

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO  
CENTRO DE CIÊNCIAS AGRÁRIAS E ENGENHARIAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA

**ANÁLISE DA DISPERSÃO DE POLUENTES EM RIOS  
VIA APROXIMAÇÃO BAYESIANA**

**RUAN DE REZENDE FARIA**

ALEGRE, ES

2019



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO  
CENTRO DE CIÊNCIAS AGRÁRIAS E ENGENHARIAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA

**ANÁLISE DA DISPERSÃO DE POLUENTES EM RIOS  
VIA APROXIMAÇÃO BAYESIANA**

**RUAN DE REZENDE FARIA**

ALEGRE, ES

2019

**Ruan de Rezende Faria**

**ANÁLISE DA DISPERSÃO DE POLUENTES EM RIOS  
VIA APROXIMAÇÃO BAYESIANA**

Dissertação de mestrado apresentada ao Programa de Pós-graduação em Engenharia Química do Centro de Ciências Agrárias e Engenharias da Universidade Federal do Espírito Santo, como parte das exigências para obtenção do Título de Mestre em Engenharia Química.

Área de Concentração: Engenharia Química

Linha de Pesquisa: Modelagem, Otimização e Análise de Processos

Orientador: Prof. D. Sc. Julio Cesar Sampaio Dutra

Coorientadores: Prof. D. Sc. Wellington Betencurte da Silva

Prof. D. Sc. Icaro Pianca Guidolini

ALEGRE, ES

Fevereiro, 2019

Ficha catalográfica disponibilizada pelo Sistema Integrado de Bibliotecas - SIBI/UFES e elaborada pelo autor

---

D278a de Rezende Faria, Ruan, 1994-  
Análise da dispersão de poluentes em rios via aproximação Bayesiana / Ruan de Rezende Faria. - 2019.  
90 f. : il.

Orientador: Julio Cesar Sampaio Dutra.

Coorientadores: Wellington Betencurte da Silva, Icaro Pianca Guidolini.

Dissertação (Mestrado em Engenharia Química) - Universidade Federal do Espírito Santo, Centro de Ciências Agrárias e Engenharias.

1. Filtro de Partículas. 2. Problemas Inversos. 3. Modelagem Ambiental. 4. Métodos Heurísticos. 5. Aproximação Bayesiana. I. Sampaio Dutra, Julio Cesar. II. Betencurte da Silva, Wellington. III. Pianca Guidolini, Icaro. IV. Universidade Federal do Espírito Santo. Centro de Ciências Agrárias e Engenharias. V. Título.

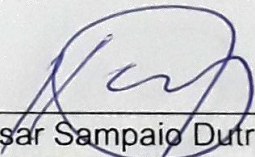
CDU: 66.0

---

**ANÁLISE DA DISPERSÃO DE POLUENTES VIA APROXIMAÇÃO BAYESIANA**

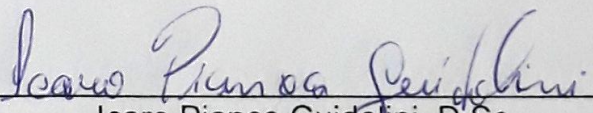
Dissertação de mestrado apresentada ao Programa de Pós-graduação em Engenharia Química do Centro de Ciências Agrárias e Engenharias da Universidade Federal do Espírito Santo, como parte das exigências para obtenção do Título de Mestre em Engenharia Química.

Aprovada em 13 de fevereiro de 2019.



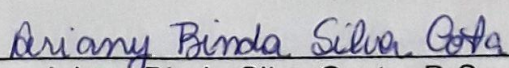
---

Julio Cesar Sampaio Dutra, D.Sc.  
Universidade Federal do Espírito Santo



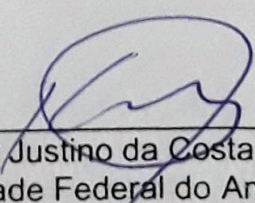
---

Icaro Pianca Guidolini, D.Sc.  
Universidade Federal do Espírito Santo



---

Ariany Binda Silva Costa, D.Sc.  
Universidade Federal do Espírito Santo



---

José Mir Justino da Costa, D.Sc.  
Universidade Federal do Amazonas

## **AGRADECIMENTOS**

Primeiramente agradeço a Deus por todas as graças concedidas durante a realização do mestrado. Este período foi muito importante para o meu crescimento profissional e pessoal. Também quero agradecer aos meus orientadores, Julio, Wellington e Icaro, por todos os questionamentos e sugestões dadas, que contribuíram muito para o melhoramento desta dissertação de mestrado. Estendo os agradecimentos também aos meus familiares, especialmente, aos meus pais, os quais são exemplos para mim. E, por fim, agradeço a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) – código de financiamento 001, a qual financiou em parte este estudo.

## RESUMO

FARIA, Ruan de Rezende. **Análise da Dispersão de Poluentes em Rios via Aproximação Bayesiana**. 2019. Dissertação (Mestrado em Engenharia Química) – Universidade Federal do Espírito Santo, Alegre, ES. Orientador: Julio Cesar Sampaio Dutra. Coorientadores: Wellington B. da Silva e Icaro Pianca Guidolini.

O estudo da dispersão de poluentes em rios está vinculado à análise da carga de poluentes conservativos lançados instantaneamente (fontes difusas de poluição) ou de maneira contínua (fontes pontuais) em um dado rio. Estes problemas foram abordados nesta dissertação de mestrado por meio de dados experimentais com traçador conservativo injetado em pulso e continuamente. Objetivou-se identificar as características do traçador (isto é, magnitude, distribuição espacial e duração do lançamento). A abordagem da dinâmica desses problemas foi realizada por meio de aproximação de problemas inversos, sendo computada pelo método *Simulated Annealing* (SA) e pela assimilação de dados baseada nos métodos Bayesianos de *Particle Filtering* (PF). O método SA permitiu obter resultados satisfatórios na representação da dinâmica do sistema de modo *off-line*. Esses resultados foram avaliados estatisticamente ao se empregar 120 repetições do algoritmo e diferentes níveis de incerteza. Contudo, uma limitação importante foi o tempo computacional para aproximar em tempo real. Por esse motivo, utilizou-se o PF para quantificação de incertezas em diferentes condições de distribuição *a priori* e número de partículas, observando sequencialmente o desempenho dos filtros SIR e SIR com suavizador de Kernel em relação ao erro quadrático médio (RMS), coeficiente de determinação ( $R^2$ ), coeficiente de autocorrelação do resíduo (SC) e coeficiente de reamostragem mínimo ( $REA_{\min}$ ). A técnica de suavização de Kernel, responsável por reduzir o empobrecimento da amostra do filtro de partículas SIR, propiciou os melhores resultados. Com isso, a utilização do filtro SIR com o suavizador de Kernel confirmou-se como uma alternativa para possíveis aplicações de controle de poluição de rios, inferência virtual e alarme.

