2 \sigma_p^2)$ $u_t = \frac{1}{k(n-1)!} \exp(-t/k) (t/k)^{n-1}$	$k, n, \alpha$

Fonte: Autor.

O problema [2] consiste na geração de séries a partir de operações (lineares) de convolução entre a variável independente  $P_t$  e funções de transformação  $u_t$ . De maneira similar ao problema [1], o objetivo desse problema foi verificar se as ferramentas propostas identificam adequadamente as defasagens em diferentes esquemas de agregação temporal e filtragem. No problema [2], a defasagem é produzida de forma distribuída segundo uma função de densidade de probabilidade com parâmetros  $n$  e  $k$  (NASH, 1957). As simulações executadas com os problemas sintéticos são apresentadas nas tabelas 4 e 5.

De forma arbitrária, a série de precipitação média diária na área de drenagem SM, referente à seção de monitoramento fluviométrico Santa Maria, foi utilizada como variável independente  $P_t$ . O período utilizado é de dezembro de 2008 a fevereiro de 2019. O uso da precipitação observada, e não de uma série sintética foi motivada pela complexidade que a série

de precipitação possui e por representar condições de aplicação em dados reais, que é o objetivo maior deste estudo.

Tabela 4. Simulações com o problema sintético [1]. A máxima defasagem adotada foi de 300 dias (d). Para a DWT, 'sym5' foi utilizada para y e 'haar' foi utilizada para P.

Simulação	Descrição	Parâmetros fixos	Parâmetro variável	Modificação da série temporal
S1-1	Avaliar o efeito do parâmetro m	$c = 1$ $\alpha = 1$ $\tau^* = 150$	$m = 1; 10; 100; 200$	Nenhuma
S1-2	Avaliar o efeito do parâmetro c	$m = 1$ $\alpha = 1$ $\tau^* = 150$	$c = 0,1; 0,5; 2; 3$	Nenhuma
S1-3	Avaliar o efeito do parâmetro $\alpha$	$c = 1$ $m = 1$ $\tau^* = 150$	$\alpha = 1; 3; 5; 10$	Nenhuma
S1-4	Avaliar o efeito do parâmetro $\tau$	$c = 1$ $m = 1$ $\alpha = 1$	$\tau^* = 10; 100; 200; 500$	Nenhuma
S1-5	Avaliar o efeito da agregação temporal	$c = 1$ $m = 1$ $\alpha = 1$ $\tau^* = 150$	--	Agregações temporais (d): $\Delta t = 7, 15, 30, 60$
S1-6	Avaliar o efeito da transformada <i>wavelet</i> discreta (DWT)	$c = 1$ $m = 1$ $\alpha = 1$ $\tau^* = 150$	--	Níveis de decomposição: 1, 2, 3 e 4

Fonte: Autor.

Tabela 5. Simulações com o problema sintético [2]. A máxima defasagem adotada foi de 300 dias (d).

Simulação	Descrição	Parâmetros fixos	Parâmetro variável	Modificação da série temporal
S2-1	Avaliar o efeito do parâmetro m	$k = 20$ $\alpha = 0,1$	$n = 1; 4; 8; 12$	Nenhuma
S2-2	Avaliar o efeito do parâmetro $\alpha$	$k = 20$ $n = 8$	$\alpha = 0,01; 0,25, 0,50; 1,00$	Nenhuma
S2-3	Avaliar o efeito da agregação temporal	$k = 20$ $n = 8$ $\alpha = 0,01$	--	Agregações temporais (d): $\Delta t = 7, 15, 30, 60$
S2-4	Avaliar o efeito da transformada <i>wavelet</i> discreta (DWT)	$k = 20$ $n = 8$ $\alpha = 0,01$	--	Níveis de decomposição: 1, 2, 3 e 4

Fonte: Autor.

Os resultados foram analisados em função de diferentes defasagens até o limite superior (defasagem máxima). A defasagem máxima foi definida em geral como 300 dias. Para cada defasagem e caso analisado, a razão de verossimilhança (LR), a probabilidade de significância (valor p) dessa razão pelo teorema de Wilks (WILKS, 1938) e a informação mútua com defasagem normalizada foram avaliadas conjuntamente. Para o problema [1], a defasagem que os métodos devem detectar está diretamente relacionada ao parâmetro  $\tau$ , enquanto que no

problema [2], a defasagem a ser detectada depende dos parâmetros  $n$  e  $k$  (o valor esperado de  $u_t$  é igual a  $nk$  e a variância,  $nk^2$ ).

### 3.3.4 Análises pareadas de processos hidrológicos observados na área de estudo

Os processos hidrológicos observados na área de estudo foram analisados de forma pareada usando o teste de Granger e a informação mútua com tempo de defasagem normalizada (NMI) como meios de quantificar as correspondentes conectividades e dependências. Os resultados da etapa de análise dos efeitos de agregação temporal e de filtragem das séries temporais foram utilizados na definição do plano de análise.

A precipitação, a evapotranspiração de referência, a evapotranspiração real estimada, a diferença entre a precipitação e a evapotranspiração real estimada, o *enhanced vegetation index*, escoamento superficial e suas componentes nas diferentes seções de monitoramento fluviométrico (CP2, SM) e os níveis freáticos de nove poços de monitoramento (4, 5, 8, 9, 13, 14, 15, 16, 19) foram avaliados em grupos de variáveis de causa e de resposta, conforme indicado na Tabela 6. De forma complementar, um indicador referente a fenômenos atmosféricos de grande escala, o *oceanic Niño index*<sup>3</sup> (média móvel trimestral da anomalia na temperatura média da superfície do mar na região do Pacífico Equatorial Niño 3.4), também foi inserido na análise. Os dados foram interpolados linearmente e analisados em escala diária.

Tabela 6. Análises de causalidade e dependência entre variáveis observadas. Variáveis referentes à área de controle SM e aos poços de monitoramento 4, 5, 9, 13, 14, 15, 16 e 19.

Variáveis de causa	Variáveis de resposta
P	Q, Q <sub>b</sub> , EVI, GWL <sub>j</sub>
ET <sub>a</sub>	GWL <sub>j</sub>
EVI, ONI	P
GWL <sub>j</sub>	Q <sub>b</sub>

P: precipitação; ET<sub>a</sub>: evapotranspiração real estimada; EVI: enhanced vegetation index; ONI: *oceanic Niño index*; GWL: nível freático; Q: vazão; Q<sub>b</sub>: vazão de base.

Fonte: Autor.

Dados no período de dezembro de 2008 a fevereiro de 2019 estavam disponíveis para as análises na área SM e no período de abril de 2017 a fevereiro de 2019, para a área CP2 (calha

<sup>3</sup> Dados obtidos no NOAA-NWS/ USA website: <https://origin.cpc.ncep.noaa.gov/>. Acesso em 05/08/2019.



Parshall 2). Devido à pequena extensão das séries de dados disponíveis na área CP2 e possível redundância ou deficiência dos resultados obtidos, as análises foram feitas apenas para os dados da área SM. A defasagem máxima utilizada nos testes foi de aproximadamente 1/3 do comprimento do período de dados disponíveis (1000 dias).

As análises foram conduzidas inicialmente com os valores estimados de cada uma das variáveis. Em seguida, as incertezas das variáveis foram incluídas, assumindo, por simplificação, uma distribuição de probabilidade normal. Embora essa simplificação seja problemática para algumas variáveis (e.g., precipitação), ela foi utilizada por permitir a análise de estabilidade dos resultados obtidos. Para a precipitação e evapotranspiração, os erros-padrão resultantes dos processos de interpolação espacial e regressão foram assumidos como aproximações da incerteza. Para a vazão, as incertezas do modelo de transformação dinâmico com parâmetros determinado pelo método de otimização global foram utilizadas. Para dados de nível freático, foi assumida uma incerteza fixa de 0,02 m, para dados de EVI, 0,06, e para dados de ONI, um valor desprezível de  $5 \times 10^{-4}$ .

$$\tilde{X}_t = \mu_{X_t} + \varepsilon_{X_t}, \quad \varepsilon_{X_t} \sim N(0, \sigma_{X_t}^2) \quad (45)$$

em que  $\mu_{X_t}$  é valor estimado da variável  $X_t$ ,  $\varepsilon_{X_t}$  é um termo aleatório com distribuição normal,  $\sigma_{X_t}^2$  é a variância amostral da variável observada  $X_t$ , e  $\tilde{X}_t$  é o valor atualizado da variável  $X_t$  afetado pelo termo aleatório. Esse foi o valor utilizado nos testes com os dados observados. De modo a seguir o mesmo padrão de aplicação dos testes sintéticos, termos aleatórios foram adicionados apenas nas ‘variáveis de causa’.

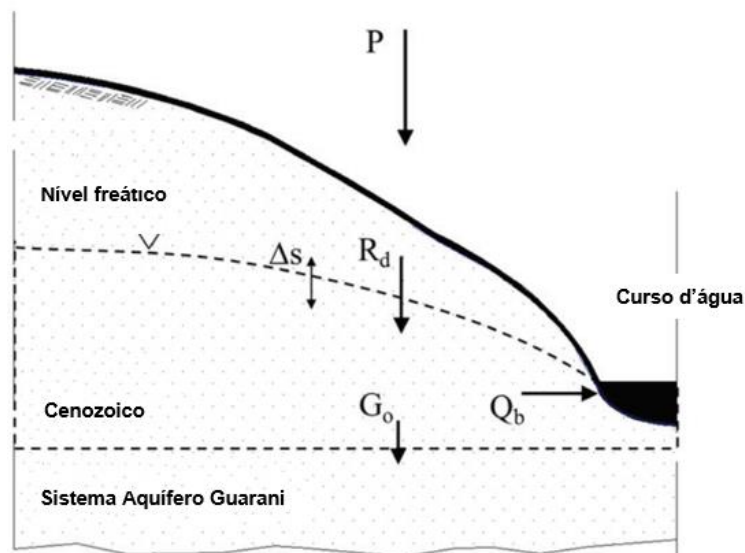
De forma similar às análises dos problemas sintéticos, os resultados do teste de Granger (razão de verossimilhança e valor p) e da informação mútua com defasagem normalizada foram analisados até a defasagem máxima e de forma conjunta. Em contraste com os problemas sintéticos, as relações de conectividade entre os dados observados são objetos de compreensão. Os resultados produzidos foram relacionados a possíveis explicações físicas, de como os processos podem ou não estar conectados.

### 3.4 Proposição do modelo conceitual hidrológico

#### 3.4.1 Conhecimentos básicos sobre o comportamento hidrológico da área de estudo

Trabalhos anteriores desenvolvidos na área de estudo constataram que o fluxo subterrâneo é o processo dominante na geração do escoamento do curso de água principal (Ribeirão da Onça). Uma estrutura que tem sido utilizada para representar o funcionamento hidrogeológico da bacia é apresentada de forma esquemática na Figura 17 (WENDLAND; GOMES; TROEGER, 2015). Esse modelo conceitual assume como volume de controle a zona saturada dos sedimentos Cenozoicos, compreendido entre a superfície freática e os arenitos do Sistema Aquífero Guarani.

Figura 17. Modelo conceitual referente aos processos da zona saturada ( $P$  = precipitação;  $R_d$  = recarga direta;  $\Delta S$  = variação no armazenamento;  $Q_b$  = escoamento de base;  $G_o$  = percolação profunda ao SAG).



Fonte: Adaptado de Wendland, Gomes e Troeger (2015).

Esse modelo foi o ponto de partida na proposição de um modelo conceitual hidrológico para a bacia, capaz de representar os processos dominantes na zona não saturada e na zona saturada e de reproduzir a dinâmica da disponibilidade hídrica em termos de fluxos e volumes armazenados.

#### 3.4.2 Modelo perceptivo da bacia

As visitas quinzenais à área de estudo para fins de monitoramento e os resultados de estudos anteriores foram utilizados para compor um conjunto de informações para a proposição

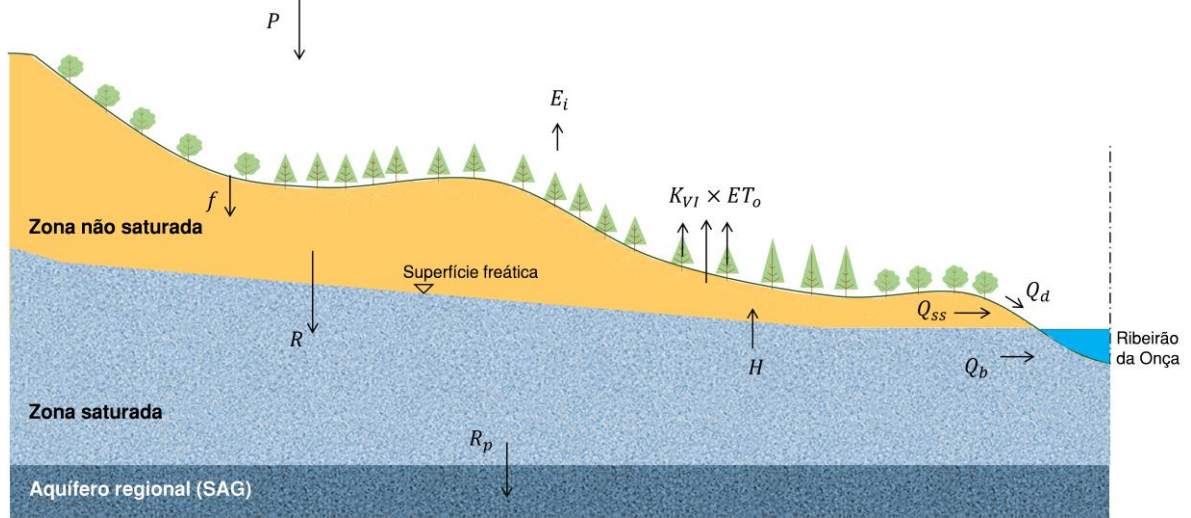
de um modelo conceitual que incorpore, além dos processos considerados no modelo proposto por Wendland, Gomes e Troeger (2015), os processos de interceptação vegetal, evapotranspiração, infiltração e formação do escoamento superficial.

O solo da área de estudo apresenta uma textura arenosa, com alta capacidade de infiltração. Nos usos de solo predominante da bacia, plantações de eucalipto e citros, há uma tendência de grande parte do total precipitado infiltrar. O escoamento superficial direto foi estimado em menos de 1% da precipitação em estudos conduzidos em parcelas experimentais instaladas em plantações de eucalipto na área de estudo (MARTINS FILHO, 2014). Do total que precipita, entre 10 e 15% contribui para a recarga do aquífero freático da bacia (LUCAS; WENDLAND, 2016). Outra informação relevante consiste na associação existente entre o comportamento do nível freático e os diferentes usos do solo na bacia, evidenciando os efeitos da evapotranspiração no sistema hidrológico (GOUVÊA; WENDLAND, 2011; LUCAS; WENDLAND, 2016).

Por outro lado, áreas com terreno mais ondulados, próximas aos pastos e à rede natural de drenagem apresentam evidências de escoamento superficial efêmero e processos erosivos intensos. Essas áreas representam, entretanto, apenas uma pequena parcela da bacia. A dinâmica sazonal de entressafra, ou entre plantios no caso de eucaliptos, pode ter um efeito nas respostas hidrológicas tão importante quanto as mudanças no uso do solo. A partir dessa percepção, um aspecto importante a ser capturado no modelo é a dinâmica temporal da vegetação para uma melhor estimativa da evapotranspiração e de seus efeitos nas respostas hidrológicas.

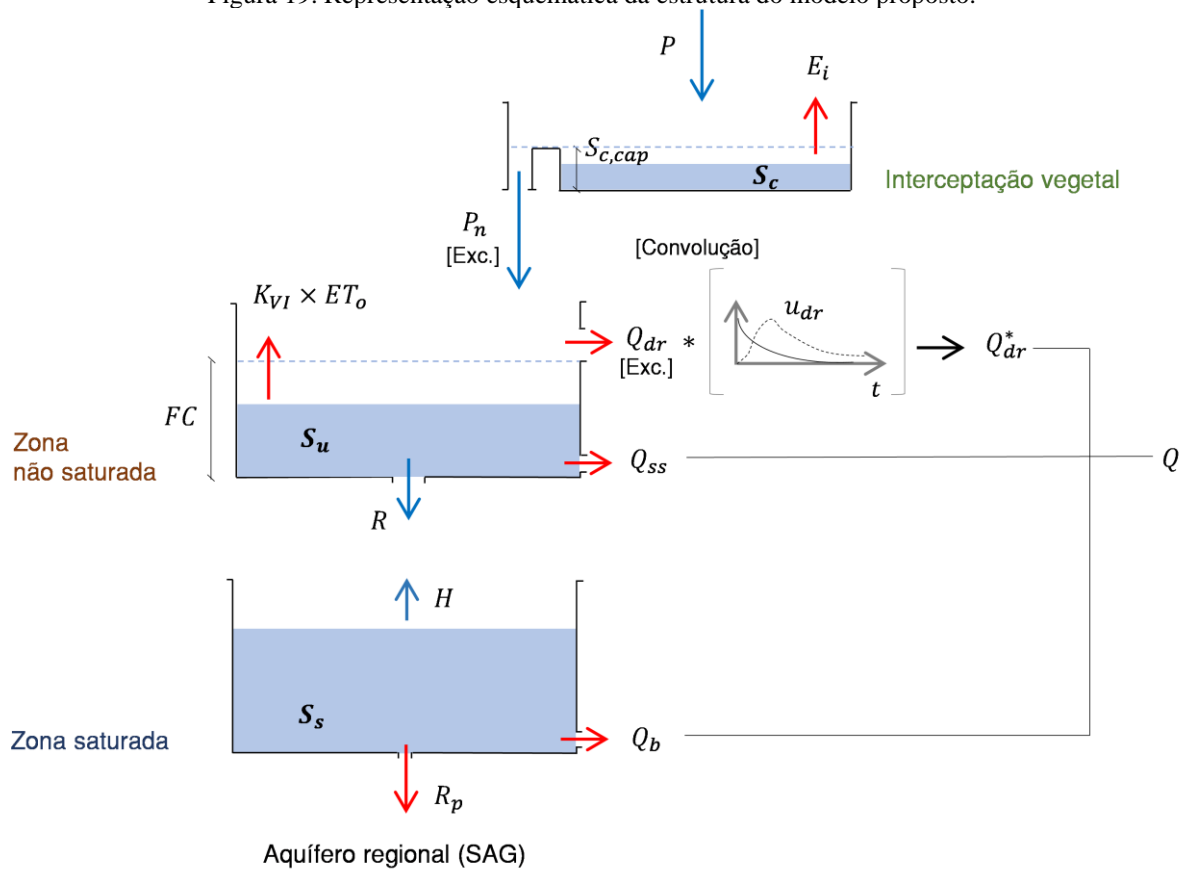
Para fins de modelagem, uma abordagem baseada em representações concentradas (não distribuídas) dos volumes de controle e dos fluxos foi seguida e buscou incluir de forma simplificada as informações levantadas sobre o comportamento da bacia. Essa decisão afetou a representação de alguns processos conhecidamente heterogêneos em superfície, como a maior geração de escoamento superficial direto em áreas de pastagem e maior evapotranspiração em zona ripária (CARVAJALINO, 2019). Entretanto, por estar compatível com o nível de complexidade suportado pelos dados disponíveis e com as incertezas envolvidas, ela foi adotada neste estudo. Quanto à estrutura do modelo, pelos resultados de trabalhos anteriores e pela existência de conectividades físicas entre solo, aquífero e rio, o modelo esquematizado nas figuras 18 e 19 foi proposto.

Figura 18. Representação de processos dominantes no modelo proposto.



Fonte: Autor.

Figura 19. Representação esquemática da estrutura do modelo proposto.



Fonte: Autor.

A estrutura apresentada na Figura 19 indica três reservatórios, que são referentes à vegetação, ao solo e ao aquífero livre da bacia. Os fluxos indicados por setas são apresentados em detalhe na seção seguinte.

### 3.4.3 Considerações sobre o funcionamento da bacia e parametrização dos fluxos

Cada um dos reservatórios considerados no modelo possui uma equação de balanço hídrico (ou conservação de massa) associada. As parametrizações indicadas a seguir foram definidas de forma iterativa, baseadas em parametrizações de outros modelos conceituais (LINDSTRÖM et al., 1997; CHIEW; PEEL; WESTERN, 2002; KNOBEN et al., 2019) e verificando os desvios entre dados observados e resultados simulados.

#### 3.4.3.1 Reservatório de interceptação vegetal [ $S_c$ ]

Interceptação vegetal é o processo de retenção da água na chuva pela vegetação antes que ela atinja o solo. Embora sua significância comparada a outros componentes do balanço hídrico seja questionável, a possibilidade de incluí-la na modelagem e prover maior fidelidade à sequência de processos que acontecem na área de estudo motivou a inclusão desse reservatório. Neste estudo, a seguinte equação simplificada de balanço foi adotada:

$$\frac{dS_c}{dt} = P_G - E_i - P_N \quad (46)$$

em que  $S_c$  é a variável de estado (volume armazenado),  $P_G$  é a precipitação média sobre a área de estudo,  $E_i$  é a evaporação da água interceptada e  $P_N$  é a precipitação efetiva média, que de alguma forma chega ao solo.

A partir dos resultados de Von Hoyningen-Huene (1981) – que relaciona a capacidade de armazenamento do reservatório de interceptação vegetal e o índice de área foliar (LAI) – e de Kang et al. (2016) – que relaciona o LAI com índices de vegetação obtidos por sensoriamento remoto, uma capacidade de armazenamento variável no tempo foi proposta da seguinte forma:

$$S_{c,cap} = 0,935 + 2,241 \times EVI - 0,116 \times EVI^2, \quad EVI \in [0,30; 0,55] \quad (47)$$

em que  $EVI$  (*enhanced vegetation index*) é um valor médio na área de estudo para um dado instante, e  $S_{c,cap}$  é a correspondente capacidade de armazenamento da interceptação vegetal referente a esse mesmo instante de tempo.

Para quantificar a evaporação da água interceptada, a seguinte parametrização foi adotada:

$$E_i = k_E \times ET_o \times \min\left(\frac{S_c}{S_{c,cap}}, 1\right) \quad (48)$$

em que  $E_i$  é a taxa de evaporação da água interceptada,  $ET_o$  é a evapotranspiração de referência,  $S_c$  é a variável de estado (volume armazenado) do reservatório de interceptação vegetal, e  $k_E$  é um coeficiente de evaporação. Esse coeficiente de evaporação foi estimado em média como 1,15 para a área de estudo ao ignorar a resistência estomática ( $r_c \cong 0$ ) na equação de evapotranspiração de Penman-Monteith FAO-56 (DAVID; VALENTE; GASH, 2005):

$$k_E = \frac{ET_o^*(r_c = 0)}{ET_o} \cong 1,15 \quad (49)$$

Por fim, para estimar a precipitação efetiva, as seguintes condições e equações foram utilizadas:

$$P_N = 0, \text{ se } S_c + P_G \leq S_{c,cap} \quad (50)$$

$$P_N = [1,07 - 0,40 \times EVI + \delta_N] \times P_G + \max(0, S_c - S_{c,cap}), \text{ se } S_c + P_G > S_{c,cap} \quad (51)$$

em que  $\delta_N \in [-0,05; 0,05]$  é um parâmetro a ser calibrado.

#### 3.4.3.2 Reservatório da zona não saturada [ $S_u$ ]

A zona não saturada é o componente principal de grande parte dos modelos hidrológicos conceituais (KNOBEN et al., 2019). Os processos relevantes nesse reservatório são indicados na equação de balanço hídrico:

$$\frac{dS_u}{dt} = f + H - Q_{ss} - ET - R \quad (52)$$

em que  $S_u$  é a variável de estado (volume armazenado na zona não saturada),  $f$  é a taxa infiltração,  $H$  é a taxa ascensão de água proveniente da zona saturada (capilaridade e absorção

por raízes),  $Q_{ss}$  é a taxa de escoamento subsuperficial,  $ET$  é a transpiração vegetal e a evaporação do solo, e  $R$  é a taxa de recarga subterrânea.

Pela falta de estudos empíricos e de modelos que funcionem de forma adequada para a área de estudo, a taxa de escoamento superficial direto ( $Q_d$ ) foi estimada de forma simplificada a partir da precipitação efetiva ( $P_N$ ), de um parâmetro relacionado à capacidade de armazenamento ( $FC$ ) e do estado do reservatório ( $S_u$ ):

$$Q_d = \max(0, S_u + P_N - FC) \quad (53)$$

A taxa de infiltração ( $f$ ) foi assumida como a diferença entre a precipitação efetiva e a taxa de escoamento superficial direto.

$$f = P_N - Q_d \quad (54)$$

A taxa de escoamento subsuperficial ( $Q_{ss}$ ) foi parametrizada como uma função linear da taxa de infiltração e do grau de saturação da zona não saturada.

$$Q_{ss} = k_{ss} \left( \frac{S_u}{FC} \right) f \quad (55)$$

em que  $k_{ss}$  é um parâmetro a ser calibrado.

Para estimar a evapotranspiração (parcela de água interceptada a ser adicionada), as seguintes relações foram propostas:

$$ET = [K_c(EVI) \times ET_o] \times \frac{S_u}{LP \times FC}, \quad \text{se } S_u \leq LP \times FC \quad (56)$$

$$ET = [K_c(EVI) \times ET_o], \quad \text{se } S_u \geq LP \times FC, \quad (57)$$

$$K_c(EVI) = \max(1,65[1 - \exp(-2,25EVI)] - c; 0,05), \quad c \in [0,125; 0,320] \quad (58)$$

em que  $K_c$  é um coeficiente estimado a partir do EVI médio da bacia (variável no tempo); LP é um parâmetro do modelo relacionado à capacidade de transpiração vegetal sob condições de estresse hídrico; e  $c$  é outro parâmetro do modelo, relacionado à curva de transformação de EVI para obtenção do coeficiente  $K_c$ .

A estimativa de recarga subterrânea (R) seguiu uma parametrização similar ao modelo HBV-96 (LINDSTRÖM et al., 1997):

$$R = \left( \frac{S_u}{FC} \right)^\beta [f - Q_{ss}] \quad (59)$$

em que  $\beta \in [1, 3]$  é um parâmetro do modelo a ser calibrado.

Sabe-se que a absorção profunda de água pode acontecer, em especial, como estratégia de adaptação a condições de estresse hídrico. Isso foi verificado em plantações de eucalipto no estado de São Paulo, em que as raízes atingiram superfícies freáticas a mais de 12 m de profundidade (CHRISTINA et al., 2017). Na área de estudo, Coutinho (2019) identificou raízes grossas de eucalipto maiores que 2 m, mas com sistema radicular predominantemente superficial. Em função da possibilidade de absorção profunda (ao menos nas partes mais baixas da bacia) e para incluir efeitos de capilaridade (mesmo que dificultados pela presença de solo arenoso), a seguinte parametrização de ascensão de água da zona saturada para a zona não saturada foi considerada:

$$H = c_H \times \left( 1 - \frac{S_u}{FC} \right) \quad (60)$$

em que  $c_H$  é um parâmetro a ser calibrado (máxima ascensão de água por unidade de tempo) e H é a taxa de ascensão de água.

### 3.4.3.3 Reservatório da zona saturada [ $S_s$ ]

A zona saturada compreende o volume entre a superfície freática e os arenitos do Sistema Aquífero Guarani. Os processos relevantes nesse reservatório são indicados na equação de balanço hídrico:



$$\frac{dS_s}{dt} = R - H - R_p - Q_b \quad (61)$$

em que  $S_s$  é a variável de estado (volume armazenado na zona saturada),  $R$  é a taxa de recarga proveniente da zona não saturada,  $H$  é a taxa ascensão para a zona não saturada,  $R_p$  é a taxa de percolação profunda para o SAG (contribuição para o aquífero regional), e  $Q_b$  é a vazão de base.

As variáveis  $R$  e  $H$  são comuns ao reservatório da zona não saturada e já foram apresentadas. A taxa de percolação profunda foi admitida como um parâmetro fixo do modelo. A vazão de base foi estimada a partir do conceito de função de sensibilidade (KIRCHNER, 2009) e da relação do escoamento de base com sua primeira derivada (TROCH et al., 2013).

$$g(Q) = \frac{dQ_b}{dS_s} = \frac{dQ_b/dt}{dS_s/dt} \cong \frac{-a Q_b^b}{-Q_b} = a Q_b^{b-1} \quad (62)$$

$$\frac{dQ_b}{dt} = g(Q) \frac{dS_s}{dt} = a Q_b^{b-1} \frac{dS_s}{dt} = a Q_b^{b-1} [R - H - R_p - Q_b] \quad (63)$$

em que  $g(Q)$  é uma função de sensibilidade e  $a$  e  $b$  são parâmetros do modelo relacionados à recessão do escoamento de base.

#### 3.4.3.4 Composição de fluxos

O fluxo total de evaporação e transpiração ( $ET_a$ ) é a soma direta das duas componentes:

$$ET_a = E_i + ET \quad (64)$$

Já para a composição do escoamento superficial ( $Q$ ), um operador de convolução (\*) foi utilizado para compensar a representação concentrada da bacia. A convolução do escoamento superficial direto utilizou uma função de distribuição de probabilidade correspondente ao hidrograma unitário de Nash (1958) ( $u_{dr}$ ), com parâmetros  $n_f$  e  $k_f$ .

$$u_{dr}(\tau) = \frac{1}{k_f \times \Gamma(n_f)} \exp\left[-\frac{\tau}{k_f}\right] \times \left[\frac{\tau}{k_f}\right]^{n_f-1} \quad (65)$$

em que  $\Gamma$  é a função gama (extensão da função fatorial para o conjunto de números reais ou complexos). Assim, a vazão total na saída da bacia é dada por  $Q_t$ :

$$Q_t = Q_{bt} + Q_{sst} + (u_d * Q_{dr})_t \quad (66)$$

### 3.4.5 Representação matemática e solução numérica

A representação no espaço de estados de um sistema é uma função matricial que relaciona variáveis de entrada, de saída e de estado por meio de equações diferenciais (SANTOS; THIREL; PERRIN, 2018; KNOBEN et al., 2019). Para o modelo proposto, cujas equações de fluxo foram previamente apresentadas, a seguinte notação pode ser adotada:

$$\mathbf{S}'(t) = \mathbf{F}(\mathbf{S}(t) | \boldsymbol{\theta}) \quad (67)$$

em que  $\mathbf{S}(t)$  indica o vetor de variáveis de estado (armazenamento dos reservatórios) ao longo do tempo,  $\mathbf{F}(t)$  é uma matriz de transformação relacionada à definição dos fluxos nos reservatórios e  $\boldsymbol{\theta}$  é o vetor de parâmetros do modelo.

$$\left[ \begin{array}{l} \frac{dS_c}{dt} = P_G - E_i - P_N \\ \frac{dS_u}{dt} = f + H - Q_{ss} - ET - R \\ \frac{dS_s}{dt} = R - H - R_p - Q_b \\ \frac{dQ_b}{dt} = a Q_b^{b-1} [R - H - R_p - Q_b] \end{array} \right. \quad (68)$$

Diferentes abordagens têm sido utilizadas para resolver sistema de equações diferenciais ordinárias em modelos hidrológicos, com destaque para a decomposição de operadores (*operator splitting*), que faz a resolução sequencial das equações (SANTOS; THIREL; PERRIN, 2018). Entretanto, inconsistências na simulação dos fluxos seguindo tal abordagem evidenciam que a resolução simultânea das equações, baseada em métodos numéricos, é a mais adequada (CLARK; KAVETSKI, 2010; FENICIA; KAVETSKI; SAVENIJE, 2011).

O método com passo de tempo adaptativo de Runge-Kutta de ordens 4 e 5 (RK45), implementado no projeto open source Scipy (VIRTANEN et al., 2019), foi utilizado como primeira opção para resolver o sistema de equações diferenciais do modelo. A implementação deste método é baseada nas fórmulas propostas por Dormand e Prince (1980). O erro absoluto e o erro relativo toleráveis foram especificados em  $1 \times 10^{-2}$  e  $1 \times 10^{-3}$  para todas as variáveis. O máximo passo de tempo permitido nos cálculos foi limitado ao passo de tempo dos dados de entrada. Os dados de entrada foram considerados constantes em seus intervalos de discretização.

A possível ocorrência de problemas numéricos com o método adaptativo de Runge-Kutta, por inadequações do modelo na etapa de testes, motivou a proposição de um método alternativo a ser utilizado em caso de falha (não convergência). Esse método alternativo foi baseado no método LSODA (HINDMARSH, 1983), que identifica e muda automaticamente o método de solução numérica de acordo com a rigidez do sistema. Ele também está implementado no Scipy. As tolerâncias de erro absoluto e relativo adotadas foram as mesmas anteriormente apresentadas.

### **3.4.5 Calibração, validação e avaliação do modelo**

Os parâmetros do modelo são apresentados em resumo na Tabela 7. Os intervalos de variação foram estimados a partir de testes com dados da área de estudo e da análise da ordem de grandeza dos resultados esperados.

Tabela 7. Parâmetros do modelo.

Domínio	Parâmetro	Processo relacionado	Intervalo de variação
Interceptação	$\delta_N$	Precipitação efetiva	$[-0,03; 0,03]$
	$\beta$	Recarga subterrânea	$[1; 2]$
Zona não saturada	$c_H$	Ascensão máxima de água subterrânea [mm/dia]	$[0,00; 0,20]$
	LP	Evapotranspiração	$[0,2; 0,8]$
	FC	Infiltração, Evapotranspiração, Recarga e outros [mm]	$[100; 1000]$
	$k_{ss}$	Escoamento subsuperficial	$[0,01; 0,10]$
	$c_{VI}$	Evapotranspiração	$[0,125; 0,320]$
Zona saturada	$R_p$	Percolação profunda [mm/dia]	$[0,08; 0,150]$
	a	Escoamento de base	$[0,0010; 0,0025]$
	b	Escoamento de base	$[1; 3]$
Escoamento superficial	$k_f$	Convolução do escoamento superficial	$[1; 10]$
	$n_f$	direto	$[1; 10]$

Fonte: Autor.

A calibração dos parâmetros seguiu uma estratégia multi-objetivo e foi conduzida de forma independente para as áreas SM e CP2. Inicialmente, diversos métodos de otimização com objetivo único, com ou sem restrição, foram testados, incluindo os métodos de gradientes conjugados, Powell, Nelder-Mead, L-BFGS (todos disponíveis na biblioteca Minimize do projeto Scipy). A não existência de soluções do sistema de equações diferenciais foi um problema enfrentado por métodos irrestritos. Com base nessa possibilidade, o algoritmo de otimização global, com restrições, *Non-dominated sorting genetic algorithm II* (NSGA-II) (DEB et al., 2002) foi utilizado por ter um desempenho computacional compatível com os propósitos deste trabalho e por estar implementado em pacotes de otimização de código aberto<sup>4</sup>. O tamanho da população foi fixado em 100 indivíduos (número de vetores de parâmetros – ou indivíduos – que são perturbados por operadores de mutação e de recombinação no sentido de buscar os pontos ótimos) e o número máximo de gerações foi fixado em 10.000, que é o número de execuções de operadores que geram diversidade e selecionam os melhores indivíduos baseado nas funções objetivo.

A reprodução das vazões observadas nos exutórios (SM ou CP2) e da dinâmica do armazenamento da zona saturada foram os objetivos do problema de otimização. A métrica não paramétrica  $KGE_{NP}$  proposta por Pool et al. (2018) foi adotada como critério de calibração para o objetivo de reprodução de vazões. Essa métrica foi selecionada por apresentar, segundo os autores, boas propriedades na reprodução de respostas relacionadas ao regime de baixas vazões.

<sup>4</sup> Pacote Platypus: <https://platypus.readthedocs.io/en/docs/algorithms.html#nsga-ii>. Acesso em 07/10/2019.

O  $KGE_{NP}$  (69) é uma função a ser maximizada (valor ótimo = 1) que avalia a correlação de postos de Spearman ( $r_s$ ) e inclui penalizações a desvios relacionados à reprodução da média ( $\mu_{sim}$ ) e à reprodução da curva de permanência ( $\alpha_{NP}$ ).

$$KGE_{NP} = 1 - \sqrt{\left(\frac{\mu_{sim}}{\mu_{obs}} - 1\right)^2 + (\alpha_{NP} - 1)^2 + (r_s - 1)^2} \leq 1 \quad (69)$$

$$\alpha_{NP} = 1 - \frac{1}{2} \sum_{k=1}^n \left| \frac{Q_{sim}(I(k))}{n\mu_{sim}} - \frac{Q_{obs}(J(k))}{n\mu_{obs}} \right| \quad (70)$$

em que  $I(k)$  e  $J(k)$  são instantes de tempo em que a  $k$ -ésima maior vazão ocorre dentro das séries de vazões simulada e observada,  $Q_{obs}$  é a vazão observada,  $\mu_{obs}$  indica a média amostral dos  $n$  valores de vazão observados,  $Q_{sim}$  indica os valores previstos pelo modelo e  $\mu_{sim}$  indica a média amostral dos  $n$  valores de vazão simulados.

O coeficiente de correlação de postos de Spearman entre a média de níveis freáticos dos poços de monitoramento da bacia hidrográfica considerada (poços 4, 5, 13, 14, 15, 16 e 19 para a área SM; e poços 4 e 5 para a área CP2) e a variável de estado  $S_s$ , do modelo conceitual, foi adotada como critério de calibração para a reprodução da dinâmica do armazenamento de água na zona saturada. Assim, o vetor de funções objetivo do problema de otimização é dado por:

$$\mathbf{F}(\mathbf{x}) = [KGE_{NP}[\mathbf{x}], r_s[\mathbf{x}]], \quad \mathbf{x} \in \mathbb{R}^d \quad (71)$$

O vetor de decisão,  $\mathbf{x}$ , é o vetor de parâmetros do modelo ( $\boldsymbol{\theta}$ ), mais um coeficiente ( $\alpha_u$ ), adotado para definir a condição inicial de estado da zona não saturada.

$$\mathbf{x} = [\boldsymbol{\theta}, \alpha_u], \quad \alpha_u = S_u(t_0)/S_{u_{max}} \quad (72)$$

em que  $S_u(t_0)$  é o armazenamento na zona não saturada no início da simulação e  $S_{u_{max}}$  é o valor máximo possível (parâmetro FC).

De modo a reduzir a influência dos valores iniciais das variáveis de estado e da vazão de base na avaliação do modelo, um trecho inicial de 7 dias foi desconsiderado no processo de

calibração, ou seja, os desvios entre modelo e observações não entraram na avaliação das funções objetivo (período de aquecimento do modelo).

Para a área SM, os períodos de calibração e validação foram fixados em 01/01/2009 a 01/09/2014 e 02/09/2014 a 18/02/2019. Para seção CP2, os períodos de calibração e validação foram fixados em 01/05/2017 a 01/09/2018 e 02/09/2018 a 17/02/2019, respectivamente.

As análises dos resultados levaram em consideração a extensão de dados disponíveis, a condição meteorológica predominante (mais seco ou mais úmido) e a instrumentação existente na bacia nos períodos.

### **3.5 Avaliação do modelo pelas métricas de causalidade e dependência**

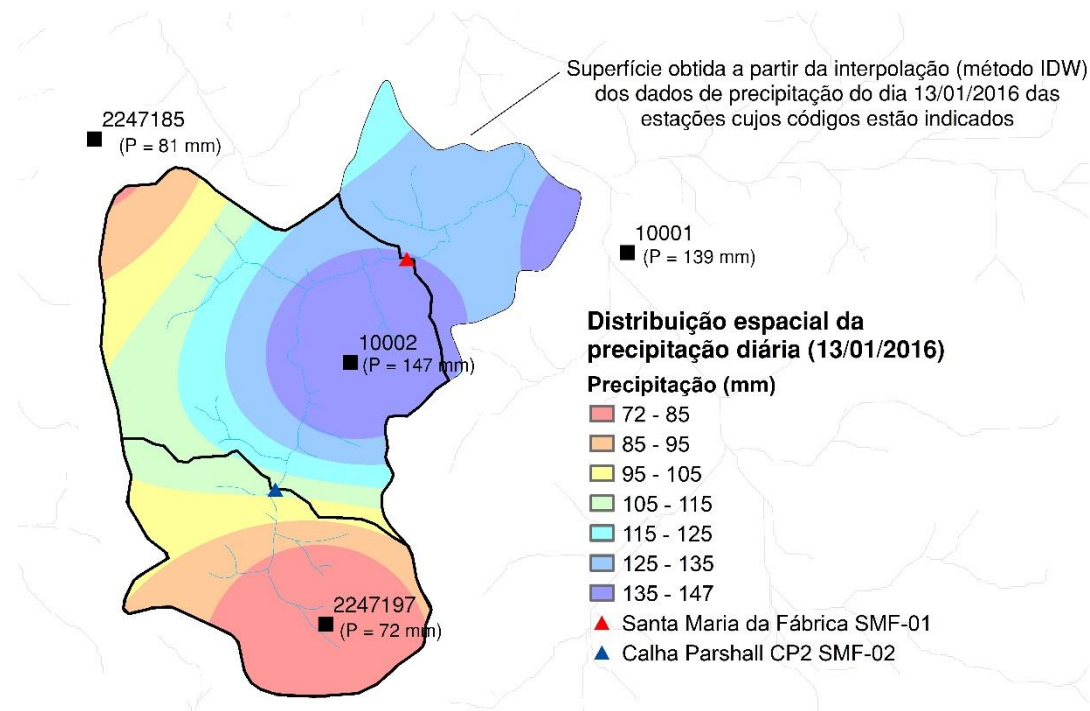
O modelo calibrado e validado foi em seguida avaliado pelas métricas de conectividade e causalidade (informação mútua com defasagem e teste de Granger). Essa avaliação adicional teve como objetivo verificar a capacidade do modelo em reproduzir as características do sistema hidrológico. As análises foram feitas em termos das relações entre precipitação ( $P$ ) e escoamento total ( $Q$ ), escoamento de base ( $Q_b$ ) e variação do armazenamento subterrâneo ( $\Delta S_s$ ), entre evapotranspiração de referência ( $ET_o$ ) e  $\Delta S_s$  e entre índice de vegetação  $EVI$  e  $\Delta S_s$ .

## 4 RESULTADOS E DISCUSSÃO

### 3.1 Precipitação média

A precipitação média horária ou diária na área de estudo foi estimada a partir da interpolação espacial de dados localmente disponíveis. Como exemplo, a distribuição espacial referente ao período das 7:00 h do dia 12/01/2016 às 7:00 h do dia 13/01/2016 de quatro estações pluviométricas é indicada na Figura 20. Pode-se perceber uma considerável variação entre estações para esse dia especificamente (mínimo = 72 mm e máximo = 147 mm), evidenciando as incertezas na estimativa da precipitação média da bacia.

Figura 20. Distribuição espacial da precipitação diária de 12/01/2016 a 13/01/2016 usando o método IDW e os dados de quatro estações pluviométricas locais. Códigos indicados fazem referência à Figura 5 (CRHEA: 10001; NMA: 10002; CAA: 2247185; FRV: 2247197).



Fonte: Autor.