Palavras-chave: Modelagem ambiental, Filtro de Partículas, Problemas Inversos, Suavizador de Kernel, Métodos Heurísticos.

## ABSTRACT

FARIA, Ruan de Rezende. **Analysis of Pollutant Dispersion in Rivers via Bayesian Approximation**. 2019. Dissertation. (Master degree in Chemical Engineering) – Federal University of Espírito Santo, Alegre, ES. Adviser: Julio Cesar Sampaio Dutra. Co-adviser: Wellington Betencurte da Silva and Icaro Pianca Guidolini.

Study of the dispersion of pollutants in rivers is related to the analysis of the load of conservative pollutants released instantaneously (diffuse sources of pollution) or of continuous way (point sources) in a given river. These problems were addressed in this master's dissertation by means of experimental data with conservative tracer injected in pulse and continuously. The objective was to identify tracer characteristics (i.e., magnitude, spatial distribution and duration of activity). The approach to the dynamics of these problems was carried out by means of approximation of inverse problems, being computed by the Simulated Annealing (SA) method and the assimilation of data based on the Bayesian methods of Particle Filtering (PF). The SA method allowed obtaining satisfactory results in the representation of the dynamics of the system off-line. These results were statistically evaluated by using 120 algorithm repetitions and different levels of uncertainty. However, one important limitation was the computational time to approximate on-line. For this reason, the PF (*particle filtering*) has been used to quantify the uncertainties in different conditions of priori distribution and number of particles, observing sequentially the performance of the SIR and SIR particle filter with Kernel smoothing in relation to the Root Mean Square (RMS), coefficient of determination ( $R^2$ ), autocorrelation coefficient of the residue (SC) and minimum resampling coefficient ( $REA_{\min}$ ). Better results were obtained through the Kernel smoothing technique, responsible for reducing sample depletion of the SIR particle filter. Therefore, the use of the SIR filter with the Kernel smoothing has been confirmed as an alternative for possible applications of river pollution control, virtual inference and alarm.

Keywords: Environmental Modeling, Particles Filter, Inverse Problems, Kernel Smoothing and Heuristic Methods.

## LISTA DE FIGURAS

Figura 1 – Volume de controle $dx dy dz$ . Fonte: BIRD <i>et al.</i> (2006), modificado pelo autor.....	20
Figura 2 – Diagrama conceitual dos pontos de monitoramento. Fonte: LIU <i>et al.</i> (2008), modificado pelo autor.....	26
Figura 3 – Diagrama conceitual do experimento com injeção em pulso. Fonte: autor. ....	27
Figura 4 – Diagrama esquemático dos componentes do sistema. Fonte: HUTTON <i>et al.</i> (2010), modificado pelo autor.....	28
Figura 5 – Pseudocódigo do PF, original para estimativas de estados. Fonte: DA SILVA (2012), modificado pelo autor.....	36
Figura 6 – Pseudocódigo do filtro de Partículas SIR. Fonte: DA SILVA (2012), modificado pelo autor. ....	38
Figura 7 - Pseudocódigo do filtro de Partículas SIR, com Kernel <i>smoothing</i> . Fonte: DA SILVA (2012), modificado pelo autor.....	40
Figura 8 - Pseudocódigo do filtro de Partículas ASIR. Fonte: DA SILVA (2012), modificado pelo autor. ....	41
Figura 9 – Representação esquemática do estudo de caso. Fonte: autor. ....	45
Figura 10 – Concentração de poluente na posição de detecção.....	55
Figura 11 – Caracterização da distribuição de probabilidade dos valores de massa (a) e distância da fonte poluidora (b), considerando 120 repetições do algoritmo <i>Simulated Annealing</i> . Fonte: autor. ....	60
Figura 12 – Análise detalhada das amostras de massa ajustada pelo algoritmo <i>Simulated Annealing</i> : (a) Histograma com ajuste pela curva normal; (b) análise <i>Box Plot</i> . Fonte: autor.....	60
Figura 13 – Análise detalhada das amostras de distância da fonte poluidora ajustada pelo algoritmo <i>Simulated Annealing</i> : (a) Histograma com ajuste pela curva normal; (b) análise <i>Box Plot</i> . Fonte: autor. ....	61
Figura 14 – Valores de RMS obtidos considerando 120 repetições do algoritmo <i>Simulated Annealing</i> em relação à condição inicial. Fonte: autor. ....	62
Figura 15 – Ajuste da concentração de poluente, estimando a massa do poluente e a posição de lançamento pelo algoritmo <i>Simulated Annealing</i> . Fonte: autor.....	63
Figura 16 – Caracterização da distribuição de probabilidade dos valores de massa (a), distância da fonte poluidora (b), dispersão longitudinal (c) e velocidade (d), considerando 120 repetições do algoritmo <i>Simulated Annealing</i> . Fonte: autor.....	64
Figura 17 – Análise detalhada das amostras de massa ajustada pelo algoritmo <i>Simulated Annealing</i> : (a) Histograma com ajuste pela curva normal; (b) análise <i>Box Plot</i> . Fonte: autor.....	65
Figura 18 – Análise detalhada das amostras de distância da fonte poluidora ajustada pelo algoritmo <i>Simulated Annealing</i> : (a) Histograma com ajuste pela curva normal; (b) análise <i>Box Plot</i> . Fonte: autor.....	65
Figura 19 – Análise detalhada das amostras de dispersão longitudinal ajustada pelo algoritmo <i>Simulated Annealing</i> : (a) Histograma com ajuste pela curva normal; (b) análise <i>Box Plot</i> . Fonte: autor. ....	66
Figura 20 – Análise detalhada das amostras de velocidade ajustada pelo algoritmo <i>Simulated Annealing</i> : (a) Histograma com ajuste pela curva normal; (b) análise <i>Box Plot</i> . Fonte: autor.....	66
Figura 21 – Ajuste da concentração de poluente, estimando a massa do poluente, a posição de lançamento, dispersão longitudinal e velocidade pelo algoritmo <i>Simulated Annealing</i> . Fonte: autor. ....	67

Figura 22 – Valores de RMS obtidos considerando 120 repetições do algoritmo <i>Simulated Annealing</i> em relação à condição inicial. Fonte: autor.....	68
Figura 23 – Estimativas da concentração em tempo real (a), carga de poluente conservativo (b) e distância da fonte poluidora (c), considerando a pior condição examinada da Tabela 4: $N_{part} = 100$ e $\sigma_p / P_0 = 25\%$ . Fonte: autor.....	70
Figura 24 – Estimativas da concentração em tempo real (a), carga de poluente conservativo (b) e distância da fonte poluidora (c), considerando a melhor condição examinada da Tabela 4: $N_{part} = 1000$ e $\sigma_p / P_0 = 60\%$ . Fonte: autor.....	71
Figura 25 – Análise do REA para $N_{part} = 100$ e $\sigma_p / P_0 = 25\%$ (a) e $N_{part} = 1000$ e $\sigma_p / P_0 = 60\%$ (b). Fonte: autor.....	72
Figura 26 – Análise da autocorrelação do resíduo de estimação para $N_{part} = 100$ e $\sigma_p / P_0 = 25\%$ (a) e $N_{part} = 1000$ e $\sigma_p / P_0 = 60\%$ (b). Fonte: autor.....	73
Figura 27 – Estimativas da concentração em tempo real (a), carga de poluente conservativo (b), distância da fonte poluidora (c), dispersão longitudinal (d) e velocidade (e), considerando a pior condição examinada da Tabela 5: $N_{part} = 100$ e $\sigma_p / P_0 = 25\%$ . Fonte: autor.....	75
Figura 28 – Estimativas da concentração em tempo real (a), carga de poluente conservativo (b), distância da fonte poluidora (c), dispersão longitudinal (d) e velocidade (e), considerando a melhor condição da Tabela 5: $N_{part} = 1000$ e $\sigma_p / P_0 = 60\%$ Fonte: autor. ....	76
Figura 29 – Análise do REA para $N_{part} = 100$ e $\sigma_p / P_0 = 25\%$ (a) e $N_{part} = 1000$ e $\sigma_p / P_0 = 60\%$ (b). Fonte: autor. ....	77
Figura 30 – Análise da autocorrelação do resíduo de estimação para $N_{part} = 100$ e $\sigma_p / P_0 = 25\%$ (a) e $N_{part} = 1000$ e $\sigma_p / P_0 = 60\%$ (b). Fonte: autor.....	77
Figura 31 – Análise do REA para o filtro de partículas SIR com Kernel <i>Smoothing</i> (a) e filtro de partículas SIR (b), considerando $N_{part} = 1000$ e $\sigma_p / P_0 = 60\%$ . Fonte: autor.....	80
Figura 32 – Concentração estimada em tempo real, considerando método de Kernel Smoothing para $E_x$ e $U$ , $N_{part} = 1000$ e $\sigma_p / P_0 = 60\%$ . Fonte: autor.....	80
Figura 33 – Ajuste em tempo real da massa (a), distância da fonte de poluição (b), velocidade (c), dispersão longitudinal (d) e tempo de duração do lançamento (e), considerando Kernel Smoothing para $E_x$ e $U$ , $N_{part} = 1000$ e $\sigma_p / P_0 = 60\%$ . Fonte: autor.....	81

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>13</b>
1.1	Apresentação da Pesquisa .....	13
1.2	Justificas da Escolha do Tema de Pesquisa.....	14
1.3	Objetivos Geral e específicos .....	15
1.4	Trabalhos Preliminares .....	16
<b>2</b>	<b>MODELAGEM DO TRANSPORTE DE POLUENTES EM RIOS .....</b>	<b>19</b>
2.1	Equação Geral do Transporte de Poluentes em Rios (EGTPR) .....	19
2.2	Advecção e Difusão em Rios .....	21
2.2.1	Transformações na Água: Reações Químicas e Biológicas.....	22
2.2.2	Contribuições Externas: Fontes Pontuais e difusas.....	23
2.3	Comentários Finais.....	24
<b>3</b>	<b>TÉCNICAS PARA ANÁLISE DA DISPERSÃO DE POLUENTES EM RIOS.....</b>	<b>25</b>
3.1	Problemas Inversos.....	25
3.2	Assimilação de dados.....	27
3.3	Fluidodinâmica Computacional .....	30
3.4	Comentários finais .....	32
<b>4</b>	<b>ESTIMAÇÃO VIA APROXIMAÇÃO BAYESIANA .....</b>	<b>33</b>
4.1	Amostragem de Monte Carlo Sequencial ( <i>Particle Filtering - PF</i> ) .....	33
4.2	Estimação conjunta de parâmetros e estados .....	39
4.3	Amostragem de Monte Carlo Sequencial utilizando importância auxiliar – ASIR.....	40
4.4	Comentários Finais.....	42
<b>5</b>	<b>METODOLOGIA.....</b>	<b>43</b>
5.1	Problema de Lançamento Instantâneo .....	43
5.1.1	Descrição do Estudo de Caso – medidas experimentais.....	43
5.1.2	Formulação Matemática e solução analítica do problema físico.....	44
5.1.3	Problema de Estimação Off-line.....	46
5.1.4	Problema de Estimação em Tempo Real.....	50
5.2	Problema de Lançamento Contínuo .....	53
5.2.1	Descrição do Estudo de Caso – medidas sintéticas.....	53
5.2.2	Sistema Físico e Formulação Matemática .....	54
5.2.3	Medidas sintéticas .....	54
5.2.4	Problema de Estimação em tempo real .....	56
5.3	Comentários Finais.....	57
<b>6</b>	<b>RESULTADOS E DISCUSSÃO .....</b>	<b>59</b>
6.1	Problema de Lançamento Instantâneo .....	59
6.1.1	Análise com Simulated Annealing (Estimação Off-line) .....	59

6.1.2	Análise com Filtro de Partículas (Estimação On-line) .....	68
6.1.2.1	Cenário 1: apenas $M$ e $x - x_0$ são desconhecidos .....	68
6.1.2.2	Cenário 2: Método Kernel smoothing para $E_x$ e $U$ .....	73
6.2	Problema de Lançamento Contínuo.....	77
6.2.1	Análise com Filtro de Partículas (Estimação On-line) .....	78
<b>7</b>	<b>CONCLUSÕES E SUGESTÕES .....</b>	<b>83</b>
	<b>REFERÊNCIAS .....</b>	<b>85</b>

# 1 INTRODUÇÃO

## 1.1 Apresentação da Pesquisa

Os rios sempre foram os receptores das descargas de água residuais de atividades humanas, sejam elas fontes domésticas, efluentes industriais ou agrícolas, ou águas de processo de mineração, entre outras fontes. Nos últimos dois séculos, desde a era da industrialização, o aumento maciço da produção industrial acompanhado pelo crescimento de grandes populações urbanas levou a graves problemas de poluição da água em muitos rios (BENEDINI; TSAKARIS, 2013; JIRKA; WEITBRECHT, 2005).