A variabilidade da precipitação na área de estudo foi abordada indiretamente por Meira Neto et al. (2018), que verificaram a existência de uma considerável influência dos esquemas de interpolação e das fontes de dados de precipitação para aplicação em modelos de transformação chuva-vazão. Neste trabalho, essa questão foi tratada de forma determinística,

fazendo um amplo uso de dados medidos em superfície, na tentativa de identificar padrões espaciais de precipitação. Os resultados das análises de interpolação IDW e regressão envolvendo múltiplas combinações de fontes de dados são apresentados nas tabelas 8 a 11.

Na escala mensal, pode-se concluir (da Tabela 8) que as estações pluviométricas Campo Alegre (CAA) e Ferroviária (FRV) são as que melhor caracterizam a distribuição de precipitações. A estação pluviométrica do CRHEA apresentou o pior desempenho quando usada desacompanhada de outras estações, com um erro padrão de 33 mm/mês.

Na escala diária e na área de drenagem SM (Tabela 9), a estação CRHEA novamente apresentou o pior desempenho nas análises de regressão, com um erro padrão de 7,3 mm/dia. Pelas outras combinações, o erro na estimativa da precipitação média diária está em grande parte das vezes entre 3 e 6 mm/dia. Na área de drenagem CP2 (Tabela 10), pode-se perceber que o erro padrão é da ordem de 10 mm/dia em esquemas em que a estação CAA ou FRV não está presente.

A análise em escala horária (Tabela 11), feita apenas para a área CP2, mostra erros entre 1 e 3 mm/h. A estação STO, localizada a cerca de 3 km do centro da bacia, é a que possui importância relativa (peso) em análises horárias. Entretanto, na escala diária, ela é a que resulta em maior erro padrão nas análises de regressão. Essa contradição pode estar associada ao método adotado, que aproxima a distribuição correta de precipitações por uma superfície interpolada obtida a partir de um conjunto reduzido de pontos monitorados.

Tabela 8. Relações de regressão da precipitação média mensal obtidas a partir dos resultados de interpolação espacial no período 01/1988 a 04/2018 na área de drenagem SM.

ID	Estações	Coeficientes de regressão					$\sigma_r$	R <sup>2</sup>	N <sub>obs</sub>
		a <sub>0</sub> [cte]	a <sub>1</sub> [CRHEA]	a <sub>2</sub> [FRV]	a <sub>3</sub> [CAA]	a <sub>4</sub> [NMA]			
1	CRHEA, FRV, CAA		0,15	0,55	0,30		0,0	1,000	323
2	FRV, CAA	1,53		0,61	0,38		5,5	0,997	323
3	CRHEA, FRV	-0,43	0,31	0,67			11,4	0,989	323
4	CRHEA, FRV, NMA	-0,93	0,29	0,55		0,14	13,7	0,979	76
5	FRV	3,7		0,94			17,8	0,973	323
6	CRHEA, CAA	4,14	0,42		0,57		22,5	0,957	323
7	CRHEA, NMA	2,17	0,50			0,44	22,7	0,942	76
8	CAA	11,3			0,96		28,6	0,930	323
9	NMA	11,8				0,91	28,7	0,906	76
10	CRHEA	5,4	0,96				33,0	0,908	323

ID: identificação da regressão;  $\sigma_r$ : erro padrão da regressão múltipla [mm/mês]; R<sup>2</sup>: coeficiente de determinação; e N<sub>obs</sub>: número de superfícies interpoladas pelo método IDW utilizadas na regressão.

Fonte: Autor.



Tabela 9. Relações de regressão da precipitação média diária obtidas a partir dos resultados de interpolação espacial no período 08/2011 a 04/2018 na área de drenagem SM.

ID	Estações	Coeficientes de regressão					$\sigma_r$	R <sup>2</sup>	N <sub>obs</sub>	
		a <sub>0</sub> [cte]	a <sub>1</sub> [CRHEA]	a <sub>2</sub> [FRV]	a <sub>3</sub> [CAA]	a <sub>4</sub> [NMA]				a <sub>5</sub> [STO]
1	CRHEA, FRV, CAA, NMA		0,05	0,16	0,36	0,43		1,00	1160	
2	CAA CRHEA, NMA,	0,08	0,27	0,30	0,39		3,30	0,91	1160	
3	STO	1,13	0,09			0,48	0,19	4,19	0,86	418
4	FRV, CAA	0,80		0,48	0,42			4,40	0,84	1160
5	CAA, CRHEA	0,76	0,47		0,41			4,67	0,82	1160
6	NMA, STO	1,74				0,53	0,21	5,01	0,79	464
7	NMA	2,34				0,70		5,23	0,78	1160
8	FRV, CRHEA	2,08	0,35	0,34				6,28	0,68	1160
9	CRHEA	2,95	0,57					7,30	0,56	1160

$\sigma_r$ : erro padrão da regressão múltipla [mm/dia].

Fonte: Autor.

Tabela 10. Relações de regressão da precipitação média diária obtidas a partir dos resultados de interpolação espacial no período 08/2011 a 04/2018 na área de drenagem CP2.

ID	Estações	Coeficientes de regressão					$\sigma_r$	R <sup>2</sup>	N <sub>obs</sub>	
		a <sub>0</sub> [cte]	a <sub>1</sub> [CRHEA]	a <sub>2</sub> [FRV]	a <sub>3</sub> [CAA]	a <sub>4</sub> [NMA]				a <sub>5</sub> [STO]
1	CRHEA, FRV, CAA, NMA		0,03	0,05	0,75	0,17		1,00	1160	
2	CRHEA, FRV, CAA	0,03	0,12	0,10	0,76		1,26	0,99	1160	
3	CRHEA, CAA	0,26	0,18		0,77		1,69	0,98	1160	
4	FRV, CAA	0,34		0,18	0,78		1,80	0,98	1160	
5	CRHEA, NMA, STO	2,15	0,08			0,24	0,19	8,63	0,38	418
6	NMA, STO	2,19				0,30	0,21	8,70	0,38	409
7	NMA	4,12				0,46		10,30	0,28	1160
8	CRHEA, FRV	3,97	0,27	0,18				10,57	0,24	1160
9	CRHEA	4,42	0,39					10,74	0,21	1160
10	STO	3,63					0,43	10,75	0,24	420

$\sigma_r$ : erro padrão da regressão múltipla [mm/dia].

Fonte: Autor.

Tabela 11. Relações de regressão da precipitação média horária obtidas a partir dos resultados de interpolação espacial no período 08/2011 a 12/2013 na área de drenagem CP2.

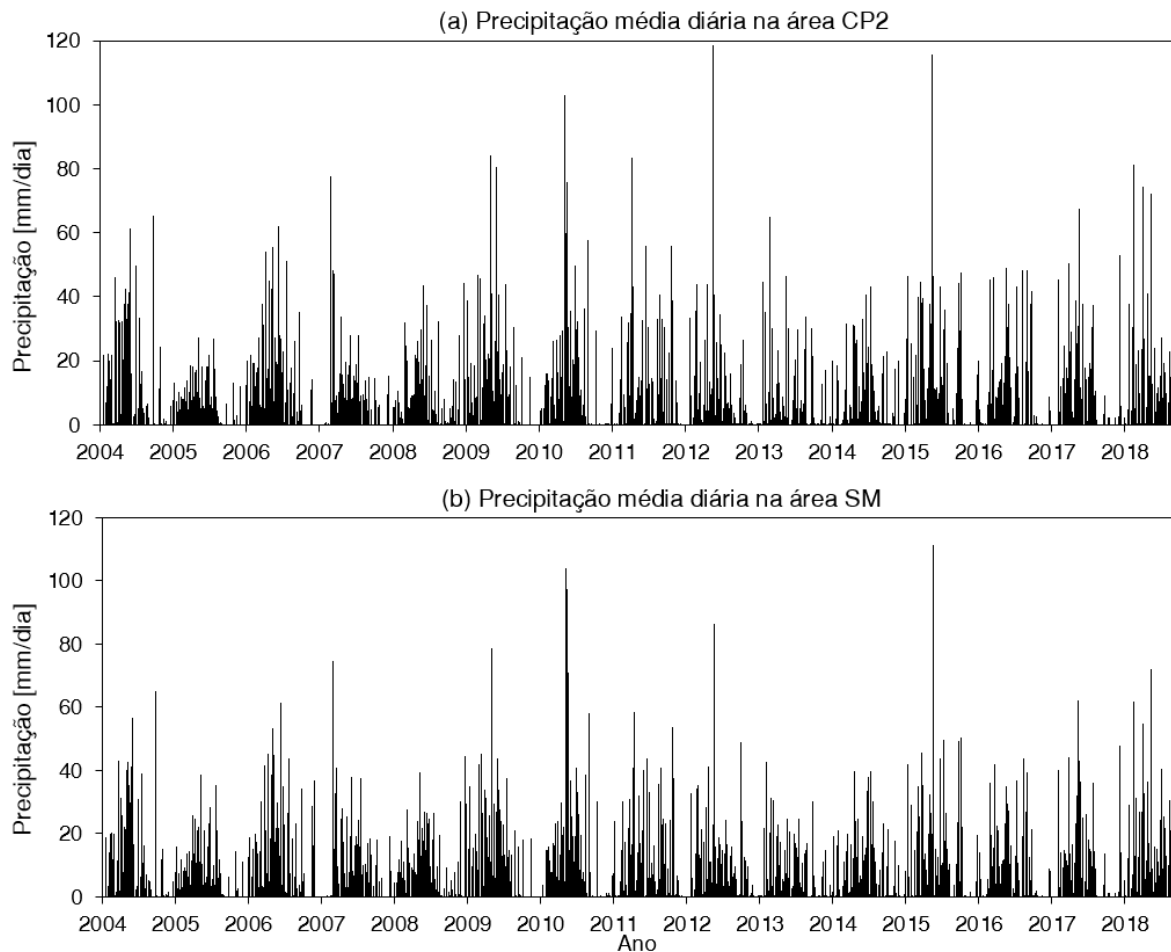
ID	Estações	Coeficientes de regressão				$\sigma_r$	R <sup>2</sup>	N <sub>obs</sub>
		a <sub>0</sub> [cte]	a <sub>1</sub> [SMF]	a <sub>2</sub> [NMA]	a <sub>3</sub> [STO]			
1	SMF, NMA, STO		0,079	0,261	0,660	0,00	1,00	1103
2	SMF, STO	0,108	0,243		0,720	0,68	0,97	1103
3	STO	0,438			0,838	1,13	0,91	1103
4	SMF, NMA	0,770	0,200	0,553		2,40	0,62	1103
5	NMA	0,898		0,706		2,47	0,60	1103
6	SMF	1,184	0,613			2,84	0,46	1103

SMF: estação Santa Maria;  $\sigma_r$ : erro padrão da regressão múltipla [mm/hora].

Fonte: Autor.

Os esquemas de interpolação que resultaram nos menores erros-padrão foram definidos a cada passo de tempo, tendo em vista a restrição de disponibilidade de dados. Esse processo resultou nas séries de precipitação média diária apresentadas na Figura 21.

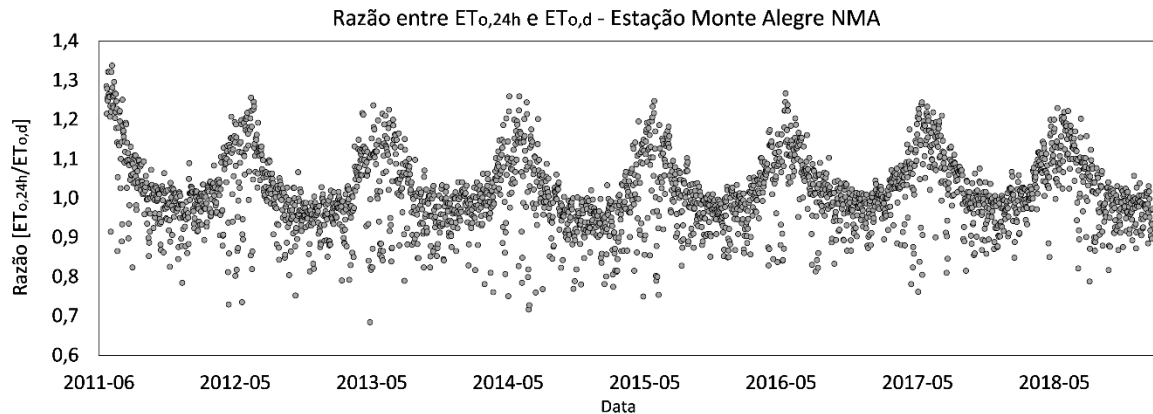
Figura 21. Séries de precipitação média sobre as duas bacias estudadas (CP2 e SM).



### 3.2 Evapotranspiração de referência

O método de cálculo de evapotranspiração de referência Penman-Monteith FAO-56 com passo de tempo diário foi utilizado para a estação CRHEA convencional, enquanto o método de cálculo horário foi adotado para as estações automáticas. Uma breve comparação entre os dois métodos é apresentada na Figura 22. O comportamento observado é similar ao obtido em trabalhos relacionados (IRMAK et al., 2005; SULEIMAN; HOOGENBOOM, 2009) e está de acordo com o esperado para áreas com clima subtropical úmido (ALLEN et al., 1998).

Figura 22. Comparação entre os métodos de cálculo horário e diário para determinação da evapotranspiração de referência. Dados da estação Monte Alegre tomados como exemplo.



Fonte: Autor.

Os valores médios de evapotranspiração de referência para cada estação são apresentados na Tabela 12 e as séries temporais, na Figura 23. Devido ao curto intervalo de operação e a anomalia observada na componente radiativa, a estação automática do CRHEA não foi considerada nas análises subsequentes, sendo utilizada apenas para fins de comparação.

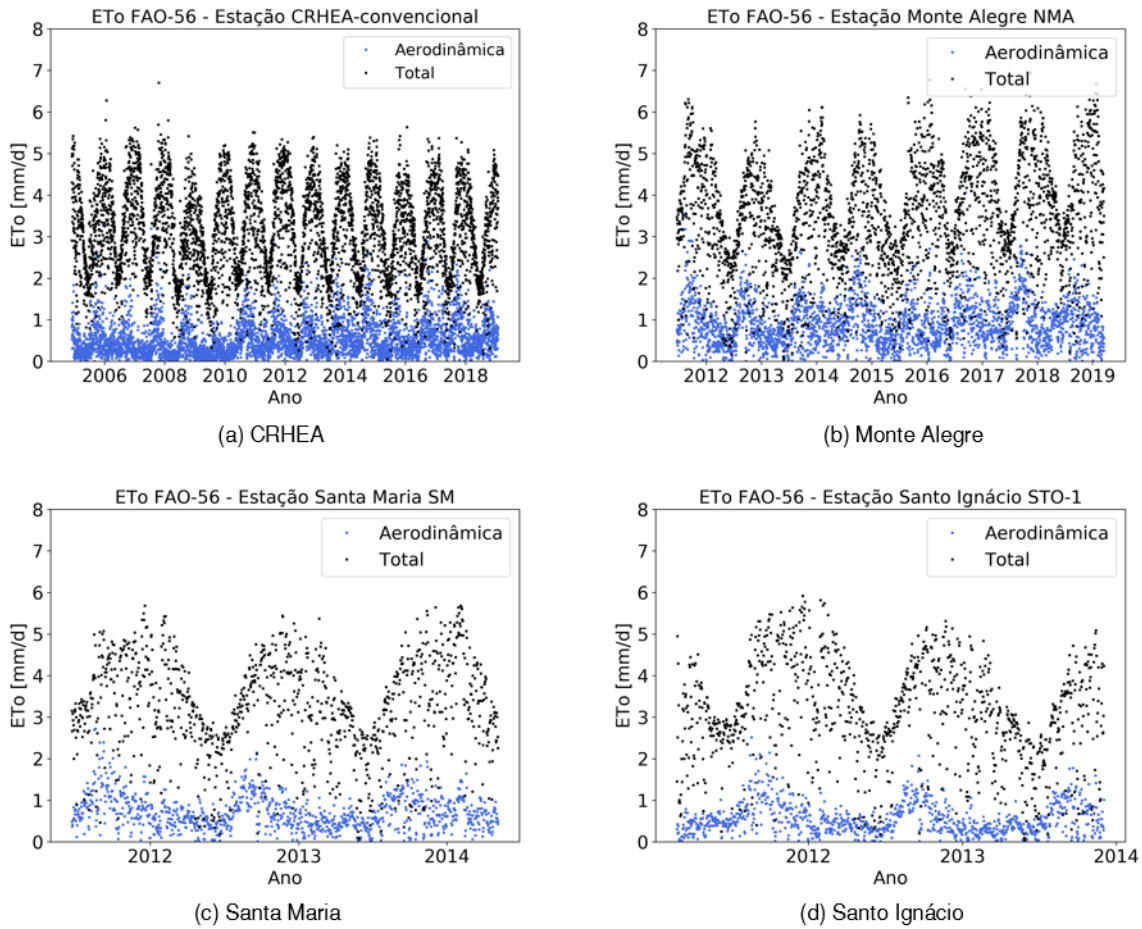
De forma geral, os resultados estão compatíveis com outros estudos: 3,3 mm/dia com dados da estação CHREA convencional no período de 2000 a 2015 (CABRERA, 2016b). As diferenças, da ordem de 0,5 mm, são controladas pela componente aerodinâmica. Fatores topográficos e a rugosidade da superfície (e.g., altura média da vegetação na vizinhança de onde as estações estão instaladas) podem justificar essas diferenças.

Tabela 12. Resultados da evapotranspiração de referência para cada estação no período total de operação. Notação: média amostral (desvio padrão amostral).

Estação	Componente aerodinâmica [mm/dia]	Componente radiativa [mm/dia]	ET <sub>o</sub> [mm/dia]
CRHEA-Automática	0,71 (0,36)	2,85 (1,01)	3,59 (1,21)
CRHEA-Convencional	0,52 (0,41)	2,52 (0,95)	3,04 (1,09)
Monte Alegre NMA	0,89 (0,53)	2,51 (1,01)	3,46 (1,31)
Santa Maria SMF	0,70 (0,41)	2,46 (0,91)	3,22 (1,16)
Santo Ignácio STO-1	0,54 (0,38)	2,48 (1,01)	3,09 (1,21)
Santo Ignácio STO-2	0,33 (0,22)	2,44 (0,94)	2,88 (1,03)

Fonte: Autor.

Figura 23. Evapotranspiração de referência diária de algumas estações utilizadas neste estudo.

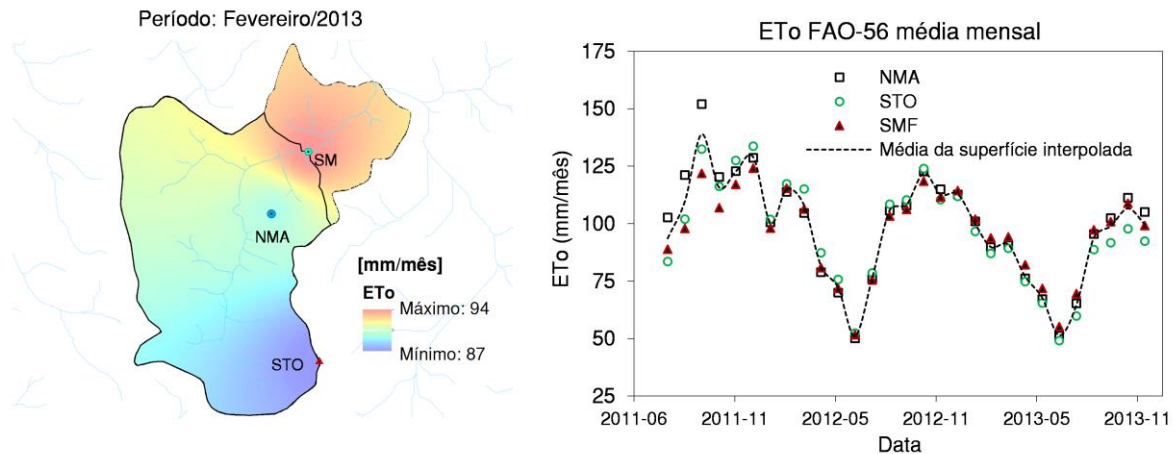


Fonte: Autor.

De forma similar à variável precipitação, esquemas de interpolação e regressão foram utilizados para estimar a série de evapotranspiração de referência média na área de estudo. Os resultados são apresentados nas tabelas 13 a 15.

As diferenças são pequenas, como exemplificado na Figura 24 (diferença entre estimativa máxima e mínima de 7 mm/mês). Tendo em vista a ordem de grandeza da incerteza de outras variáveis hidrológicas (e.g., precipitação, evapotranspiração real), esse processo de interpolação espacial pode não fazer sentido na escala da área de estudo ( $\sim 50 \text{ km}^2$ ), entretanto, contribui para reduzir os erros decorrentes da hipótese de uniformidade espacial da  $ET_o$  e evita o uso de dados meteorológicos fora dos padrões de variabilidade esperados.

Figura 24. Evapotranspiração de referência acumulada mensal nas estações e média da superfície resultante do processo de interpolação na área SM (NMA: Monte Alegre; Santo Ignácio STO; SM: Santa Maria).



Fonte: Autor.

Tabela 13. Relações de regressão da evapotranspiração de referência média mensal obtidas a partir dos resultados de interpolação espacial no período 07/2011 a 12/2013.

ID	Estações	Coeficientes de regressão					$\sigma_r$	R <sup>2</sup>	N <sub>obs</sub>
		a <sub>0</sub> [cte]	a <sub>1</sub> [NMA]	a <sub>2</sub> [SMF]	a <sub>3</sub> [STO]	a <sub>4</sub> [CRHEA]			
1	NMA, SMF, STO		0,47	0,24	0,30		0,00	1,000	29
2	NMA,STO,CRHEA	2,88	0,52		0,30	0,12	0,87	0,999	29
3	NMA, STO	4,09	0,54		0,41		1,17	0,997	29
4	NMA, SMF	-2,65	0,55	0,47			1,85	0,993	29
5	NMA, CRHEA	2,10	0,61			0,30	1,84	0,993	29
6	NMA	7,14	0,91				3,40	0,977	29
7	STO	6,16			0,95		4,61	0,958	29
8	CRHEA	1,33				0,84	5,77	0,934	29

$\sigma_r$ : erro padrão da regressão múltipla [mm/mês].

Fonte: Autor.

Tabela 14. Relações de regressão da evapotranspiração de referência média diária obtidas a partir dos resultados de interpolação espacial no período 06/2011 a 12/2013.

ID	Estações	Coeficientes de regressão					$\sigma_r$	R <sup>2</sup>	N <sub>obs</sub>
		a <sub>0</sub> [cte]	a <sub>1</sub> [NMA]	a <sub>2</sub> [SMF]	a <sub>3</sub> [STO]	a <sub>4</sub> [CRHEA]			
1	NMA, SMF, STO		0,466	0,239	0,295			1,000	895
2	NMA,STO,CRHEA	0,051	0,565		0,388	0,029	0,058	0,998	895
3	NMA, STO	0,065	0,569		0,408		0,060	0,998	895
4	NMA, SMF	-0,025	0,584	0,420			0,085	0,995	895
5	NMA, CRHEA	0,039	0,849			0,128	0,134	0,988	895
6	NMA	0,111	0,950				0,151	0,985	895
7	STO	0,138			0,971		0,208	0,971	895
8	CRHEA	0,206				0,959	0,536	0,807	895

$\sigma_r$ : erro padrão da regressão múltipla [mm/dia].

Fonte: Autor.

Tabela 15. Relações de regressão da evapotranspiração de referência média horária obtidas a partir dos resultados de interpolação espacial no período 06/2011 a 09/2012.

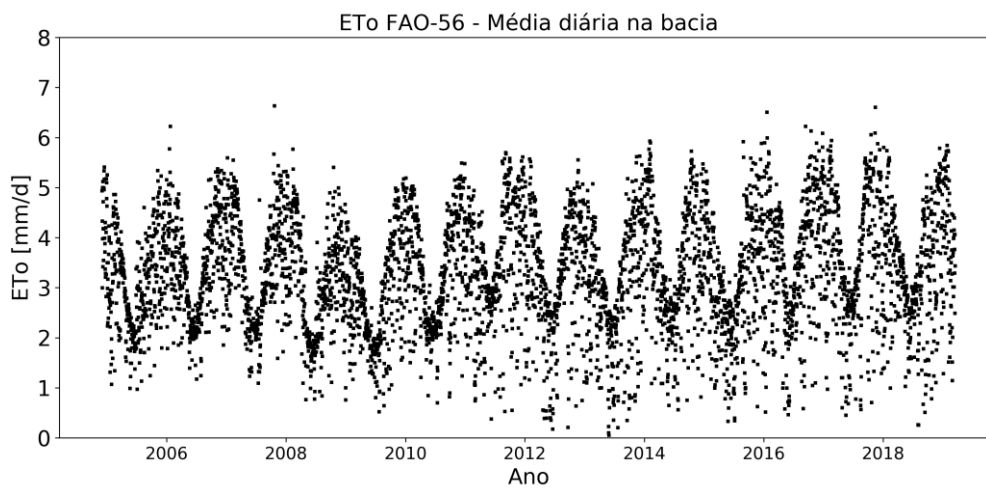
ID	Estações	Coeficientes de regressão				$\sigma_r$	R <sup>2</sup>	N <sub>obs</sub>
		a <sub>0</sub> [cte]	a <sub>1</sub> [NMA]	a <sub>2</sub> [SMF]	a <sub>3</sub> [STO]			
1	NMA, SMF, STO		0,503	0,203	0,294		1,000	7047
2	NMA, STO		0,597		0,390	0,007	0,999	7047
3	NMA, SMF		0,603	0,405		0,013	0,996	7047
4	SMF, STO	0,004		0,541	0,464	0,022	0,988	7047
5	NMA	0,001	0,979			0,022	0,988	7047
6	SMF	0,005		1,023		0,031	0,977	7047
7	STO	0,008			0,959	0,033	0,974	7047

$\sigma_r$ : erro padrão da regressão múltipla [mm/hora].

Fonte: Autor.

A série de evapotranspiração de referência média diária resultante dos esquemas de interpolação com menores erros-padrão é apresentada na Figura 25. Os valores foram calculados para a área SM e foram assumidos como representativos também da área CP2. A diferença de comportamento antes e depois de 2011 é justificada pelas diferentes fontes de dados utilizados. Antes de 2011, apenas dados da estação CRHEA-convencional estavam disponíveis, e após, dados de estações automáticas começaram a fazer parte do conjunto de dados disponíveis para estimativa da  $ET_o$ .

Figura 25. Série de evapotranspiração de referência média diária na área de estudo.

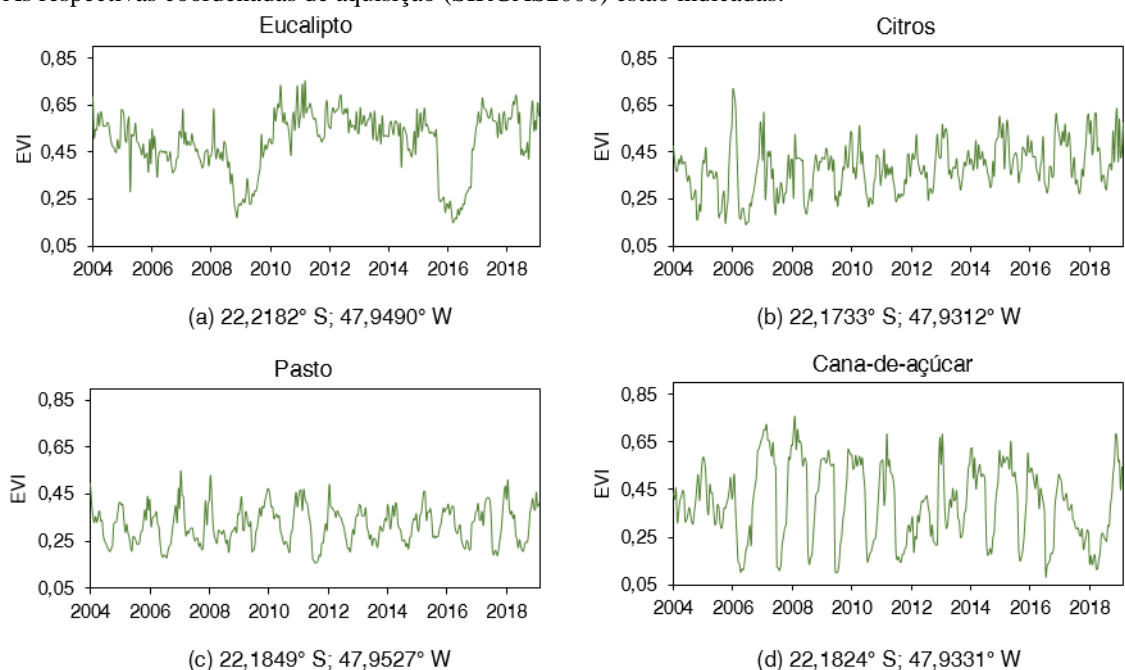


Fonte: Autor.

### 3.3 Índices de vegetação e estimativa da evapotranspiração real

As séries temporais de alguns pontos representativos das coberturas vegetais da bacia são apresentadas na Figura 26 para o período de 12/2004 a 02/2019. Pode-se perceber a possibilidade de uso do índice na detecção de mudanças na cobertura do solo e para acompanhar as variações sazonais que ocorrem na vegetação.

Figura 26. Séries temporais de EVI, entre o ano de 2004 e 2019, em pontos com coberturas representativas da área de estudo. As respectivas coordenadas de aquisição (SIRGAS2000) estão indicadas.



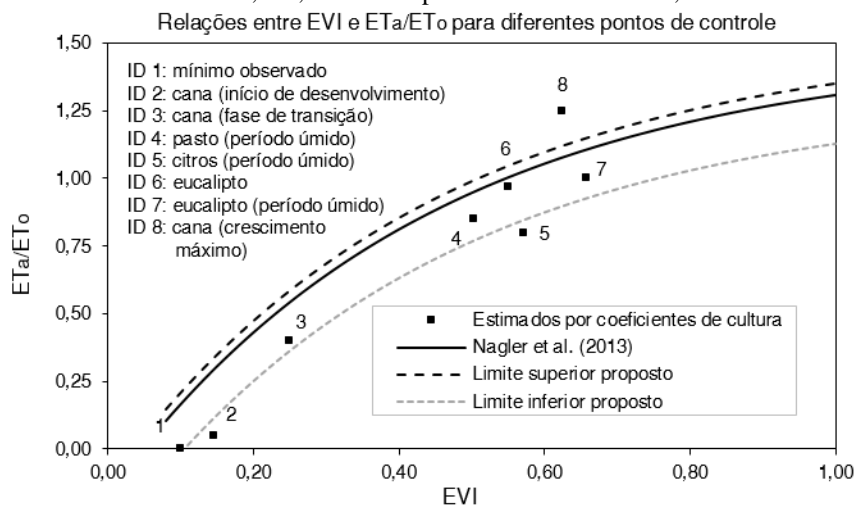
Fonte: Autor.

Nos meses mais chuvosos da região de estudo (de dezembro a fevereiro), os índices de vegetação atingem valores máximos. Assim, é razoável esperar que as relações entre  $ET_a$  e  $ET_o$  nesse período estejam próximas aos valores dos coeficientes de cultura ( $K_c$ ) tabelados. Por exemplo, nas áreas de citros, com ampla presença de plantas daninhas, o valor indicado por Allen et al. (1998) foi adotado para ser relacionado com a média dos máximos valores de EVI.

Na situação contrária, nos meses de julho e agosto, a indisponibilidade de água no solo pode limitar consideravelmente a evapotranspiração. Anache et al. (2019), utilizando o método do balanço hídrico do solo em uma área experimental localizada a 10 km da bacia de estudo, identificaram que tanto a cana-de-açúcar quanto áreas de pastagem apresentam transpiração praticamente nula nos períodos mais secos do ano.

Para áreas de eucalipto, o trabalho de Cabral et al. (2010) desenvolvido a cerca de 80 km da área de estudo, usando torre de fluxo (método *eddy-covariance*), foi tomado como referência. Nesse trabalho, a evapotranspiração média ao longo de três anos (entre 03/2006 e 03/2008) foi estimada em 1179 mm. No mesmo período, a estação do CRHEA (indicada na Figura 3), localizada a 70 km do local do experimento, registrou uma evapotranspiração de referência (FAO-56) de 1220 mm anuais. Assim, uma relação entre  $ET_a/ET_o$  de aproximadamente 1179/1220 é esperada para o EVI médio do pixel centrado na torre de fluxo ao longo do período do estudo. Com base nesses critérios, os valores indicados na Figura 27 foram determinados. Na Figura 27, os pontos obtidos em referência às diferentes coberturas da área da bacia foram plotados em conjunto com três curvas de referência. A curva central reproduz os resultados de Nagler et al. (2013), que adota um coeficiente  $c = 0,169$  na equação 20. As curvas inferior e superior delimitam uma faixa de variação esperada para a relação.

Figura 27. Relações estimadas entre EVI e coeficiente de cultura em diferentes condições. Limite inferior obtido com  $c = 0,320$ , e limite superior obtido com  $c = 0,125$ .



Fonte: Autor.

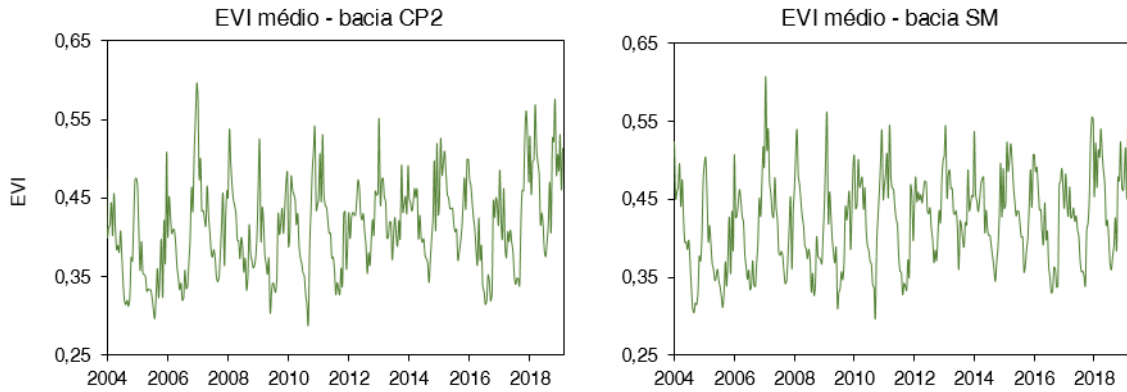
A curva do limite superior proposto não incluiu a cana-de-açúcar, em seu crescimento máximo, porque seu cultivo está presente em apenas uma pequena parcela da área de bacia (< 1%). Caso contrário, o comportamento singular e pouco representativo da cana-de-açúcar prejudicaria de forma significativa a relação para as demais coberturas vegetais.

As análises das etapas seguintes deste trabalho foram feitas a partir das séries de valores médios de EVI, que foram calculados para as áreas CP2 e SM (Figura 28). Considerando os critérios anteriormente apresentados, os valores de EVI médio e os correspondentes  $K_{VI}$  foram calculados e são apresentados na Figura 29. O padrão observado reflete as alterações estruturais



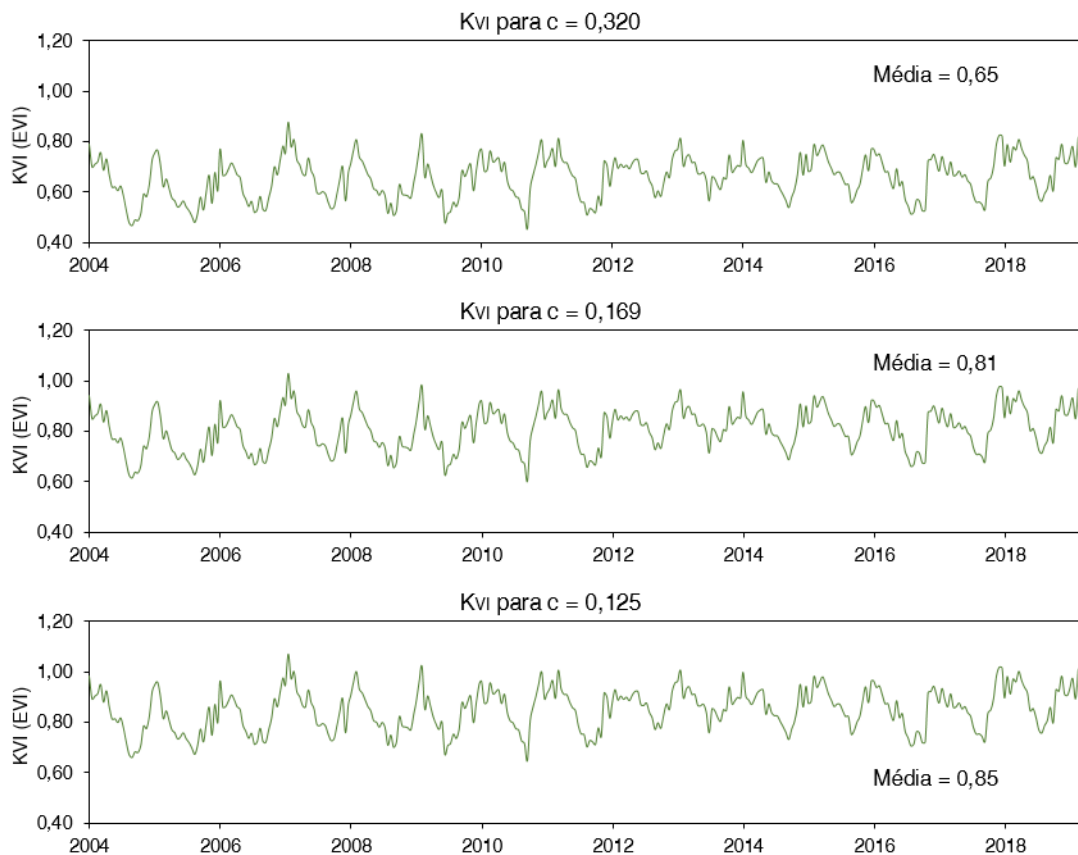
azonais da vegetação e as mudanças de uso do solo. Uma possível não estacionariedade na série é evidenciada pelo aumento de valores mínimos.

Figura 28. Séries de valores médios de EVI (a cada 16 dias) na área de estudo.



Fonte: Autor.

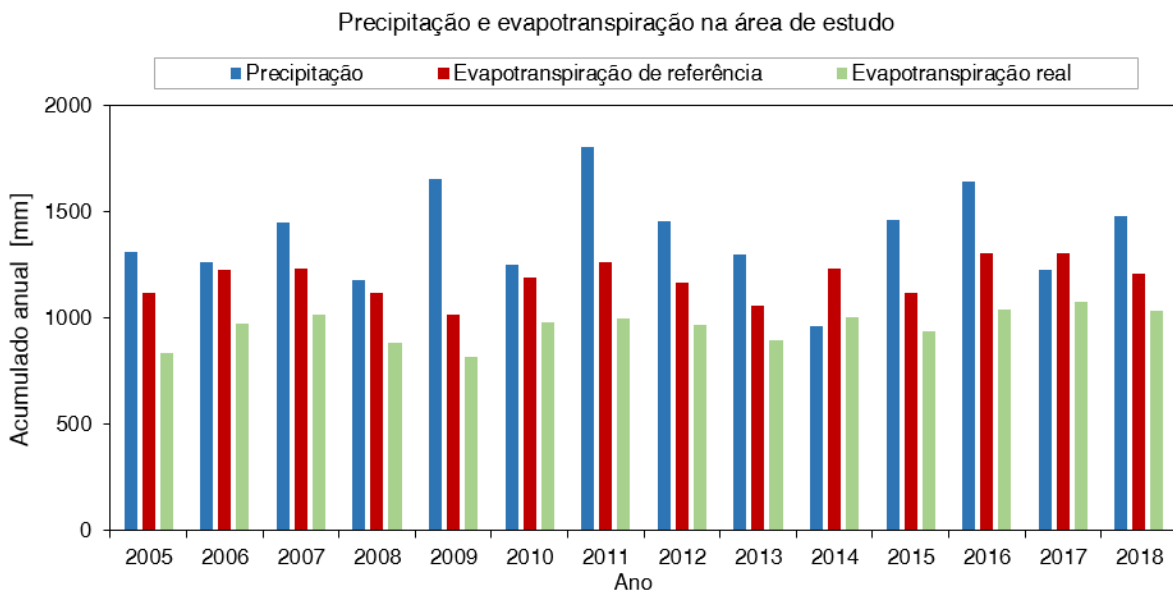
Figura 29. Séries de coeficiente  $K_{VI} = ETa/ETo$  (a cada 16 dias) para a área SM. Gráficos apresentam resultados de diferentes valores do parâmetro 'c'.



Fonte: Autor.

Uma ideia da proporção entre as variáveis precipitação, evapotranspiração de referência e evapotranspiração é indicada na Figura 30. Em termos anuais, a precipitação é em média 1390 mm, a evapotranspiração de referência, 1180 mm, e a evapotranspiração real, 975 mm. Portanto, a  $ET_a$  está entre 65% e 85% da  $ET_o$  e  $ET_a$  é da ordem de 55% a 72% da precipitação, tomando como referência o período de 2005 a 2018. Informações sobre os dados de escoamento superficial e do nível freático, abordados na sequência, possibilitam reduzir as incertezas sobre essas quantidades.

Figura 30. Totais anuais em mm na área SM: precipitação (média = 1390 mm), evapotranspiração de referência (média = 1180 mm) e de evapotranspiração estimada (média = 975 mm).



Fonte: Autor.

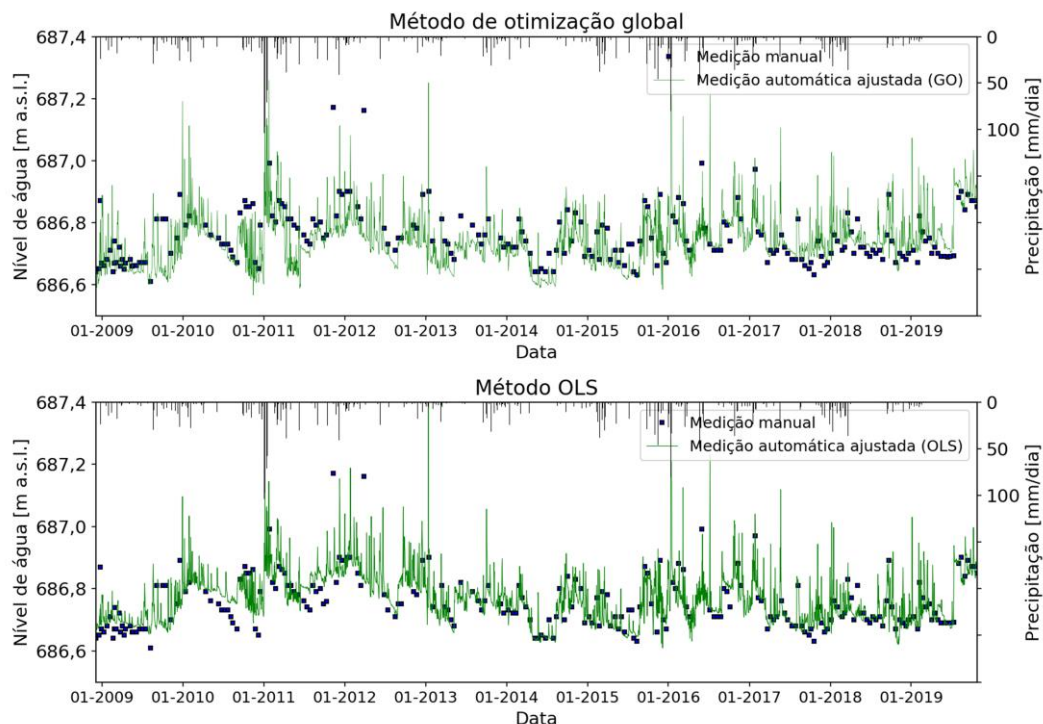
## 4.4 Estimativa das séries de vazão total e de vazão de base

### 4.4.1 Seção de monitoramento fluviométrico Santa Maria (SM)

A Tabela 16 apresenta as relações de regressão utilizadas para preenchimento de falhas dos períodos de monitoramento automático (09/2010 a 11/2010; 05/2011 a 06/2011; 11/2015 a 12/2015; 04/2016 a 06/2016). Para os meses de maio e julho, relações estatisticamente não significantes (valor  $p \geq 0,05$ ) entre precipitação e nível foram encontradas.

Os resultados dos ajustes pelo método de otimização global e pelo método de mínimos quadrados ordinários (OLS) entre dados manuais instantâneos e os dados automáticos observados ou preenchidos são apresentados na Figura 31. Pode-se perceber que o método de otimização global oferece um ajuste com desvios ligeiramente menores (e.g., ano de 2010) quando comparado ao método OLS. Os desvios podem estar associados a erros de medição manual, problemas com os sensores automáticos e com a conectividade hidráulica entre o poço de monitoramento, onde o sensor está instalado, e o curso d'água.

Figura 31. Métodos de ajuste por otimização global e por mínimos quadrados ordinários (OLS). Pontos indicam medição manual instantânea, e linhas, médias diárias de medições automáticas ajustadas a partir das medições manuais.



Fonte: Autor.

Tabela 16. Regressões lineares entre precipitação média diária P [mm] (&gt; 0) e nível d'água [m a.s.l.].

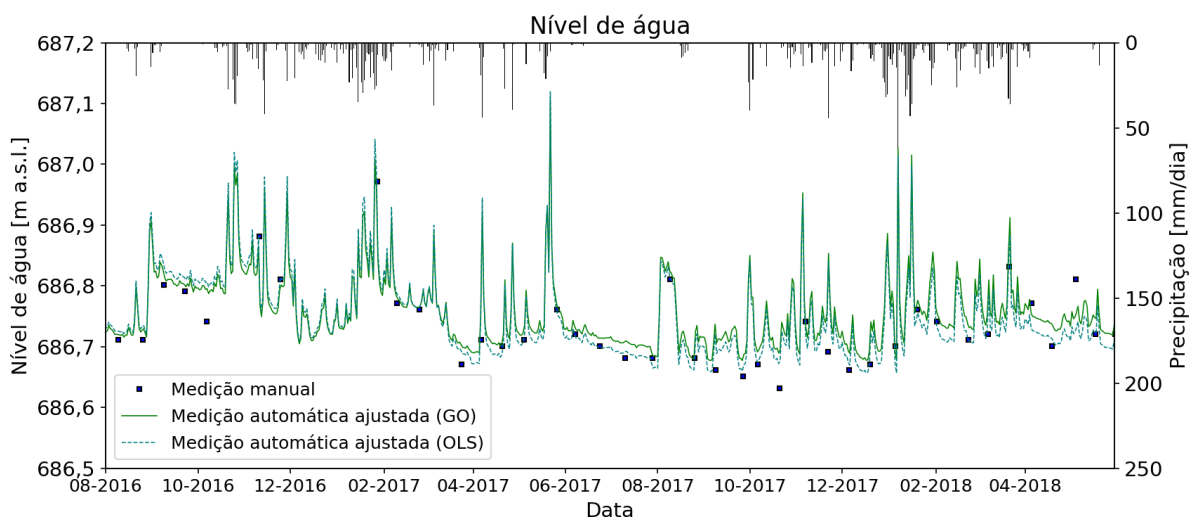
Mês	Nível médio	Desvio padrão	$N_{obs}$	$R^2$	Valor p	$b$	$a$	Erro padrão
Janeiro	686,741	0,091	172	0,572	0,00	686,728	0,0054	0,074
Fevereiro	686,736	0,066	130	0,257	0,00	686,725	0,0039	0,055
Março	686,727	0,055	143	0,402	0,00	686,716	0,0045	0,050
Abril	686,691	0,055	78	0,449	0,00	686,684	0,0045	0,049
Mai	686,683	0,063	56	0,067	0,05	686,721	0,0022	0,085
Junho	686,687	0,071	66	0,420	0,00	686,701	0,0052	0,056
Julho	686,679	0,049	50	0,002	0,79	686,682	0,0005	0,048
Agosto	686,697	0,054	30	0,257	0,00	686,721	0,0056	0,043
Setembro	686,761	0,057	52	0,247	0,00	686,772	0,0031	0,057
Outubro	686,768	0,050	90	0,380	0,00	686,761	0,0038	0,053
Novembro	686,761	0,046	95	0,331	0,00	686,761	0,0034	0,047
Dezembro	686,760	0,064	146	0,495	0,00	686,721	0,0043	0,046

Nível médio diário estimado:  $\hat{h} = aP + b$ .

Fonte: Autor.

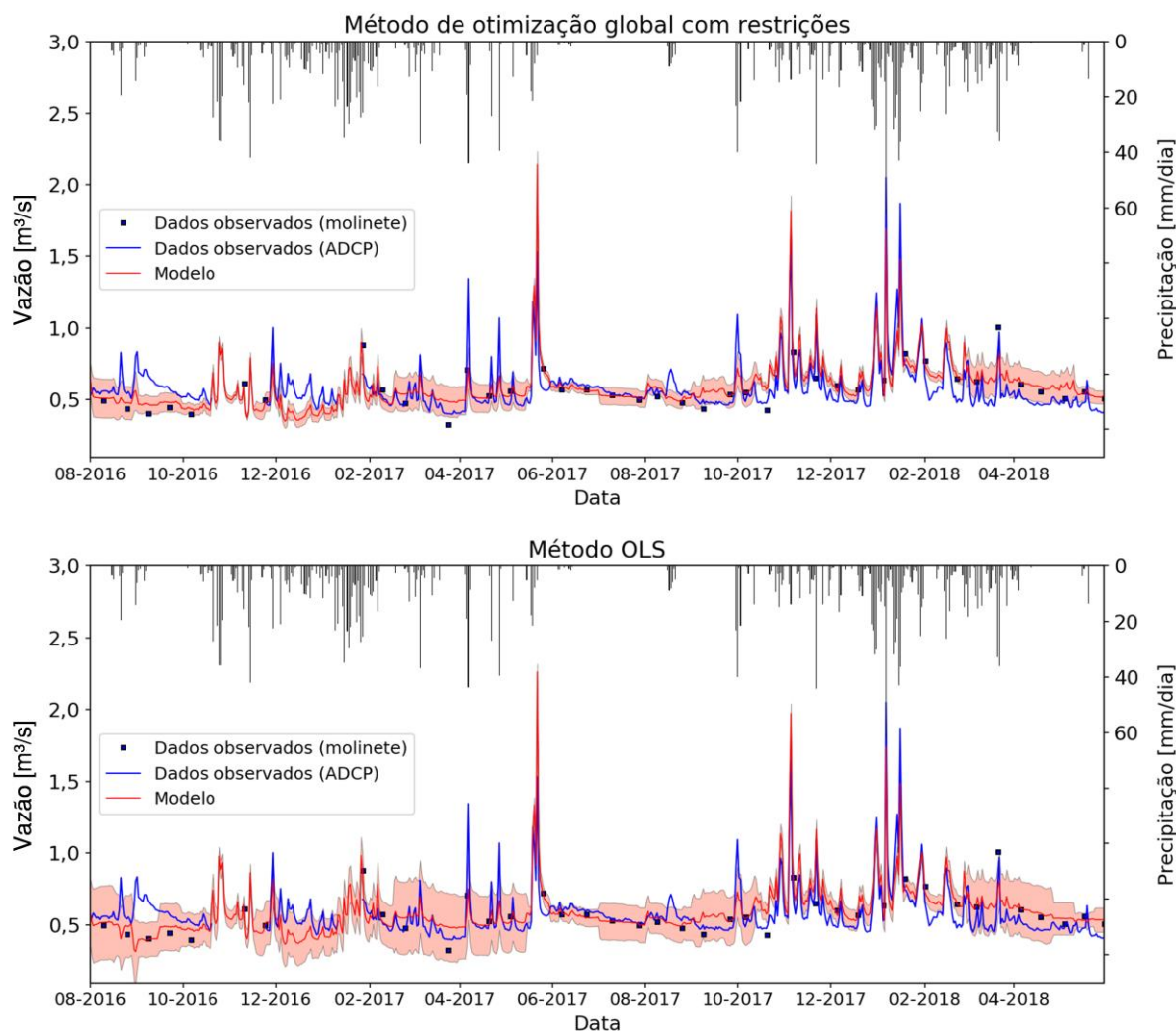
A etapa seguinte consistiu na avaliação da estratégia de determinação dinâmica dos parâmetros de transformação de nível d'água em vazão. Novamente, os métodos de otimização global e OLS foram comparados. Para o período comum com dados de vazão medidos pelo ADCP SonTek IQ-Plus e sem falhas (08/2016 a 05/2018), os dados de nível d'água ajustados são apresentados em detalhe na Figura 32. Os resultados das vazões estimadas pelos métodos com suas respectivas incertezas são apresentados na Figura 33. Métricas de erro e de desempenho são sintetizadas na Tabela 17 nas escalas horária e diária.

Figura 32. Nível de água ajustado entre 2016 e 2018, período em que o método de estimativa de vazão foi avaliado.



Fonte: Autor.

Figura 33. Resultados do método proposto de transformação de nível de água em vazão e dados observados pelo Sontek IQ-Plus (ADCP). Sombreamento indica o desvio do valor estimado pelo modelo por mais ou menos a raiz quadrada da variância.



Fonte: Autor.

Tabela 17. Métricas de avaliação do método proposto.

Métrica	Valor ótimo	Global	OLS	Global	OLS
		Dados horários		Dados diários	
RMSE [ $\text{m}^3/\text{s}$ ]	0	0,12	0,13	0,11	0,12
MAPE	0	15,8	15,8	15,2	15,2
PBIAS	0	-2,07	-2,11	-2,14	-2,22
NSE	1	0,48	0,44	0,55	0,50
KGE-NP	1	0,37	0,35	0,38	0,36

RMSE: raiz quadrada da soma dos erros quadráticos; MAPE: erro percentual absoluto médio; PBIAS: percentual de viés; NSE: coeficiente de eficiência de Nash-Sutcliffe; coeficiente de Kling-Gupta não paramétrico (POOL et al., 2018).

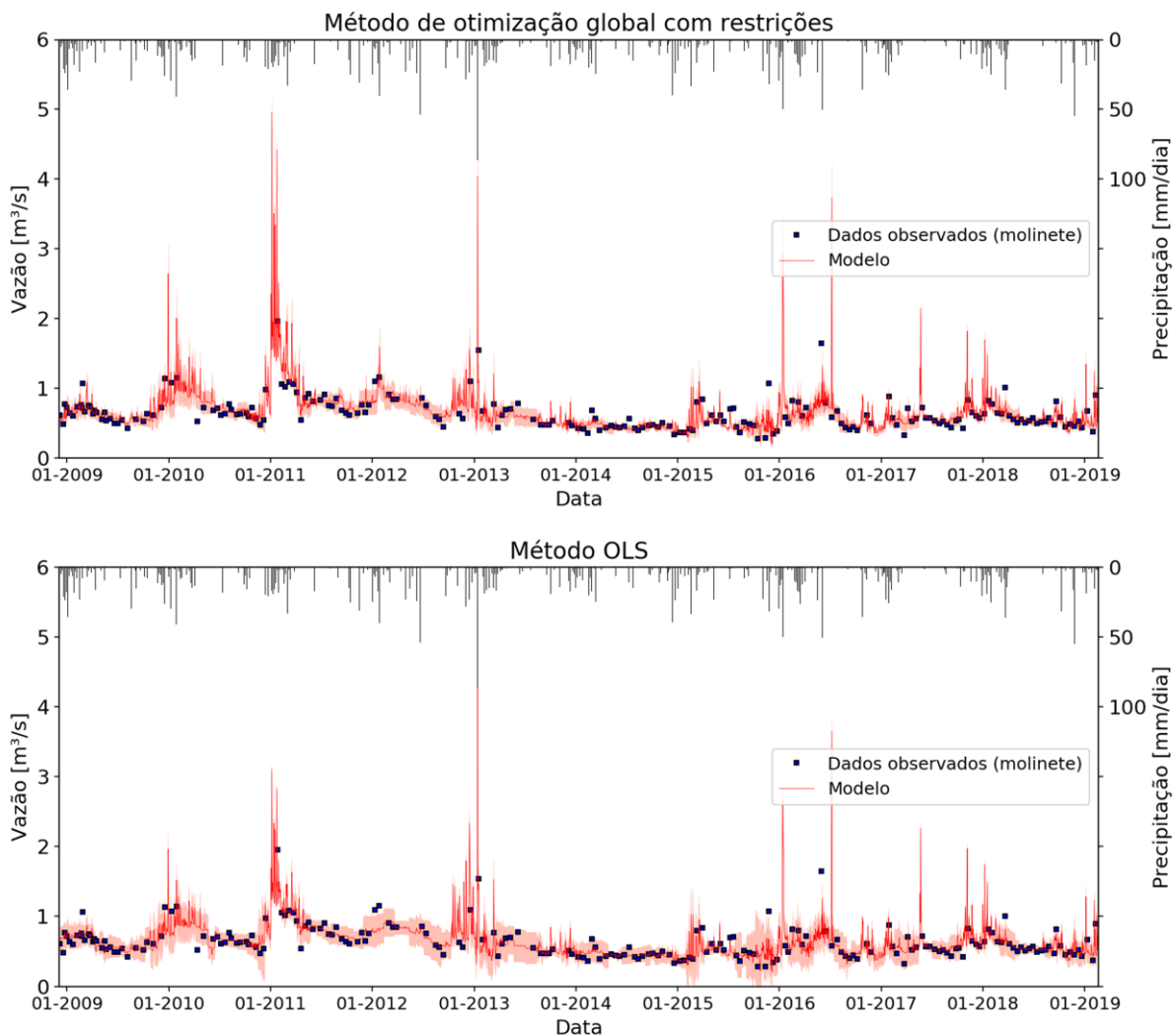
Fonte: Autor.

Para fins de avaliação gráfica geral, a Figura 34 apresenta os resultados do método proposto para o período completo de dados disponíveis. Em termos absolutos, os resultados são similares e a principal diferença está relacionada à variância das estimativas. O método OLS

leva a variâncias maiores porque, embora não restrinja os parâmetros a intervalos de variação, garante a normalidade e independência dos resíduos, produzindo resultados idênticos à estimativa de máxima verossimilhança (EMV). Do ponto de vista estatístico, as incertezas estimadas pelo método OLS são mais adequadas. Outra importante vantagem do método OLS quando comparado ao método de otimização global (baseado no *Differential Evolution*, aqui adotado) é o pequeno tempo de processamento requerido, centenas de vezes menor.

A série de referência utilizada nas etapas seguintes deste trabalho foi a determinada pelo método de otimização global devido ao ligeiro melhor desempenho apresentado (Tabela 17). Os dados do Sontek IQ Plus, quando disponíveis (a partir de 07/2015, com períodos de falha) substituíram as estimativas no período comum de monitoramento.

Figura 34. Resultados do método proposto de transformação de nível de água em vazão para o período completo de dados de nível de água disponível.

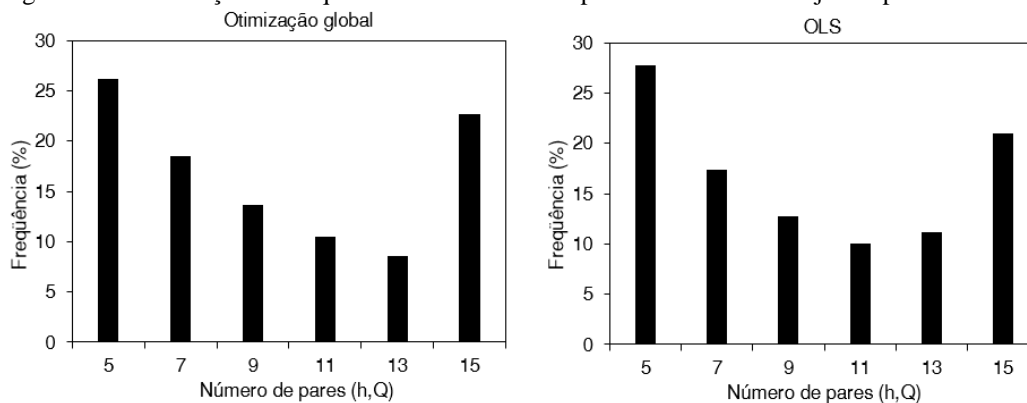


Fonte: Autor.

A seleção do número de pares ( $h, Q$ ) utilizados nos ajustes foi feita a partir do critério de maximização da função de verossimilhança e a distribuição resultante é apresentada na Figura 35. Este resultado indica que há uma tendência de usar ou o mínimo ou o máximo de pontos disponíveis. Isso parece estar coerente com as possíveis condições de não permanência para as quais os parâmetros das funções de transformação de nível em vazão são determinados: enquanto não há mudanças significativas, o número máximo de pontos é utilizado; por outro lado, quando uma mudança significativa é identificada, o número mínimo de pontos é utilizado.

Para avaliar isso de forma mais objetiva, a Tabela 18 apresenta a moda e a mediana (estatísticas de tendência central) dos números de pares selecionados de cada mês. Nos meses mais chuvosos (janeiro a março), longe do período em que há represamento a jusante da seção de monitoramento, há uma tendência de haver seleção do número máximo de pontos. No período mais seco, a situação inversa tende a ocorrer. Essas características evidenciam que o critério de máxima verossimilhança, aplicado em condições atípicas (espaço amostral variável), pode levar a resultados razoáveis.

Figura 35. Distribuição de frequência dos números de pontos adotados nos ajustes por cada método.



Fonte: Autor.

Tabela 18. Distribuição sazonal do número de pares utilizados para a determinação dos parâmetros.

Mês	OLS		Global	
	Moda	Mediana	Moda	Mediana
Janeiro	15	9	15	9
Fevereiro	7	7	5	7
Março	15	13	15	13
Abril	15	11	15	11
Mai	5	7	5	9
Junho	5	7	7	7
Julho	9	9	15	9
Agosto	13	9	5	9
Setembro	7	9	5	7
Outubro	5	7	5	7
Novembro	5	9	5	9
Dezembro	15	9	5	9

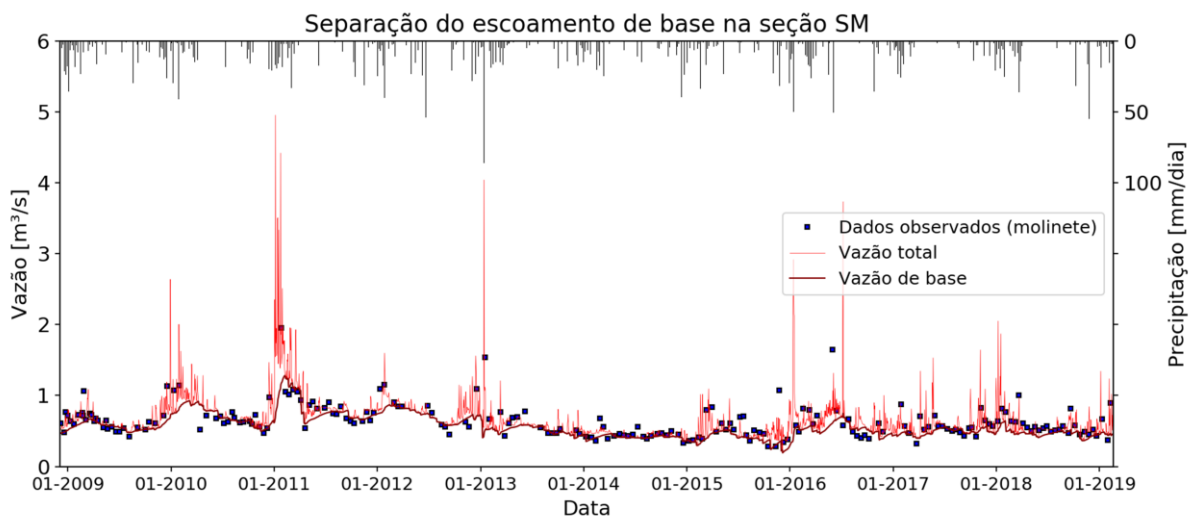
Fonte: Autor.

Para a separação do escoamento de base, o método de Duncan (2019) foi adotado com constantes  $k_1 = k_2 = 0,983$ . O resultado gráfico é apresentado na Figura 36.

O valor de  $BFI_{max}$  está entre os valores estimados por Gómez (2016) (0,93), que usou um filtro recursivo e os dados da área de estudo, e por Batista et al. (2018) (0,83), que fez a separação por balanço de massa com isótopos em cursos de água em condições hidrogeológicas similares e próximos à área de estudo.

As irregularidades observadas no gráfico são em geral produzidas pelas estimativas de vazão total, que, conforme anteriormente discutido, estão sujeitas a múltiplos erros. Com exceção disso, o método de separação parece estar razoável tomando como critério de avaliação os períodos de recessão e o índice de escoamento de base.

Figura 36. Separação do escoamento de base pelo método proposto por Duncan (2019).



Fonte: Autor.



#### 4.4.2 Seção de monitoramento fluviométrico calha Parshall 2 (CP2)

Devido a possíveis alterações nas condições de instalação do transdutor de pressão que permite o monitoramento automático do nível de água, três equações de ajuste (Tabela 19) foram utilizadas para relacionar a pressão compensada com o nível d'água na seção de montante da calha 2 ( $H_a$ ), que é quinzenalmente registrado de forma manual (régua).

Tabela 19. Ajustes para o cálculo da carga hidráulica de montante na calha CP2.

Calha	Período	K [mm]	$\sigma_r$ [m]
CP2	Antes de 05-01-2018	-20	10
	Entre 05-01-2018 e 17-05-2018	+56	18
	Após 17-05-2018	-86	16

Fonte: Autor.

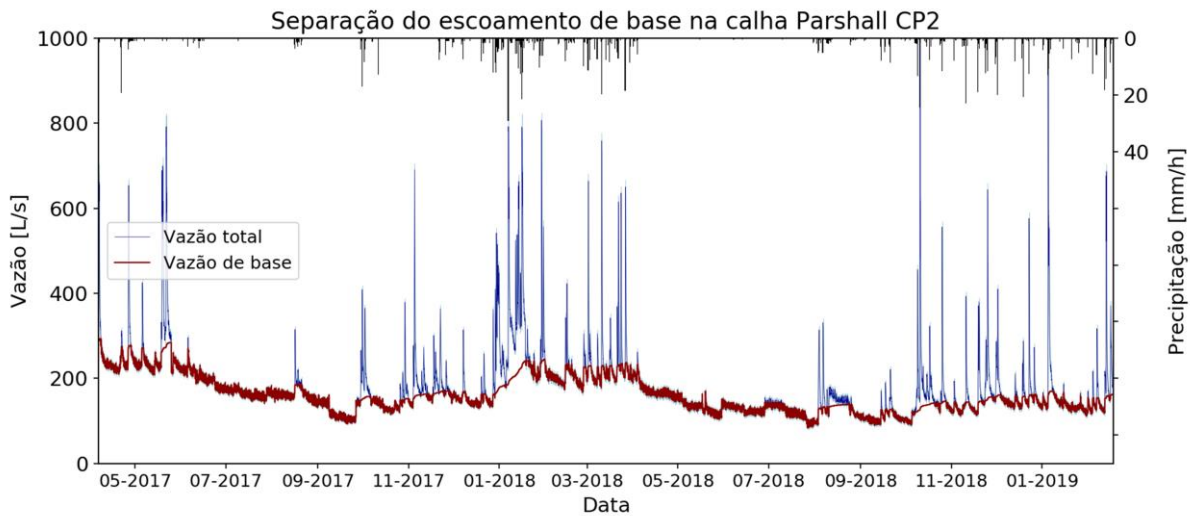
As vazões médias horárias são apresentadas na Figura 37. De forma geral, os dados são compatíveis com os dados da calha Parshall 1, mais a montante (dados não apresentados aqui), e apresentam como limitação a capacidade de medição da calha, que é da ordem de 700 L/s. Assim valores acima disso estão associados a elevadas incertezas.

A variabilidade espacial da precipitação na área de estudo fica evidenciada em períodos em que o monitoramento fluviométrico indica picos de vazão, enquanto a média espacial da precipitação indica valores praticamente nulos (e.g., entre maio e julho de 2017). A separação do escoamento de base, indicada na Figura 37, também foi feita pelo método de Duncan (2019) e usou os parâmetros  $k_1 = k_2 = 0,999$  (na escala horária; na escala diária,  $k^{24} \cong 0,976$ ). O índice de escoamento de base resultou em  $BFI_{\max} = 0,95$ .

Esse resultado em contraste com o obtido na seção SM (0,86) pode ser motivado pela limitação de medição de vazão na calha Parshall. Como a máxima vazão medida é da ordem de 700 L/s, vazões acima disso acabam sendo subestimadas, afetando a estimativa do índice  $BFI_{\max}$ . Por conta dessa subestimativa de vazões totais, certamente o índice de escoamento de base deve ser inferior a 0,95. Ao analisar a Figura 37, pode-se verificar que o escoamento de base estimado apresenta consistentemente uma resposta lenta, e que a mudança no parâmetro  $k$  pouco afetará a estimativa do índice de escoamento de base.

Uma ligeira superestimativa do índice de escoamento de base (devido a uma subestimação de vazões totais) também pode estar presente na seção SM. Isso porque para altas vazões (situação pouco frequente), parte do escoamento ocorre contornando os bueiros pelo terreno natural.

Figura 37. Estimativa de vazão total e escoamento de base na calha Parshall 2 (CP2).



Fonte: Autor.

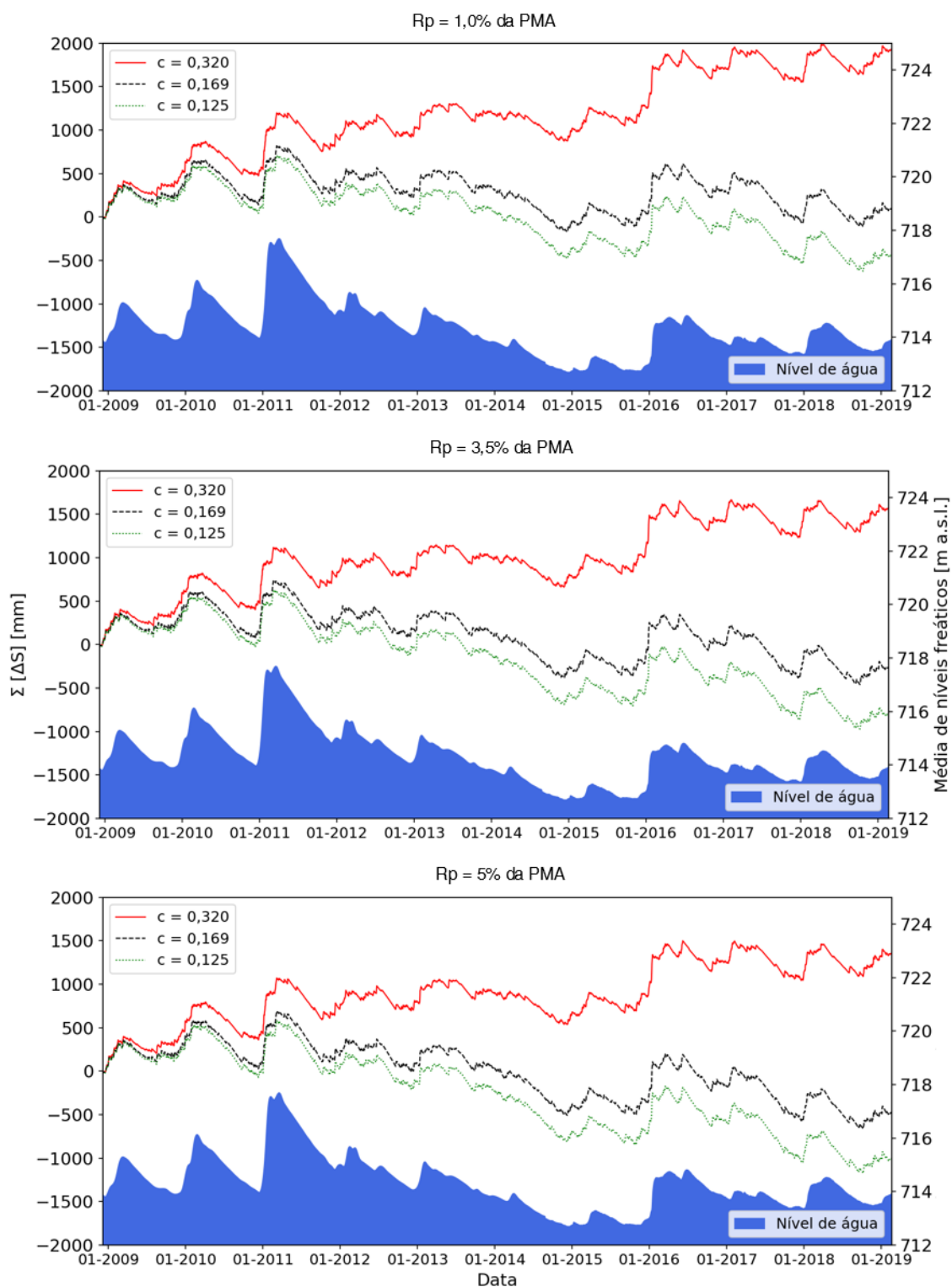
#### 4.5 Avaliação da consistência das estimativas dos processos hidrológicos

A variação das componentes do balanço hídrico em ciclos anuais deve corresponder à variação do armazenamento total na bacia. Como a umidade do solo na área de estudo segue ciclos anuais de variação aproximadamente nula (POMPEO, 1990), a variação do armazenamento total da bacia em ciclos anuais pode ser aproximada pela variação do armazenamento na zona saturada.

A partir dessa ideia e aproximando a variação do armazenamento na zona saturada pela variação da média dos níveis freáticos observados (dados dos poços de monitoramento 4, 5, 13, 14, 15, 16 e 19), algumas análises foram feitas com as estimativas dos processos hidrológicos, sobretudo com aqueles que envolvem as maiores incertezas: a evapotranspiração e a percolação profunda no Sistema Aquífero Guarani (SAG).

Os resultados dessas análises são apresentados na Figura 38, com variação da curva de transformação de índice de vegetação EVI em coeficiente de cultura  $K_{VI}$  (especificamente do coeficiente  $c$  da equação 20) e com variação da percolação profunda, tomando como referência o resultado estimado por Wendland, Barreto e Gomes (2007) de 3,5% da precipitação média anual (PMA).

Figura 38. Balanço hídrico ( $\Delta S = P - ET - Q - R_p$ ) correspondente à área SM para diferentes condições de evapotranspiração real baseadas na alteração do parâmetro  $c$  da equação (20). PMA: precipitação média anual.  $R_p$ : percolação profunda. Nível de água indicado é a média dos níveis freáticos observados.



Fonte: Autor.

Como não tem sentido o armazenamento estar aumentando (em contraposição ao comportamento do nível médio freático observado), certamente valores de  $c$  próximos a 0,300

não podem ser adotados, contrariando a curva inicialmente estimada a partir dos diferentes usos do solo na área de estudo. Isso pode indicar que há saídas não consideradas (e.g., fluxo lateral para bacias adjacentes, onde há plantações de eucalipto) que estejam sendo compensadas pela evapotranspiração. As altas taxas de evapotranspiração na zona ripária (CARVAJALINO, 2019) também podem colaborar nesse sentido de elevar a evapotranspiração média real da bacia.

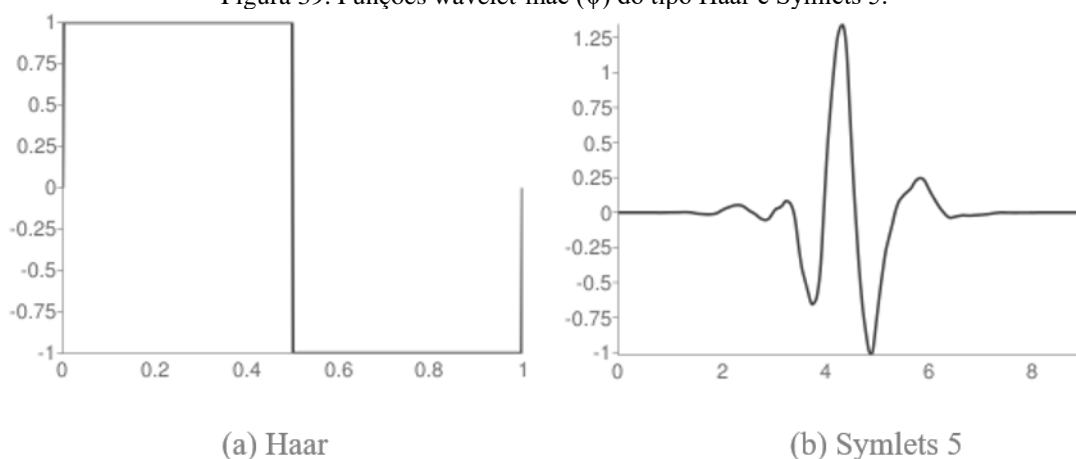
A combinação de  $R_p = 3,5\%$  e  $c = 0,169$ , por exemplo, parece estar adequada. Entretanto, a existência de dependência entre essas variáveis indica a importância de deixar tanto  $R_p$  quanto  $c$  como parâmetros livres do modelo a serem calibrados. A principal contribuição desta etapa foi, portanto, direcionar a definição do intervalo de variação do parâmetro  $c$  (certamente, maior que 0,125 e menor que 0,300) a partir de estimativas de  $R_p$  feitas em trabalhos anteriores.

#### **4.6 Análise de multirresolução de séries temporais**

A análise de multirresolução a partir de transformadas *wavelet* discreta é uma técnica que permite escrever uma função qualquer como a soma de uma função de aproximação de baixa frequência e termos de detalhe, de alta frequência. A Figura 39 apresenta as funções *wavelet*-mãe adotadas nas exemplificações de decomposição de séries temporais apresentadas nas figuras 40 e 41.

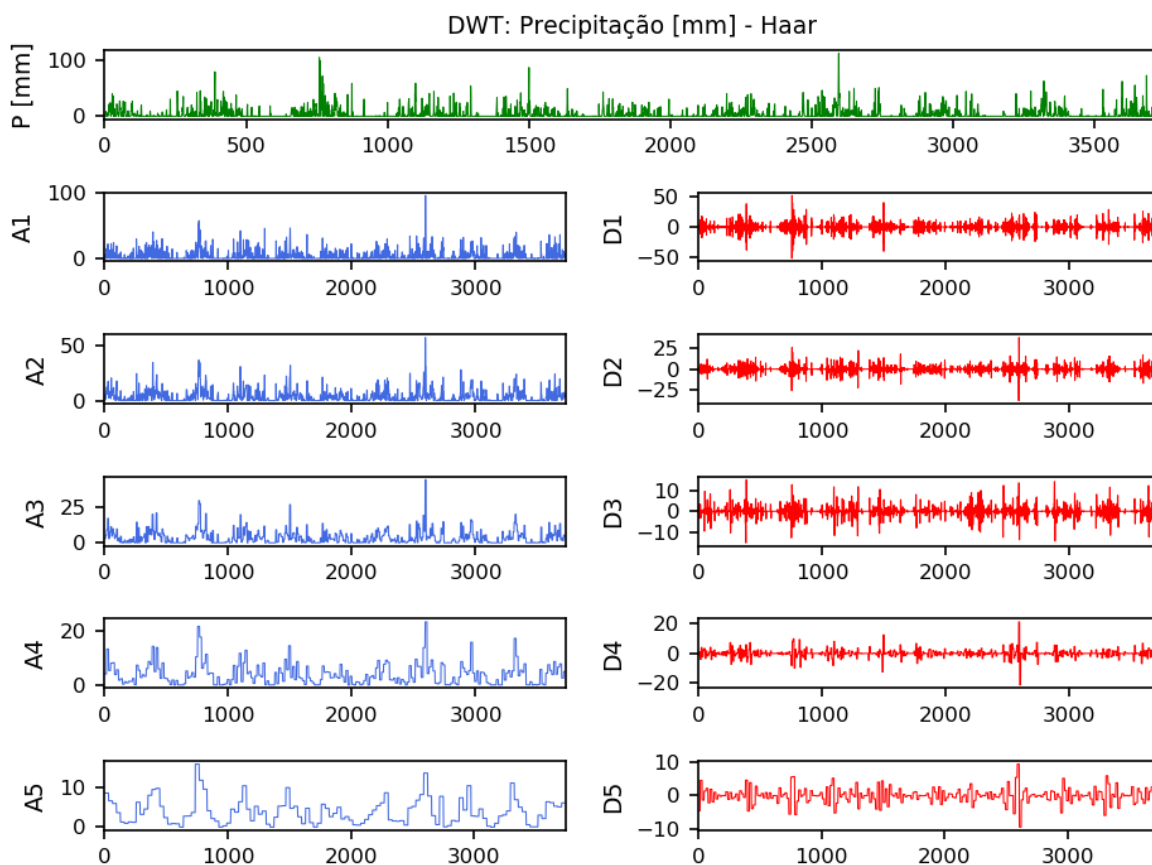
Devido à capacidade de remoção de ruídos, o uso dessa ferramenta foi considerado, a princípio, pela possibilidade de melhorar a identificação de algumas características do sistema hidrológico da área de estudo. Por outro lado, alguns trabalhos sugerem que operações de suavização podem prejudicar essa caracterização (BARNETT; SETH, 2015).

Essa questão foi avaliada de forma objetiva em problemas sintéticos apresentados na seção seguinte. Algumas variáveis foram modificadas de forma similar ao apresentado nas figuras 40 e 41. O termo ‘nível de decomposição’ adotado na seção 3.3.3, nas tabelas 4 e 5 e nos resultados da seção seguinte faz referência aos termos de aproximação (A1, ..., A5) apresentados em azul. Os termos de detalhe (D1, ..., D5) não foram utilizados em situação alguma.

Figura 39. Funções wavelet-mãe ( $\psi$ ) do tipo Haar e Symlets 5.

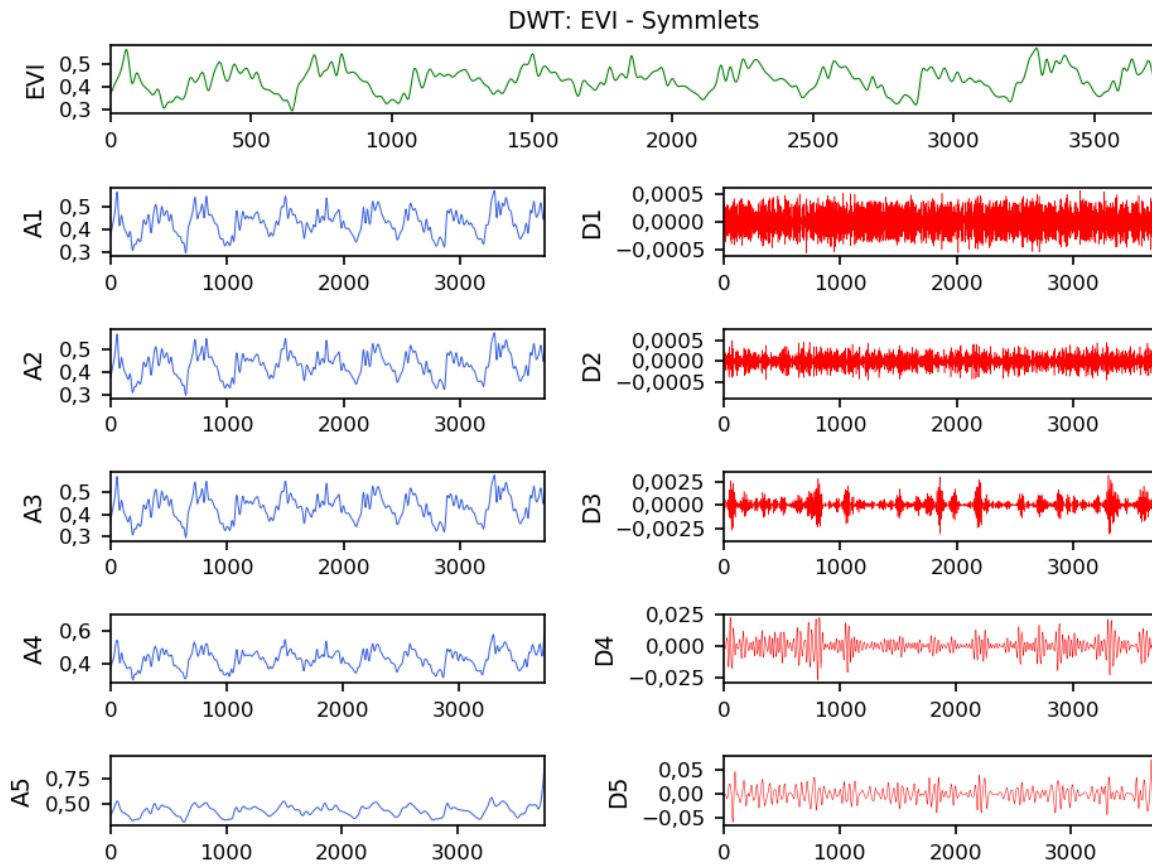
Fonte: <http://wavelets.pybytes.com/>. Acesso em 14-março-2020.

Figura 40. Decomposição da série de precipitação média diária na seção SM usando ‘Haar’ como *wavelet*-mãe e até o nível cinco. Eixo das abscissas corresponde ao número de dias após 05-12-2008. ‘A-j’ e ‘D-j’ indicam as aproximações (baixa frequência) e detalhes (alta frequência) no nível de decomposição ‘j’.



Fonte: Autor.

Figura 41. Decomposição da série de evapotranspiração real média na área da bacia usando ‘Symmlets 5’ como *wavelet*-mãe e até o nível cinco. Eixo das abscissas corresponde ao número de dias após 05-12-2008. ‘A-j’ e ‘D-j’ indicam as aproximações (baixa frequência) e detalhes (alta frequência) no nível de decomposição ‘j’.