A qualidade da água de um ponto de vista local, com seus atributos de distribuição espacial e temporal, é o resultado de vários processos físicos, químicos e biológicos que ocorrem devido às múltiplas cargas de poluentes lançados em um trecho do rio. Uma sólida compreensão e descrição desses processos são necessárias para que os engenheiros de projeto e as autoridades de água possam desenvolver soluções de descarga apropriadas e estratégias de controle que atendam às leis e objetivos de qualidade de água aplicáveis (SOCOLOFSKY; JIRKA, 2005; JIRKA; WEITBRECHT, 2005).

É útil distinguir dois tipos de problemas de descarga de poluentes, uma vez que estes possuem diferentes características de resposta física, bem como diferentes estratégias de controle de poluição que são usadas pelas autoridades de água. São os problemas de descarga contínua e problemas de liberação instantânea. A distinção básica entre os dois problemas é a forma como o efeito da duração da descarga afeta como o poluente se mistura ao longo do rio (JI, 2017).

As diferentes metodologias possíveis para analisar os fenômenos mencionados acima apresentam em comum o fato de dependerem do entendimento da dispersão de poluentes em rios. Esse fenômeno é o mais importante, tendo interferência significativa na dinâmica do poluente. Por esse motivo, diversas são as alternativas apresentadas na literatura para sua modelagem, sendo a utilização de traçadores conservativos para a representação do poluente as mais utilizadas (LAUNAY et al., 2015; LUGON et al., 2008; RODRIGUES et. al, 2013; ZENG et al., 2014).

Normalmente, o emprego de tal técnica objetiva estimar o coeficiente de dispersão conhecida a distribuição da grandeza em um ponto à jusante de um rio particular. Apesar de uma alternativa interessante, constando em diversos trabalhos na literatura (métodos determinísticos e heurísticos), não é relatada a possibilidade de identificação da magnitude do traçador injetado, posição injetada e a dinâmica via monitoramento em tempo real (ou online). Deve-se somar a tudo isso, a identificação de incertezas nos parâmetros dos modelos, estrutura, entradas e dados de calibração. Diante deste cenário inovador, as técnicas Bayesianas para monitoramento têm sido cada vez mais aplicadas por sua capacidade de lidar com a incerteza e incorporar informações a priori, o que permite o uso de experiências anteriores e fornece a avaliação probabilística necessária para apoiar a tomada de decisões. Em comparação com as abordagens tradicionais de modelagem, os parâmetros estimados do modelo na abordagem Bayesiana são apresentados em termos de uma distribuição conjunta posterior; assim, podem-se prever melhor as violações do padrão de qualidade da água e fornecer mais informações sobre a incerteza. Além disso, os intervalos de credibilidade Bayesianos (correlatos aos intervalos de confiança da estatística frequentista) resultantes podem então ser transmitidos diretamente para o supervisor do sistema de gestão da qualidade da água com certo nível de risco probabilístico (YONG *et al.*, 2008; MANSOURI *et al.*, 2013).

## 1.2 Justificas da Escolha do Tema de Pesquisa

As justificativas para abordagem da análise da dispersão de poluentes em rios podem ser melhor contextualizadas ao mencionar o relatório mundial da Organização das Nações Unidas (ONU - 2017) sobre Desenvolvimento dos Recursos Hídricos. Destacou-se, em conformidade com este relatório: um terço dos rios do planeta está poluído; mais de 80% das águas residuais vão parar nos rios e oceanos do planeta; mais de 800 mil pessoas morreram em decorrência da falta de saneamento básico em 2012; de 181 países pesquisados, apenas 55 tinham dados a respeito do tratamento e utilização de águas residuais; e, no Brasil, 60% do esgoto não é tratado.

Neste contexto, também vale ressaltar a ocorrência de desastres ambientais, por exemplo, resultantes de: (1) Atividades de mineração de minério de ferro - O colapso de uma barragem de mineração em Mariana, Brasil, em 2015: foi responsável por lançar uma quantidade estimada de  $60 \times 10^3 \text{ m}^3$  de rejeitos da extração de minério de ferro, resultando na destruição total de Bento Rodrigues e destruição parcial de algumas outras aldeias, vitimando um total de 19 pessoas (Marta-Almeida et al., 2016; Neves et al. , 2016; Segura et al., 2016); (2) Exploração de petróleo - derrame de petróleo no Delta do Níger, Nigéria: foi relatado que desde 1958, a quantidade total que foi derramada no ambiente do Delta do Níger é entre 9 milhões e 13 milhões de barris (Abii e Nwosu, 2009; Allison et al., 2017); e (3) descarga de esgoto a céu aberto - A poluição do rio em Guangzhou, China: foi descrita a poluição do rio Donghaochong em espaços urbanos por transbordamento de esgoto e desastre de inundação (Wu et al., 2016; Zhou et al., 2016) .

Todas essas informações destacam a relevância e multidisciplinaridade dos problemas envolvendo recursos hídricos, encaixando-se nesse contexto os assuntos abordados nesta Dissertação de Mestrado.

### 1.3 Objetivos Geral e específicos

O objetivo geral dessa Dissertação de Mestrado é propor uma nova metodologia, baseada na técnica Bayesiana de filtros de partículas com amostragem e reamostragem, para a avaliação da dispersão de poluentes em rios. Uma vez realizada uma revisão bibliográfica sobre o tema, percebeu-se a ausência de materiais que descrevessem a dinâmica do poluente em tempo real. Isso é extremamente necessário, principalmente para o gerenciamento de cargas de poluentes lançados de modo instantâneo ou contínuo em rios, nos quais a presença do poluente deve ser prontamente identificada.

Para que o objetivo geral descrito acima possa ser alcançado, os seguintes objetivos específicos devem ser contemplados, a saber:

1. Descrição de técnicas pertinentes à análise da carga de poluentes lançados instantaneamente e continuamente em rios;

2. Modelagem da dinâmica do fenômeno com fins de monitoramento off-line e em tempo real, bem como a obtenção das medidas necessárias para a determinação dos parâmetros do modelo matemático;
3. Implementação computacional de um método off-line, como a técnica heurística *Simulated Annealing*, para fins de comparação e dos filtros de partículas SIR e SIR com suavizador de Kernel;
4. Análise de desempenho das duas técnicas mencionadas acima quanto aos parâmetros associados à dinâmica da dispersão de poluentes em rios, e comparação dos resultados obtidos.

Um cenário de investigações teórico-computacionais é explorado nesta Dissertação, cuja apresentação é estruturada nos seguintes capítulos: Revisão Bibliográfica, Fundamentos Teóricos, Metodologia, Resultados e Discussões, Conclusão e Referências.

#### 1.4 Trabalhos Preliminares

Esta pesquisa de dissertação foi desenvolvida no Laboratório de Modelagem e Otimização de Processos (MOP), no Centro de Ciências Agrárias e Engenharias da UFES. Os trabalhos deste laboratório têm se dedicado a análise de processos químicos empregando diferentes metodologias para modelagem, otimização e análise de processos, o que propiciou bases teóricas e computacionais importantes para esta dissertação.