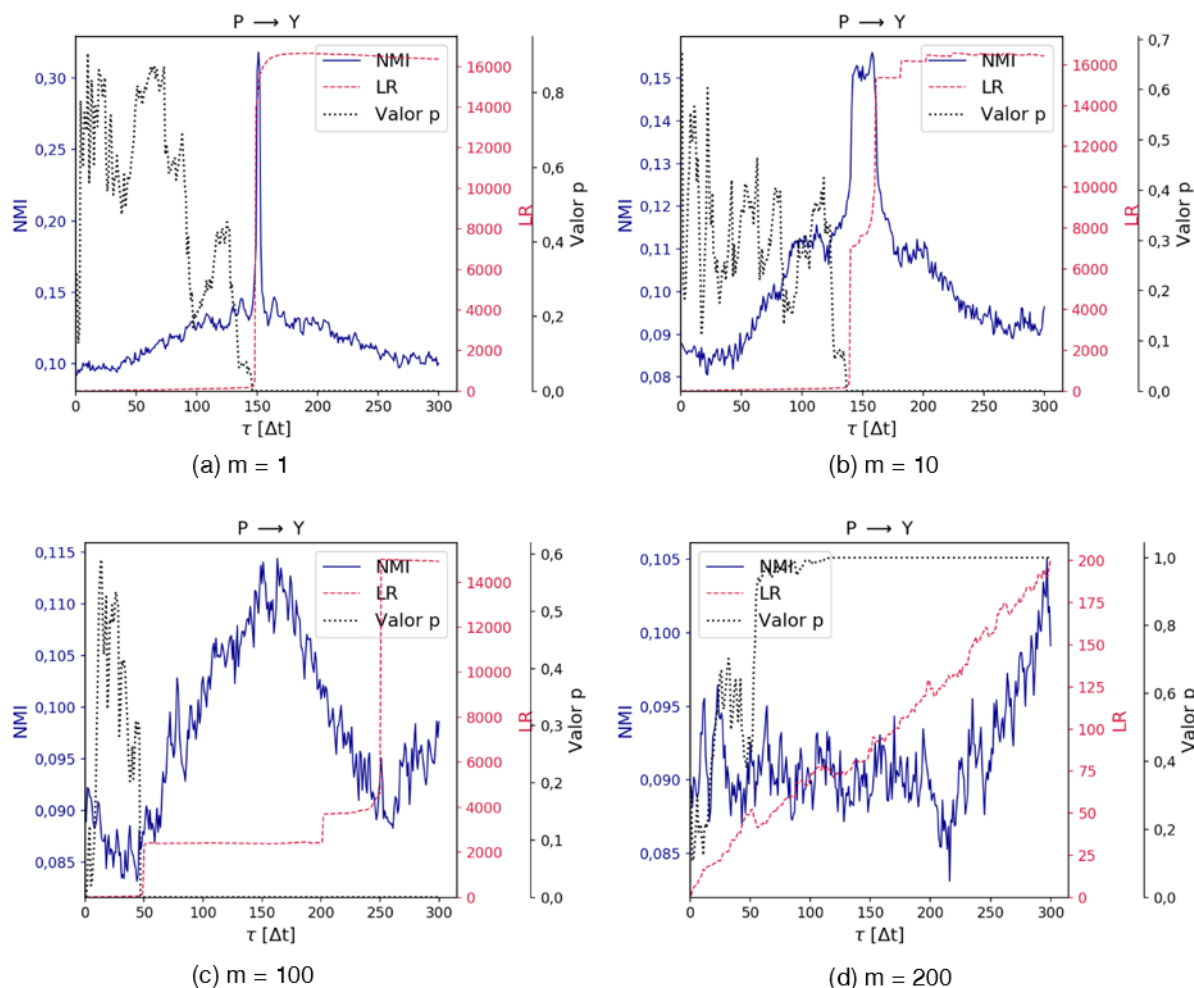
Fonte: Autor.

#### 4.7 Identificação de conectividades e dependências em problemas sintéticos

As ferramentas de detecção de causalidade entre processos foram aplicadas nos problemas sintéticos apresentados e detalhados nas tabelas 3, 4 e 5, e os resultados gráficos estão indicados nas figuras 42 a 47 e no Apêndice D. Cada gráfico apresenta três resultados: um referente à informação mútua normalizada (NMI); um referente à razão de verossimilhança (LR) do teste de Granger, e outro referente ao nível descritivo (valor-p) relacionado à LR (valor-p < 0,05 indica evidência estatística de existência de relação causal). Picos de NMI e aumentos de LR (acompanhados de valor-p  $\rightarrow$  0) indicam a defasagem (ou a defasagem mínima) entre processos que potencialmente apresentam relação de causa e efeito. A variável  $\tau$  indicada no eixo horizontal representa a defasagem avaliada na unidade de tempo  $\Delta t$  (=1 dias para a maioria dos casos). De forma geral, pode-se perceber que as ferramentas foram capazes de identificar sob uma diversidade de condições, incluindo não linearidades e elevados ruídos, os mecanismos

responsáveis pela geração das séries de dados. Além disso, pode-se concluir que a suavização e agregação temporal prejudicam a aplicação das ferramentas aqui testadas.

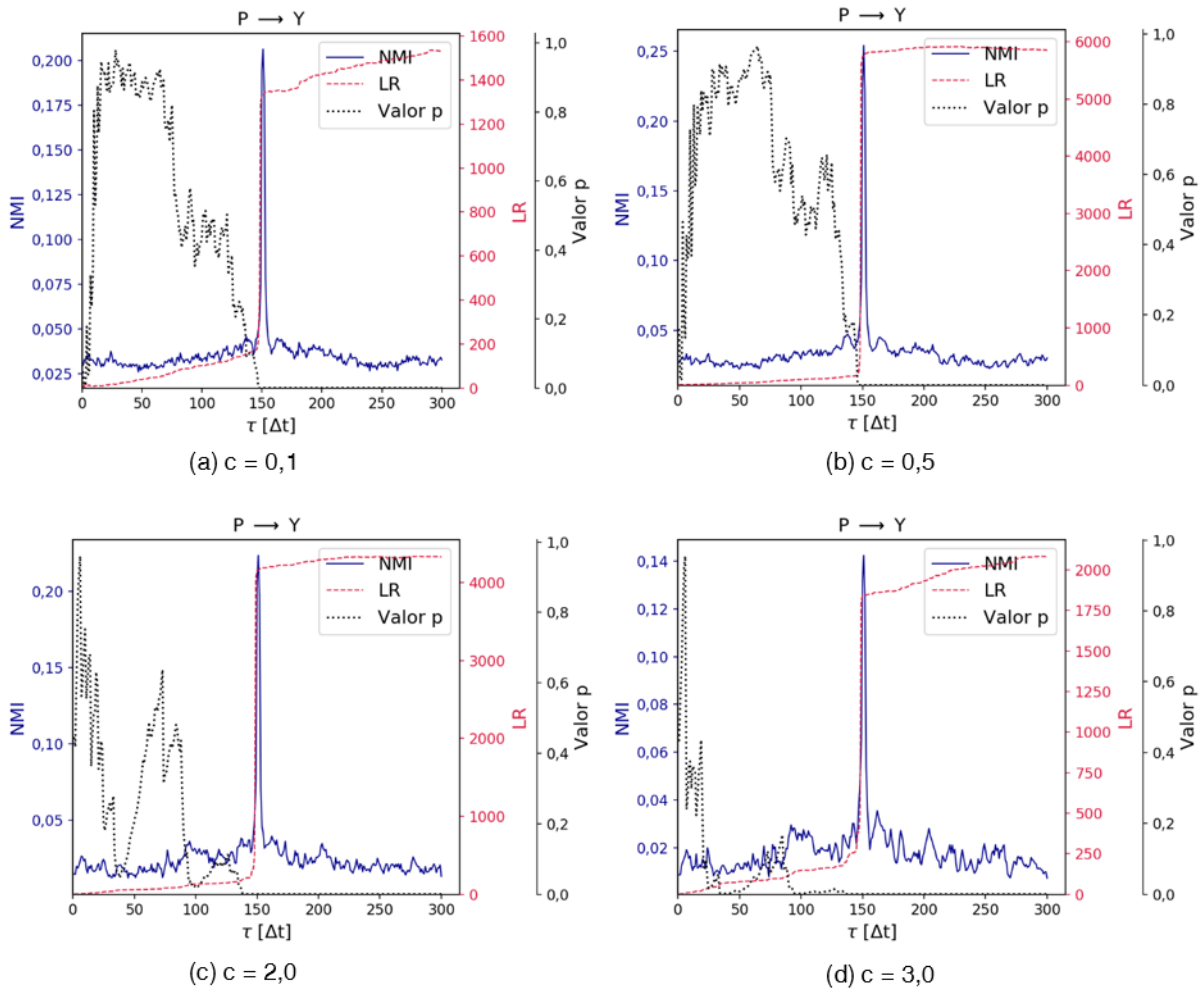
Figura 42. Resultados da simulação referentes ao problema sintético S1-1. NMI: informação mútua com defasagem; LR e valor p: razão de verossimilhança e probabilidade de significância relacionadas ao teste de Granger.  $\Delta t = 1$  dia.



Fonte: Autor.

O problema S1-1 (resultados indicados na Figura 42) avaliou a influência da variação da quantidade de termos da série de causa utilizada na geração da série de resposta (tamanho da janela do somatório, relacionado ao parâmetro  $m$ ). Quando um único termo é utilizado, os métodos conseguem identificar exatamente a defasagem  $\tau = 150$ . À medida que  $m$  aumenta, essa identificação é prejudicada. O método da informação mútua (NMI) é um pouco menos sensível que o teste de Granger (LR e valor p). Para  $m = 200$ , nenhum dos métodos é capaz de identificar o mecanismo responsável pela geração dos dados.

Figura 43. Resultados da simulação referentes ao problema sintético S1-2.  $\Delta t = 1$  dia.

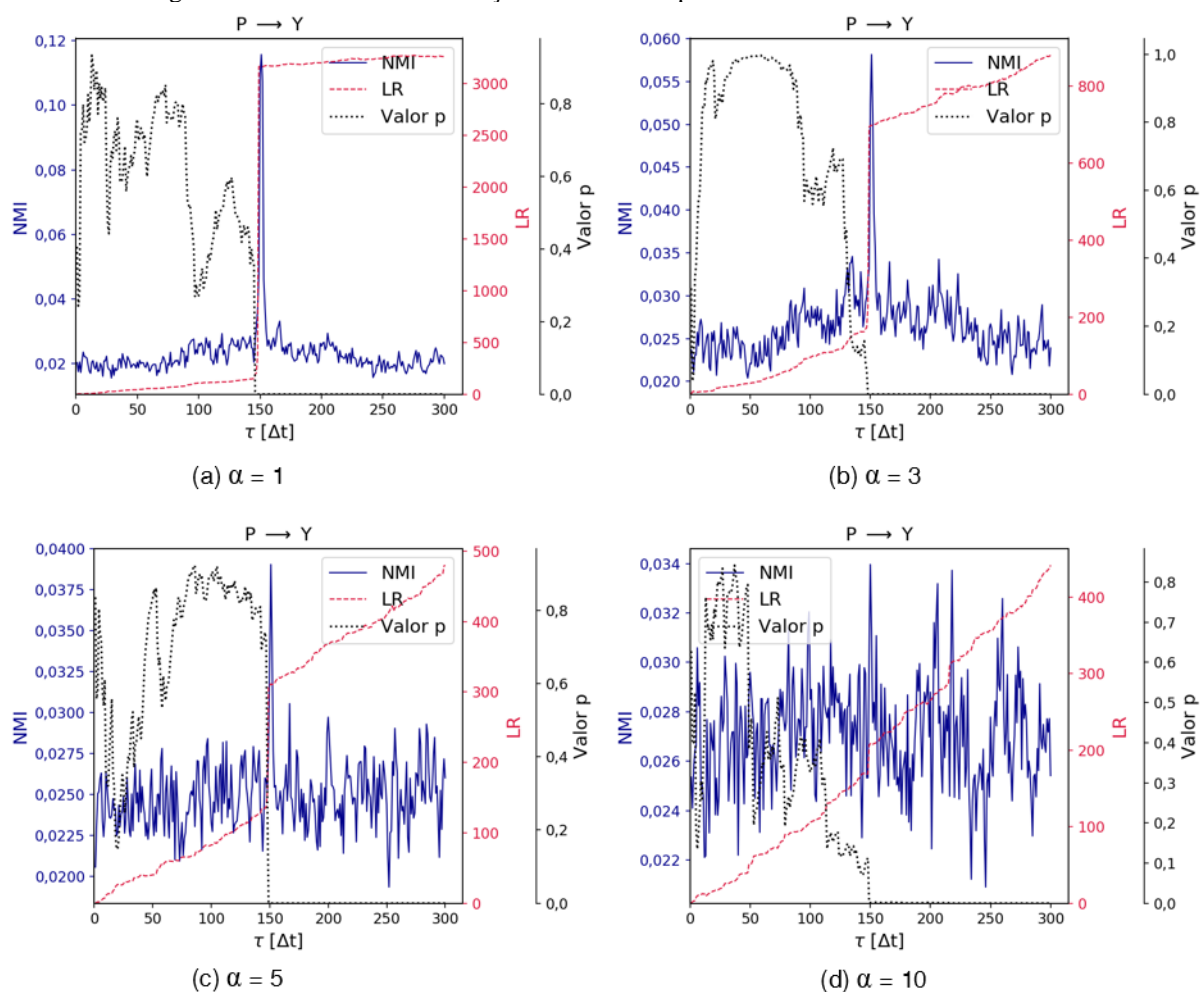


Fonte: Autor.

O problema S1-2 (resultados na Figura 43) avaliou a influência do expoente  $c$  da variável precipitação. Os métodos identificaram a defasagem de forma consistente e razoável para expoentes entre 0,1 e 3,0, demonstrando que possuem capacidade para avaliar problemas não lineares. O método de Granger (LR e valor p) antecipou a defasagem em alguns dias para expoentes maiores que 1,0. Já a informação mútua (NMI) identificou exatamente as defasagens para qualquer condição.



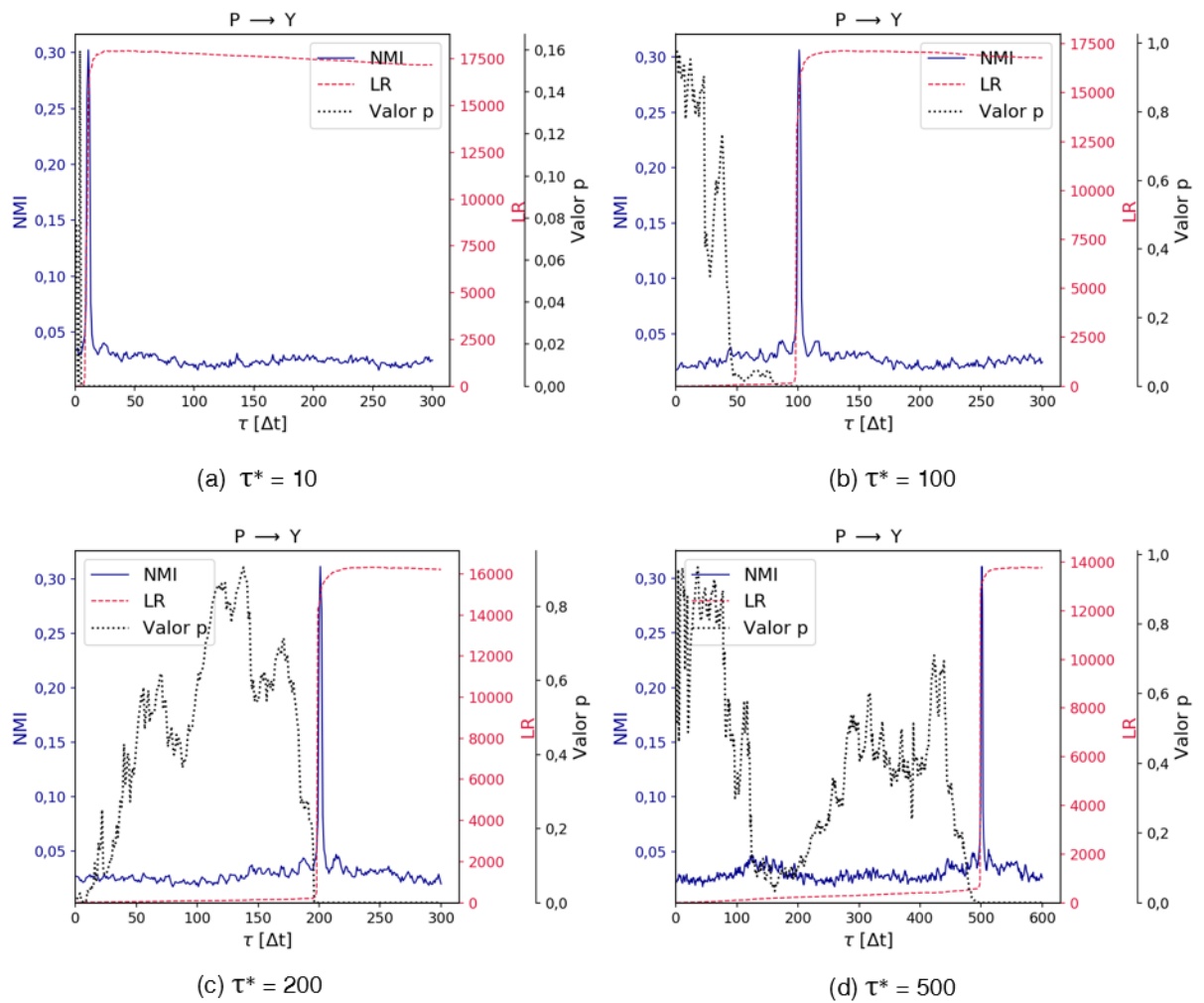
Figura 44. Resultados da simulação referentes ao problema sintético S1-3.  $\Delta t = 1$  dia.



Fonte: Autor.

O problema S1-3 (resultados na Figura 44) avaliou a influência da variância do termo aleatório  $\alpha$ . Para variâncias entre  $1^2$  e  $10^2$  vezes a variância amostral da precipitação, os métodos conseguiram identificar os mecanismos de geração das respostas. Ou seja, a presença de ruídos brancos gaussianos parece não ser um obstáculo para estas ferramentas.

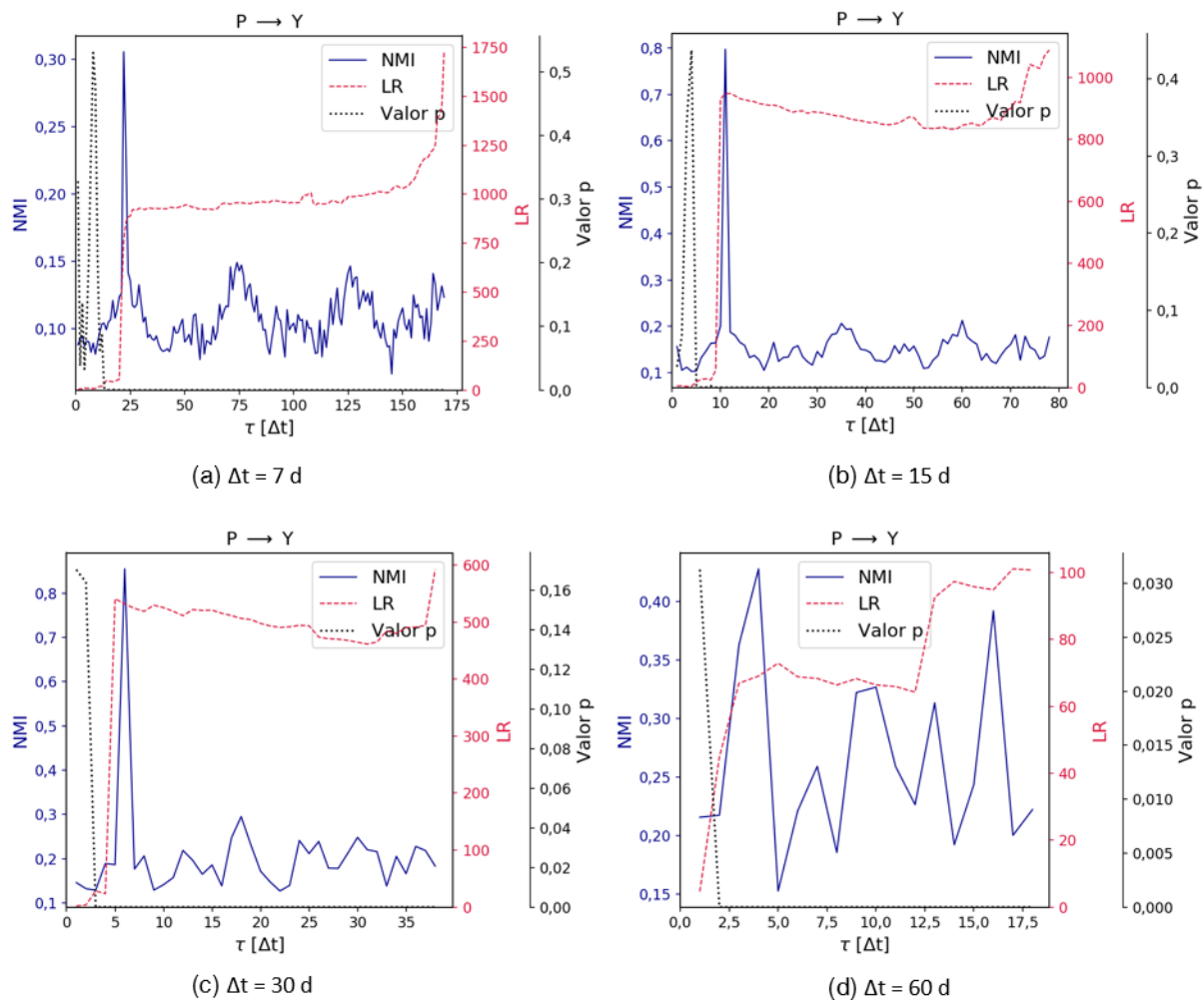
Figura 45. Resultados da simulação referentes ao problema sintético S1-4.  $\Delta t = 1$  dia.



Fonte: Autor.

O problema S1-4 (resultados na Figura 45) avaliou a influência da posição da defasagem  $\tau^*$ . Independentemente de estar no início, no meio ou no fim do período, os métodos foram capazes de identificar as defasagens criadas. O efeito de memória da variável da precipitação, que possui um ciclo de variação anual, não causou interferência na identificação das defasagens.

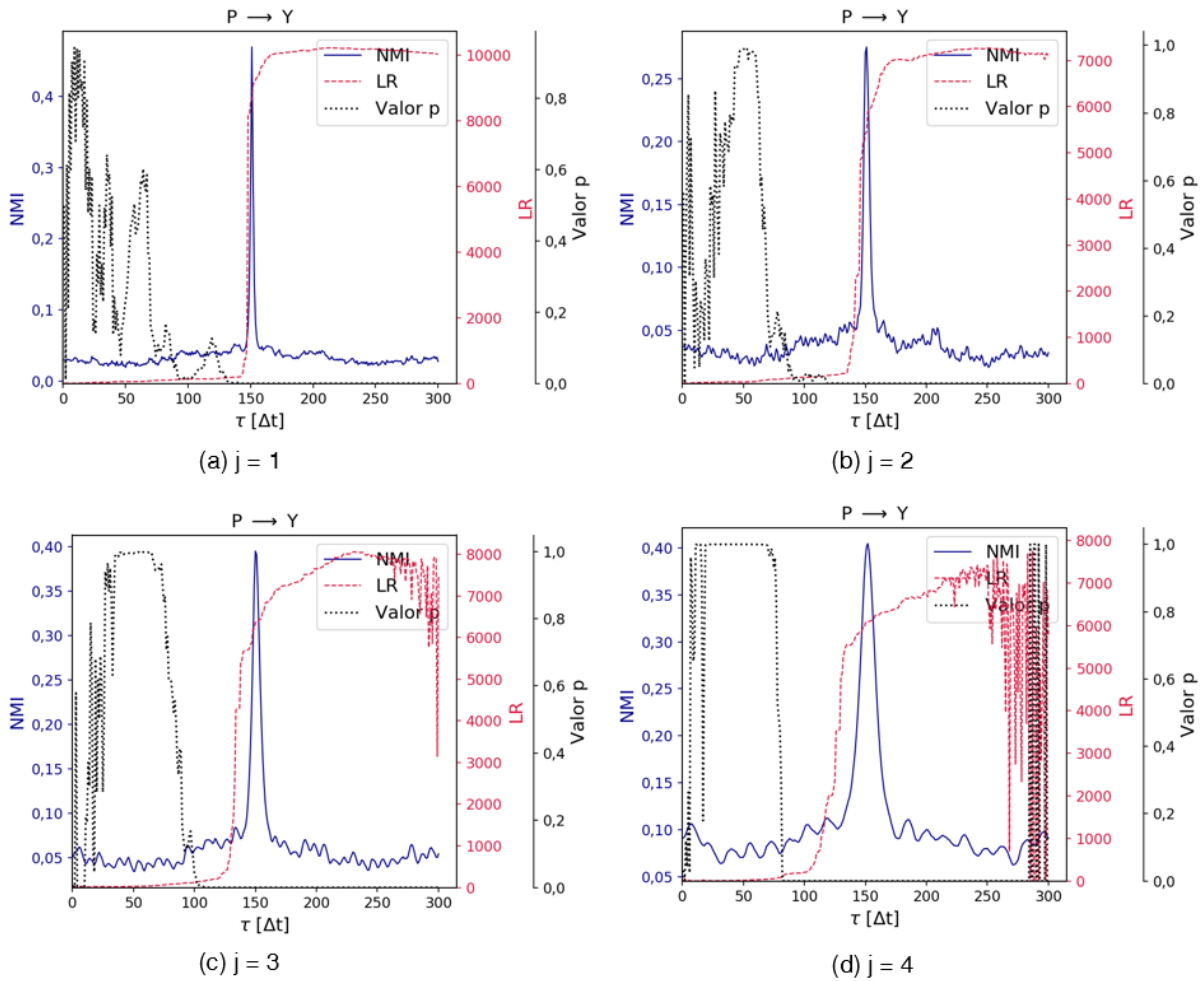
Figura 46. Resultados da simulação referentes ao problema sintético S1-5 (agregação temporal).



Fonte: Autor.

O problema S1-5 (resultados na Figura 46) avaliou explicitamente a influência da agregação temporal. Similarmente ao problema S1-1, os métodos são afetados, entretanto, conseguem identificar de forma aproximada as defasagens de  $150/\Delta t$ , em que  $\Delta t$  é a agregação temporal utilizada (e.g.,  $\Delta t = 7$  d se dados são agregados em escala semanal).

Figura 47. Resultados da simulação referentes ao problema sintético S1-6 (análise de multirresolução).  $\Delta t = 1$  dia.



Fonte: Autor.

O problema S1-6 (resultados na Figura 47) avaliou a influência da suavização das séries de resposta, que foram aproximadas pela componente de baixa frequência em diferentes níveis de decomposição (decomposição por transformada *wavelet* discreta - DWT). O teste de Granger foi consideravelmente afetado, identificando defasagens de até metade da que deveria, e com instabilidades no cálculo da LR. A informação mútua apresentou-se como uma ferramenta pouco sensível a variações nos graus de suavização. Além disso, ela foi capaz de identificar corretamente as defasagens em todos os casos.

O problema sintético 2 teve objetivos similares ao problema sintético 1, de avaliar o funcionamento das ferramentas, porém com uma estrutura de geração de resposta diferente, baseada em operador de convolução. Os resultados são apresentados de forma similar ao problema anterior no Apêndice D (a defasagem a ser identificada é igual ao produto entre os parâmetros  $n$  e  $k$ ). Os problemas S2-1 a S2-4 (que avaliaram a influência da localização da

defasagem com a variação do parâmetro  $n$ ), da variância do termo aleatório, os efeitos da agregação temporal e a análise de multirresolução, respectivamente, levaram a conclusões similares às apresentadas para o problema sintético 1. A principal diferença foi em relação à análise de multirresolução. Como o processo de convolução leva a resultados mais suaves, a remoção de termos de alta frequência pouco influenciou as ferramentas estatísticas.

Em síntese, o método baseado na informação mútua com defasagem normalizada (NMI) apresentou um comportamento adequado sob as diversas condições nas quais foi testada. O teste de Granger apresentou em geral um comportamento também adequado, porém levou a algumas conclusões equivocadas e a instabilidades em situações de não linearidade ou em casos de suavização (redução de ruídos). Entre os equívocos comumente observados destaca-se a antecipação da identificação da existência de relação causal. Em termos computacionais, o tempo de processamento requerido pelo teste de Granger cresce consideravelmente à medida que o número máximo de defasagens aumenta.

Adicionalmente, esta etapa de teste das ferramentas estatísticas permitiu concluir que as análises com dados hidrológicos observados devem ser conduzidas, preferencialmente, sem eliminação de ‘ruídos’, uma vez que esse processo além de não contribuir, pode prejudicar na identificação das conectividades e dependências entre processos.

#### **4.8 Identificação de conectividade e de dependências nas séries de processos hidrológicos observados na área de estudo**

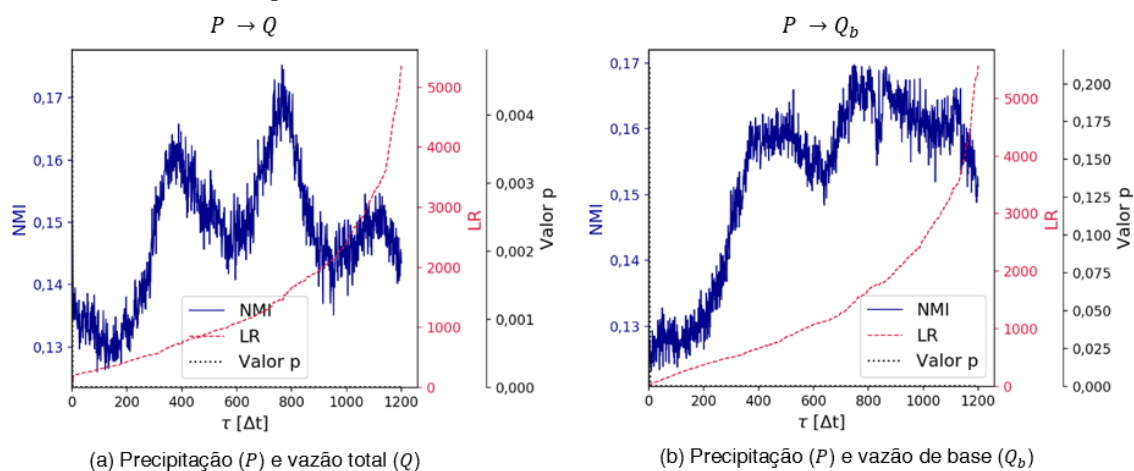
A avaliação exploratória da conectividade e dependência nas séries de processos hidrológicos observados foi conduzida apenas para a área SM em função da maior extensão das séries disponíveis (2009 a 2018). Os resultados obtidos nos problemas sintéticos foram utilizados para definir que essa avaliação fosse feita com os dados quantificados na escala diária, sem qualquer procedimento de eliminação de ruídos ou agregação temporal.

A Figura 48 apresenta as análises de conectividade entre as séries diárias estimadas de precipitação, vazão e vazão de base. Os resultados do teste de Granger apontam evidência estatística (valor- $p = 0$ ) de que os dados possuam relação causal entre si, com defasagem mínima de 0, ou seja, o teste indica que a precipitação continuamente influencia a variável de vazão ou vazão de base. Do ponto de vista físico, esse resultado podia ser esperado.

A NMI indica dois picos para defasagens da ordem de 400 e 800 dias. Como esta ferramenta não elimina os efeitos estáticos de memória ou persistência, como o teste de Granger

ou a entropia de transferência (SCHREIBER, 2000), o primeiro pico pode ser resultado do efeito de memória da própria precipitação, que apresenta ciclos de variação anual. O segundo pico coincide com o identificado na análise de espectro de potência global de *wavelet* da série de vazão de base (GOMÉZ et al., 2018). Os autores desse trabalho justificam esta defasagem como um tempo de regulação do aquífero, ou seja, o tempo médio requerido para que a água precipitada infiltre e seja transformada em fluxo de base no curso d'água.

Figura 48. Resultados das relações entre precipitação e escoamento superficial e escoamento de base. Dados do período entre 12/2008 e 02/2019.  $\Delta t = 1$  dia.



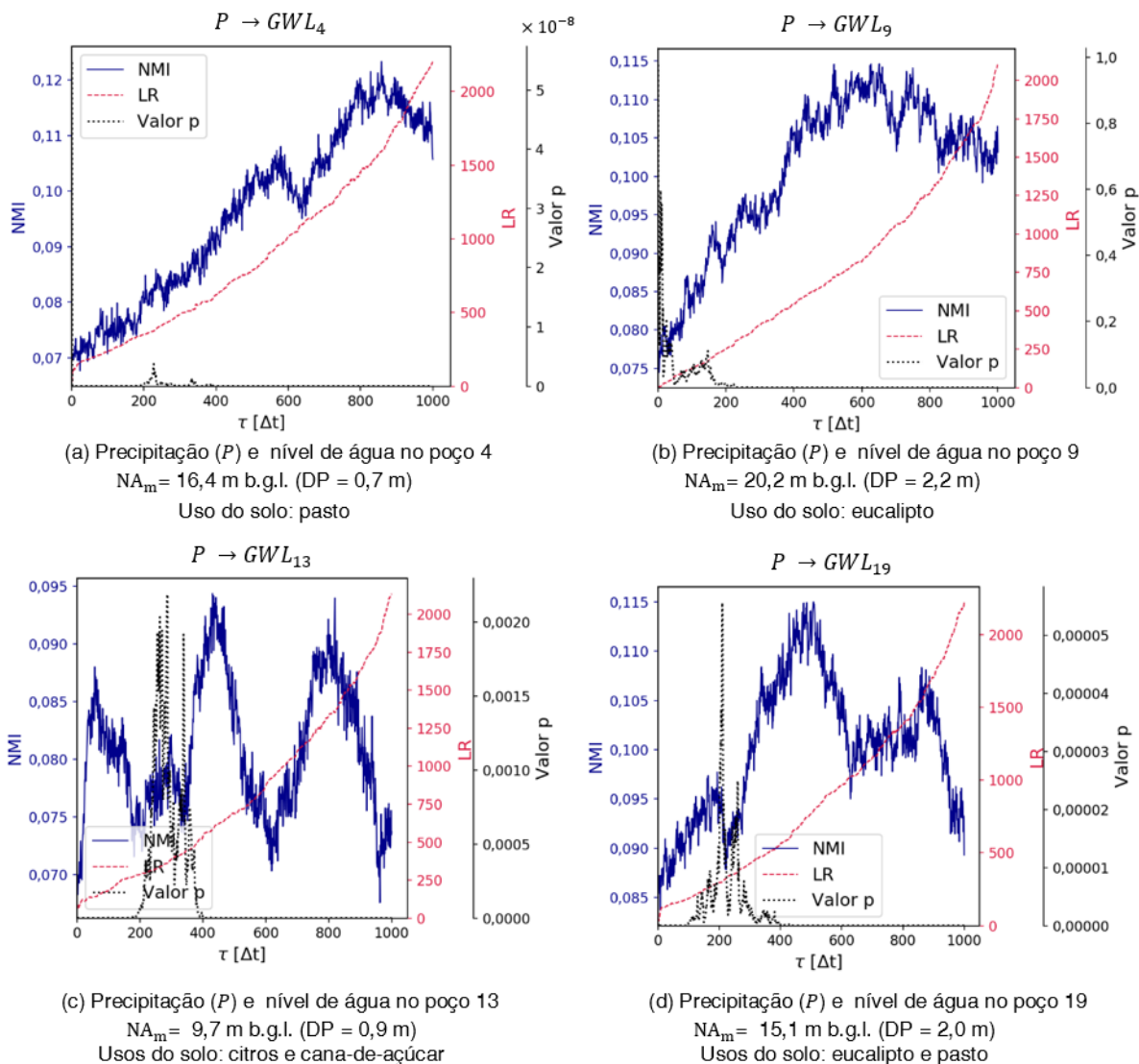
Fonte: Autor.

Análises similares com a precipitação e níveis freáticos são apresentadas nas figuras 49 (poços mais profundos) e 50. De forma geral, pode-se perceber a existência de três picos. O primeiro possivelmente relacionado ao tempo médio entre a precipitação e a elevação do nível freático. Nem todos apresentam um padrão bem estabelecido, como pode ser observado com os poços de monitoramento P4 e P9, que são mais profundos que os demais. O segundo pico possivelmente está relacionado à característica sazonal da precipitação. E o terceiro pico evidencia alguma relação entre as defasagens entre o processo de precipitação, geração do escoamento superficial (e escoamento de base) e variação de longo termo dos níveis freáticos. Esses resultados coincidem com os apresentados na Figura 48.

Junto com o poço P9 (eucalipto), o poço P15 (em que os cultivos predominantes foram cana-de-açúcar e recentemente citros) apresenta um comportamento singular, distinto dos demais. O poço P15, por estar próximo aos poços P13 e P14 (ver Figura 3) e sujeito a condições hidrogeológicas similares, permite identificar a existência de influência do histórico do uso do solo nas respostas observadas. As respostas dos poços P4 e P5 (pasto), P13 e P14 (citros e

recentemente cana-de-açúcar) e P16 e P19 (alternância entre pasto e eucalipto) são comparáveis, e isso pode ser consequência tanto do solo comum do solo quanto da semelhança em características hidrogeológicas (pela proximidade geográfica).

Figura 49. Resultados das relações entre precipitação e níveis freáticos (poços P4, P9, P13 e P19). [m b.g.l.: metros abaixo do nível do solo;  $NA_m$ : nível de água médio; DP: desvio padrão].  $\Delta t = 1$  dia.



Fonte: Autor.

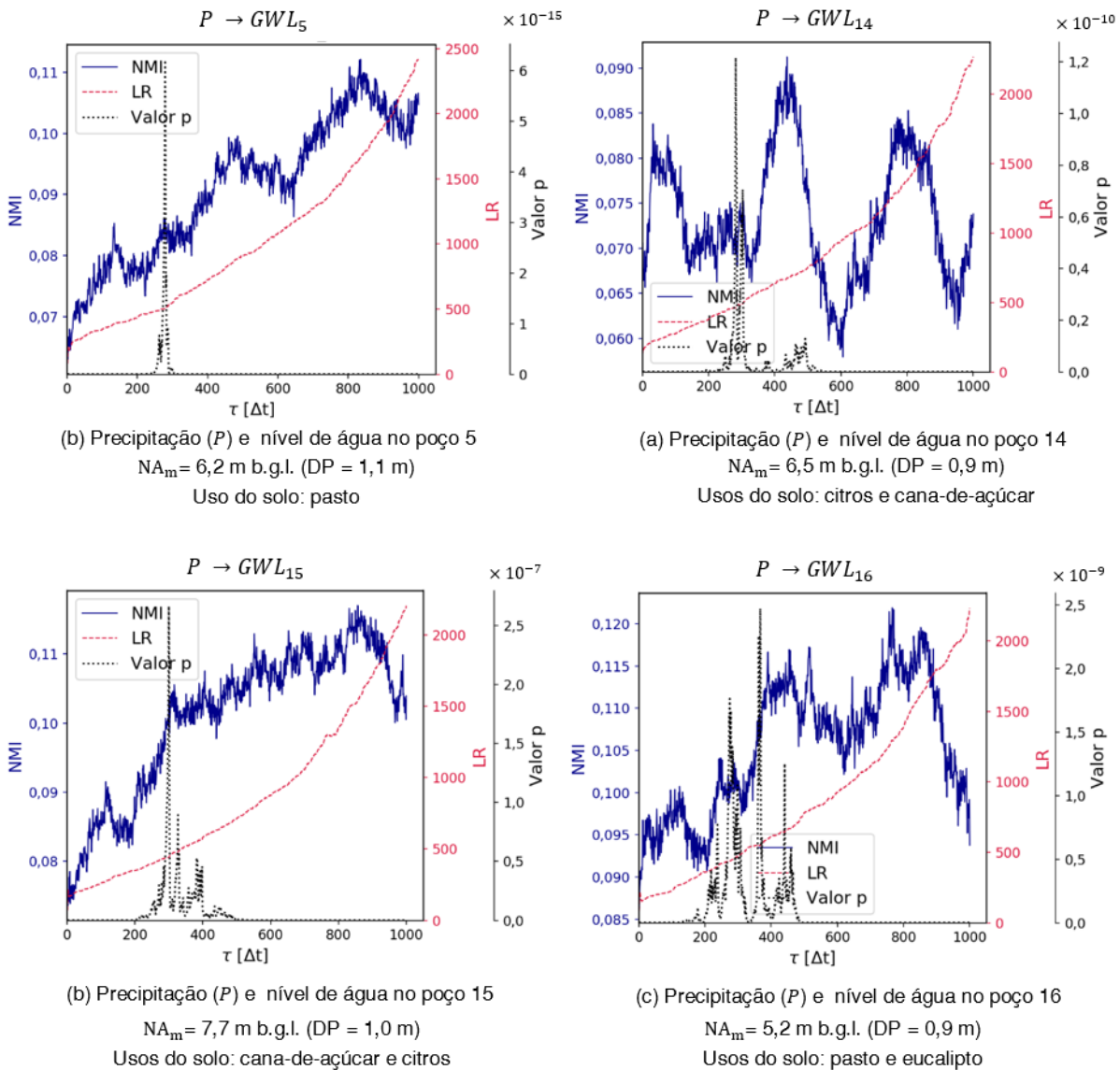
De forma a unificar os resultados das variações dos níveis freáticos, a análise foi feita com a média aritmética dos níveis freáticos observados nos poços de monitoramento, e a Figura 51 (a) apresenta o resultado obtido. O padrão com três picos bem definidos fica claro nessa abordagem.

A influência entre precipitação e o índice espectral de vegetação EVI também foi explorada e está apresentada na Figura 51 (b e d). Ambas as direções foram avaliadas. Pelo



gráfico (b) pode-se perceber que a partir de aproximadamente 200 dias a precipitação começa a apresentar uma relação de causa estatisticamente significativa com o índice EVI e há o primeiro pico da NMI. Essa resposta pode estar relacionada às características estruturais da vegetação local. Já o gráfico (d) revela que a relação inversa, de EVI influenciando a formação de precipitação, possui defasagem da ordem de 500 dias (primeiro pico de NMI com valor  $p = 0$ ). A complexidade de mecanismos locais responsáveis pela geração da precipitação torna difícil avaliar qual o significado desse valor. A existência de *feedbacks* entre a disponibilidade hídrica, a dinâmica da vegetação e o processo de precipitação é, no entanto, evidenciado nessas análises.

Figura 50. Resultados das relações entre precipitação e níveis freáticos (poços P14, P15, P16 e P19). [m b.g.l.: metros abaixo do nível do solo;  $NA_m$ : nível de água médio; DP: desvio padrão].  $\Delta t = 1$  dia.

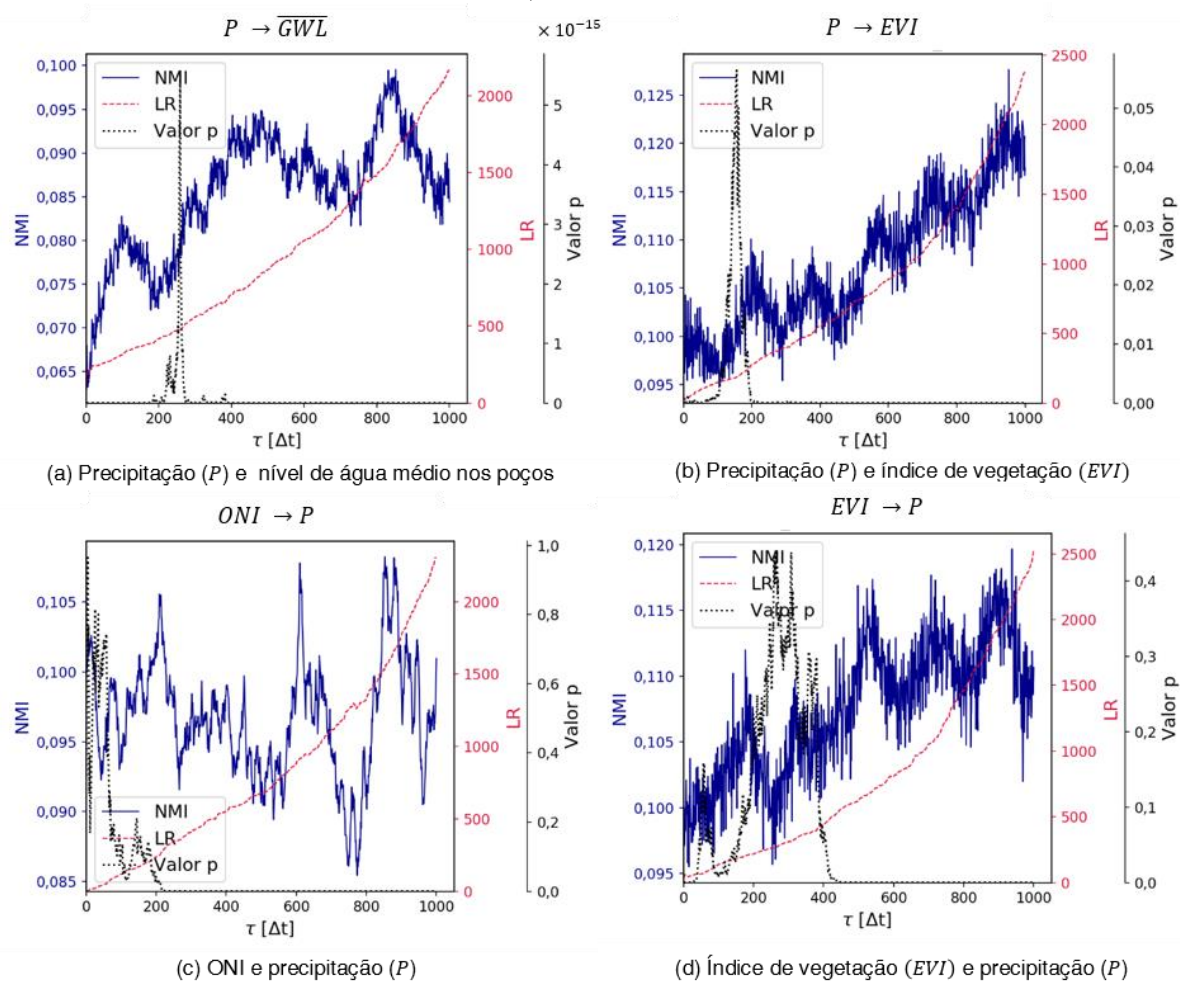


Fonte: Autor.



Outro aspecto avaliado foi a influência de fenômenos de grande escala no regime local de precipitações. O *oceanic Niño index*, um dos índices utilizados para monitorar a ocorrência do *El Niño* Oscilação Sul (ENOS) (evento relacionado ao aquecimento atípico da temperatura da superfície da água do Oceano Pacífico), foi interpolado linearmente em escala diária e sua relação com a precipitação média é apresentada no gráfico (c) da Figura 51. No período comum de monitoramento da bacia (a partir de 2009), eventos *El Niño* de diferentes intensidades ocorreram três vezes. O teste de Granger sugere a existência de relação causal a partir de 200 dias. O uso de séries mais longas e indicadores estatísticos para avaliar essas influências é uma oportunidade para aprofundar a compreensão da propagação dos efeitos de fenômenos oceânicos e continentais na escala local.

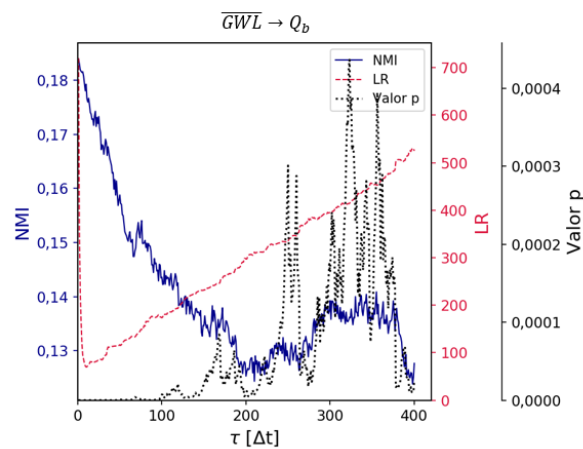
Figura 51. Resultados de relações entre precipitação, níveis freáticos, índice de vegetação e anomalias de grande escala.  $\overline{GWL}$ : média aritmética de níveis freáticos dos poços 4, 5, 13, 14, 15, 16, 19. EVI: *Enhanced Vegetation Index* médio da área SM; ONI: *Oceanic Niño Index*.  $\Delta t = 1$  dia.



Fonte: Autor.

O nível de água subterrâneo e a vazão de base possuem relações de causa e efeito praticamente simultâneas. Tanto o teste de Granger quanto a informação mútua foram capazes de identificar isso (Figura 52), com o pico de NMI acontecendo para  $\tau = 0$ . Esse resultado está coerente com as avaliações das relações entre precipitação e vazão e precipitação e nível freático. Também está coerente com as relações propostas por Wendland, Barreto e Gomes (2007), em que o escoamento de base é estimado a partir de níveis freáticos observados nos mesmos instantes de tempo (método da curva-chave).

Figura 52. Efeitos simultâneos entre média dos níveis freáticos e vazão de base.  $\Delta t = 1$  dia.

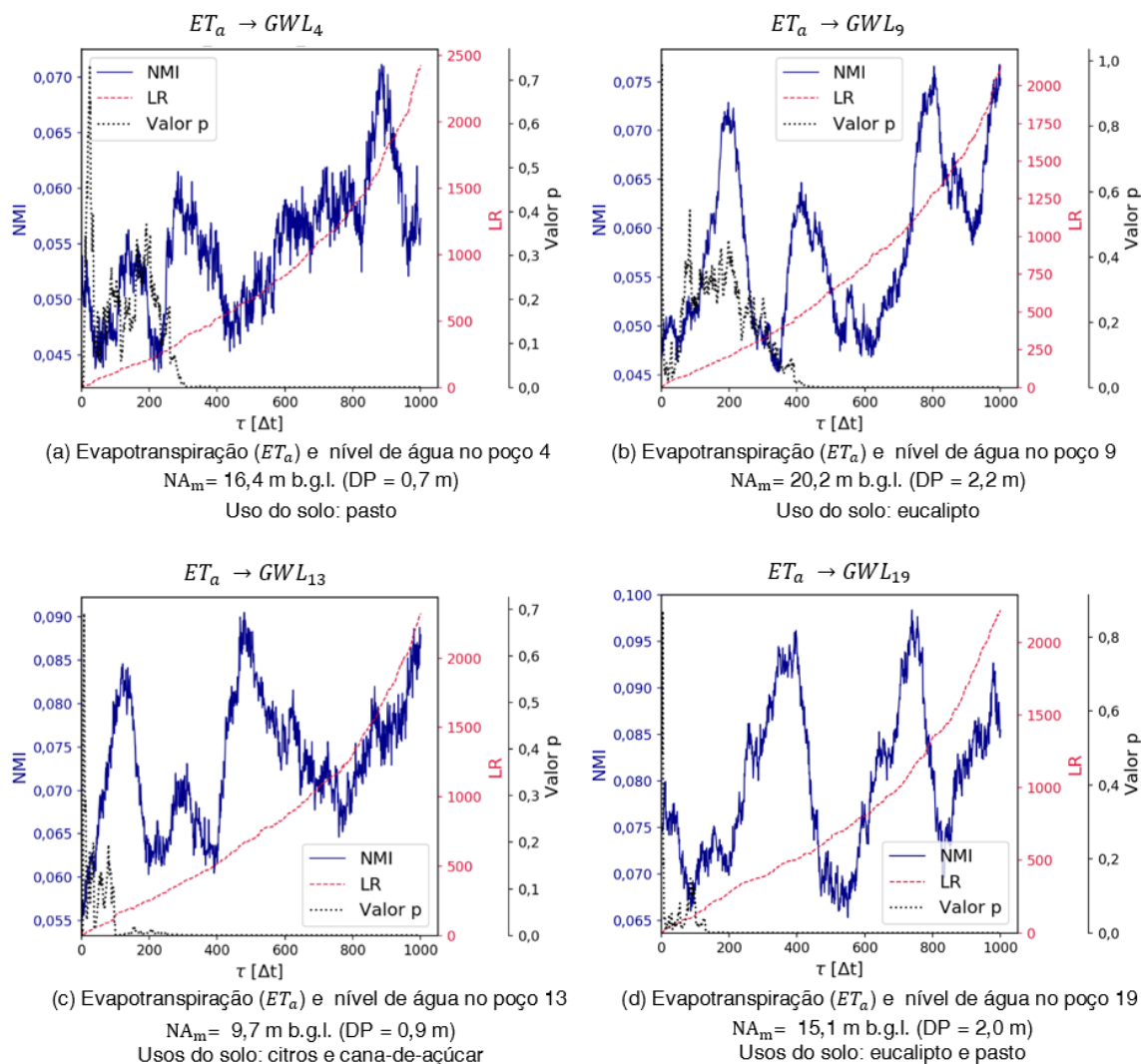


Fonte: Autor.

Para avaliar a influência do tipo de uso do solo, séries de EVI próximas aos poços de monitoramento foram utilizadas para estimar a evapotranspiração real local. Por simplificação, o parâmetro  $c$  da equação (20) foi adotado igual a 0,169 em todos os casos (nessas análises, as variações importam mais que os valores absolutos).

Os resultados das avaliações entre a evapotranspiração real estimada e os níveis freáticos são apresentados nas figuras 53 (poços mais profundos) e 54. As relações de causa são quase sempre estatisticamente significantes no sentido de Granger. As exceções são os poços mais profundos (P4 e P9), em que a relação é confirmada apenas para defasagens acima de 300 ou 400 dias. Pela análise da NMI, alguns padrões podem ser identificados, como picos próximos a defasagens de 100 dias, entre 300 e 400 dias e entre 700 e 900 dias. Poços de mesma cultura apresentam comportamentos similares (P4 e P5 – pasto, P13 e P14 – citros e cana-de-açúcar, e P9, P16, P19 – eucalipto ou pasto e eucalipto). Esses resultados indicam que tanto a profundidade do poço monitorado quanto o histórico de cultivo local levam a respostas únicas. Possíveis aplicações desse tipo de caracterização precisam ser avaliadas futuramente.

Figura 53. Resultados das relações entre evapotranspiração e níveis freáticos (poços P4, P9, P13 e P19). [m b.g.l.: metros abaixo do nível do terreno;  $NA_m$ : nível de água médio; DP: desvio padrão].  $\Delta t = 1$  dia.



Fonte: Autor.

Em síntese, esta seção avaliou as relações entre processos hidrológicos estimados na área de estudo pelo teste de Granger e pela informação mútua com defasagem. O teste de Granger em geral não contribuiu com informações originais, servindo mais para confirmar que algumas variáveis contribuem para melhorar a previsão de outras. Além disso, a não-linearidade dos processos e a não-normalidade dos dados são aspectos que prejudicam os resultados de uma forma geral. A informação mútua, por outro lado, contribuiu com a quantificação de algumas defasagens no sistema hidrológico.

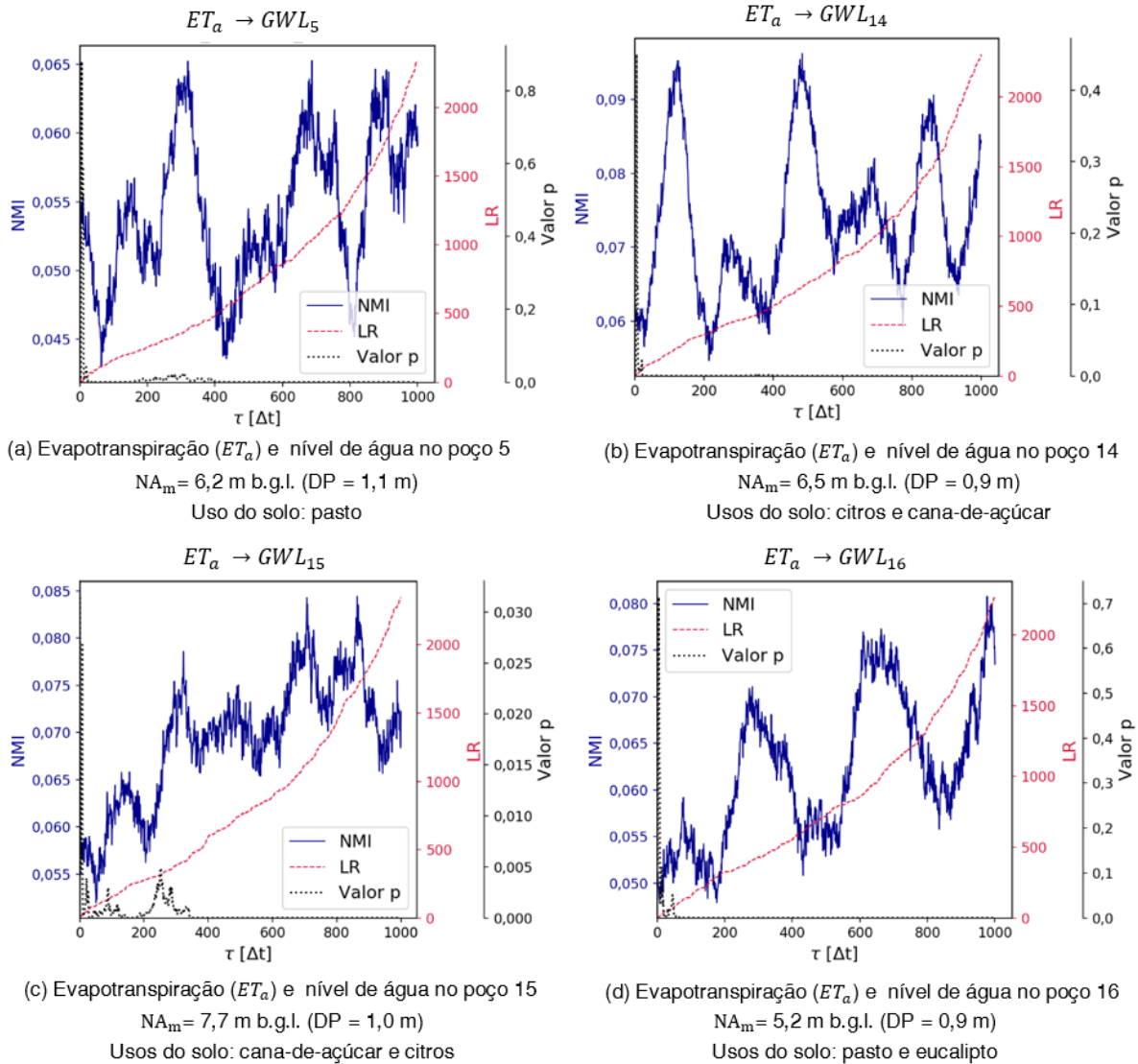
As defasagens identificadas não necessariamente possuem algum sentido físico. As análises (pelo teste de Granger) entre dados de EVI e precipitação, por exemplo, indicaram que a precipitação de hoje pode estar sendo influenciada pelo desenvolvimento vegetal da área de estudo de cerca de 400 dias atrás. A atribuição da possibilidade física desse valor parece

improvável. Assim, os valores apresentados aqui nas diversas análises devem ser entendidos e avaliados com cuidado tendo em vista as limitações dos métodos utilizados.

Alguns padrões de comportamento não foram claramente identificados, e isso pode ser resultado da insuficiência de observações ou da inadequada estimativa dos processos ou das probabilidades pelo método dos histogramas. O número de blocos dos histogramas em geral não influenciou de forma significativa a distribuição da curva de NMI, mas influenciou seus valores absolutos.

Na etapa seguinte, de modelagem dos processos hidrológicos, foi verificado se alguns desses comportamentos estatísticos são reproduzidos.

Figura 54. Resultados das relações entre evapotranspiração e níveis freáticos (poços P5, P14, P15 e P16). [m b.g.l.: metros abaixo do nível do terreno;  $NA_m$ : nível de água médio; DP: desvio padrão].  $\Delta t = 1$  dia.

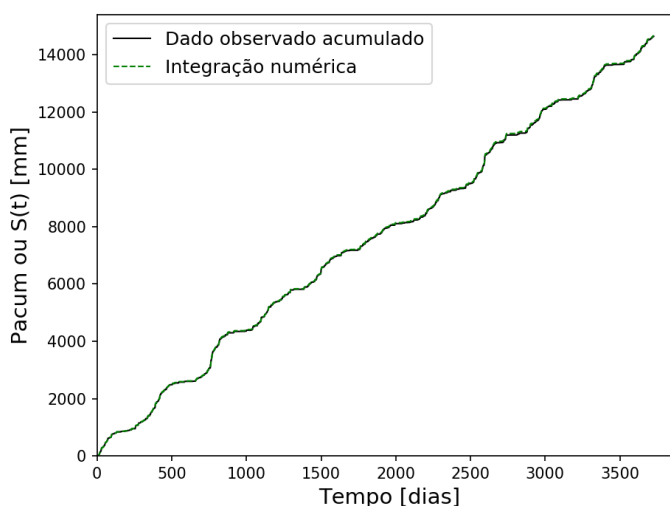


Fonte: Autor.

## 4.9 Calibração e validação do modelo conceitual proposto

O primeiro aspecto avaliado no modelo foi o esquema de integração numérica RK45. Como o esquema utilizado possui passo de tempo adaptativo e os dados são admitidos constantes no intervalo de discretização (diário), houve um interesse em entender se isso poderia gerar alguma inconsistência. A integração da equação diferencial  $dS/dt = P_t$ , em que  $P_t$  é a precipitação média diária, foi feita para analisar isso e o resultado é apresentado na Figura 55. Pode-se perceber que o resultado obtido está de acordo com o esperado, havendo sobreposição com os valores acumulados da série de precipitação ao longo de todo o domínio. A conclusão é a não necessidade de adaptação nos dados de entrada para a utilização do método de integração com passo de tempo adaptativo.

Figura 55. Avaliação do esquema de integração numérica RK45 com passo de tempo adaptativo e com dados de entrada admitidos constantes no intervalo de discretização.



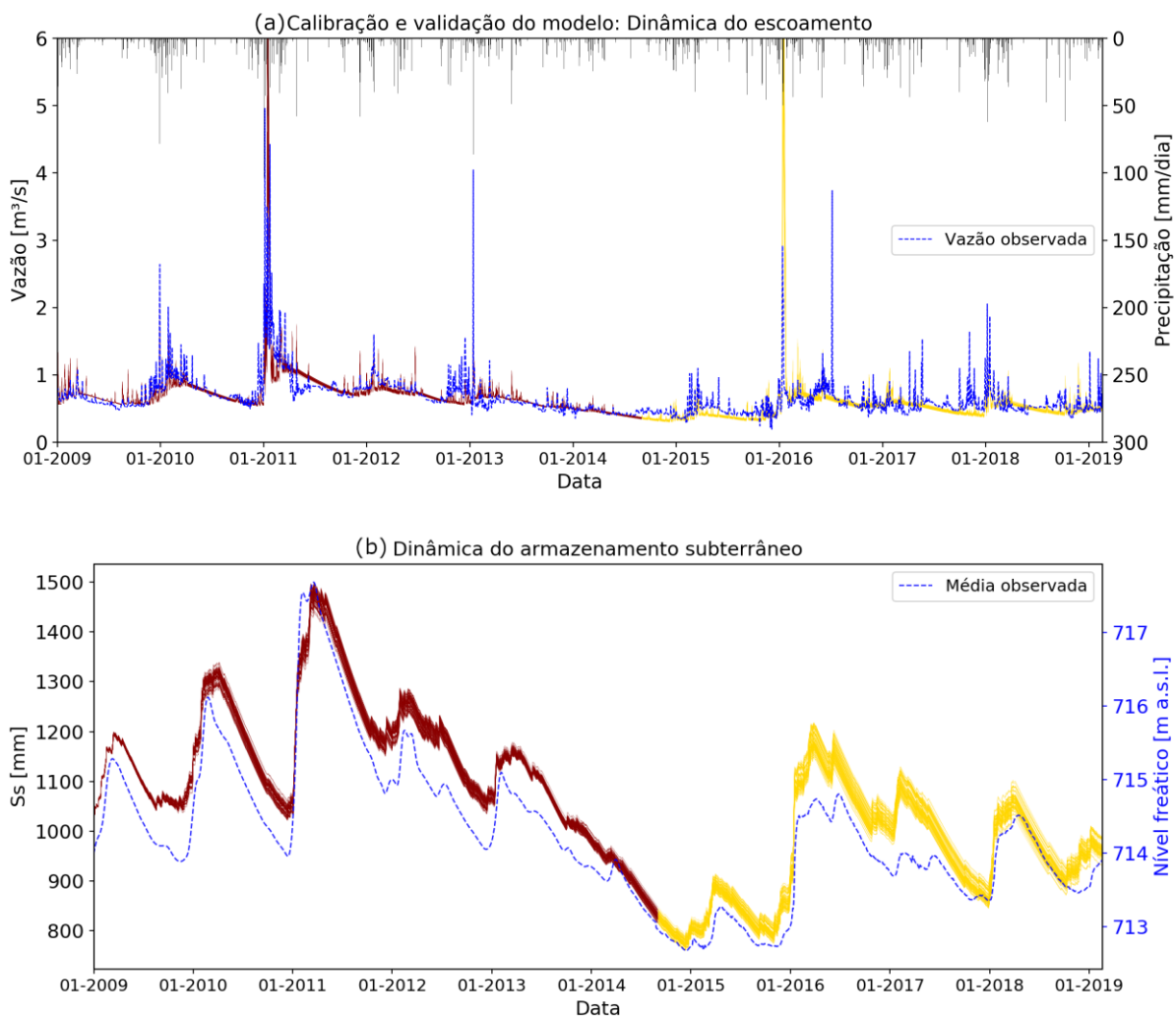
Fonte: Autor.

Após testes iniciais com o método numérico, a etapa seguinte foi de calibração e validação do modelo conceitual proposto para as duas áreas de controle (SM e CP2). Inicialmente são apresentados os resultados gráficos para a área SM (figuras 56 a 59) e depois para a área CP2 (figuras 60 a 62). Métricas de desempenho de ajuste do modelo são apresentadas em síntese na Tabela 20.

Inicialmente, tomando como referência os resultados para a área SM, a resposta simulada em termos de vazão no exutório seguiu razoavelmente bem o regime de vazões mínimas e os períodos de recessão (Figura 56a). Há uma notável deficiência no acompanhamento de picos de vazão. Em 01/2011 e 01/2016, o modelo prevê picos de vazão

maiores que os observados, possivelmente em função de condições antecedentes de precipitação, e a partir de 2015 há uma dificuldade em reproduzir as variações de vazão. Essa deficiência pode estar associada à forma de representação da bacia (concentrada), à discretização temporal (diária), à variabilidade da precipitação (e limitado monitoramento) ou ainda à alguma deficiência estrutural do modelo. Para os propósitos deste estudo, que objetiva avaliar questões relacionadas à gestão dos recursos hídricos e à previsão de disponibilidade hídrica, esse aspecto não foi considerado de grande relevância. Por isso, a função objetivo adotada na calibração ( $KGE_{NP}$ ) foi uma função não paramétrica, que não faz suposições sobre a natureza dos erros (e.g., normalidade, ausência de *outliers*) e que não atribui igual importância a elevadas vazões ou a correspondentes elevados desvios (e.g., NSE) (POOL et al., 2018).

Figura 56. Resultados referentes à área SM correspondente à população da última geração do algoritmo NSGA-II. Linhas vermelhas e amarelas diferenciam os períodos utilizados para calibração (01/2009 a 09/2014) e validação (09/2014 a 02/2019), respectivamente. Barras verticais invertidas indicam precipitação média diária. Cada linha (total de 100) indica um resultado da simulação para um conjunto de parâmetros diferente.



Pelo critério de armazenamento subterrâneo (Figura 56b), o modelo conseguiu reproduzir a variabilidade, em termos de correlação, de forma adequada (notar escalas diferentes para a variável simulada e variável observada). A avaliação em valores absolutos de armazenamento, que envolveria uma etapa adicional para estimar valores médios de rendimento específico, não foi conduzida, e é uma possibilidade a ser considerada em trabalhos futuros.

Tabela 20. Síntese de métricas de avaliação de desempenho do modelo conceitual. Resultados referentes à população final (100 soluções) de parâmetros calibrados. IQR: intervalo interquartil.

Critérios		Seção SM		Seção CP2	
		Calibração	Validação	Calibração	Validação
<i>KGE<sub>NP</sub></i> [Q]	Mediana	0,853	0,637	0,743	0,475
	IQR	0,010	0,004	0,045	0,117
<i>PBIAS</i> [Q]	Mediana	0,348	4,17	-0,580	12,43
	IQR	1,090	2,47	1,640	16,43
<i>RMSE</i> [Q] [m <sup>3</sup> /s]	Mediana	0,211	0,338	0,052	0,049
	IQR	0,014	0,062	< 0,001	0,005
<b>Varição de armazenamento subterrâneo, S</b>		Calibração	Validação	Calibração	Validação
<i>r<sub>s</sub></i> [S]	Mediana	0,983	0,953	0,952	0,414
	IQR	0,003	0,006	0,045	0,172

Fonte: Autor.

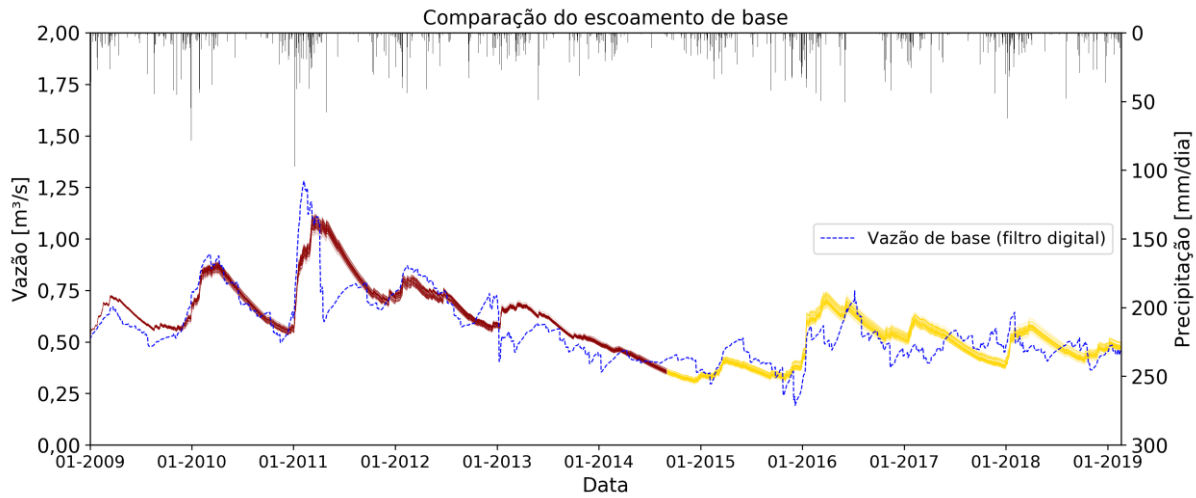
O desempenho na calibração foi consideravelmente melhor que na validação para o critério de vazão, e similar para o critério de variação de armazenamento (Tabela 20). A diferença para o critério de vazão pode ser atribuída, por exemplo, às diferentes condições meteorológicas dos períodos (e.g., o período de calibração possui períodos de recessão melhor definidos, contribuindo para uma melhor avaliação da métrica) e às diferentes fontes de dados de vazão (inclusão parcial dos dados do ADCP SonTek IQ Plus a partir de 07/2015). Graficamente, não é detectável uma piora considerável da qualidade do modelo.

A comparação entre o escoamento de base estimado pelo filtro digital e o determinado pelo modelo é apresentada na Figura 57. Um comportamento mais suave é identificado como a resposta do modelo. As oscilações produzidas pelo filtro digital são resultantes de problemas do próprio dado de vazão total (ou mais precisamente, problemas do dado de nível de água e do modelo de transformação).

A reprodução da variabilidade de vazões em termos da curva de permanência é um dos critérios incluídos na função *KGE<sub>NP</sub>*. A Figura 58 apresenta o comportamento de uma solução possível do modelo. O desempenho observado demonstra que essa métrica permite atingir esse objetivo de forma satisfatória.

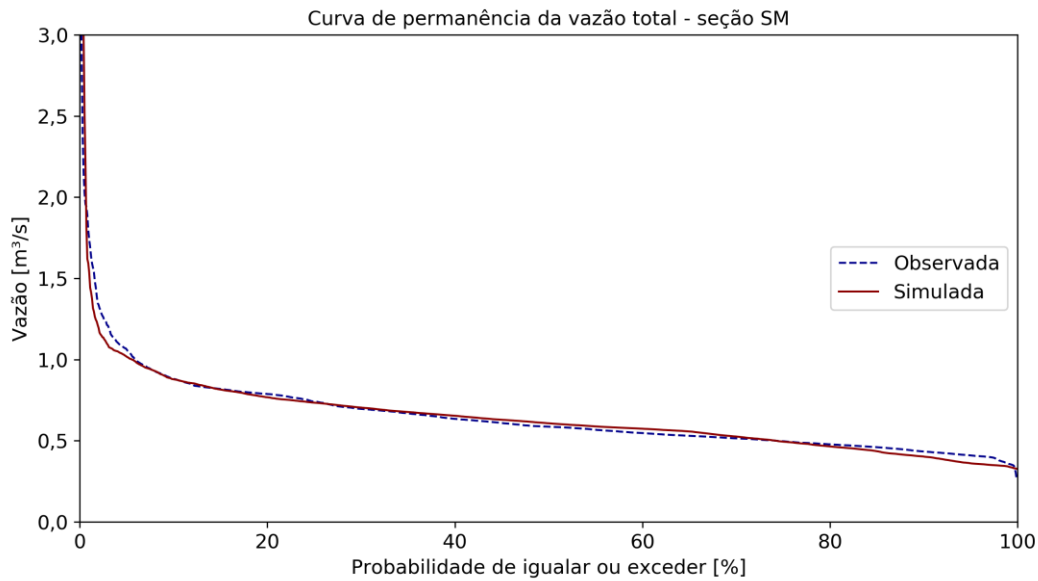


Figura 57. Escoamento de base simulado e estimado pelo filtro digital referente à área SM. As cores vermelha e amarela indicam os períodos de calibração e validação, respectivamente. Cada linha (total de 100) indica um resultado da simulação para um conjunto de parâmetros diferente.



Fonte: Autor.

Figura 58. Reprodução da curva de permanência de vazões na seção SM exemplificada para um vetor de parâmetros calibrados.



Fonte: Autor.

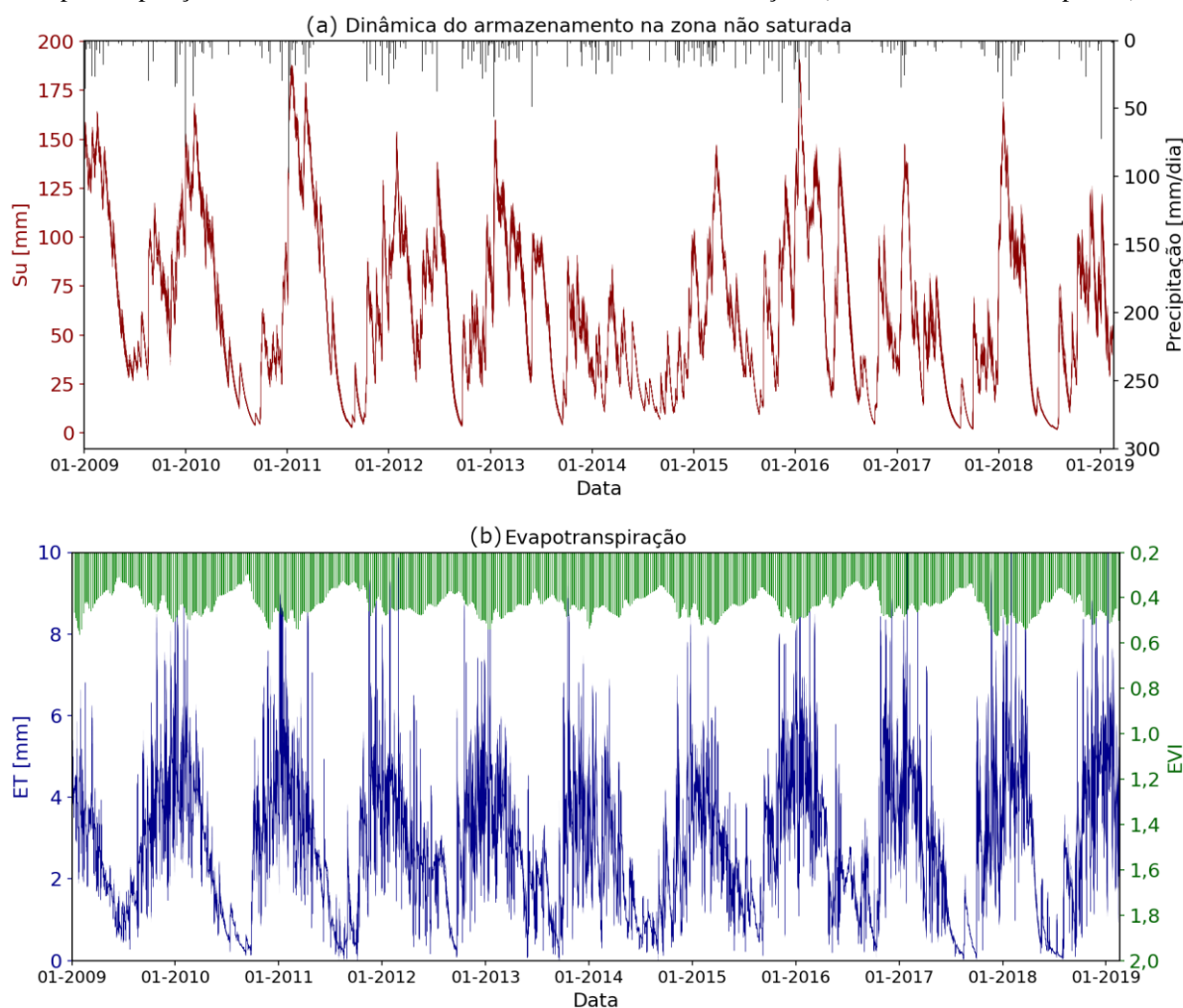
Outras características do modelo, embora sem medições em campo para comparação, foram avaliadas para fins de entendimento da dinâmica e do comportamento do modelo. O armazenamento na zona não saturada e a evapotranspiração total (ET) (contando a interceptação) são apresentadas na Figura 59. Os ciclos anuais, com verões chuvosos (entre dezembro e março), ficam bem demarcados nos gráficos e está compatível com resultados experimentais que estimaram a variação do teor de umidade no solo em alguns pontos da bacia



(POMPEO, 1990). O ano de 2014, que registrou um volume de precipitações abaixo da média, indica uma redução considerável do armazenamento, porém com a taxa de evapotranspiração apenas ligeiramente alterada. Em geral, os valores de armazenamento na zona não saturada,  $S_u$ , aproximam de zero ao final de períodos de estiagem (entre agosto e setembro). No período de 2013 a 2015, uma redução considerável no valor médio de  $S_u$  é observada. Porque neste período a aproximação a zero não ocorre pode indicar algum problema na parametrização do modelo e é uma questão a ser investigada em trabalhos futuros para reprodução de respostas hidrológicas a eventos extremos relacionados à seca.

O índice de vegetação EVI também é indicado junto com a ET, evidenciando o efeito de sazonalidade e as dependências criadas no modelo. Pelo modelo, a taxa de evapotranspiração é em média 2,77 mm/d (01/2009 a 02/2019), das quais aproximadamente 17% correspondem a perdas por interceptação.

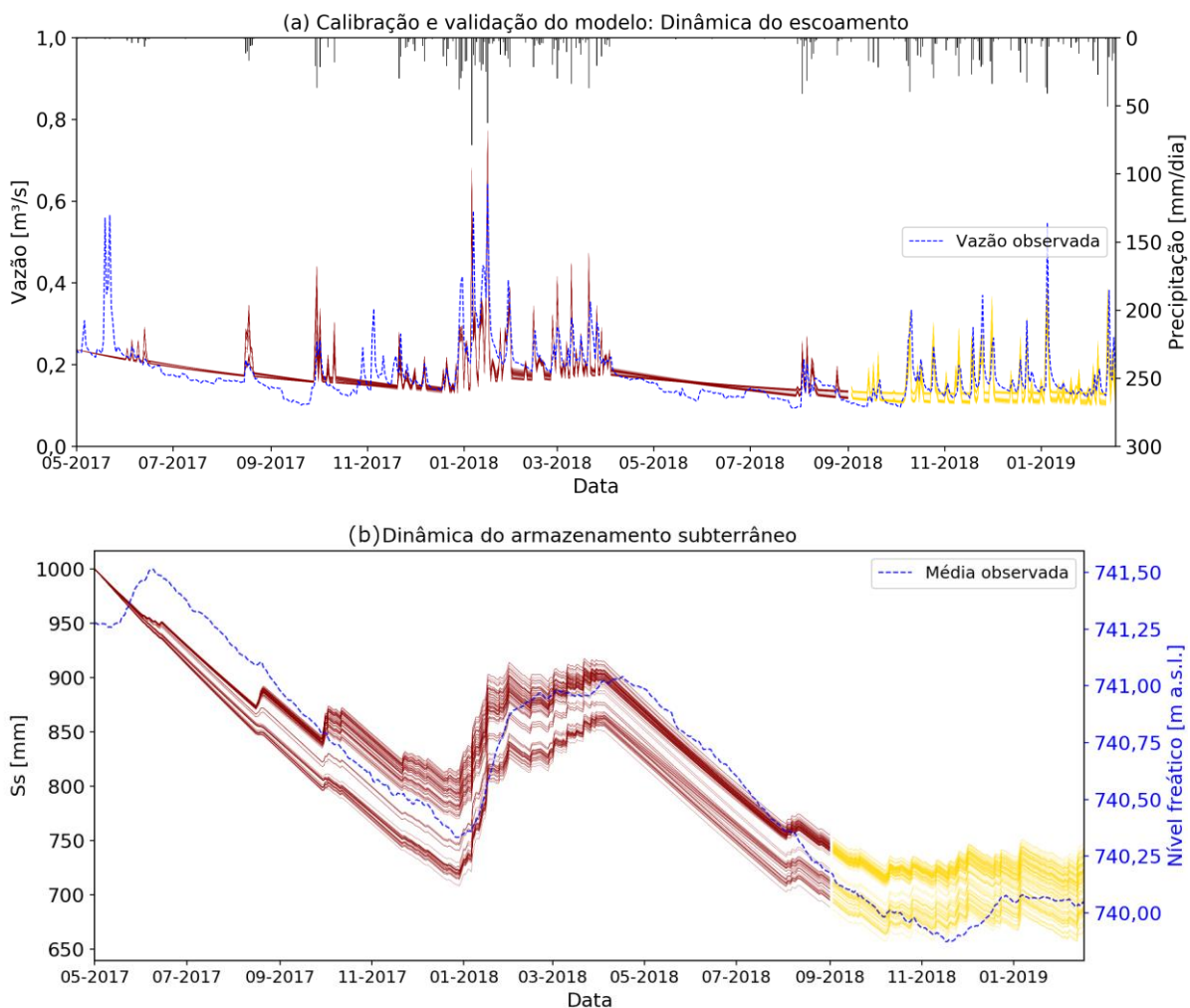
Figura 59. Avaliação de características simuladas na área SM relacionadas à disponibilidade de água no solo e à evapotranspiração. Curvas indicam os resultados referentes a 100 soluções (resultados estão sobrepostos).



Fonte: Autor.

Os resultados referentes à área CP2 não foram muito diferentes dos anteriormente apresentados para a área SM. Ao avaliar as métricas de desempenho do modelo (Tabela 20), a avaliação dos critérios de vazão e de armazenamento subterrâneo traz resultados consideravelmente piores na etapa de validação. Novamente, questões meteorológicas e a pequena extensão dos dados utilizados para validação (período chuvoso entre 2018 e 2019) podem ter motivado esses resultados. A avaliação gráfica indica problemas relacionados à deficiência no monitoramento da precipitação. Entre maio e junho de 2017, por exemplo, o monitoramento fluviométrico registrou picos de vazões, enquanto não houve registro de precipitação. Isso deixa claras as dificuldades envolvidas na simulação de pequenas bacias, e a dependência de um monitoramento adequado.

Figura 60. Resultados referentes à área CP2 correspondentes à população (100 indivíduos) da última geração do algoritmo de otimização (NSGA-II). Linhas vermelhas e amarelas diferenciam os períodos utilizados para calibração (05/2017 a 09/2018) e validação (09/2018 a 02/2019). Cada linha (total de 100) indica um resultado da simulação para um conjunto de parâmetros diferente.

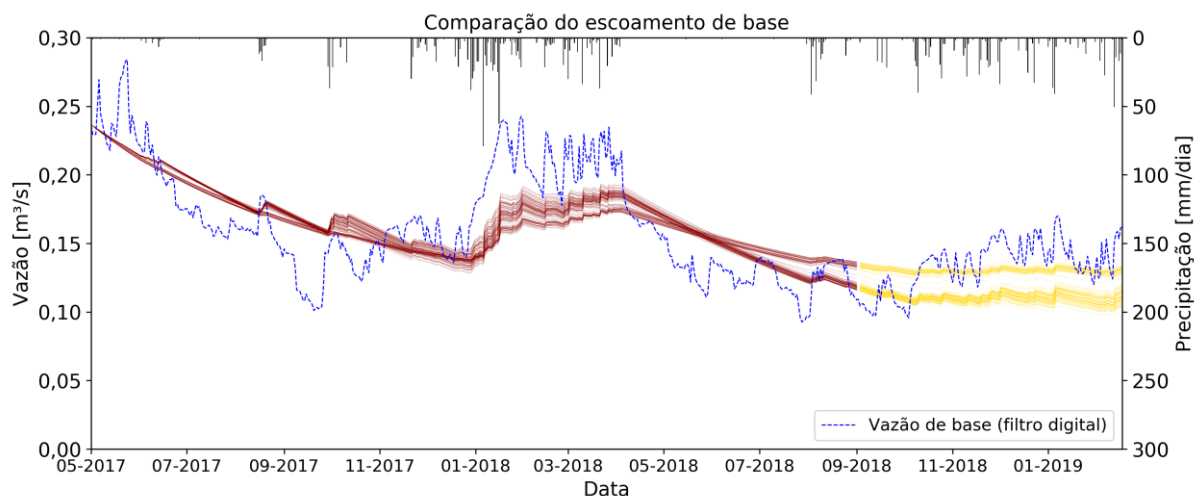


Fonte: Autor.