A revisão da literatura proporcionada por esta pesquisa de dissertação possibilitou a realização de dois trabalhos acadêmicos. O primeiro focou-se no ajuste dos parâmetros de dispersão do Rio Macaé utilizando o método *Simulated Annealing*, empregado com propósito de investigar novas opções de avaliação da dispersão de poluentes em rios. O outro trabalho investigou a análise de desempenho do filtro de partículas SIR usando a autocorrelação do resíduo, termo esse ainda pouco explorado em problemas solucionados sequencialmente. Ambos os trabalhos resultaram em artigos que foram submetidos e apresentados no XXI Encontro Nacional de Modelagem Computacional realizado em Búzios (FARIA et al., 2018). Todo este aprendizado, juntamente com os resultados obtidos nesta Dissertação de

Mestrado, viabilizou a submissão de um novo artigo em uma importante revista internacional.



## 2 MODELAGEM DO TRANSPORTE DE POLUENTES EM RIOS

A modelagem da dispersão de poluentes em rios é uma maneira matemática de interpretar as possíveis modificações que o poluente poderá sofrer ao entrar em contato com a água do rio. Tal análise com o modelo requer o entendimento dos fenômenos de transporte do rio e transferência de massa do poluente, os quais são influenciados pelas características meteorológicas, geográficas e biológicas do ambiente em que se encontram, tornando dessa forma a modelagem de rios uma tarefa complexa (JI, 2017; BENEDINI; TSAKARIS, 2013).

Além disso, a variedade de poluentes e rios existentes a serem modelados é extensa, em virtude disso o objetivo deste capítulo é apresentar a equação geral do transporte de poluentes em rios mais utilizada na literatura consultada (JI, 2017; BENEDINI; TSAKARIS, 2013; SOCOLOFSKY; JIRKA, 2005).

### 2.1 Equação Geral do Transporte de Poluentes em Rios (EGTPR)

A EGTPR é uma equação de transferência de massa que depende de dois mecanismos: a transferência de massa molecular ( $-\rho E \nabla \alpha_i$ ) e a transferência de massa convectiva ( $\rho \alpha_i v$ ), em que  $E$  é o coeficiente de difusão,  $\rho$  é a densidade,  $\alpha_i$  é a fração mássica do componente e  $v$  é a velocidade. A equação de transferência de massa é obtida através da conservação do fluxo total ( $n_i = \rho \alpha_i v - \rho E \nabla \alpha_i$ ) sobre um volume elementar de água do rio  $dx dy dz$  (Figura 1), fixo no espaço, através do qual circula o poluente (JI, 2017).

O aumento de massa no interior do volume de controle da Figura 1 é igual à taxa de entrada menos a taxa de saída de massa do poluente através da superfície de controle. Considera-se a variação de massa na direção  $x$ , como:  $(n)_x dy dz$  na entrada (face  $x$ ) e  $(n)_{x+dx} dy dz$  na saída (face  $x + dx$ ). As taxas de massa nas direções  $y$  e  $z$  são definidas de maneira análoga. Já o aumento de massa com o tempo ( $t$ ) no interior do volume de controle é  $dx dy dz (\partial \rho / \partial t)$ . Conseqüentemente, o balanço de matéria para o volume de controle é dado pela Equação 1.

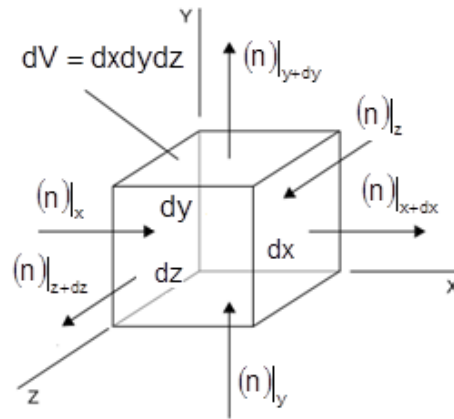


Figura 1 – Volume de controle  $dx dy dz$ . Fonte: BIRD *et al.* (2006), modificado pelo autor.

$$dx dy dz \left( \frac{\partial \rho}{\partial t} \right) = dy dz [(n)|_x - n|_{x+dx}] + dx dz [(n)|_y - n|_{y+dy}] + dx dy [(n)|_z - n|_{z+dz}] \quad (1)$$

Dividindo toda a equação por  $dx dy dz$  e aplicando o limite quando  $dx dy dz$  tende a zero, obtém-se a partir da definição das derivadas parciais a Equação 2.

$$\frac{\partial \rho_i}{\partial t} = - \left( \frac{\partial n_{i,x}}{\partial x} + \frac{\partial n_{i,y}}{\partial y} + \frac{\partial n_{i,z}}{\partial z} \right) \quad (2)$$

Substituindo o termo  $n_i$  na Equação 2, obtém-se a Equação 3 que relaciona os efeitos difusivos e advectivos para todas as dimensões espaciais.

$$\frac{\partial (\rho \alpha_i)}{\partial t} = - \left( \frac{\partial (\rho \alpha_i v_x)}{\partial x} + \frac{\partial (\rho \alpha_i v_y)}{\partial y} + \frac{\partial (\rho \alpha_i v_z)}{\partial z} \right) + \left( \frac{\partial}{\partial x} \left( \rho E_x \frac{\partial (\alpha_i)}{\partial x} \right) + \frac{\partial}{\partial y} \left( \rho E_y \frac{\partial (\alpha_i)}{\partial y} \right) + \frac{\partial}{\partial z} \left( \rho E_z \frac{\partial (\alpha_i)}{\partial z} \right) \right) \quad (3)$$

A Equação 3 descreve os processos de entrada e saída de massa no interior do volume de controle. Outras variantes ainda podem ser mencionadas, a saber: a carga de fontes externas de poluição (S); e geração e consumo por reações que

podem ser de origem química ou biológica (R). Dessa forma, a expressão completa é descrita pela Equação 4.

$$\frac{\partial(\rho\alpha_i)}{\partial t} = -\left(\frac{\partial(\rho\alpha_i v_x)}{\partial x} + \frac{\partial(\rho\alpha_i v_y)}{\partial y} + \frac{\partial(\rho\alpha_i v_z)}{\partial z}\right) + \left(\frac{\partial}{\partial x}\left(\rho E \frac{\partial(\alpha_i)}{\partial x}\right) + \frac{\partial}{\partial y}\left(\rho E \frac{\partial(\alpha_i)}{\partial y}\right) + \frac{\partial}{\partial z}\left(\rho E \frac{\partial(\alpha_i)}{\partial z}\right)\right) + S + R \quad (4)$$

A Equação 4 é o modelo base para o desenvolvimento desta Dissertação de Mestrado, assim é necessário um estudo sobre o que cada termo significa e como associá-los a caracterização dos principais tipos de poluentes.