O escoamento de base estimado pelo filtro digital apresenta um comportamento mais variável e que o modelo não consegue reproduzir (Figura 61). Possivelmente, essa alta variabilidade produzida pelo filtro digital ocorre pela inclusão de parte da componente do escoamento sub-superficial, prejudicando a comparação entre os resultados. Assim, métodos alternativos de separação do escoamento de base (e.g., balanço de massa isotópico ou geoquímico) devem ser utilizados em trabalhos futuros para buscar uma estimativa mais adequada e orientar a reprodução dessa componente (uma dificuldade a ser abordada nesses eventuais trabalhos é a necessidade de amostragem de água durante eventos moderados de precipitação).

Outra diferença na aplicação do modelo na área CP2 está relacionada a maior dispersão dos resultados, possivelmente em função da menor extensão das séries de dados observados, o que dificulta a identificação dos parâmetros do modelo (maior incerteza paramétrica).

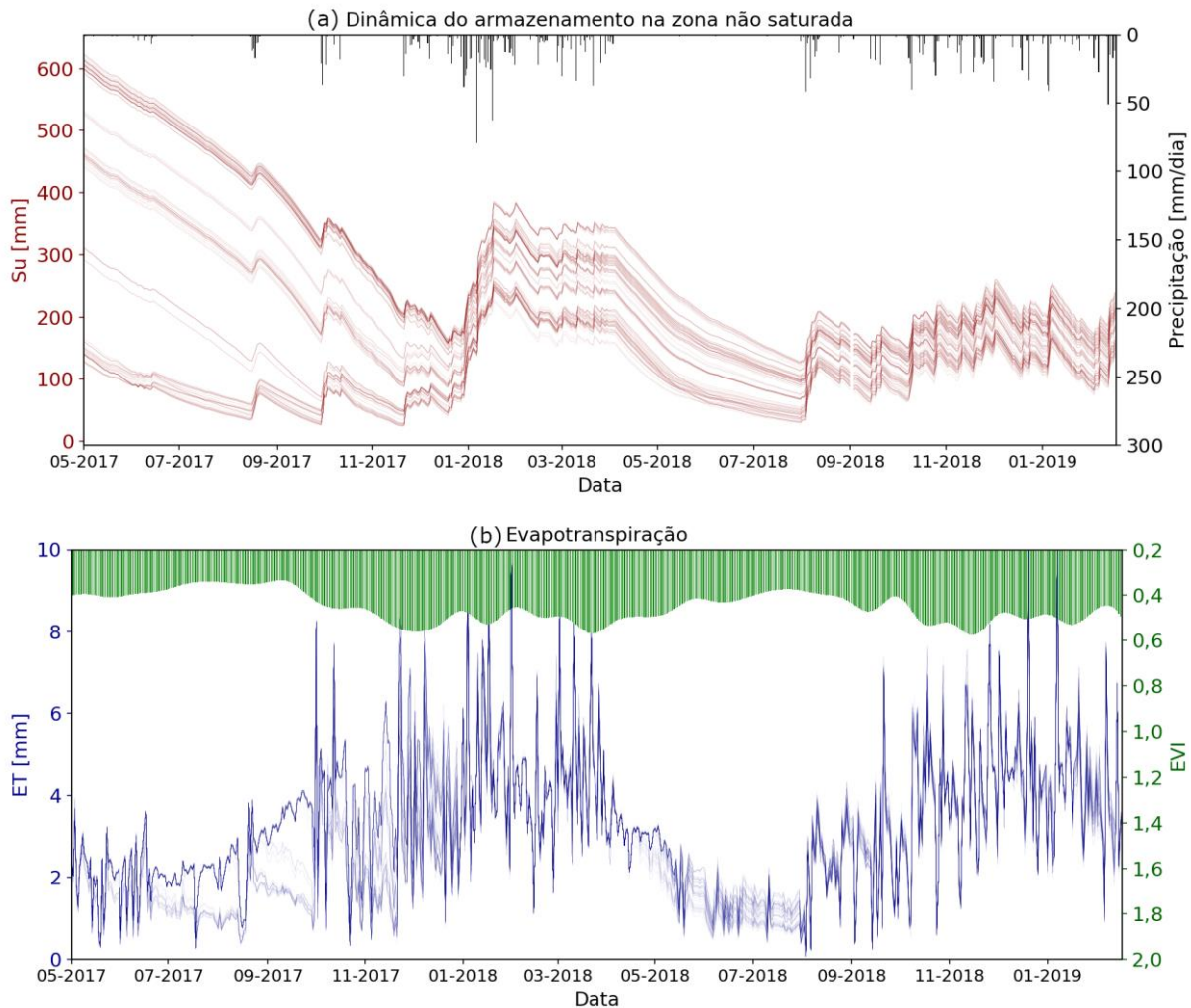
Figura 61. Escoamento de base simulado e estimado pelo filtro digital referente à área CP2. As cores vermelha e amarela indicam os períodos de calibração e validação, respectivamente. Cada linha (total de 100), vermelha ou amarela, indica um resultado da simulação para um conjunto de parâmetros diferente.



Fonte: Autor.

Ao avaliar a dinâmica do armazenamento na zona não saturada e de estimativas de evapotranspiração, a questão de incerteza paramétrica e das condições iniciais fica mais evidente (Figura 62) (resultados não sobrepostos, como ocorreu com dados da seção SM). A taxa de evapotranspiração estimada pelo modelo (05/2017 a 02/2019) resultou em 3,27 mm/d (média), das quais, aproximadamente, 16% correspondem a perdas por interceptação.

Figura 62. Avaliação de características simuladas na área CP2 relacionadas à disponibilidade de água no solo e à evapotranspiração. Cada linha (total de 100) indica um resultado da simulação para um conjunto de parâmetros diferente.



onte: Autor.

Os valores dos parâmetros calibrados para as áreas SM e CP2 são apresentados na Tabela 21. Grandes incertezas estão presentes nos processos dependentes dos parâmetros  $c_H$ ,  $c_{VI}$  e  $R_P$ , dada as diferenças entre os parâmetros calibrados de forma independente com diferentes conjuntos de dados observados. Na área SM, os parâmetros calibrados indicam que a evapotranspiração tende a ser maior, a percolação profunda, menor, e a ascensão máxima de água subterrânea, menor, quando comparadas aos processos correspondentes que ocorrem na área CP2 (área de montante da SM). Os parâmetros  $\beta$ ,  $LP$ ,  $FC$ ,  $k_{SS}$ ,  $a$  e  $b$  apresentaram convergência ou comportamento dentro do esperado dadas as características físicas de cada uma das áreas. O parâmetro  $\delta_N$  influencia pouco nos resultados e os demais possuem significado local ( $k_f$ ,  $n_f$ ,  $\alpha$ ).

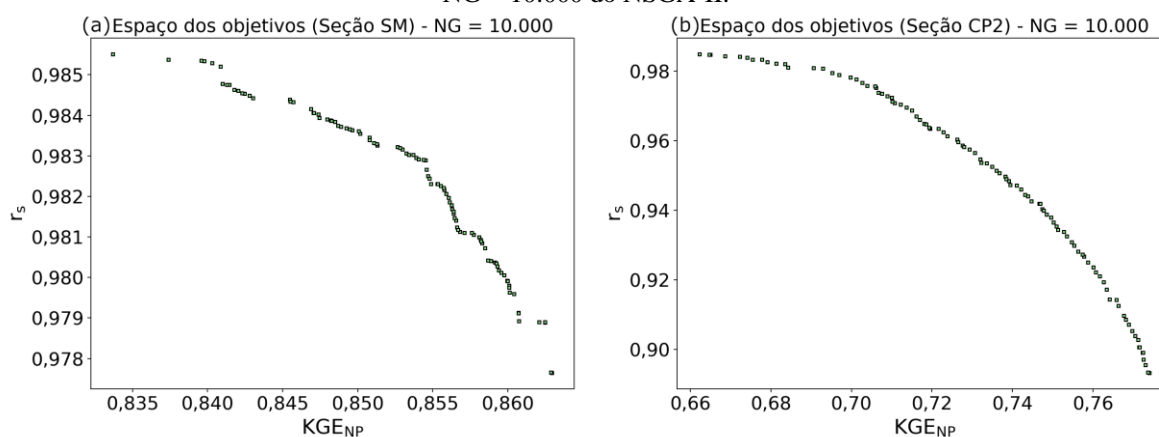
Tabela 21. Resultados da calibração. Valores indicados são as medianas dos 100 vetores de parâmetros calibrados resultantes de uma execução de otimização com 10.000 gerações.

Parâmetro	Processo relacionado	Mediana - Área SM	Mediana - Área CP2
$\delta_N$	Precipitação efetiva	0,01	-0,02
$\beta$	Recarga subterrânea	1,22	1,22
$c_H$	Ascensão máxima de água subterrânea [mm/d]	0,02	0,37
LP	Evapotranspiração	0,23	0,26
FC	Infiltração, Evapotranspiração, Recarga e outros [mm]	222	589
$k_{ss}$	Escoamento subsuperficial	0,052	0,090
$c_{VI}$	Evapotranspiração	0,134	0,216
$R_p$	Percolação profunda [mm/dia]	0,156	0,248
$a$	Escoamento de base	0,00177	0,00239
$b$		1,47	1,58
$k_f$	Convolução do escoamento superficial	2,15	4,69
$n_f$	direto	1,58	4,03
$\alpha$	Umidade inicial na zona não saturada	0,58	0,77

Fonte: Autor.

O espaço das funções objetivo ao final da etapa de calibração (100 soluções, com número de gerações  $NG = 10.000$ ) é apresentado na Figura 63. Os padrões de dominação observados para a SM e CP2 (mais regular) sugerem que o modelo não apresenta problemas evidentes na estrutura ou na parametrização das equações de fluxo (EFSTRATIADIS; KOUTSOYIANNIS, 2010).

Figura 63. Valores das funções objetivo (fase de calibração) referentes às soluções não dominadas na geração  $NG = 10.000$  do NSGA-II.



Fonte: Autor.

#### 4.9.1. Comparação com trabalhos anteriores de modelagem na área de estudo, avaliação geral do modelo proposto e limitações

A simplicidade de representação dos processos no modelo, o uso combinado de critérios baseados na dinâmica do escoamento no exutório da bacia e do armazenamento subterrâneo, a inclusão de dados que informam sobre a dinâmica temporal do desenvolvimento da vegetação, e a consistência em diferentes escalas espaciais (19,2 km<sup>2</sup> e 52,1 km<sup>2</sup>) são algumas características que diferenciam o modelo proposto de trabalhos anteriores na área de estudo.

Machado, Wendland e Krause (2016) adotaram um modelo fisicamente baseado e distribuído, J2000 (KRAUSE, 2001), para simular a vazão de saída da área SM em escala diária e melhorar as avaliações do balanço hídrico na área de estudo. Com 26 parâmetros, 976 unidades de resposta hidrológica, e usando dados observados entre 2008 e 2011, o modelo apresentou bom desempenho, com NSE de 0,76 e 0,81 e RMSE de 0,140 e 0,180 m<sup>3</sup>/s para períodos de calibração e validação, respectivamente. Pequenas oscilações no hidrograma observado também não são reproduzidas, assim como no modelo aqui proposto. Esse problema pode estar relacionado à variabilidade da precipitação e a deficiências no sistema de monitoramento pluviométrico da área de estudo.

Meira Neto et al. (2018) utilizaram a ferramenta SWAT (*Soil and Water Assessment Tool*) (ARNOLD et al., 1998) para avaliar a influência de diferentes conjuntos de dados de precipitação na reprodução das vazões no exutório da bacia SM em escala diária. Para o caso com melhor desempenho geral, obtido por meio da ponderação Bayesiana de diferentes resultados de simulação, foram encontrados NSE (e percentual de viés) de 0,82 (0,1%) e 0,65 (-6,8%) para os períodos de calibração e validação, respectivamente (dados entre 10/2008 e 05/2012). As possíveis inadequações do modelo criado no SWAT não foram abordadas e os erros e incertezas paramétricas foram atribuídos exclusivamente à série de precipitação utilizada.

Coutinho (2019) simulou o escoamento subterrâneo na bacia em escala mensal com um modelo numérico discretizado em elementos finitos e implementado no pacote computacional SPA (Simulador de Processos em Aquíferos) (WENDLAND, 2003). Por representar a variabilidade dos níveis freáticos observados, esse modelo é o único, em comparação com os demais, que considera a variabilidade do armazenamento subterrâneo. Para a reprodução dos níveis freáticos dos poços, as métricas de desempenho  $KG_{ENP}$  ficaram entre 0,33 e 0,75 na fase de calibração e 0,72 a 0,92 na validação, com dados de 2004 a 2018. Para as condições em que os níveis freáticos são calibrados, o escoamento de base simulado, comparado com o estimado a partir das vazões observadas, apresentou um  $KG_{ENP}$  de 0,66 na área SM (período de 12/2008

a 09/2018) e de 0,93 na área CP2 (período de 02/2018 a 08/2018; pequena ocorrência de precipitação). Os resultados indicam que o modelo fisicamente baseado de escoamento subterrâneo é adequado para simular o processo de geração do escoamento na bacia, mesmo com os problemas de heterogeneidade espacial, na definição da geometria dos volumes de controle e as incertezas paramétricas e nas condições de contorno.

Quanto à avaliação das defasagens, os resultados aqui obtidos concordam com os resultados de Gómez et al. (2018) e Gómez, Wendland e Melo (2019), que estimaram uma defasagem da ordem de 2 a 3 anos entre a precipitação e o escoamento de base na área de estudo e em bacias maiores aninhadas. Nesses trabalhos, as defasagens entre as séries temporais (de precipitação e escoamento de base) foram estimadas a partir da análise do espectro de potência global *wavelet*.

O desempenho do modelo conceitual proposto (Tabela 20) é compatível com o de trabalhos anteriores. A fraca capacidade de identificação de alguns parâmetros (e.g.,  $\delta_N$ ) e as incertezas paramétricas (e.g.,  $c_H$ ,  $c_{VI}$ ,  $R_P$ ) são algumas deficiências do modelo. Os parâmetros relacionados à percolação profunda e à evapotranspiração, em especial, apresentaram uma clara condição de equifinalidade (diferentes combinações levando a resultados similares), que garantem o balanço hídrico dos reservatórios da zona não saturada e da zona saturada da bacia, mas representam processos diferentes. Para a área de estudo, que é uma área de afloramento do Sistema Aquífero Guarani, uma quantificação menos incerta do processo de recarga profunda e a possibilidade de extrapolação do método são de grande interesse para fins práticos de gestão de longo prazo e para desenvolvimento de políticas de uso dos recursos hídricos dessa importante reserva hídrica regional.

Para avançar nessa questão, a avaliação de variáveis de estado tanto na zona não saturada quanto na zona saturada parece ser fundamental. Neste estudo, a avaliação de variações do armazenamento subterrâneo foi feita a partir do coeficiente de correlação de Spearman (com a variável de estado da zona saturada do modelo e a média aritmética dos níveis freáticos observados na bacia). Em trabalhos futuros, essa avaliação pode tentar incluir, além da correlação, quantidades absolutas. Características hidrogeológicas medidas em campo, como o rendimento específico médio, entre 8% e 17% na área de estudo (GOMES, 2008), podem contribuir para isso. Por outro lado, o elevado intervalo de variação desse coeficiente pode indicar que a correlação seja, de fato, uma métrica mais adequada a ser utilizada.

A umidade do solo está relacionada à distribuição de grande parte dos fluxos que entra e sai do sistema hidrológico. Assim, estimativas de umidade do solo são essenciais para reduzir as incertezas do modelo e para informar sobre a variabilidade a ser seguida pela variável de

estado  $S_u$ . A ausência do monitoramento de longo termo e a falta de instrumentação local para coleta de dados de umidade do solo dificultaram a avaliação da adequação do modelo na zona não saturada. Nessas condições, técnicas alternativas de transferência de informação de estudos experimentais realizados em áreas próximas e uso de dados de sensoriamento remoto precisam ser exploradas. Simular o período para o dados de variação de umidade estão disponíveis em 13 pontos da área de estudo em profundidades de 2,5 m a 5,0 m, em um estudo experimental conduzido por Pompeo (1990) entre 1985 e 1988 na área de estudo, é outra opção para tentar aperfeiçoar a proposição de um modelo conceitual para a bacia.

Outro problema está relacionado à definição da geometria do volume de controle e dos fluxos existentes. Com auxílio de poços de monitoramento a leste do divisor topográfico da bacia, estudos anteriores (MELO, 2013; COUTINHO, 2019) indicaram a existência de inversões no sentido do fluxo subterrâneo, para dentro ou fora da bacia, onde há uma continuidade de plantações de eucalipto (Figura 9). Essa situação pode ocorrer em outros pontos e levar a compensações entre os processos hidrológicos modelados, evidenciando a complexidade e as incertezas dos resultados simulados mesmo de uma pequena bacia.

#### **4.10 Avaliação do comportamento do modelo**

As etapas anteriores, de caracterização das relações entre os processos e de modelagem, foram conduzidas de forma independente. Esta etapa teve como objetivo verificar se o modelo proposto é adequado para reproduzir alguns dos comportamentos identificados na etapa de caracterização do sistema hidrológico, especificamente as dependências entre variáveis ao longo do tempo.

Como o teste de Granger mostrou-se pouco informativo na etapa de caracterização (em geral, apenas confirmando a existência de relações causais para qualquer defasagem), apenas a informação mútua normalizada com defasagem (NMI) foi adotada para essa análise. Pela maior disponibilidade de dados, apenas a área SM foi considerada. Os resultados das relações observadas e simuladas são apresentados nas figuras 64 e 65.

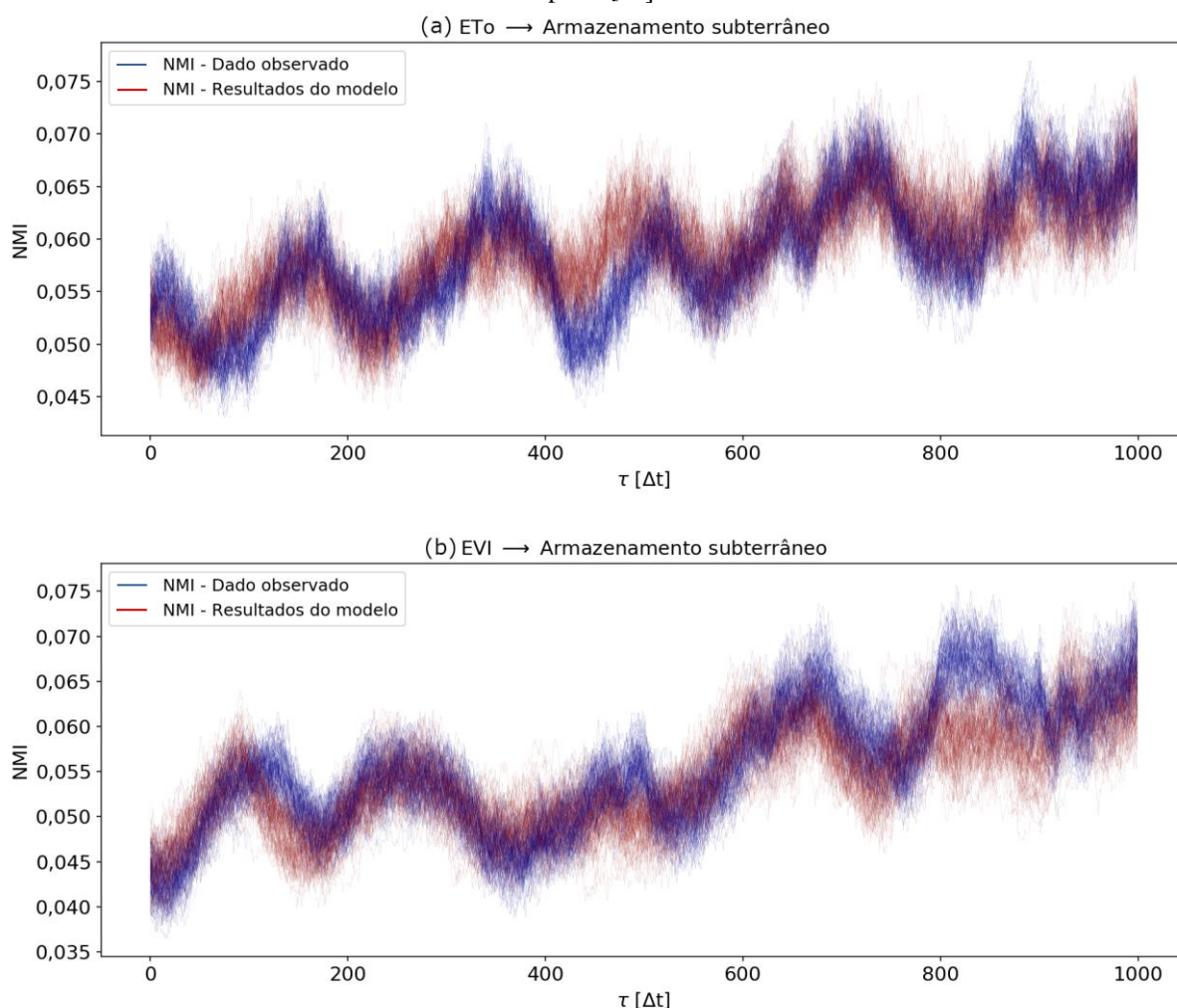
A proximidade das curvas referente às conectividades entre dados observados e resultados simulados sugere que o modelo está reproduzindo as dependências temporais de forma adequada.

As relações entre índice de vegetação (EVI), evapotranspiração de referência, precipitação e armazenamento subterrâneo são bem reproduzidas pelo modelo proposto



(Figura 64). Pela conexão existente entre os processos atmosféricos e os processos que ocorrem em superfície e em subsuperfície, a evapotranspiração, assim como a precipitação, possui grande importância para reproduzir a dinâmica do sistema hidrológico. Portanto, essa importância vai além de contribuir na estimativa de processos isolados, como indicado em trabalhos anteriores na área de estudo (LUCAS; WENDLAND, 2016; MATTOS et al., 2019).

Figura 64. Avaliação do modelo na área SM para relações entre evapotranspiração de referência, índice de vegetação e armazenamento subterrâneo. Gráficos indicam as 100 soluções da última geração da otimização. Escala temporal  $[\Delta t]$ : diária.



Fonte: Autor.

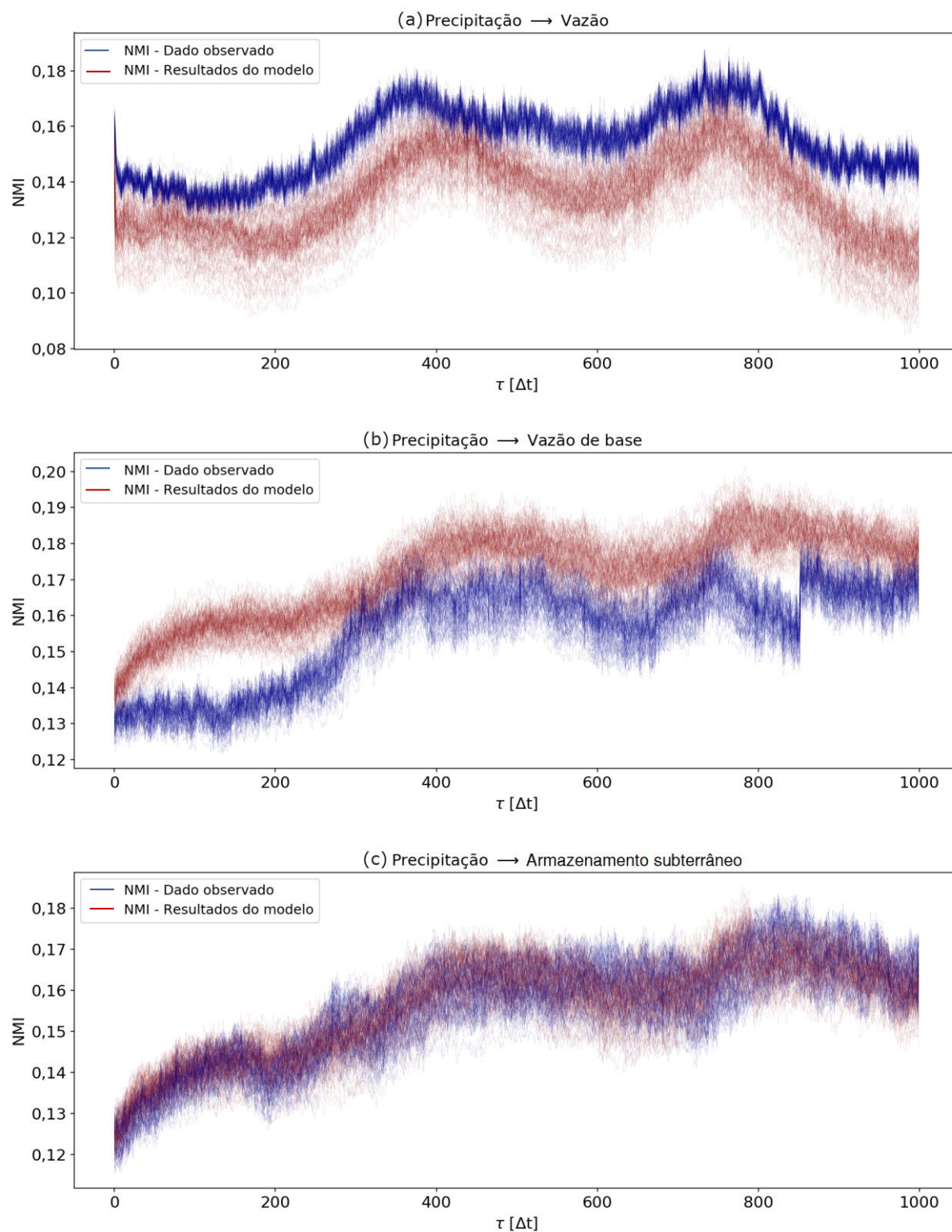
As diferenças em valores absolutos de NMI das análises entre precipitação e vazão observada e vazão simulada podem indicar algumas das deficiências do modelo, porém os padrões de variação são aproximadamente os mesmos. As relações entre precipitação e armazenamento subterrâneo são bem reproduzidas (Figura 65). Por outro lado, o processo de

geração do escoamento de base parece não corresponder ao de geração dos dados observados. Isso sugere a existência de inconsistências em métodos de separação de escoamento de base baseados na separação de resposta rápida e lenta (filtros digitais), que foi o método aqui utilizado para estimar o escoamento de base, e o método proposto no modelo conceitual para estimar essa componente. Essa constatação abre possibilidades de proposição e avaliação de métodos alternativos que preservem a dinâmica esperada do escoamento de base. Esse aspecto contraditório, de boa reprodução do armazenamento subterrâneo e má reprodução do escoamento de base, deve ser melhor explorado futuramente.

A entropia de transferência (com defasagem de uma unidade, *lag-1*) entre componentes do balanço hídrico aplicando diferentes modelos em bacias hidrográficas localizadas no Noroeste dos Estados Unidos foi quantificada por Bennett et al. (2019). Esses autores concluíram que a análise permitiu um entendimento mais completo do comportamento do modelo, e que essa estratégia possui potencial para guiar a seleção e o desenvolvimento de modelos hidrológicos, e reduzir incertezas estruturais e preditivas. Futuros trabalhos podem explorar e comparar a efetividade da ferramenta aqui proposta, baseada na informação mútua com defasagem, e da entropia de transferência para justificar algumas decisões de modelagem.

Pela Figura 65, pode-se perceber que os picos da NMI são reproduzidos praticamente com as mesmas defasagens dos dados observados (vazão e armazenamento subterrâneo). Isso indica que a dinâmica do sistema hidrológico, incluindo fluxos e armazenamentos, possivelmente está sendo reproduzida de forma adequada pelo modelo conceitual proposto. Mais importante, as origens dos efeitos de memória ou persistência reportados em alguns estudos hidrológicos (PELLETIER; TUCOTTE, 1997; ZIMMERMANN; ELSENBEER; MORAES, 2006; TOMASELLA et al., 2008; DEY; MUJUMDAR, 2018) parecem ficar esclarecidas. Esses efeitos são produtos da dinâmica hidrológica, e não requerem estruturas específicas dentro do modelo que forcem a reprodução de dependências no domínio temporal. Isso ocorre possivelmente porque o histórico de condições anteriores é considerado e atualizado pelas variáveis de estado do modelo.

Figura 65. Avaliação do modelo na área SMF para relações entre precipitação, vazão, vazão de base e armazenamento subterrâneo. Gráficos indicam as 100 soluções da última geração da otimização. Escala temporal: diária.



Fonte: Autor.

#### **4.11 Contribuição para prever a dinâmica da disponibilidade hídrica**

O modelo proposto é uma ferramenta que a partir de um conjunto de condições iniciais e de dados de entrada (precipitação, temperatura e umidade relativa do ar, radiação solar, velocidade do vento, EVI) permite estimar as respostas hidrológicas em termos de vazão e de armazenamento subterrâneo. Portanto, a previsão da dinâmica da disponibilidade hídrica depende da proposição de cenários de simulação e das respectivas estimativas desses dados de entrada. Eventuais mudanças nos regimes de precipitação, temperatura média do ar ou no uso do solo, por exemplo, podem ser consideradas em simulações de interesse no contexto da gestão de recursos hídricos.

Gómez, Wendland e Melo (2019) propuseram um modelo empírico linear capaz de prever o escoamento de base médio mensal a partir da média móvel da precipitação e do tempo médio de resposta de bacias. Esse modelo apresenta vantagens em situações de anomalias ou de não-estacionariedades em séries de precipitação quando comparado a métodos tradicionais, baseados na frequência de uma vazão ser igualada ou superada (e.g., curva de permanência). Entretanto, por não incluir outros dados, associados ao uso do solo, à dinâmica da vegetação ou a condições de evapotranspiração, por exemplo, o modelo pode apresentar inconsistências.

Duethmann, Blöschl e Parajka (2020), por exemplo, abordam as limitações de um modelo hidrológico conceitual (HBV) em simular respostas para condições diferentes das quais foram calibrados, incluindo, em especial, diferentes condições climáticas. A deficiência relacionada a não inclusão da dinâmica da vegetação no modelo, em especial, mostrou-se relevante nos casos analisados. Em adição, os resultados obtidos confirmaram a importância de considerar no modelo as inter-relações entre mudanças no clima, na vegetação e nas respostas hidrológicas.

Ao abordar algumas dessas deficiências, a modelagem dinâmica do sistema hidrológico proposta neste trabalho tem potencial para contribuir na produção de estimativas mais acuradas da disponibilidade hídrica em situações variadas, envolvendo tanto mudanças no uso do solo quanto variações em padrões meteorológicos. Estas questões envolvendo mudanças, entretanto, precisam ser exploradas em estudos específicos.

## 5 CONCLUSÕES

Uma estratégia flexível de modelagem hidrológica conceitual concentrada foi conduzida com foco no aproveitamento da diversidade de dados disponíveis na área de estudo. A possibilidade de unir informações de campo e de modelagem, resultantes de trabalhos anteriores, motivou o desenvolvimento de um modelo próprio, que inclui alguns aspectos relevantes, como, por exemplo, a recarga regional do Sistema Aquífero Guarani (SAG) e a dinâmica da vegetação.

Previamente ao desenvolvimento do modelo, as componentes do balanço hídrico foram estimadas em escala diária em duas superfícies de controle da área de estudo (SM, CP2). Os dados de precipitação e de evapotranspiração de referência (FAO-56) foram interpolados espacialmente usando o método IDW e os valores médios utilizados. Valores médios do índice de vegetação EVI foram determinados para a área de estudo e disponibilizados como dado de entrada do modelo para estimar os processos de interceptação vegetal e evapotranspiração.

As vazões nas seções fluviométricas foram estimadas a partir da transformação de dados de nível de água. Essa transformação demandou métodos dinâmicos, variáveis no tempo, para lidar com questões de não-estacionariedade, condicionadas por alterações no regime de escoamento. O método proposto é um resultado de interesse na área de hidrometria, que pode ser explorado e mais bem avaliado com dados de outros locais que apresentam situações complexas de escoamento (e.g., alterações nas condições geométricas e de contorno), visando adicionais verificações, adequações e aperfeiçoamento.

A consistência física do modelo conceitual proposto para a bacia de estudo foi avaliada a partir de métricas estatísticas não paramétricas que informam sobre o fluxo de informação e dependência entre variáveis do sistema hidrológico no domínio temporal. Na etapa de calibração e validação convencional, o modelo apresentou um desempenho adequado na reprodução de regimes de baixas vazões e na reprodução da variabilidade do armazenamento subterrâneo. Após essas etapas, o teste de Granger e a informação mútua com defasagem foram utilizados para comparar o comportamento do sistema observado e do sistema simulado. O modelo proposto conseguiu reproduzir o comportamento do sistema observado de forma razoável (picos de informação mútua em defasagens aproximadas), evidenciando estatisticamente estar adequado para reproduzir a dinâmica da disponibilidade hídrica da bacia em termos de vazão e em termos de armazenamento subterrâneo. Isso atende ao objetivo principal do trabalho e à contribuição pretendida de reproduzir as interdependências no domínio temporal.

O modelo matemático baseado em um sistema de equações diferenciais ordinárias e em condições iniciais mostrou-se suficiente para reproduzir as defasagens entre variáveis do sistema hidrológico. Essa, em conjunto com a estrutura do modelo e a forma de calibração multi-objetivo, talvez seja a contribuição mais importante do trabalho, indicando que uma adequada representação dos processos e calibração dos parâmetros, pode resultar em uma adequada representação de outras características do sistema. Indiretamente, o trabalho demonstra que não é necessário propor uma estrutura que force a reprodução de interdependências entre variáveis, uma vez que isso é atingido de forma natural.

O teste de Granger e a informação mútua com defasagem foram implementados e testados em problemas sintéticos, o que permitiu tanto garantir a adequada implementação computacional quanto explorar o comportamento dessas ferramentas estatísticas sob diferentes condições de aplicação. O principal destaque desta etapa é que a aplicação de técnicas de suavização e de agregação temporal podem prejudicar a capacidade de identificação e caracterização de relações causais. Essas ferramentas também foram empregadas para caracterizar relações de conectividade e dependência entre variáveis observadas ou estimadas, fornecendo informações adicionais sobre o comportamento do sistema hidrológico da área de estudo e que poderão ser exploradas em trabalhos futuros. O teste de Granger confirmou a causalidade entre os processos hidrológicos na maioria dos casos. Com isso, o teste de Granger resultou pouco informativo e a informação mútua com defasagem foi utilizada como critério na etapa final de avaliação do modelo.

Entre as principais limitações deste trabalho destaca-se a ausência de comparações de modelos alternativos para testar a eficácia da avaliação baseada na caracterização de conectividades e dependências entre variáveis hidrológicas no domínio temporal. O uso de modelos intencionalmente inadequados, por exemplo, poderia comprovar, ou não, a consistência desse tipo de avaliação. Além disso, métodos não-paramétricos (com menos exigências quanto ao comportamento estatístico dos dados) e métricas mais robustas de caracterização de sistemas dinâmicos, como a entropia de transferência (informação mútua condicionada), podem ser comparadas como critério de avaliação de modelos.

De natureza mais específica, outras limitações estão relacionadas à incapacidade do modelo proposto em reduzir as incertezas relacionadas à recarga do aquífero regional (SAG) e à evapotranspiração. A avaliação de variáveis de estado, em especial das variações dos teores de umidade do solo e das variações de níveis freáticos, parece ser essencial para reduzir essas e outras incertezas do modelo. Em adição, esquemas de calibração conjuntos, envolvendo as

duas seções de monitoramento fluviométrico, e alterando os intervalos de variação dos parâmetros mais incertos, podem contribuir na redução de incertezas paramétricas.

O modelo aqui proposto é uma pequena contribuição, que pode ser aproveitada, aperfeiçoada e adaptada para diversos fins, incluindo o uso em modelos semi-distribuídos, e acoplado a modelos de escoamento em superfície ou subsuperfície. Novas concepções de parametrização e de esquemas numéricos certamente seriam necessárias. Para estudos em grande escala, a calibração multi-objetivo aqui adotada poderia enfrentar alguns problemas de disponibilidade de dados de níveis freáticos. Para superar isso, dados do GRACE (*Gravity Recovery and Climate Experiment*), por exemplo, poderiam ser utilizados para avaliar as variações do armazenamento total de água. Por fim, a redução de incertezas na quantificação de componentes do balanço hídrico e nas parametrizações das equações de fluxo certamente depende da continuidade das atividades de monitoramento na bacia, da otimização da instrumentação existente e da instalação de novos instrumentos para avaliar as variações do teor de umidade em diferentes pontos da área de estudo. Em futuros trabalhos na área de estudo, a influência do método de interpolação espacial de dados de precipitação precisa ser avaliada a fim de verificar a existência ou não de um processo mais adequado. Além disso, estudos específicos envolvendo traçadores isotópicos e geoquímicos na bacia é outra demanda, que pode contribuir na redução de incertezas relacionadas à dinâmica do escoamento de base e à estimativa das origens espaciais e temporais da água que escoam nos cursos de água.

O aperfeiçoamento de ferramentas de previsão e a redução das incertezas de predição frente às mudanças globais, e ameaças associadas, é um passo importante para garantir uma exploração sustentável dos recursos disponíveis em escalas locais e regionais. Como transformar os avanços técnicos em benefícios práticos ao desenvolvimento de políticas de uso e gestão dos recursos hídricos e de uso e ocupação do solo é outro desafio, que precisa ser abordado e considerado ao longo do processo de desenvolvimento de novos modelos.





## REFERÊNCIAS BIBLIOGRÁFICAS

- ABTEW, W.; MELESSE, A. **Evaporation and Evapotranspiration: measurements and estimations**. Heidelberg: Springer, 2012.
- ADDOR, N.; MELSEN, L. A. Legacy, Rather Than Adequacy, Drives the Selection of Hydrological Models. **Water Resources Research**, v. 55, p. 378–390. 2019.
- AGÊNCIA NACIONAL DE ÁGUAS (ANA). **Plano Nacional de Segurança Hídrica**. Brasília: ANA, 2019.
- AKANSU, A. N.; HADDAD, R. A. **Multiresolution Signal Decomposition**. 2nd ed., Cambridge: Academic Press, 2001.
- ALCÂNTARA, G. M. **Modelagem dos impactos hidrológicos decorrentes de alterações do uso e ocupação do solo em bacia hidrográfica rural**. 157 f. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2018.
- ALLEN, R. G. et al. **Crop evapotranspiration**. FAO Irrigation and Drainage Paper n. 56, Rome: FAO - Food and Agriculture Organization of the United Nations, 1998.
- ALVARES, C. A. et al. Koppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p.711–728, 2014.
- AMBLARD, P. O.; MICHEL, O. J. J. The relation between Granger causality and directed information theory: a review. **Entropy**, v. 15, p. 113–143, 2013.
- ANACHE, J. A. A. et al. Hydrological trade-offs due to different land covers and land uses in the Brazilian Cerrado. **Hydrology and Earth System Sciences**, v. 23, p. 1263–1279, 2019.
- ANDERSON, M. C. et al. A two-source time-integrated model for estimating surface fluxes using thermal infrared remote sensing. **Remote Sensing of Environment**, v. 60, n. 2, p. 195–216, 1997.
- ALVES SILVA, A. S. et al. Comparison of Interpolation Methods for Spatial Distribution of Monthly Precipitation in the State of Pernambuco, Brazil. **Journal of Hydrological Engineering**, v. 24, n. 3, p. 1–11, 2019.
- ARAÚJO, L. M. N. **Identificação de padrões hidrológicos de precipitação e de umidade do solo na bacia hidrográfica do rio Piabanha/RJ**. 250 f. Tese (Doutorado) – COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2016.
- ARNOLD, J. G. et al. Large area hydrologic modeling and assessment – Part I: model development. **Journal of the American Water Resources Association**, v. 34, n. 1, p. 73–89, 1998.
- BARNETT, L.; BARRETT, A. B.; SETH, A. K. Granger causality and transfer entropy are equivalent for gaussian variables. **Physical Review Letters**, v. 103, n. 23, p. 2–5, 2009.
- BARNETT, L.; SETH, A. K. Granger causality for state-space models. **Physical Review E - Statistical, Nonlinear, and Soft Matter Physics**, v. 91, n. 4, p. 1–5, 2015.
- BATISTA, L. V. et al. Groundwater and surface water connectivity within the recharge area

of Guarani aquifer system during El Niño 2014–2016. **Hydrological Processes**, v. 32, n. 16, p. 2483–2495, 2018.

BENNETT, A. et al. Quantifying Process Connectivity With Transfer Entropy in Hydrologic Models. **Water Resources Research**, v. 55, n. 6, p. 4613–4629, 2019.

BERGSTRÖM, S.; FORSMAN, A. Development of a conceptual deterministic rainfall-runoff model. **Nordic Hydrology**, v. 4, p. 147–170, 1973.

BEVEN, K. Towards integrated environmental models of everywhere: uncertainty, data and modelling as a learning process. **Hydrology and Earth System Sciences**, v. 11, p. 460–467, 2007.

\_\_\_\_\_. Causal models as multiple working hypotheses about environmental processes. **Comptes Rendus - Geoscience**, v. 344, n. 2, p. 77–88, 2012.

\_\_\_\_\_. Towards a methodology for testing models as hypotheses in the inexact sciences. **Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 475, n. 2224, 2019.

BIONDI, D. et al. Validation of hydrological models: Conceptual basis, methodological approaches and a proposal for a code of practice. **Physics and Chemistry of the Earth**, v. 42–44, p. 70–76, 2012.

BLÖSCHL, G. et al. Twenty-three unsolved problems in hydrology (UPH) – a community perspective. **Hydrological Sciences Journal**, v. 64, n. 10, p. 1141–1158, 2019.

BURT, T. P.; MCDONNELL, J. J. Whither field hydrology? The need for discovery science and outrageous hydrological hypotheses. **Water Resources Research**, v. 51, n. 8, p. 5919–5928, 2015.

CABRAL, O. M. R. et al. The energy and water balance of a Eucalyptus plantation in southeast Brazil. **Journal of Hydrology**, v. 388, p. 208–216, 2010.