## 2.2 Advecção e Difusão em Rios

A advecção e difusão em rios são os mecanismos de transporte de poluentes mais estudados na literatura (BENEDINI; TSAKARIS, 2013; JI, 2017) e estão presentes na equação anterior. O primeiro termo à direita da igualdade refere-se aos efeitos advectivos ( $\nabla \cdot (\rho\alpha_i v)$ ), que caracterizam o transporte do poluente devido ao movimento macroscópico da água do rio. Entre as variáveis que o compõe, o vetor de velocidade  $v(x, y, z, t) = (v_x, v_y, v_z)$  é o componente mais difícil de ser interpretado em decorrência da geometria irregular dos rios, indicando-se para os casos mais complexos (fluxo turbulento) a utilização de equações de conservação de momento acopladas à equação geral do transporte de poluentes em rios (SOCOLOFSKY; JIRKA, 2005).

O outro termo da equação refere-se ao transporte por difusão  $\nabla \cdot (E\nabla \cdot (\rho w_i))$ . Neste caso, o principal componente estudado é o vetor do coeficiente de difusão  $E(x, y, z, t) = (E_x, E_y, E_z)$  devido à impossibilidade de obtenção de medidas diretamente no campo e variedades de fenômenos que podem influenciá-lo. Estes fenômenos em ordem decrescente de magnitude, segundo Benedini e Tsakiris (2013), englobam os efeitos combinados de mistura turbulenta, difusão molecular e mistura devido ao cisalhamento vertical e lateral. Quando combinados, de acordo com a literatura consultada (ABDERREZZAK *et al.*, 2015; ANI *et al.*, 2009; BAEK;

SEO, 2016; DUARTE; BOAVENTURA, 2008), redefinem a nomenclatura de coeficiente de difusão para coeficiente de dispersão.

### 2.2.1 Transformações na Água: Reações Químicas e Biológicas

Interpretar o termo R de uma forma matemática para incorporar na Equação 4 depende do entendimento das características do poluente e a forma como interage com a água. Tratando-se de transformações na água, o símbolo matemático R é apenas uma abreviação representando os mais variados tipos de modificações químicas que podem ocorrer e descrevem a variação da concentração do poluente, seguido um mecanismo que é assumido ser uma função do tempo e ser proporcional à concentração inicial (BENEDINI; TSAKIRIS, 2013). Por exemplo, supondo uma cinética simples para a transformação de um poluente na água, a representação para o termo R pode ser dada pela Equação 5.

$$\frac{dC}{dt} = R = -kC^m \quad (5)$$

Nesta equação, m é a ordem da reação, C é a concentração do poluente e k é a constante de decaimento da reação de ordem m. Segundo Ji (2017), as ocasiões mais comuns em que pode ser aplicada a Equação 5 quanto à ordem de reação e sua solução matemática são sumarizadas de acordo com o Quadro 1.

Os processos afetando o destino e o decaimento de substâncias tóxicas são resultados de transformações químicas e biológicas. Além desses fatores, incluem-se os processos de mineralização e decomposição, hidrólise, fotólise, biodegradação, bioconcentração e volatilização. Segundo Ji (2017), esses processos de transformação podem ser descritos por reações de cinética de primeira ordem, e as constantes de decaimento para cada um dos processos mencionados acima, respectivamente, são aditivas e podem ser linearmente combinadas para formar uma nova constante de decaimento conforme a Equação 6.

$$k = k_m + k_h + k_p + k_{bd} + k_{bc} + k_v \quad (6)$$

Quadro 1 – Caracterização das reações cinéticas mais comuns em rios.

Ordem	Solução	Exemplos
0	$C = C_0 - kt$	Produção de metano e lançamento de produtos de hidrólise de sedimentos anaeróbicos.
1	$C = C_0 \exp(-kt)$	Taxa de morte e respiração de bactérias, demanda de oxigênio bioquímico e reação de produção de algas, além de ser a mais utilizada para a maioria dos processos.
2	$\frac{1}{C} = \frac{1}{C_0} + kt$	Reações de gases atmosféricos, substâncias tóxicas, metais pesados e taxa de morte de zooplâncton.

Fonte: Ji, 2017, modificado pelo autor.

Ao final dessa subseção, percebe-se que a complexidade de transformação dos materiais, que podem degradar o ambiente aquático, é muito variada. Algumas sugestões podem ser dadas, porém, apenas através da aquisição de uma quantidade significativa de dados, pode-se explorar o problema de uma forma adequada. E, isso é explicado pelo fato de que as condições em um ambiente real podem mudar a qualquer momento, gerando uma grande incerteza associada ao problema. Resumindo, entende-se que não é possível afirmar que um processo descrito na literatura como de primeira ordem, será dessa forma para todas as condições e em todos os ambientes (BENEDINI; TSAKARIS, 2013; JI, 2017).

### 2.2.2 Contribuições Externas: Fontes Pontuais e difusas

As contribuições externas são possíveis fontes de poluição adicionadas no rio (volume de controle) que contribuem para o aumento da massa do poluente. Quando se originam de locais em que a incidência e periodicidade de lançamento do poluente no rio são conhecidas, como acontece no destino final de plantas de tratamento de esgoto e resíduos industriais, nomeiam-se fontes pontuais de forma diferente das fontes de poluição resultantes de derramamentos, vazamentos de atividades industriais e exploração de recursos minerais, que são as fontes difusas. O Quadro 2, adaptado de Ji (2017), sumariza as principais diferenças entre elas

quanto às seguintes características importantes para modelagem: variação do tempo, magnitude do fluxo, impacto e previsibilidade, e indicam que as fontes difusas são as mais difíceis de serem identificadas e modeladas devido ao desconhecimento do que possa estar influenciando o processo (JI, 2017; BENEDINI; TSAKIRIS, 2013).

Quadro 2 – Diferenças entre as contribuições externas, importantes para modelagem.

<b>Características</b>	<b>Fontes pontuais</b>	<b>Fontes difusas</b>
Variação no tempo	Relativamente estacionário	Altamente variável
Magnitude do fluxo	Variação menor do que uma ordem de magnitude	Variação a algumas ordens de magnitude
Impacto	Mais severos durante fluxo baixo	Mais severos durante ou após eventos de tempestade
Previsibilidade	Bastante previsível	Menos previsível

Fonte: Ji, 2017, modificado pelo autor.

Finalizando, fontes difusas e pontuais podem ser modeladas utilizando a Equação 4, no entanto, os efeitos predominantes em ambas são, respectivamente, difusão e advecção (SOCOLOFSKY; JIRKA, 2005).

### 2.3 Comentários Finais

Ao final deste capítulo, percebeu-se que apesar de relativamente simples a EGTPR possa parecer, deve-se interpretar a influência das condições de escoamento e os processos de transformação nas diferentes variáveis que a compõe. No próximo capítulo, denominado como técnicas para análise da dispersão de poluentes em rios, é feita uma revisão da literatura sobre sua utilização em problemas inversos, fluidodinâmica computacional e assimilação de dados.

### 3 TÉCNICAS PARA ANÁLISE DA DISPERSÃO DE POLUENTES EM RIOS

As TADPR em rios conceituam-se como maneiras de analisar a dispersão de poluentes em rios com suporte da EGTPR. Isto significa dizer que essas técnicas podem incluir abordagens como Problemas Inversos, Fluidodinâmica Computacional e Assimilação de Dados, e depender diretamente da sua utilização. Por meio de uma revisão da literatura, podem-se perceber as inúmeras aplicações possíveis com essas técnicas. Neste capítulo, cada uma é analisada em uma seção respectiva, objetivando-se demonstrar a técnica, aplicações e modelagem. De qualquer modo, o objetivo principal da revisão de literatura é contextualizar a aplicação conjunta de problemas inversos e assimilação de dados.

#### 3.1 Problemas Inversos

Os problemas inversos são técnicas de estimação de parâmetros e estados desenvolvidos para análise de problemas cuja caracterização por instrumentos de medidas pode ser inviável ou custoso. As primeiras aplicações referem-se aos problemas de transferência de calor para lidar com a estimação de fluxos de calor nas fronteiras do sistema, usando medidas de temperatura tomadas em outro ponto do domínio. Dessa forma, trata-se de uma técnica que permite predizer, a partir de um efeito, a causa por trás de um problema (ÖZİŞİK, 1993; ORLANDE *et al.*, 2012; TARANTOLA, 1987).

Tendo como referência a EGTPR, encontram-se aplicações semelhantes na análise da dispersão de poluentes em rios, que incluem a calibração (ou ajuste) de parâmetros e estimação de estados. Para a calibração de parâmetros da EGTPR, os problemas de estimação se baseiam no histórico de medidas de concentração do poluente em diferentes locais de monitoramento ou em dados obtidos em experimentos com perturbação em pulso (BENEDINI; TSAKARIS, 2013; JI, 2017).

Neste contexto, é muito comum o gerenciamento sazonal de fontes de poluição múltipla em segmentos de rios (histórico de medidas). Os trabalhos de Chen *et al.* (2012), Liu *et al.* (2008) e Zhao *et al.* (2014) utilizaram aproximação bayesiana, método de Monte Carlo via cadeias de Markov, como técnica para solução inversa do problema e obtiveram estimativas das cargas de fontes externas de poluição (S)

e constante de decaimento ( $k$ ) para todos os segmentos do rio. A modelagem foi realizada para o domínio abstrato representado pela Figura 2, em que se realizaram as seguintes simplificações na EGTPR: regime de fluxo estacionário do rio (velocidade média na direção de escoamento  $x$  igual a  $U$ ), fontes de poluição pontuais e reagente simples com dinâmica de primeira ordem, resultando na Equação 7.

$$U \frac{dC}{dx} = -kC + S \quad (7)$$

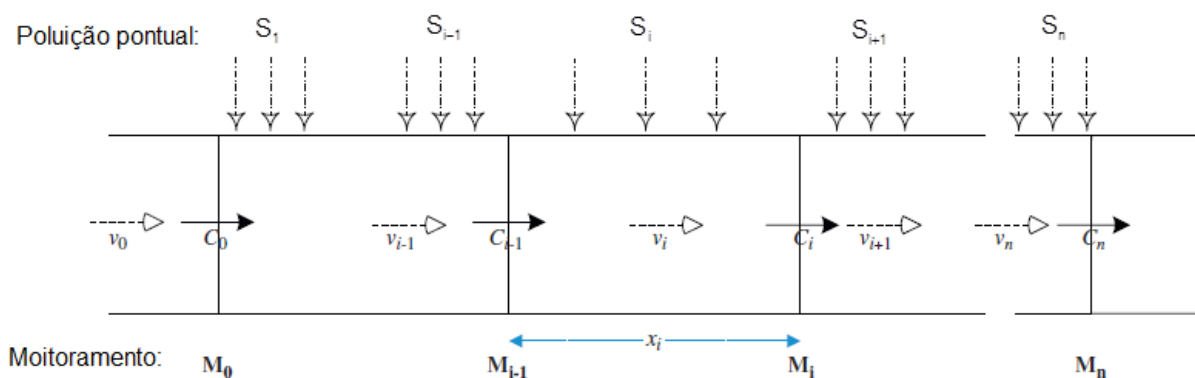


Figura 2 – Diagrama conceitual dos pontos de monitoramento. Fonte: LIU *et al.* (2008), modificado pelo autor

O outro problema de calibração da EGTPR depende de dados obtidos em campo através de testes com traçador. A Figura 3 ilustra as modificações do perfil de concentração do poluente entre o ponto de injeção e detecção. Esta metodologia é utilizada para validar aplicações de diferentes modelos de transporte de poluentes em rios (ANI *et al.*, 2009; BAEK; SEO, 2016; ABDERREZAK *et al.*, 2015), aproximações de problemas inversos usando métodos determinísticos e meta-heurísticas (RODRIGUES *et al.*, 2013; LUGON *et al.*, 2008) e amostragem de Monte Carlo Sequencial (LUGAO *et al.*, 2015).

Entre os diversos trechos de rios examinados por meio desta técnica, vale ressaltar os testes realizados no Rio Macaé, localizado na cidade do Rio de Janeiro, alvo de publicações de Rodrigues *et al.* (2013), Lugon *et al.* (2008), Telles *et al.* (2013) e Parolin *et al.* (2014), por subsidiar atividades que vão desde a agricultura até a extração de petróleo, altamente alarmantes para a qualidade dos recursos hídricos.

Apesar de todas as aplicações desta metodologia, segundo Weitbrecht e Jirka (2005), as medições de campo podem ser dispendiosas, muitas vezes difíceis de executar e geralmente limitadas a determinadas condições ambientais. Por esses motivos, na próxima seção são abordadas as técnicas baseadas em assimilação de dados, as quais permitem abordar sequencialmente o problema, sem necessidade do planejamento de um ensaio com traçador conservativo.