CABRERA, M. C. M. et al. Performance of evaporation estimation methods compared with standard 20 m<sup>2</sup> tank. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 20, n. 10, p. 874–879, 2016a.

CABRERA, M. C. M. **Valoração do impacto ambiental de cavas de mineração de areia e argila com base na evapotranspiração**. 170 f. Tese (Doutorado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2016b.

CARTWRIGHT, I.; MORGENSTERN, U. Using tritium and other geochemical tracers to address the “old water paradox” in headwater catchments. **Journal of Hydrology**, v. 563, p. 13–21, 2018.

CARVAJALINO, J. C. D. **Avaliação de estimativas de evapotranspiração dos métodos baseados na flutuação diária do nível piezométrico comparando com o método de balanço de energia - Razão de Bowen em zona ripária**. 137 f. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2019.

CASTRO JÚNIOR, P. R.; WENDLAND, E. **Mapeamento morfopedológico aplicado à bacia-piloto do Ribeirão da Onça (SP) em áreas de recarga do Aquífero Guarani**. São

Carlos: EESC/USP, 2015. Relatório de Pós-Doutoramento do Programa Nacional de Cooperação Acadêmica (CAPES).

CHIEW, F. H. S.; PEEL, M. C.; WESTERN, A. W. Application and testing of the simple rainfall–runoff model SIMHYD. In: SINGH, V. P.; FREVERT, D. K. (Ed.). **Mathematical Models of Small Watershed Hydrology and Applications**. Littleton (EUA): Water Resources Publications, 2002, cap. 11, p. 335–367.

CHRISTINA, M. et al. Importance of deep water uptake in tropical eucalypt forest. **Functional Ecology**, v. 31, n. 2, p. 509–519, 2017.

CLARK, M. P. et al. A unified approach for process-based hydrologic modeling: 1. Modeling concept. **Water Resources Research**, v. 51, n. 4, p. 2498–2514, 2015.

CLARK, M. P.; KAVETSKI, D. Ancient numerical daemons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes. **Water Resources Research**, v. 46, n. 10, p. 1–23, 2010.

COELHO, C. A. S. et al. The 2014 southeast Brazil austral summer drought: regional scale mechanisms and teleconnections. **Climate Dynamics**, v. 46, n. 11–12, p. 3737–3752, 2016.

COLLISCHONN, W.; FAN, F. M. Defining parameters for Eckhardt’s digital baseflow filter. **Hydrological Processes**, v. 27, p. 2614–2622, 2013.

COUTINHO, J. V. **Caracterização geofísica e modelagem do escoamento subterrâneo em área de afloramento do Sistema Aquífero Guarani**. 151 f. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2019.

COVER, T. M.; THOMAS, J. A. **Elements of Information Theory**. 2nd ed. New Jersey: John Wiley & Sons, 2006.

DAVID, J. S.; VALENTE, F.; GASH, J. H. Evaporation of Intercepted Rainfall. **Encyclopedia of Hydrological Sciences**, p. 1–8, 2005.

DEB, K. et al. A fast and elitist multiobjective genetic algorithm: NSGA-II. **IEEE Transactions on Evolutionary Computation**, v. 6, n. 2, p. 182–197, 2002.

DETTO, M. et al. Causality and Persistence in Ecological Systems: A Nonparametric Spectral Granger Causality Approach. **The American Naturalist**, v. 179, n. 4, p. 524–535, 2012.

DETZEL, D. H. M.; FERNANDES, C.V.S.; MINE, M.R.M. Não Estacionariedade na Construção de Curvas de Permanência com Vistas à Outorga de Recursos Hídricos. **Revista Brasileira de Recursos Hídricos**, v. 21, p. 80–87, 2016.

DEY, P.; MUJUMDAR, P. P. Multiscale evolution of persistence of rainfall and streamflow. **Advances in Water Resources**, v. 121, p. 285–303, 2018.

DIDAN, K. **MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250 m SIN Grid V006**. Sioux Falls: NASA EODIS Land Processes DAAC, 2015.

DINGMAN, S.L. **Physical Hydrology**. 3<sup>a</sup> ed., Long Grove (IL): Waveland Press, 2015.

DIONISIO, A.; MENEZES, R.; MENDES, D. A. Mutual information: A measure of dependency for nonlinear time series. **Physica A: Statistical Mechanics and its Applications**, v. 344, n. 1–2, p. 326–329, 2004.

DIRKS, K. N. et al. High-resolution studies of rainfall on Norfolk Island Part II : Interpolation of rainfall data. **Journal of Hydrology**, v. 208, p. 187–193, 1998.

DONOHUE, R. J.; RODERICK, M. L.; MCVICAR, T. R. On the importance of including vegetation dynamics in Budyko's hydrological model. **Hydrology and Earth System Sciences**, v. 11, n. 2, p. 983–995, 2007.

DORMAND, J. R.; PRINCE, P. J. A family of embedded Runge-Kutta formulae. **Journal of Computational and Applied Mathematics**, v. 6, n. 1, p. 19–26, 1980.

DUETHMANN, D.; BLÖSCHL, G.; PARAJKA, J. Why does a conceptual hydrological model fail to predict discharge changes in response to climate change? **Hydrology and Earth System Sciences**, Discussion, 2020. Disponível em: < <https://doi.org/10.5194/hess-2019-652>>. Acesso em: 15 jan 2020.

DUNCAN, H. P. Baseflow separation – A practical approach. **Journal of Hydrology**, v. 575, p. 308–313, 2019.

DUVERT, C. et al. Correlation and spectral analyses to assess the response of a shallow aquifer to low and high frequency rainfall fluctuations. **Journal of Hydrology**, v. 527, p. 894–907, 2015.

ECKHARDT, K. How to construct recursive digital filters for baseflow separation. **Hydrological Processes**, v. 19, p. 507–515, 2005.

EFSTRATIADIS, A.; KOUTSOYIANNIS, D. One decade of multi-objective calibration approaches in hydrological modelling: a review. **Hydrological Sciences Journal**, v. 55, n. 1, p. 58–78.

EVARISTO, J.; MCDONNELL, J. J. Global analysis of streamflow response to forest management. **Nature**, v. 570, n. 7762, p. 455–461, 2019.

FAURÈS, J. M. et al. Impact of small-scale spatial rainfall variability on runoff modeling. **Journal of Hydrology**, v. 173, n. 1–4, p. 309–326, 1995.

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

FIENER, P.; AUERSWALD, K. Spatial variability of rainfall on a sub-kilometre scale. **Earth Surface Processes and Landforms**, v. 34, n. 6, p. 848–859, 2009.

FRASER, A. M.; SWINNEY, H. L. Independent coordinates for strange attractors from mutual information. **Physical Review A**, v. 33, n. 2, p. 1134–1140, 1986.

FREEDMAN, D.; DIACONIS, P. On the histogram as a density estimator: L2 theory. **Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete**, v. 57, n. 4, p. 453–476, 1981.

GAIN, A. K.; GIUPPONI, C.; WADA, Y. Measuring global water security towards sustainable development goals. **Environmental Research Letters**, v.11, n.12, p. 1–13, 2016.

GENÇAĞA, D. Transfer Entropy. **Entropy**, v. 20, n. 4, p. 1–4, 2018.

- GHIL, M.; ZALIAPIN, I.; THOMPSON, S. A delay differential model of ENSO variability: parametric instability and the distribution of extremes. **Nonlinear Processes in Geophysics**, v. 15, p. 417–433, 2008.
- GLENN, E. P. et al. Vegetation index-based crop coefficients to estimate evapotranspiration by remote sensing in agricultural and natural ecosystems. **Hydrological Processes**, v. 25, p. 4050–4062, 2011.
- GLENN, E.P. et al. Scaling sap flux measurements of grazed and ungrazed shrub communities with fine and coarse-resolution remote sensing. **Ecohydrology**, v.1, p. 316–329, 2008.
- GOMES, L. H. **Determinação da recarga profunda na bacia piloto do Ribeirão da Onça em zona de afloramento do Sistema Aquífero Guarani a partir do balanço hídrico em zona saturada**. 167 f. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2008.
- GÓMEZ, D. **Modelo empírico linear para a previsão da disponibilidade hídrica integrada em função da média móvel da precipitação**. 132 f. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2016.
- GÓMEZ, D. et al. Aquifer Responses to Rainfall through Spectral and Correlation Analysis. **Journal of the American Water Resources Association**, v. 54, n. 6, p.1341–1354, 2018.
- GÓMEZ, D.; WENDLAND, E.; MELO, D. C. D. Empirical rainfall-based model for defining baseflow and dynamical water use rights. **River Research and Applications**, p. 1–10, 2019.
- GOUVÊA, T. H.; WENDLAND, E. Influência de características do solo na variação do nível d’água em região de recarga do Aquífero Guarani. **Revista Brasileira de Recursos Hídricos**, v. 16, n. 1, p. 55–65, 2011.
- GRANGER, C. W. J. Investigating causal relations by econometric models and cross-spectral methods. **Econometrica**, v. 37, n. 3, p. 424–438, 1969.
- GRANGER, C. W. J.; JOYEUX, R. An introduction to long-memory time series models and fractional differencing. **Journal of Time Series Analysis**, v. 1, n. 1, p. 15–29, 1980.
- GRAVES, T. et al. A brief history of long memory: Hurst, Mandelbrot and the road to ARFIMA, 1951-1980. **Entropy**, v. 19, n. 9, p. 1–21, 2017.
- GUERSCHMAN, J. P. et al. Scaling of potential evapotranspiration with MODIS data reproduces flux observations and catchment water balance observations across Australia. **Journal of Hydrology**, v. 369, n. 1–2, p. 107–119, 2009.
- GUPTA, H. V. et al. Towards a comprehensive assessment of model structural adequacy. **Water Resources Research**, v. 48, n. 8, p. 1–16, 2012.
- HADKA, D. **Platypus Documentation**, 2015. Disponível em: <<https://readthedocs.org/projects/platypus/downloads/pdf/docs/>>. Acesso em: 15 jan. 2020.
- HARMAN, C.; TROCH, P. A. What makes Darwinian hydrology “darwinian”? Asking a different kind of question about landscapes. **Hydrology and Earth System Sciences**, v.18, n. 2, p. 417–433, 2014.

- HAWKINS, R. H. Curve Number Method: time to think anew? **Journal of Hydrologic Engineering**, v. 19, n. 6, p. 1059, 2014.
- HERSCHY, R. The velocity-area method. **Flow Measurement and Instrumentation**, v. 4, n. 1, p. 7–10, 1993.
- HINDMARSH, A. C. ODEPACK – A systematized collection of ODE solvers. **IMACS Transactions on Scientific Computation**, v. 1, p. 55–64, 1983.
- HLAVÁČKOVÁ-SCHINDLER, K. et al. Causality detection based on information-theoretic approaches in time series analysis. **Physics Reports**, v. 441, p. 1–46, 2007.
- HOFSTRA, N. et al. Comparison of six methods for the interpolation of daily, European climate data. **Journal of Geophysical Research**, v. 113, p. 1–19, 2008.
- HOSKING, J. R. M. Fractional differencing. **Biometrika**, v. 68, n. 1, p. 165–176, 1981.
- HRACHOWITZ, M. et al. Transit times-the link between hydrology and water quality at the catchment scale. **WIREs Water**, v.3, p. 629–657, 2016.
- HRACHOWITZ, M. et al. What can flux tracking teach us about water age distribution patterns and their temporal dynamics? **Hydrology and Earth System Sciences**, v. 17, n. 2, p. 533–564, 2013.
- HUANG, X. et al. Hydrological natural inflow and climate variables: Time and frequency causality analysis. **Physica A: Statistical Mechanics and its Applications**, v. 516, p. 480–495, 2019.
- HURST, H. E. Long-Term Storage of Reservoirs: An Experimental Study. **Transactions of the American Society of Civil Engineers**, v. 116, p. 770–799, 1951.
- INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (IPCC). **Climate Change and Land**: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems, 2019. Disponível em: <<https://www.ipcc.ch/srccl/>>. Acesso em: 05 jan. 2020.
- IRMAK, S. et al. Standardized ASCE Penman-Monteith: impact of sum-of-hourly vs. 24-h-timestep computations at reference weather stations. **Transactions of the ASAE**, v. 48, n. 2, p. 1067–1077, 2005.
- JASECHKO, S. et al. Global aquifers dominated by fossil groundwaters but wells vulnerable to modern contamination. **Nature Geoscience**, v. 10, n. 6, p. 425–429, 2017.
- KAMBLE, B.; KILIC, A.; HUBBARD, K. Estimating crop coefficients using remote sensing-based vegetation index. **Remote Sensing**, v. 5, n. 4, p. 1588–1602, 2013.
- KANG, Y. et al. How Universal Is the Relationship between Remotely Sensed Vegetation Indices and Crop Leaf Area Index? A Global Assessment. **Remote Sensing**, v. 8, n. 7, p. 1–29, 2016.
- KIM, B.; LEE, T.; OUARDA, T. B. N. J. Total least square method applied to rating curves. **Hydrological Processes**, v. 28, p. 4057–4066, 2014.
- KIM, B. et al. Exploring wavelet applications in civil engineering. **KSCE Journal of Civil Engineering**, v. 21, n.4, p.1076-1086, 2017.

- KIRCHHEIM, R. E. et al. The use of isotopes in evolving groundwater circulation models of regional continental aquifers: The case of the Guarani Aquifer System. **Hydrological Processes**, v. 33, n. 17, p. 2266–2278, 2019.
- KIRCHNER, J.W. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. **Water Resources Research**, v. 42, p. 1-5, 2006.
- KIRCHNER, J. W. Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward. **Water Resources Research**, v. 45, n. 2, p. 1–34, 2009.
- KLEMEŠ, V. The Hurst Phenomenon: A Puzzle? **Water Resources Research**, v. 10, n. 4, p. 675–688, 1974.
- KNOBEN, W. J. M. et al. Modular Assessment of Rainfall-Runoff Models Toolbox (MARRMoT) v1.2: An open-source, extendable framework providing implementations of 46 conceptual hydrologic models as continuous state-space formulations. **Geoscientific Model Development**, v. 12, n. 6, p. 2463–2480, 2019.
- KRAFT, P. et al. CMF: A Hydrological Programming Language Extension For Integrated Catchment Models. **Environmental Modelling and Software**, v. 26, n. 6, p. 828–830, 2011.
- KRAUSE, P. **The hydrological modelling system J2000 – Documentation and application in large river basins**. Schriften des Forschungszentrums Jülich, Reihe Umwelt/Environment, Band 29, 2001.
- KŘEČEK, J.; HAIGH, M. Land use policy in headwater catchments. **Land Use Policy**, v. 80, p. 410–414, 2019.
- KUMAR, P. Hydrocomplexity: addressing water security and emergent environment risks. **Water Resources Research**, v. 51, p. 5827–5838, 2015.
- KUPPEL, S. et al. Ech2O-iso 1.0: water isotopes and age tracking in a process-based, distributed ecohydrological model. **Geoscientific Model Development**, v. 11, p. 3045–3069, 2018.
- LABAT, D. Cross wavelet analyses of annual continental freshwater discharge and selected climate indices. **Journal of Hydrology**, v. 385, n. 1–4, p. 269–278, 2010.
- LE COZ, J. et al. Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: A Bayesian approach. **Journal of Hydrology**, v. 509, p. 573–587, 2014.
- LE COZ, J. et al. Uncertainty in open-channel discharges measured with the velocity – area method. **Flow Measurement and Instrumentation**, v. 26, p.18–29, 2012.
- LEE, G. R. et al. PyWavelets: A Python package for wavelet analysis. **The Journal of Open Source Software**, v. 4, n. 36, p. 1–2, 2019.
- LEVY, B.S.; SIDEL, V.W. Water Rights and Water Fights: Preventing and Resolving Conflicts Before They Boil Over. **American Journal of Public Health**, v. 101, n.5, p. 778-780, 2011. Editorial.

- LI, C. Z. et al. The transferability of hydrological models under nonstationary climatic conditions. **Hydrology and Earth System Sciences**, v. 16, n. 4, p. 1239–1254, 2012.
- LI, S. et al. Causal inference in nonlinear systems: Granger causality versus time-delayed mutual information. **Physical Review E**, v. 97, n. 5, p. 1–9, 2018.
- LI, S. et al. The characterization of hippocampal theta-driving neurons-A time-delayed mutual information approach. **Scientific Reports**, v. 7, n. 1, p. 1–12, 2017.
- LINDSTRÖM, G. et al. Development and test of the distributed HBV-96 hydrological model. **Journal of Hydrology**, v. 201, p. 272–288, 1997.
- LIUZZO, L. et al. Modifications in Water Resources Availability Under Climate Changes : A Case Study in a Sicilian Basin. **Water Resources Management**, v. 29, n. 4, p. 1117–1135, 2015.
- LUCAS, M.; WENDLAND, E. Recharge estimates for various land uses in the Guarani aquifer system outcrop area. **Hydrological Sciences Journal**, v. 61, n. 7, p.1253–1262, 2016.
- LUTZ, W.; SAMIR, K. C. Dimensions of global population projections: What do we know about future population trends and structures? **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 365, n. 1554, p. 2779–2791, 2010.
- LY, S.; CHARLES, C.; DEGRÉ, A. Geostatistical interpolation of daily rainfall at catchment scale : the use of several variogram models in the Ourthe and Ambleve catchments, Belgium. **Hydrology and Earth System Sciences**, v.15, p. 2259–2274, 2011.
- LYNE, V.; HOLLICK, M. Stochastic time-Variable Rainfall-Runoff Modelling. **Institution of Engineers Australia National Conference**, p. 88–92, 1979.
- MACHADO, A. R.; WENDLAND, E.; KRAUSE, P. Hydrologic Simulation for Water Balance Improvement in an Outcrop Area of the Guarani Aquifer System. **Environmental Processes**, v. 3, n. 1, p. 19–38, 2016.
- MAIER, H. R. et al. Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions. **Environmental Modelling and Software**, v. 62, p. 271–299, 2014.
- MALDONADO, L.H.M. **Validação experimental de metodologias para quantificação contínua de vazão no Ribeirão da Onça**. 2009. 175 f. Dissertação (Mestrado) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2009.
- MANDELBROT, B. B.; VAN NESS, J. W. Fractional Brownian Motions, Fractional Noises and Applications. **SIAM Review**, v. 10, n. 4, p. 422–437, 1968.
- MARENGO, J. A. et al. Changes in climate and land use over the Amazon region: current and future variability and trends. **Frontiers in Earth Science**, v. 6, p. 1-21, 2018.
- MARENGO, J. A.; ALVES, L. M. Crise Hídrica em São Paulo em 2014: Seca e Desmatamento. **GEOUSP: Espaço e Tempo** (Online), v. 19, n. 3, p. 485, 2015.
- MARINAZZO, D.; PELLICORO, M.; STRAMAGLIA, S. Kernel method for nonlinear Granger causality. **Physical Review Letters**, v. 100, n.14, p.1–4, 2008.



- MARTINS FILHO, F. M. **Balço hídrico e erosão do solo em uma plantação de eucaliptos novos**. 138 f. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2014.
- MATTOS, T. S. et al. Groundwater Recharge Decrease Replacing Pasture by Eucalyptus Plantation. **Water**, v. 11, n. 6, p. 1–13, 2019.
- MAYS, L.W. **Water Resources Engineering**. 2nd ed., Tempe (AZ): John Wiley & Sons, 2010.
- MCDONALD, J. E. On the Ratio of Evaporation to Precipitation. **Bulletin of the American Meteorological Society**, v. 42, p. 185–189, 1961.
- MCDONNELL, J. J. et al. Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology. **Water Resources Research**, v. 43, n. 7, p. 1–6, 2007.
- MCDONNELL, J. J.; BEVEN, K. Debates: The future of hydrological sciences: A (common) path forward? A call to action aimed at understanding velocities, celerities and residence time distributions of the headwater hydrograph. **Water Resources Research**, v. 50, p. 5342–5350, 2014.
- MCGRAW, M. C.; BARNES, E. A. Memory Matters: A Case for Granger Causality in Climate Variability Studies. **Journal of Climate**, v. 31, p. 3289–3300, 2018.
- MCGUIRE, K. J.; MCDONNELL, J. J. Tracer advances in catchment hydrology. **Hydrological Processes**, v. 29, n. 25, p. 5135–5138, 2015.
- MCMAHON, T. A.; PEEL, M. C. Uncertainty in stage–discharge rating curves: application to Australian Hydrologic Reference Stations data. **Hydrological Sciences Journal**, v. 64, n. 3, p. 255–275, 2019.
- MCMILLAN, H. K.; WESTERBERG, I. K. Rating curve estimation under epistemic uncertainty. **Hydrological Processes**, v. 1882, p. 1873–1882, 2015.
- MEIRA NETO, A.A. **Ponderação bayesiana de modelos utilizando diferentes séries de precipitação aplicada à simulação chuva-vazão na Bacia do Ribeirão da Onça**. 151 f. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2013.
- MELO, D. C. D. **Estimativa de impacto de mudanças climáticas nos níveis do aquífero livre em zona de recarga do Sistema Aquífero Guarani**. 190 f. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2013.
- MORETTIN, P.A. **Ondas e ondaletas: da análise de Fourier à Análise de Ondaletas de Séries Temporais**. 2 ed. São Paulo: EDUSP, 2014.
- MOSTAFAIE, A. et al. Comparing multi-objective optimization techniques to calibrate a conceptual hydrological model using in situ runoff and daily GRACE data. **Computational Geosciences**, v. 22, n. 3, p. 789–814, 2018.
- NAGLER, P. L. et al. Estimating riparian and agricultural actual evapotranspiration by reference evapotranspiration and MODIS enhanced vegetation index. **Remote Sensing**, v. 5, n. 8, p. 3849–3871, 2013.

- NAGLER, P. L. et al. Wide-area estimates of evapotranspiration by red gum (*Eucalyptus camaldulensis*) and associated vegetation in the Murray-Darling River Basin, Australia. **Hydrological Processes**, v. 30, n. 9, p. 1376–1387, 2016.
- NASCIMENTO, L. S. V.; REIS JR., D. S.; MARTINS, E. S. P. R. Avaliação do Algoritmo Evolutivo Mopso na Calibração Multiobjetivo do Modelo SMPA no Estado do Ceará. **Revista Brasileira de Recursos Hídricos**, v. 14, n. 1, p. 85–97, 2009.
- NASH, J. E. The form of the instantaneous unit hydrograph. **IAHS Publication**, v. 45, n. 3, p. 114–121, 1957.
- NATHAN, R. J.; MCMAHON, T. A. Evaluation of automated techniques for base flow and recession analyses. **Water Resources Research**, v. 26, n. 7, p. 1465–1473, 1990.
- NOVÁK, V. **Evapotranspiration in the Soil-Plant-Atmosphere System**. Heidelberg, Springer, 2012.
- ORGANIZAÇÃO DOS ESTADOS AMERICANOS (OEA). **Aquífero Guarani: Programa Estratégico de Ações**. Washington: OEA, 2009.
- PAPAGIANNPOULOU, C. et al. A non-linear Granger-causality framework to investigate climate-vegetation dynamics. **Geoscientific Model Development**, v.10, n. 5, p.1945–1960, 2017.
- PELLETIER, J. D.; TURCOTTE, D. L. Long-range persistence in climatological and hydrological time series: analysis, modeling and application to drought hazard assessment. **Journal of Hydrology**, v. 203, n. 1–4, p. 198–208, 1997.
- PETZOLD, L. Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations. **SIAM Journal on Scientific and Statistical Computing**, v. 4, n. 1, p. 136–148, 1983.
- PIRES, A. et al. Sustainability Assessment of indicators for integrated water resources management. **Science of the Total Environment**, v. 578, p.139–147, 2017.
- POMPEO, C. A. **Balanço hídrico da zona não-saturada do solo na bacia do Ribeirão da Onça (SP)**. 177 f. Tese (Doutorado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 1990.
- POOL, S.; VIS, M.; SEIBERT, J. Evaluating model performance: towards a non-parametric variant of the Kling-Gupta efficiency. **Hydrological Sciences Journal**, v. 63, n. 13–14, p. 1941–1953, 2018.
- RATHINASAMY, M. et al. Wavelet-based multiscale performance analysis: An approach to assess and improve hydrological models. **Water Resources Research**, v. 50, p. 9721–9737, 2014.
- RODRIGUES, R. R. et al. Common cause for severe droughts in South America and marine heatwaves in the South Atlantic. **Nature Geoscience**, v. 12, n. 8, p. 620–626, 2019.
- ROSSI, M. **Mapa Pedológico do Estado de São Paulo: Revisado e Ampliado**. Secretaria do Meio Ambiente: Instituto Florestal. São Paulo, 2017.
- RUNGE, J. et al. Inferring causation from time series in Earth system sciences. **Nature Communications**, v. 10, n. 1, p. 1–13, 2019.

- RUNGE, J.; PETOUKHOV, V.; KURTHS, J. Quantifying the strength and delay of climatic interactions: The ambiguities of cross correlation and a novel measure based on graphical models. **Journal of Climate**, v. 27, n. 2, p. 720–739, 2014.
- RUSSELL, B. On the notion of cause. **Proceedings of the Aristotelian Society**, v. 13, n. 1, p. 1–26, 1913.
- SANTOS, L.; THIREL, G., PERRIN, A. Continuous state-space representation of a bucket-type rainfall-runoff model: a case study with the GR4 model using state-space GR4 (version 1.0). **Geoscientific Model Development**, v. 11, p. 1591–1605, 2018.
- SCHREIBER, T. Measuring Information Transfer. **Physical Review Letters**, v. 85, n. 2, p. 461–464, 2000.
- SEABOLD, S.; PERKTOLD, J. Statsmodels: Econometric and Statistical Modeling with Python. In: PYTHON IN SCIENCE CONFERENCE, 9<sup>th</sup>, 2010, Austin. *Proceedings...* p. 57–61.
- SEIBERT, J. **Conceptual runoff models – fiction or representation of reality?** 52f. Tese (Doutorado) – Department of Earth Sciences, Uppsala University, Uppsala, 1999.
- SINDICO, F.; HIRATA, R.; MANGANELLI, A. The Guarani Aquifer System: From a Beacon of hope to a question mark in the governance of transboundary aquifers. **Journal of Hydrology: Regional Studies**, v. 20, p. 49–59, 2018.
- SINGH, N. K.; BORROK, D. M. A Granger causality analysis of groundwater patterns over a half-century. **Scientific Reports**, v. 9, n. 1, p. 1–8, 2019.
- SKLASH, M. G.; FARVOLDEN, R. N. The role of groundwater in storm runoff. **Journal of Hydrology**, v. 43, p. 45-65, 1979.
- SOPHOCLEOUS, M. Interactions between groundwater and surface water: the state of the science. **Hydrogeology Journal**, v.10, p. 52–67, 2002.
- STORN, R.; PRICE, K. Differential Evolution – A simple and efficient heuristic for global optimization over continuous spaces. **Journal of Global Optimization**, v. 11, n. 4, p. 341–359, 1997.
- SULEIMAN, A. A.; HOOGENBOOM, G. A comparison of ASCE and FAO-56 reference evapotranspiration for a 15-min time step in humid climate conditions. **Journal of Hydrology**, v. 375, n. 3–4, p. 326–333, 2009.
- TEGOS, A.; MALAMOS, N.; KOUTSOYIANNIS, D. A parsimonious regional parametric evapotranspiration model based on a simplification of the Penman – Monteith formula. **Journal of Hydrology**, v. 524, p. 708–717, 2015.
- THOMAS, B. F.; CAINETA, J.; NANTEZA, J. Global assessment of groundwater sustainability based on storage anomalies. **Geophysical Research Letters**, v. 44, p. 11445–11455, 2017.
- TOMASELLA, J. et al. The water balance of an Amazonian micro-catchment: the effect of

interannual variability of rainfall on hydrological behaviour. **Hydrological Processes**, v.22, p. 2133-2147, 2008.

TROCH, P. A. et al. The importance of hydraulic groundwater theory in catchment hydrology: The legacy of Wilfried Brutsaert and Jean-Yves Parlange. **Water Resources Research**, v. 49, n. 9, p. 5099–5116, 2013.

UN GENERAL ASSEMBLY. **Transforming our world: the 2030 Agenda for Sustainable Development**. New York, 2015. Disponível em: <<https://www.refworld.org/docid/57b6e3e44.html>>. Acesso em: 05 jan. 2020.

UNISDR. **Impacts of disasters since the 1992 Rio de Janeiro Earth Summit**. Geneva: The United Nations Office for Disaster Risk Reduction, 2012. Disponível em: <[https://www.unisdr.org/files/27162\\_infographic.pdf](https://www.unisdr.org/files/27162_infographic.pdf)>. Acesso em: 05 jan. 2020.

VAN LANEN, H. A. J. et al. Hydrological drought across the world: Impact of climate and physical catchment structure. **Hydrology and Earth System Sciences**, v. 17, n.5, p.1715–1732, 2013.

VIRTANEN, P. et al. **Scipy 1.0 – Fundamental Algorithms for Scientific Computing in Python**. Preprint (arXiv), 2019. Disponível em: <<https://arxiv.org/pdf/1907.10121.pdf>>. Acesso em: 15 jan. 2020.

VON HOYNINGEN-HUENE, J. **Die interzeption des Niederschlages in landwirtschaftlichen Pflanzenbeständen Kulturbau** (A interceptação de precipitação em culturas agrícolas). Braunschwig: DVWK, 1981.

WAGENER, T., et al. A framework for development and application of hydrological models. **Hydrology and Earth System Sciences**, v. 5, n. 1, p. 13–26, 2001.

WEGEHENKEL, M. Modeling of vegetation dynamics in hydrological models for the assessment of the effects of climate change on evapotranspiration and groundwater recharge. **Advances in Geosciences**, v. 21, p. 109–115, 2009.

WENDLAND, E. **Implementação de pacote computacional SPA (Simulação de Processos em Aquíferos)**. São Paulo: USP, 2003. (FAPESP relatório de pesquisa 02/09696-3).

WENDLAND, E.; BARRETO, C.; GOMES, L. H. Water balance in the Guarani Aquifer outcrop zone based on hydrogeologic monitoring. **Journal of Hydrology**, v. 342, p. 261–269, 2007.

WENDLAND, E.; GOMES, L. H.; TROEGER, U. Recharge contribution to the Guarani aquifer system estimated from the water balance method in a representative watershed. **Anais da Academia Brasileira de Ciências**, v. 87, n. 2, p. 595–609, 2015.

WHEATER, H.S.; GOBER, P. Water security and the science agenda. **Water Resources Research**, v. 51, p. 5406-5424, 2015.

WHITEHEAD, P. G.; ROBINSON, M. Experimental Basin Studies - an International and Historical-Perspective of Forest Impacts. **Journal of Hydrology**, v. 145, n. 3 -4, p. 217–230, 1993

WILKS, S. S. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. **The Annals of Mathematical Statistics**, v. 9, n. 1, p. 60–62, 1938.

WU, B. et al. Optimizing water resources management in large river basins with integrated surface water-groundwater modeling: A surrogate-based approach. **Water Resources Research**, v. 51, p. 2153–2173, 2015.

ZIMMERMANN, B.; ELSENBEER, H.; MORAES, J. M. The influence of land-use changes on soil hydraulic properties: Implications for runoff generation. **Forest Ecology and Management**, v. 222, n. 1–3, p. 29–38, 2006.



## APÊNDICE A - VARIÁVEIS METEOROLÓGICAS

Este apêndice trata do controle de qualidade de variáveis meteorológicas utilizadas para calcular a evapotranspiração de referência.

### Análises de regressão entre variáveis meteorológicas

Numa primeira etapa, os dados meteorológicos das estações foram comparados entre si. De forma a reduzir o número de comparações, a estação Monte Alegre foi tomada como referência. Em cada comparação, os instantes que apresentaram diferenças de temperatura (T), umidade relativa (RH), velocidade do vento (WS) e radiação solar (Rs) superiores a 12°C, 48%, 3 m/s e 1000W/m<sup>2</sup>, respectivamente, não foram considerados nas análises de regressão de modo a evitar a inclusão de dados que potencialmente apresentaram problemas de medição. A partir disso, as equações apresentadas nas tabelas A1 a A4 foram obtidas.

Tabela A1 – Análises de regressão linear com a variável temperatura média horária do ar (T).

Variável independente X [°C]	Variável dependente Y [°C]	Equação de ajuste linear [médias horárias]	R <sup>2</sup>
T <sub>IAB</sub>	T <sub>NMA</sub>	Y = 1,020X – 0,018	0,88
T <sub>STO-1</sub>		Y = 1,034X – 0,492	0,93
T <sub>SMF</sub>		Y = 0,801X + 4,442	0,92
T <sub>CRHEA-A</sub>		Y = 0,987X + 0,119	0,97
T <sub>STO-2</sub>		Y = 1,007X + 0,878	0,87
T <sub>ZRIP</sub>		Y = 0,756X + 7,684	0,83

Fonte: Autor.

Tabela A2– Análises de regressão linear com a variável umidade relativa do ar (RH).

Variável independente X [%]	Variável dependente Y [%]	Equação de ajuste linear [médias horárias]	R <sup>2</sup>
RH <sub>IAB</sub>	RH <sub>NMA</sub>	Y = 0,926X + 13,8	0,87
RH <sub>STO-1</sub>		Y = 0,923X + 8,3	0,87
RH <sub>SMF</sub>		Y = 0,882X + 5,8	0,83
RH <sub>CRHEA-A</sub>		Y = 0,985X + 1,9	0,92
RH <sub>STO-2</sub>		Y = 1,062X – 7,4	0,87
RH <sub>ZRIP</sub>		Y = 0,945X + 1,3	0,77

Fonte: Autor.

Tabela A3 – Análises de regressão linear com a variável velocidade do vento a 2 m do solo (WS).

Variável independente X [m/s]	Variável dependente Y [m/s]	Equação de ajuste linear [médias horárias]	R <sup>2</sup>
WS <sub>IAB</sub>	WS <sub>NMA</sub>	Y = 0,79X + 0,015	0,61
WS <sub>STO-1</sub>		Y = 1,20X + 0,35	0,57
WS <sub>SMF</sub>		Y = 1,04X + 0,30	0,63
WS <sub>CRHEA-A</sub>		Y = 1,67X – 0,21	0,74
WS <sub>STO-2</sub>		Y = 1,05X + 0,96	0,43
WS <sub>ZRIP</sub>		Y = 1,64X + 0,82	0,62

Fonte: Autor.

Tabela A4 – Análises de regressão linear com a variável radiação solar global (R).

Variável independente X [W/m <sup>2</sup> ]	Variável dependente Y [W/m <sup>2</sup> ]	Equação de ajuste linear [médias horárias]	R <sup>2</sup>
R <sub>IAB</sub>		Y = 0,63X + 37,2	0,74
R <sub>STO-1</sub>		Y = 0,93X + 8,9	0,93
R <sub>SMF</sub>	R <sub>NMA</sub>	Y = 1,00X + 3,4	0,96
R <sub>CRHEA-A</sub>		Y = 1,00X + 7,2	0,93
R <sub>STO-2</sub>		Y = 1,22X + 4,0	0,96

Fonte: Autor.

Em termos horários, as estações da área de estudo apresentaram concordâncias razoáveis entre os valores registrados. A única anormalidade foi observada com os dados de radiação solar da estação IAB, que, portanto, não foram utilizados neste trabalho para fins de preenchimento ou análise de consistência.

Uma série meteorológica Y foi dita preenchida e consistida após a verificação das seguintes condições: 1) para todo instante de tempo pertencente ao período de monitoramento da estação existe um valor, observado ou estimado, da variável Y; e 2) a diferença entre o valor observado e o valor estimado é inferior a um erro máximo.

$$\forall t \in I_M, \exists y_t \in \mathbb{R} \text{ e } |y_t - y_{est,t}(x_t)| < D \quad (A1)$$

Preenchimentos foram feitos sempre que dados não foram registrados por uma estação durante seu período de operação. O critério adotado para efetuar eventuais substituições de valores observados foi baseado no erro máximo admissível ( $D_{max}$ ). Esse erro é uma função do erro padrão ( $\sigma_r$ ) do ajuste entre variáveis aleatórias independentes (genericamente, X e Y), e de um coeficiente (z) que depende do nível de confiança com o qual se substitui um valor observado por uma aproximação condicionada por outra observação. Quando a diferença entre o valor observado ( $y_t$ ) e o valor estimado ( $y_{est,t}$ ) é superior a  $D_{max}$ , o valor observado é substituído pelo estimado (essa situação ocorre em uma pequena quantidade de casos)

$$y_{est,t} = ax_t + b + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_r^2) \quad (A2)$$

em que  $a$  e  $b$  são coeficientes determinados pelo método dos mínimos quadrados,  $x_t$  é uma observação no instante  $t$  de uma variável aleatória independente X, e  $\varepsilon_t$  é uma variável aleatória normal independente e identicamente distribuída com média 0 e variância  $\sigma_r^2$ ,



$$\sigma_r^2 = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2 = \frac{1}{n} \sum_{t=1}^n (y_t - (ax_t + b))^2 \quad (\text{A3})$$

em que  $\hat{y}_t$  é o valor da variável Y aproximado pela regressão linear considerando a observação  $x_t$ , e n é o número total de observações simultâneas das variáveis X e Y. O erro máximo admissível foi definido como

$$D_{\max} = z \sigma_r = z \sigma_Y \sqrt{1 - R^2} \quad (\text{A4})$$

em que  $\sigma_r$  é o desvio padrão dos resíduos (erro padrão da regressão),  $\sigma_Y$  é o desvio padrão da variável Y durante o período 'n' observações simultâneas com a variável X,  $R^2$  é o coeficiente de determinação da regressão entre X e Y, e z é um fator associado ao intervalos de confiança da regressão.

Os valores de z foram definidos com auxílio de análises gráficas (identificação dos dados sujeitos a substituição) e resultaram nos valores indicados na Tabela A5. A definição de intervalos de confiança fixos, mesmo que superiores a 99% (e.g.,  $z > 2,4$ ), não pareceu adequada por resultar em eliminações (ou falta de eliminações) não coerentes com o esperado. Apenas dados de radiação solar não passaram por análise de consistência horária, dadas as condições de elevada variabilidade espacial (entre as estações) nessa escala de tempo.

Tabela A5 – Definição dos máximos erros admissíveis para realizar substituição de valores observados. [H] indica valores referentes a análises em escala horária, enquanto [D], a análises em escala diária.

Variável	z	$D_{\max}$
Temperatura [H]	$4\sqrt{1 - R^2}$	$4 \sigma_{T,Y} [1 - R^2]$
Umidade relativa [H]	$4\sqrt{1 - R^2}$	$4 \sigma_{RH,Y} [1 - R^2]$
Velocidade do vento [H]	$10\sqrt{1 - R^2}$	$10 \sigma_{WS,Y} [1 - R^2]$
Radiação solar global [D]	$8\sqrt{1 - R^2}$	$8 \sigma_{RS,Y} [1 - R^2]$

Fonte: Autor.

Após consistir e preencher os dados da estação NMA (até 12/2018), esta estação foi utilizada como referência para replicar o processo às demais estações. Os parâmetros e coeficientes resultantes das análises de regressão estão indicados nas Tabelas A6 a A9.

Tabela A6 – Relações para preenchimento e análise de consistência da estação STO-1 usando dados da estação NMA como referência (relações em termos horários).

Variável independente X	Variável dependente Y	Unidade	Equação de ajuste linear	R <sup>2</sup>	Erro máximo admissível
T <sub>NMA</sub>	T <sub>STO1</sub>	[°C]	$Y = 0,902X + 1,78$	0,93	1,5
RH <sub>NMA</sub>	RH <sub>STO1</sub>	[%]	$Y = 0,957X + 2,24$	0,90	8,5
WS <sub>NMA</sub>	WS <sub>STO1</sub>	[m/s]	$Y = 0,476X + 0,26$	0,57	3,4
R <sub>NMA</sub>	R <sub>STO1</sub>	[W/m <sup>2</sup> ]	$Y = 0,996X + 4,48$	0,93	N/A

Fonte: Autor.

Tabela A7 – Relações para preenchimento e análise de consistência da estação STO-2 usando dados da estação NMA como referência (relações em termos horários).

Variável independente X	Variável dependente Y	Unidade	Equação de ajuste linear	R <sup>2</sup>	Erro máximo admissível
T <sub>NMA</sub>	T <sub>STO2</sub>	[°C]	$Y = 0,904X + 1,30$	0,90	2,4
RH <sub>NMA</sub>	RH <sub>STO2</sub>	[%]	$Y = 0,859X + 15,33$	0,93	5,3
WS <sub>NMA</sub>	WS <sub>STO2</sub>	[m/s]	$Y = 0,412X - 0,06$	0,43	3,7
R <sub>NMA</sub>	R <sub>STO2</sub>	[W/m <sup>2</sup> ]	$Y = 0,783X + 4,19$	0,96	N/A

Fonte: Autor.

Tabela A8 – Relações para preenchimento e análise de consistência da estação SMF usando dados da estação NMA como referência (relações em termos horários).

Variável independente X	Variável dependente Y	Unidade	Equação de ajuste linear	R <sup>2</sup>	Erro máximo admissível
T <sub>NMA</sub>	T <sub>SMF</sub>	[°C]	$Y = 1,141X - 3,32$	0,91	2,4
RH <sub>NMA</sub>	RH <sub>SMF</sub>	[%]	$Y = 0,905X + 12,69$	0,76	19,8
WS <sub>NMA</sub>	WS <sub>SMF</sub>	[m/s]	$Y = 0,610X + 0,26$	0,63	3,5
R <sub>NMA</sub>	R <sub>SMF</sub>	[W/m <sup>2</sup> ]	$Y = 0,954X + 4,40$	0,96	N/A

Fonte: Autor.

Tabela A9 – Relações para preenchimento e análise de consistência da estação CRHEA-convencional usando dados da estação NMA como referência (relações em termos diários).

Variável independente X	Variável dependente Y	Unidade	Equação de ajuste linear	R <sup>2</sup>	Erro máximo admissível
T <sub>NMA</sub>	T <sub>CRHEA-C</sub>	[°C]	$Y = 0,980X + 1,49$	0,83	2,4
RH <sub>NMA</sub>	RH <sub>CRHEA-C</sub>	[%]	$Y = 0,757X + 15,67$	0,73	12,2
WS <sub>NMA</sub>	WS <sub>CRHEA-C</sub>	[m/s]	$Y = 0,681X - 0,06$	0,65	2,2
R <sub>NMA</sub>	R <sub>CRHEA-C</sub>	[W/m <sup>2</sup> ]	$Y = 0,729X + 2,06$	0,75	9,3

Fonte: Autor.

Falhas nas séries de dados diários da estação CRHEA-convencional foram preenchidas usando os dados da estação INMET-São Carlos como suporte (pela grande distância entre essas estações, 23 km, os erros associados são consideravelmente maiores). As relações utilizadas estão apresentadas na Tabela A10. Como dados de radiação solar não estavam disponíveis no banco de dados históricos do INMET, o preenchimento dessa variável foi feito de forma estocástica, a partir da geração de números aleatórios com distribuição normal de média e variância iguais ao do conjunto de dados observados de radiação solar global diária no CHREA.

Valores que resultaram fora de valores fisicamente possíveis (e.g., maiores que a radiação extraterrestre) foram substituídos por estimativas determinadas a partir de uma regressão linear com dados de temperatura máxima diária registradas pelo CRHEA.

Tabela A10 – Relações para preenchimento da estação CRHEA-convencional usando dados da estação INMET como referência (relações em termos diários).

Variável independente X	Variável dependente Y	Unidade	Equação de ajuste linear	R <sup>2</sup>
T <sub>INMET</sub>	T <sub>CRHEA-C</sub>	[°C]	Y = 1,010X + 0,92	0,74
RH <sub>INMET</sub>	RH <sub>CRHEA-C</sub>	[%]	Y = 0,625X + 31,16	0,54
WS <sub>INMET</sub>	WS <sub>CRHEA-C</sub>	[m/s]	Y = 0,273X + 0,57	0,16
T <sub>MAX,CRHEA-C</sub>	R <sub>CRHEA-C</sub>	[°C], [MJ/m <sup>2</sup> /d]	Y = 0,776X – 6,20	0,30

Fonte: Autor.

### Síntese das variáveis meteorológicas

Após preenchimento e análises de consistência, as séries de dados meteorológicos de todas as estações apresentaram as estatísticas amostrais média e desvio padrão indicadas na Tabela A11.

Tabela A11 – Síntese das variáveis consistidas. Notação: média amostra (desvio padrão amostral).

Estação	T [°C]	RH [%]	WS [m/s]	R <sub>s</sub> [W/m <sup>2</sup> ]	Escala de tempo	Período de referência
CRHEA-Convencional	21,4 (3,3)	75 (12)	0,98 (0,65)	170 (46)	Diária	12/2004 a 01/2019
Santo Ignácio STO-1	19,8 (5,3)	77 (20)	0,99 (0,78)	183 (261)		02/2011 a 12/2013
Monte Alegre NMA	21,2 (5,3)	75 (21)	1,61 (1,21)	192 (270)	Horária	06/2011 a 03/2019
Santa Maria SMF	20,9 (6,0)	79 (21)	1,21 (0,95)	181 (251)		06/2011 a 05/2014
CRHEA-Automática	21,7 (4,9)	74 (20)	1,19 (0,68)	202 (281)		09/2016 a 05/2018
Santo Ignácio STO-2	21,7 (5,3)	79 (18)	0,60 (0,66)	165 (231)		02/2018 a 02/2019

Fonte: Autor.



## APÊNDICE B – CÓDIGOS COMPUTACIONAIS

Trechos de códigos computacionais (linguagem Python) implementados para geração de alguns resultados deste trabalho são apresentados a seguir.

### B1 Determinação de parâmetros variáveis da curva-chave

#### B1.1 Parâmetros determinados pelo método de mínimos quadrados ordinários

```

# -*- coding: utf-8 -*-
"""
Autor: Kalyl Gomes Calixto
"""
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
import statsmodels.api as sm
from scipy.optimize import differential_evolution
from numpy.linalg import inv
import time

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

# Importação de dados de vazão e de nível de água (simultâneos) medidos
flow_meter =
pd.read_csv("flow_meter_data.csv", sep=",", na_values="#N/D", infer_datetime_format=True).set_index('timestamp')

"""
INPUT
df: dataframe com dados de vazão e de nível observados (baixa frequência).
Primeira coluna como índice de data (timestamp); uma coluna com dados de vazão
(flow_obs) e nível observado (level_obs).
num: número de pares (h,Q) a ser utilizado para determinação dos parâmetros
p: ordem de autocorrelação dos erros. Para mínimos quadrados ordinários, p =
0
h0 = elevação do nível de água para vazão = 0

RETURN
Dataframe com os coeficientes e métricas estatísticas para cada período
"""

# Função para determinação dos parâmetros da transformação nível em vazão
para cada período
def least_squares_flow(df, num, p, h0 = 686.273):

    start = time.time()

    shape = len(df) - num + 1
    constant = np.zeros(shape)
    coeff_level = np.zeros(shape)
    aic = np.zeros(shape)
    bic = np.zeros(shape)
    loglike = np.zeros(shape)
    mse_resid = np.zeros(shape)
    correl_lag1 = np.zeros(shape)

    for i in range(shape):

        # Dados de nível de água (variável independente)

```



```

flow_meter =
pd.read_csv("flow_meter_data_nov.csv", sep=",", na_values="#N/D", infer_datetime_format='True').set_index('timestamp')

# Definir o intervalo de variação dos parâmetros
# x[0] = a; x[1] = b
bounds_flow = [(0.5,4.5), (0.5,2.5)]

# número de pares (h, Q) a ser utilizado na estimativa dos parâmetros
num = 5

# Funcao objetivo para a otimizacao global - considerando as variaveis
transformadas
"""
INPUT
X: vetor de parâmetros a ser otimizado (a, b)
df: dataframe com dados de vazão e de nível observados (baixa frequência).
Primeira coluna como índice de data (timestamp); uma coluna com dados de vazão
(flow_obs) e nível observado (level_obs).
num: número de pares (h,Q) a ser utilizado para determinação dos parâmetros
RETURN
Escalar, valor da função objetivo a ser utilizado na função de otimização
"""

def func_obj_transf_flow(x,df,num):

    aux = df
    # Elevação do nível de água para vazão = 0
    h_reference = 686.273

    lim = int((num + 1)/2)

    aux['log_Q'] = np.log(x[0]) + x[1]*np.log(aux.level_obs - h_reference)

    aux['sqerror'] = (aux.log_Q - np.log(aux.flow_obs))**2

    for i in range(-lim+1, lim, 1):
        aux[i] = aux.sqerror.shift(i)

    temporario = pd.DataFrame(index = aux.index)
    for i, col in enumerate(aux.columns):
        temporario[i+100] = np.where(np.isreal(col),aux[col],0)

    temporario['sum'] = temporario.sum(axis = 1,skipna = False)

    aux['objetivo'] = temporario['sum']

    return aux.objetivo

# Rotina para otimização dinâmica, usando o algoritmo differential
evolution Scipy
"""
INPUT
df:dataframe com dados de vazão e de nível observados (baixa frequência).
Primeira coluna como índice de data (timestamp); uma coluna com dados de vazão
(flow_obs) e nível observado (level_obs).
num:número de pares (h,Q) a ser utilizado para determinação dos parâmetros
p: ordem de autocorrelação dos erros. Para mínimos quadrados ordinários, p =
0
h0 = elevação do nível de água para vazão = 0

RETURN
Dataframe com os coeficientes e métricas estatísticas para cada período
"""

def optimization_flow(x, df, num):

```

```

start = time.time()

coeff_a, coeff_b, func_evaluation = [], [], []

# Número de parâmetros do modelo (a, b, variância)
par = 3

adj_1 = int((num-1)/2)
adj_2 = int(num-1)

for i in range(len(df)-adj_2):

    # Função objetivo a ser minimizada
    result = differential_evolution(lambda x:
func_obj_transf_flow(x,df,num)[i+adj_1],bounds_flow)
    coeff_a.append(result.x[0])
    coeff_b.append(result.x[1])
    func_evaluation.append(result.fun)

    optimization_results = pd.DataFrame({'coeff_a': coeff_a, 'coeff_b':
coeff_b, 'sse': func_evaluation})

    for i in range(len(optimization_results)):
        # estimativa da variancia do erro
        optimization_results['variance'] = (1/(num-
2))*optimization_results.sse

        optimization_results['AIC'] = num*(np.log(2*np.pi) + 1) +
num*np.log(optimization_results.variance) + 2*par

        optimization_results['BIC'] = num*(np.log(2*np.pi) + 1) +
num*np.log(optimization_results.variance) + par*np.log(num)

    end = time.time()

    return optimization_results

#####

```

## B2 Teste de causalidade de Granger e informação mútua

```

"""
Autor: Kalyl Gomes Calixto
A versão do Statsmodels utilizada foi a 0.9.0. e a da biblioteca Scipy 1.2.0
"""
import numpy as np
import pandas as pd
import statsmodels.api as sm
from statsmodels.tsa.tsatools import lagmat, lagmat2ds, add_trend
from scipy import stats
import matplotlib.pyplot as plt
import time
import pywt

#####

## Importacao dos dados observados (df) e das estimativas de incertezas (du).
Os nomes das colunas devem ser iguais nesses dois dataframes e uma coluna deve ter
a data (timestamp).
e.g.:

df =
pd.read_csv("data.csv",sep=',',infer_datetime_format=True').set_index('timestamp')

```



```

du =
pd.read_csv("data_unc.csv",sep=',',infer_datetime_format=True).set_index('timestamp')

"""
Função básica que executa o teste de causalidade de Granger.
Ela foi adaptada da função original disponível no Statsmodels versão 0.9.0.
A modificação foi motivada pelo elevado consumo de memória na proposição
original.
x: vetor de duas colunas de mesmo tamanho [primeira coluna: variável de
resposta; segunda coluna: variável de causa]
maxlag: valor da defasagem a ser analisada
"""

def grangercausalitytests(x, maxlag, addconst=True, verbose=True):

    # Código fonte completo disponível em
    https://www.statsmodels.org/stable/_modules/statsmodels/tsa/stattools.html#grangercausalitytests

    # [...]

    # Linha original da função Statsmodels
    #for mlg in range(1, maxlag + 1):

    # Modificação realizada no código original
    # Ao invés de fazer todas as defasagens até maxlag, apenas a defasagem maxlag
    é analisada
    for mlg in range(maxlag, maxlag + 1):

        # [...]

    return resli

# Lista de 'variáveis de causa'
def cause_paired_data(df):

    list_1 = [df.col_name_c1,
              ...
              df.col_name_cN]

    return list_1

# Lista de 'variáveis de resposta'
def response_paired_data(df):

    list_1 = [df.col_name_r1,
              ...
              df.col_name_rN]

    return list_1

# Rotina para fazer a análise de todas as combinações causa -> efeito
incluídas na list_1.
def main_obs():

    for i in range(len(cause_paired_data(df))):
        name = str(i+100)
        maxlag = 1000 # A ser definido com base na extensão de dados
disponíveis

        data_cause = cause_paired_data(df)[i] +
np.random.normal(0,cause_paired_data(du)[i],size = len(cause_paired_data(du)[i]))
        cause = data_cause.name

```

```

        data_response = response_paired_data(df)[i] +
np.random.normal(0,response_paired_data(du)[i],size =
len(response_paired_data(du)[i]))
        response = data_response.name
        granger_causality(data_cause, data_response, maxlag,name, cause,
response)

    return

# Função para análise de multirresolução (aproximações ou detalhes das séries
temporais)
"""
data: dataframe df (dados)
w: wavelet mãe
decomp_level: nível de decomposição a ser utilizado
"""
def decomp(data, w, decomp_level, time=False):
    """
     $F = A_n + [D_n + D_{n-1} + \dots + D_1] = A_n + S[D_n]$ 
    """
    if (decomp_level == 0):
        df = data
    else:
        w = pywt.Wavelet(w)
        a = np.asarray(data)
        ca = []
        for i in range(decomp_level):
            (a, d) = pywt.dwt(a, w, mode)
            ca.append(a)
        rec_a = []
        for i, coeff in enumerate(ca):
            coeff_list = [coeff, None] + [None] * i
            rec_a.append(pywt.waverec(coeff_list, w))
        diff = len(rec_a[decomp_level-1]) - len(data)
        #print("Diff =",diff)
        for k in range(diff):
            rec_a[decomp_level-1] = np.delete(rec_a[decomp_level-1],-1)
        if time is True:
            df = pd.DataFrame(index = data.index)
            df['a'] = rec_a[decomp_level-1]
            df['d'] = data - df.a
        else:
            df = pd.DataFrame()
            df['a'] = rec_a[decomp_level-1]
            df['d'] = data - df.a
    return df

# Função para compor o vetor x, que é argumento do teste de Granger. A
primeira coluna é a variável de resposta, e a segunda, a de causa.
def decomp_cr(data_cause, data_response, time = True):
    """ Definir aqui se serao considerados apenas a aproximação ou apenas os
detalhes - Análise de multirresolução """

    if time is True:
        x = pd.DataFrame(index = data_response.index)
    else:
        x = pd.DataFrame()

    if decomp_level_response == 0:
        x['response_a'] = data_response
    else:
        x['response_a'] = decomp(data_response, wavelet_r,
decomp_level_response).a

    if decomp_level_cause == 0:

```

```

        x['cause_a'] = data_cause
    else:
        x['cause_a'] = decomp(data_cause, wavelet_c, decomp_level_cause).a

    return x

# Função que executa o teste de Granger com a lista de variáveis fornecida.
def granger_causality(data_cause, data_response, maxlag, name, cause,
response):

    start = time.time()
    X = decomp_cr(data_cause, data_response)

    p_value, lr, lag_index = [], [], []
    for key in range(1, maxlag+1, 1):
        lag_index.append(key)
        lr.append(grangercausalitytests(X, key, addconst=True,
verbose=False)[key][0]['lrtest'][0])
        p_value.append(grangercausalitytests(X, key, addconst=True,
verbose=False)[key][0]['lrtest'][1])

    gcausality_result = pd.DataFrame(index = lag_index)
    gcausality_result.index.names = ['Lag']
    gcausality_result['lr'] = lr
    gcausality_result['p'] = p_value

    end = time.time()
    print("Execution time", end-start)

gcausality_result.to_csv(result_id_{}_{}_causa_{}_maxlag_{}.csv'.format(name, cause,
response, maxlag), sep = ',')

    return

# Funcao da biblioteca NoLiTSA (https://github.com/manu-mannattil/nolitsa)
adaptada para determinação da informação mútua normalizada entre duas variáveis
"""
Função para calcular a informação mútua normalizada entre duas variáveis (x)
e (y), com número de blocos do histograma (para estimar probabilidades) igual
a 'bins'.
"""

def mi(x, y, bins):
    p_x = np.histogram(x, bins)[0]
    p_y = np.histogram(y, bins)[0]
    p_xy = np.histogram2d(x, y, bins)[0].flatten()

    # Conversão de frequências em probabilidades
    # Quando p -> 0, p*log(p) -> 0.
    p_x = p_x[p_x > 0] / np.sum(p_x)
    p_y = p_y[p_y > 0] / np.sum(p_y)
    p_xy = p_xy[p_xy > 0] / np.sum(p_xy)

    # Entropia de Shannon
    h_x = np.sum(p_x * np.log2(p_x))
    h_y = np.sum(p_y * np.log2(p_y))
    h_xy = np.sum(p_xy * np.log2(p_xy))

    # Retorna a informação mútua normalizada

    Return (h_xy - h_x - h_y)/-np.min([h_x, h_y])

```

```

# Funcao adaptada da biblioteca NoLiTSA para determinar a informação mútua
normalizada entre duas variáveis diferentes com tempos de defasagem até um valor
máximo
"""
Retorna a informação mútua normalizada com tempo de defasagem t entre as
variáveis x e y, com t variando até maxtau; bins é o número de blocos do
histograma.
"""

def dmi(x, y, maxtau, bins):

    N = len(x)
    maxtau = min(N, maxtau)

    dii = np.empty(maxtau)
    dii[0] = mi(x, y, bins)

    for tau in range(1, maxtau):
        dii[tau] = mi(x[:-tau], y[tau:], bins)

    return dii

# Função para determinacao do number of bins a ser utilizado na construcao
dos histogramas p/ determinacao das prob. conjuntas e marginais [fórmula de
Freedman-Diaconis]

def number_bins(data_cause, data_response):

    n = len(data_cause)
    X = decomp_cr(data_cause, data_response)
    bin_width_x = 2*stats.iqr(X.cause_a, interpolation =
'midpoint')/np.power(n,1/3)
    bin_width_y = 2*stats.iqr(X.response_a, interpolation =
'midpoint')/np.power(n,1/3)
    bx = (max(X.cause_a) - min(X.cause_a))/bin_width_x
    by = (max(X.response_a) - min(X.response_a))/bin_width_y
    b = (np.ceil((bx*by)**0.5))
    nbins = max(int(1+np.log2(n)), int(b))

    return nbins

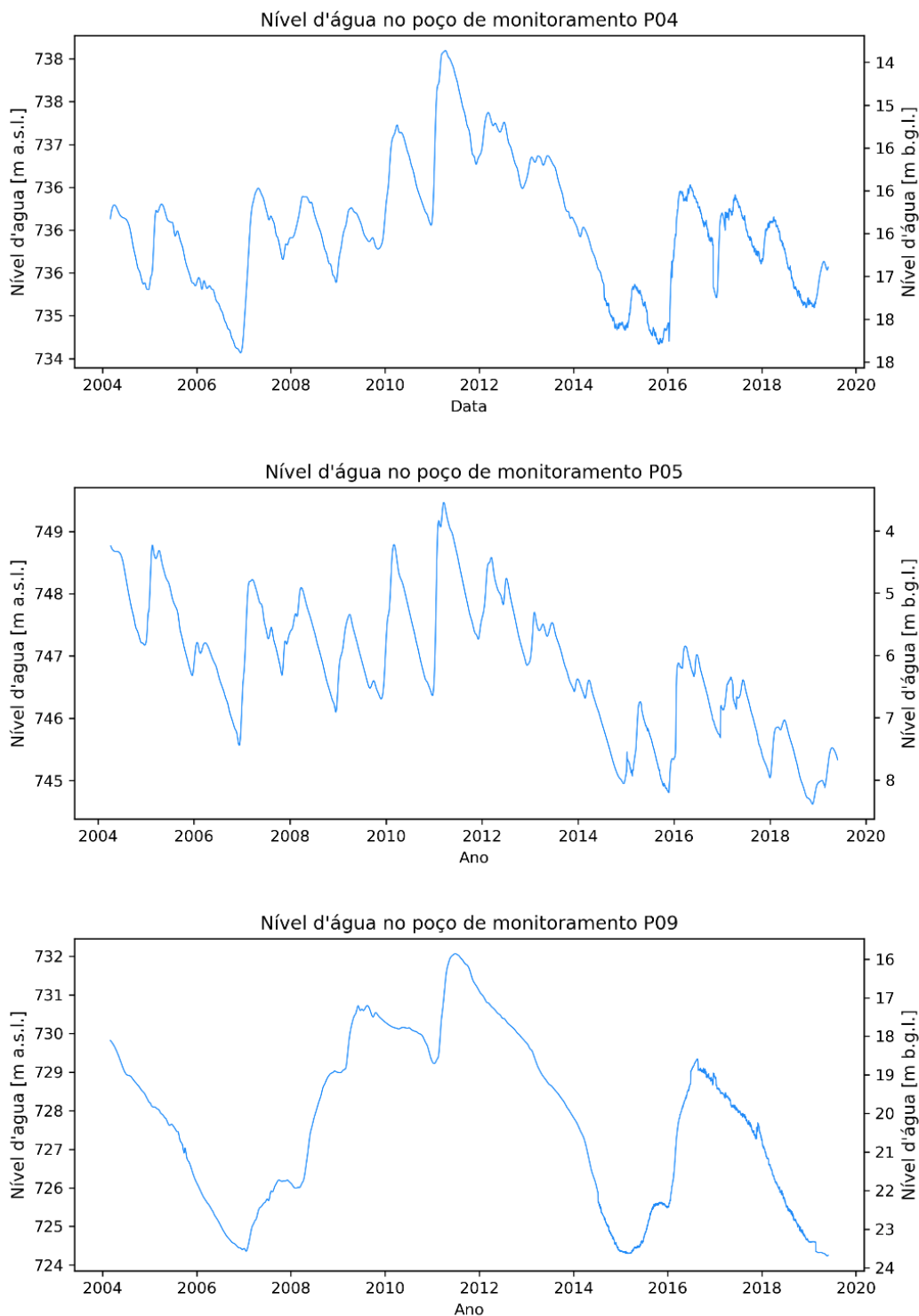
#####

```

## APÊNDICE C – NÍVEIS FREÁTICOS

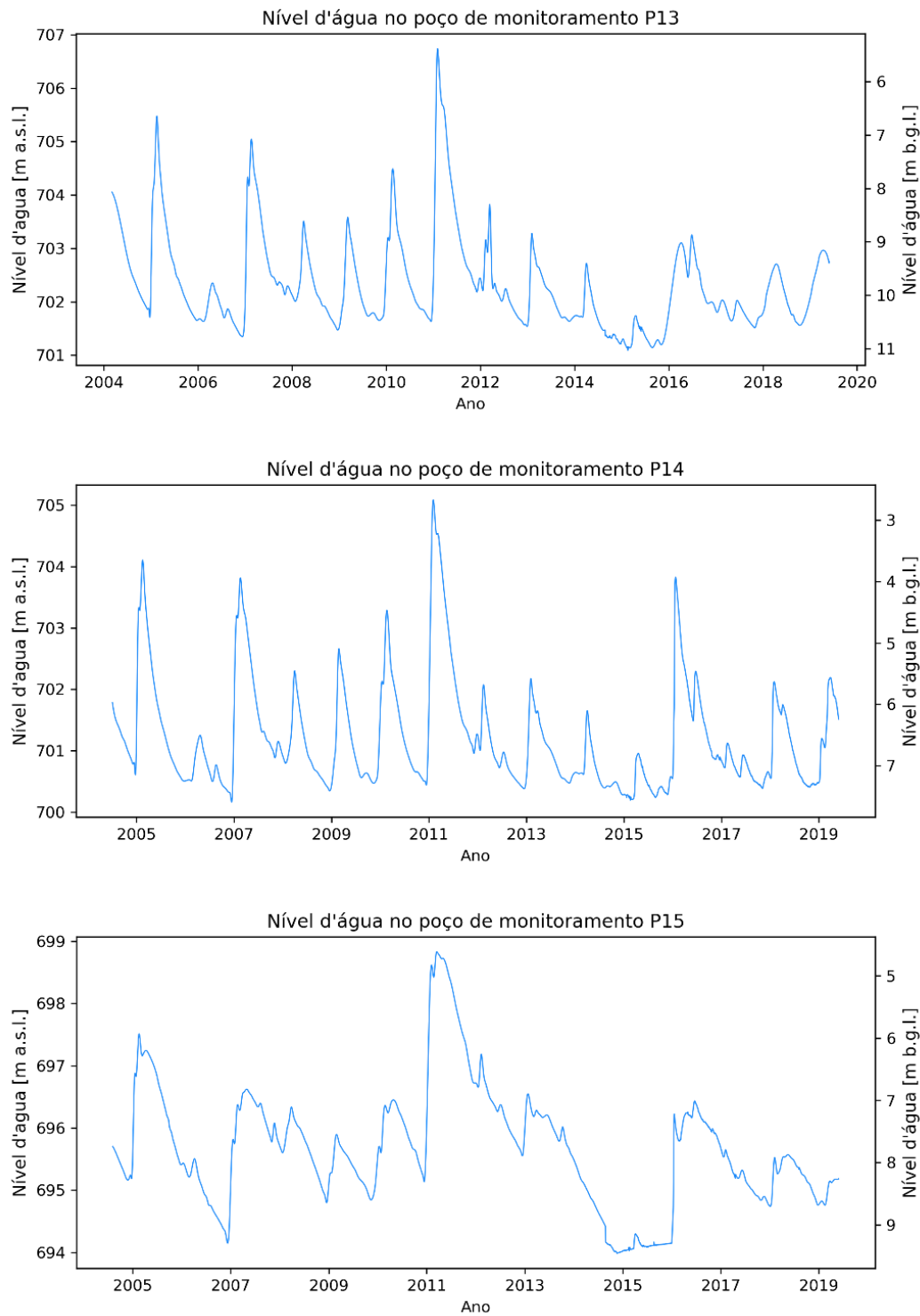
Dados de nível de água médio diário nos poços de monitoramento são apresentados nas figuras B1 a B3 [m a.s.l.: acima do nível do mar; m b.g.l.: abaixo no nível do terreno].

Figura B1. Nível freático dos poços P04 (pasto), P05 (pasto) e P09 (eucalipto).



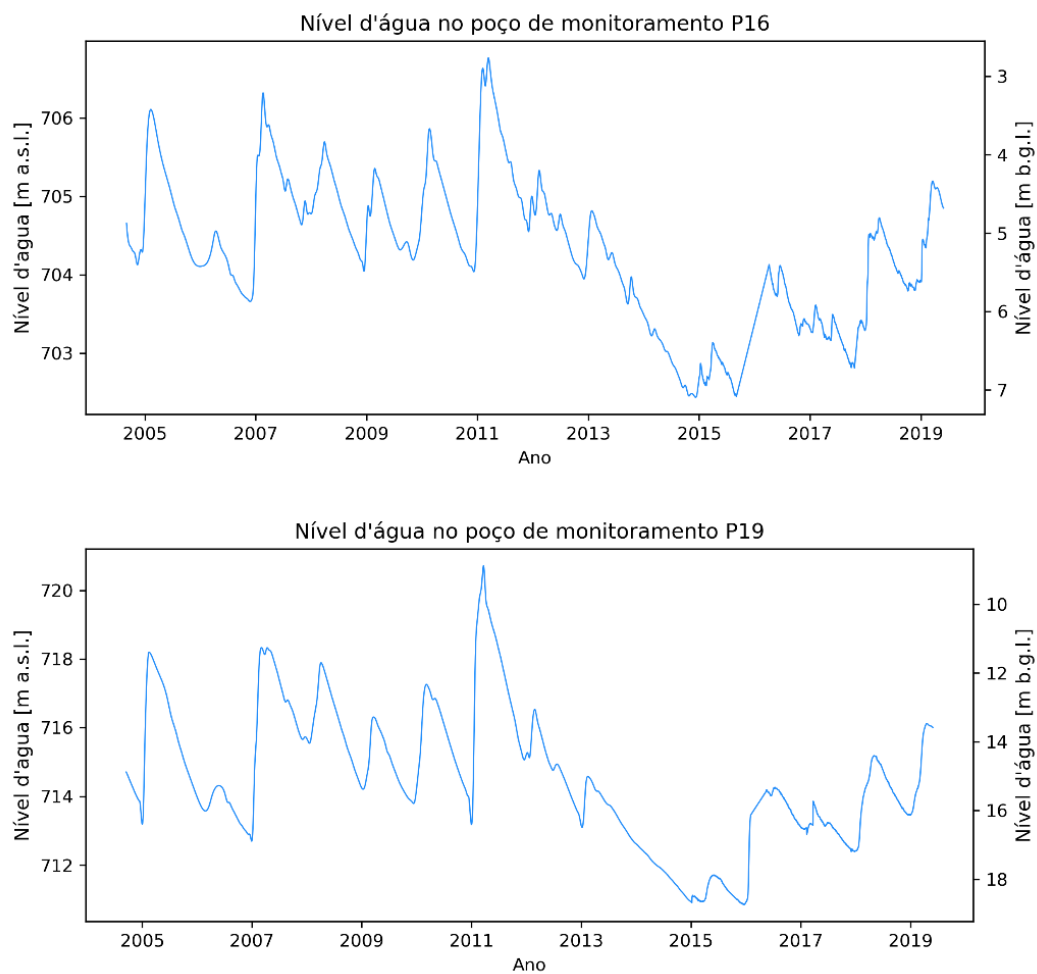
Fonte: Autor.

Figura B2. Nível freático dos poços P13 (citros e cana-de-açúcar), P14 (citros e cana-de-açúcar) e P15 (cana-de-açúcar e citros).



Fonte: Autor.

Figura B3. Nível freático dos poços P16 (pasto e eucalipto) e P19 (pasto e eucalipto).



Fonte: Autor.

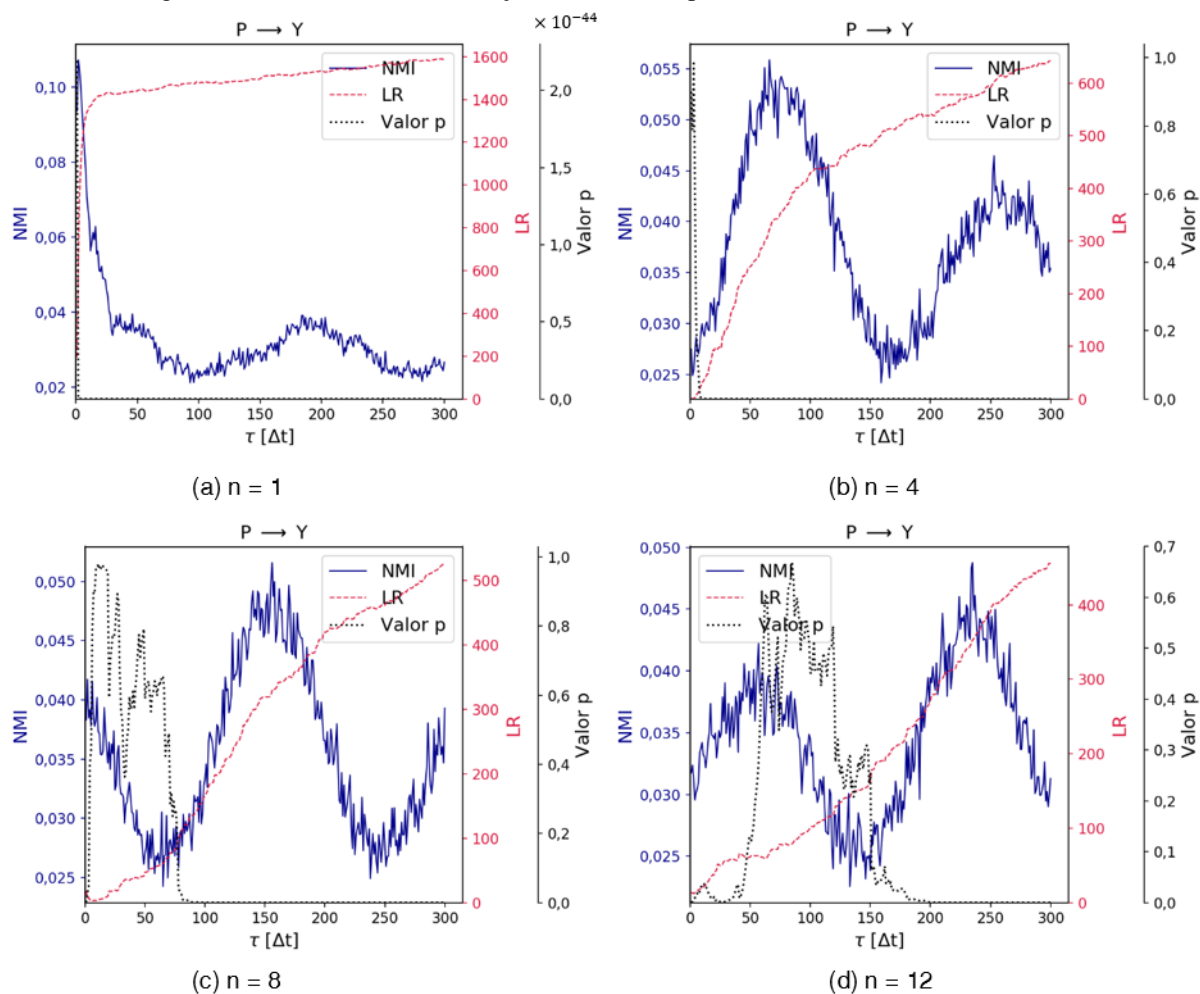




## APÊNDICE D – RESULTADOS DO PROBLEMA SINTÉTICO 2

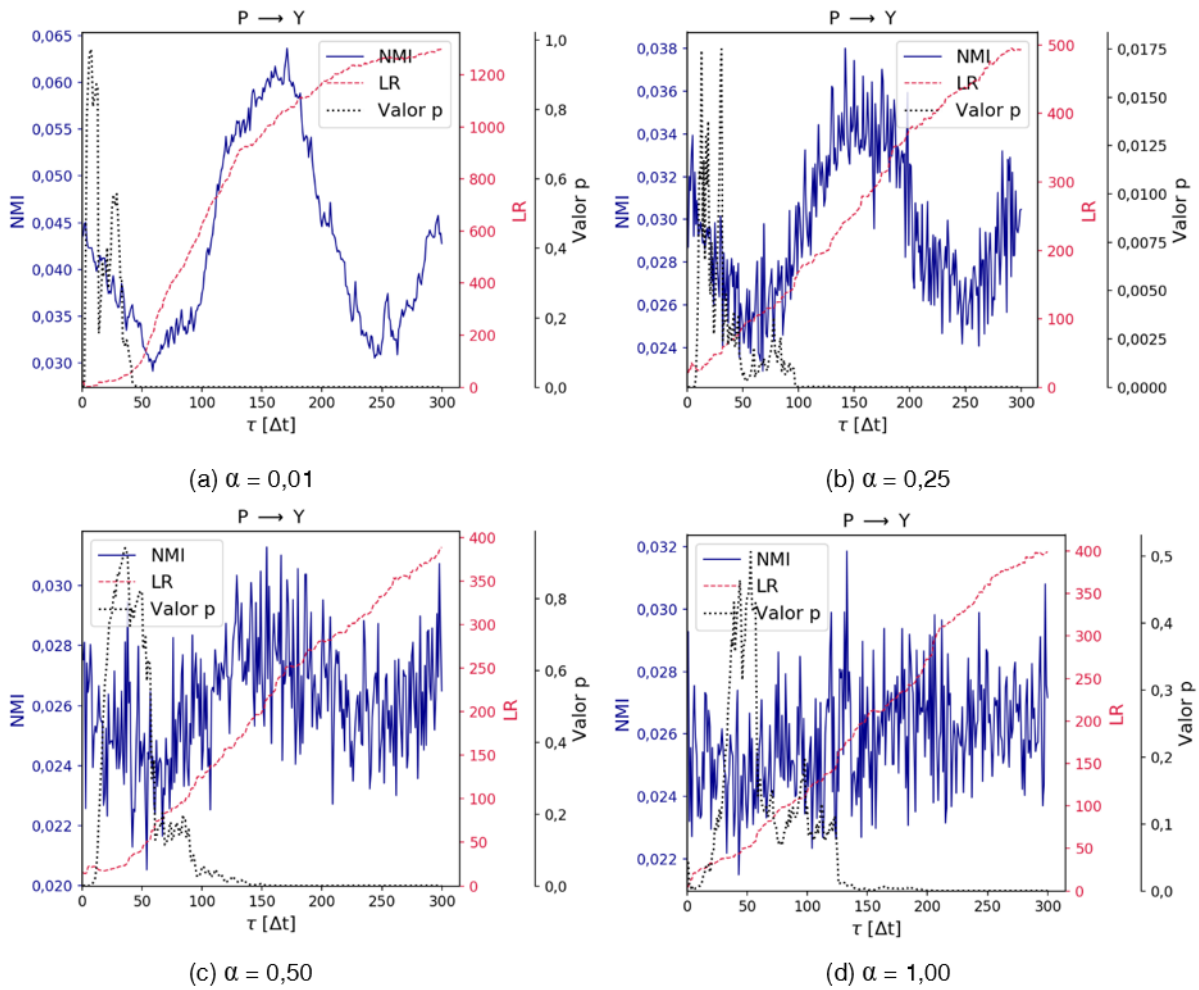
Os resultados gráficos do problema sintético 2 são apresentados nas Figuras D1 a D4. Onde não especificado,  $\Delta t = 1$  dia.

Figura D1 – Resultados da simulação referentes ao problema sintético S2-1.  $\Delta t = 1$  dia.



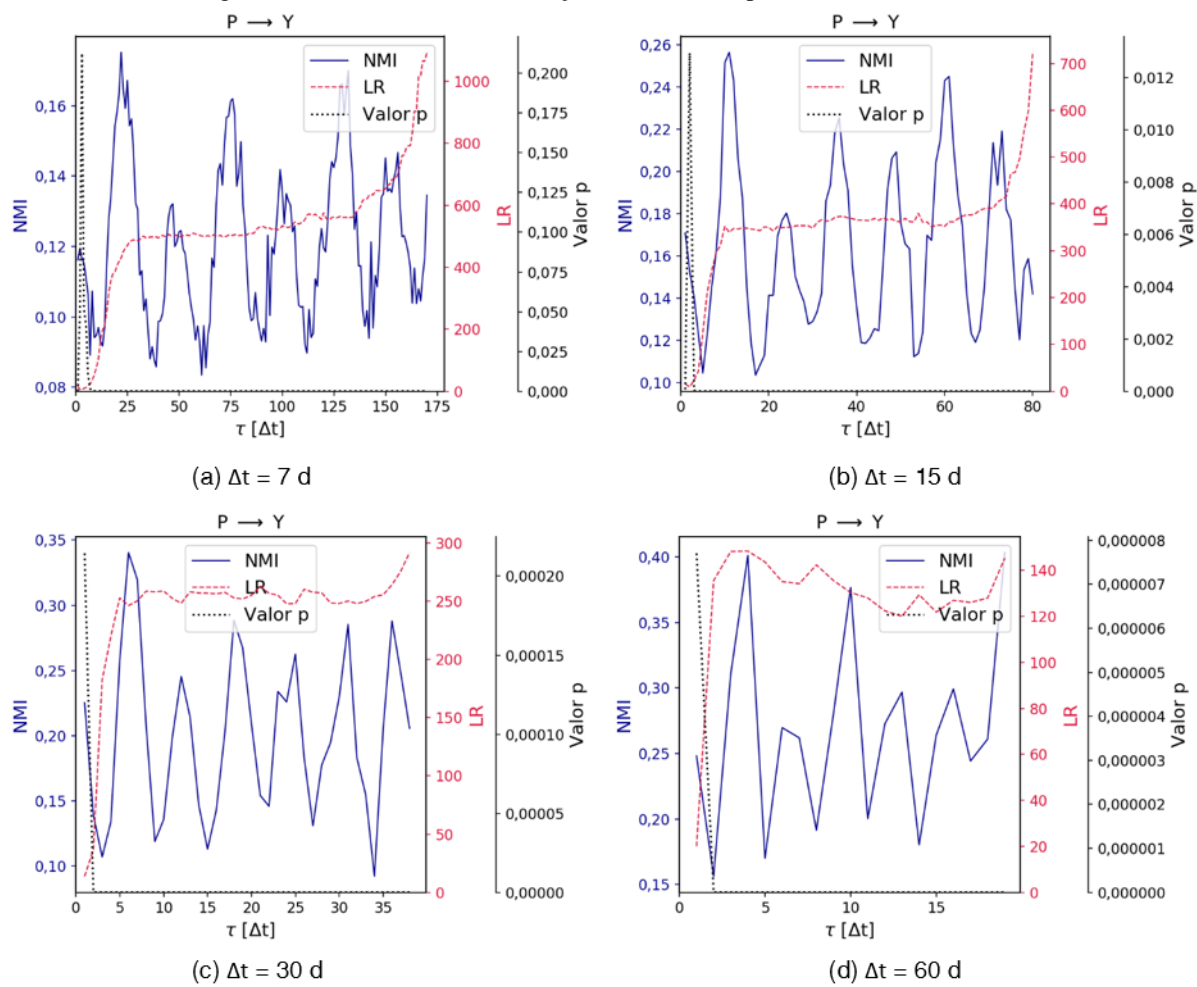
Fonte: Autor.

Figura D2 – Resultados da simulação referentes ao problema sintético S2-2.  $\Delta t = 1$  dia.



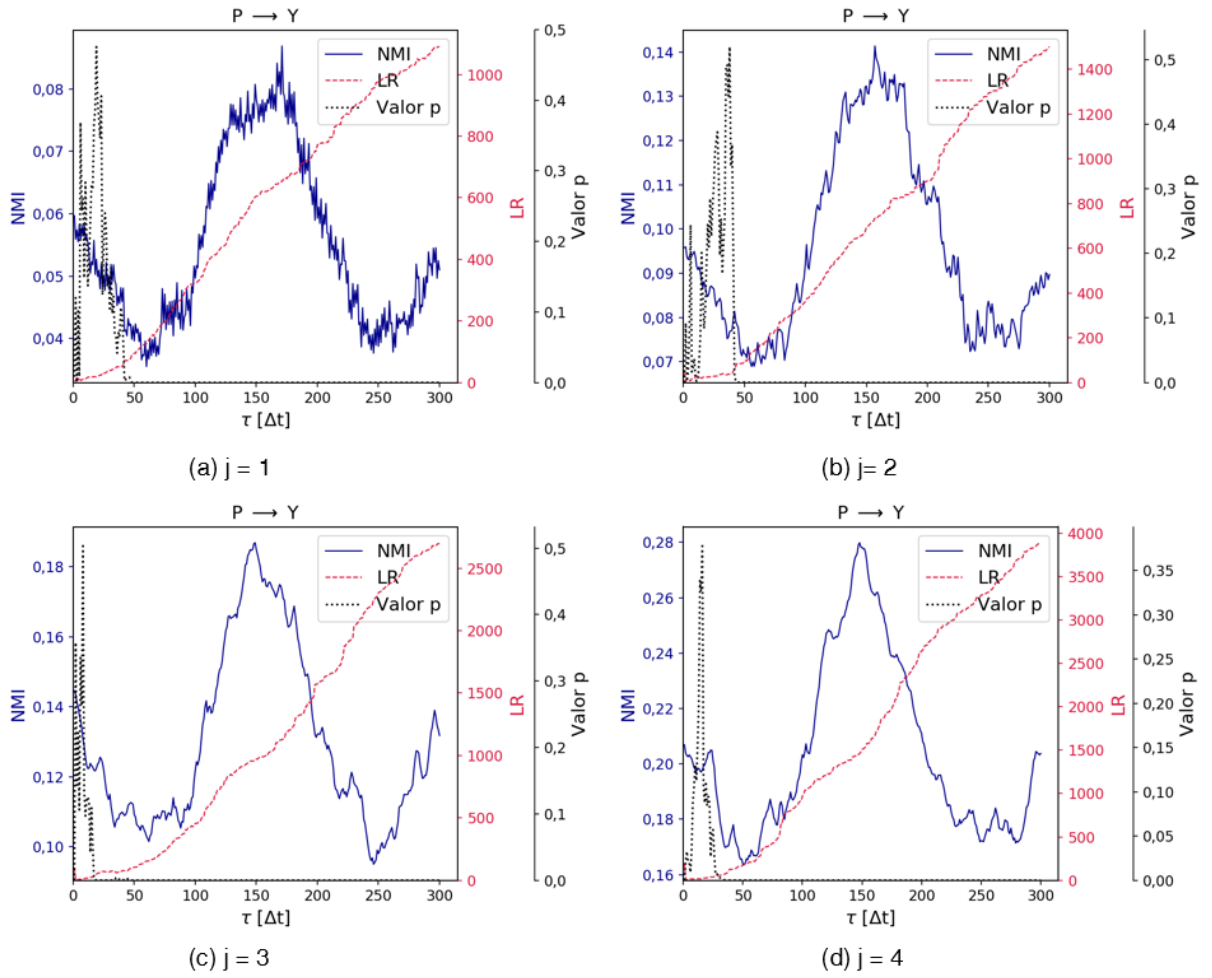
Fonte: Autor.

Figura D3 – Resultados da simulação referentes ao problema sintético S2-3.



Fonte: Autor.

Figura D4 – Resultados da simulação referentes ao problema sintético S2-4.  $\Delta t = 1$  dia.



Fonte: Autor.