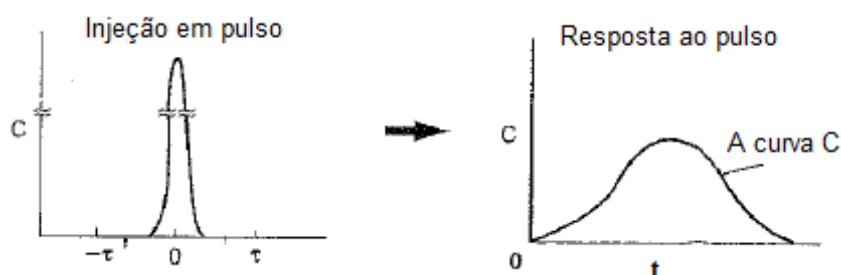


Figura 3 – Diagrama conceitual do experimento com injeção em pulso. Fonte: autor.

### 3.2 Assimilação de dados

Segundo Hutton *et al.* (2010), a assimilação de dados é um nome dado a uma classe de métodos que busca combinar a incerteza do modelo às incertezas dos dados para fornecer a melhor estimativa do estado do sistema em um determinado instante do tempo no qual as observações estão disponíveis. Isto se trata de uma análise em tempo real, em que dados referentes a um sistema são analisados e atualizados à taxa na qual são recebidos. Podem ser encontradas aplicações em monitoramento, melhorando a acurácia das previsões dos estados do sistema ( $X$ ) e das saídas ( $Y$ ) baseando-se nos dados recebidos (*real-time modelling*).

De uma perspectiva teórica, a Figura 4 ilustra os diferentes componentes da modelagem em tempo real e possíveis fontes de incertezas para a assimilação de dados, em que:  $B$  é o contorno de sistema;  $U = (u_1, \dots, u_n)$  e  $Y = (y_1, \dots, y_n)$  representam as entradas e saídas do modelo, com dimensão  $n$ ;  $x_0$  representa as condições iniciais do modelo;  $\theta = (\theta_1, \dots, \theta_m)$  são os parâmetros do modelo, com dimensão  $m$ ;  $X = (x_1, \dots, x_n)$  representam os estados do modelo, os quais dependem da dinâmica do sistema, e juntamente com  $Y$ , evoluem no tempo quando as

equações representando a modelagem do sistema (f) são condicionadas aos parâmetros do modelo e às entradas.

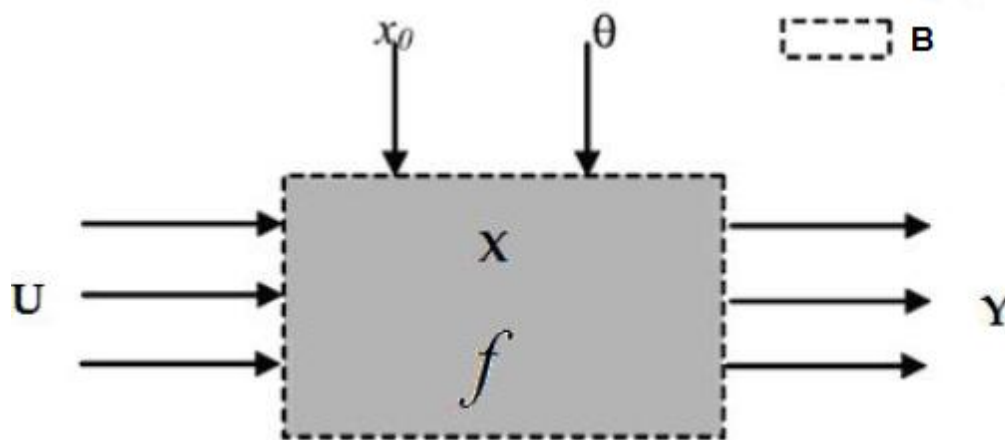


Figura 4 – Diagrama esquemático dos componentes do sistema. Fonte: HUTTON *et al.* (2010), modificado pelo autor.

Para análise da dispersão de poluentes em rios, nas situações em que se deseja estimar os parâmetros, estados ou o próprio valor da grandeza em tempo real, a assimilação de dados sequenciais baseada em métodos Bayesianos é uma alternativa para identificar as incertezas nos parâmetros dos modelos, estrutura, entradas e dados de calibração (MANSOURI *et al.*, 2013). As técnicas de filtragem de dados mais utilizadas para esse fim são conhecidas como filtro de Kalman (*Kalman Filtering - KF*) e amostragem de Monte Carlo sequencial (*Particle Filtering - PF*) (HUTTON *et al.*, 2010; BENEDINI; TSAKIRIS, 2013; JI, 2017; HUTTON *et al.*, 2011).

Nesta conjuntura, Leisenring e Moradkhani (2012) analisaram as incertezas da predição de fontes de poluição sedimentar em um trecho da bacia do lago Tahoe, localizado na Califórnia, Estados Unidos da América (EUA), através de assimilação de dados sequencial com o filtro de partículas amostragem e reamostragem por importância sequencial (*Sequential Importance Resampling - SIR*). Optaram pela utilização desta metodologia devido à possibilidade de avaliar a fonte de poluição em tempo real ao invés de apenas eventos particulares. Ao final do trabalho, concluíram tratar-se de uma técnica de modelagem eficiente, permitindo controlar e entender os impactos do poluente na região estudada. Além disso, também mencionaram que os

métodos computacionais desenvolvidos e aplicados na sua pesquisa poderiam servir como base para estudos similares.

Samuel *et al.* (2014) analisaram a umidade do solo e variações de fluxo na bacia hidrográfica *Spencer Creek*, localizada no sudoeste de Ontario, Canadá, através de assimilação de dados sequencial usando o *Ensemble Kalman Filter* (EnKF) para estimação conjunta de parâmetros e estados. Esta análise foi feita devido à impossibilidade de se obter medidas da umidade do solo diretamente, combinando assim a estimação do parâmetro (fluxo) e do estado do sistema (umidade do solo). Foi concluído que, apesar de estimativas em tempo real da umidade do solo ser difíceis de obter quando o desconhecimento sobre o seu valor é grande, resultados satisfatórios ainda podem ser alcançados.

Dechant e Moradkhani (2012) examinaram a efetividade e robustez de métodos de assimilação sequencial de dados, filtro de partículas (SIR) e o filtro de Kalman (EnKF), para quantificação da incerteza de predição de fontes de poluição sedimentar na bacia hidrográfica do Rio Leaf, localizado no sudeste de Mississipi, EUA. Justificaram essa escolha devido à ausência de trabalhos que comparassem as metodologias e apontassem qual o melhor momento de utilizá-las. Ao final do trabalho concluíram que, se a quantificação da incerteza da predição for importante, o filtro de partículas SIR é uma escolha melhor.

Salamon e Feyen (2009) analisaram incertezas na modelagem hidrológica distribuída usando modelos de erro multiplicativos e assimilação sequencial de dados, através do filtro de partículas SIR combinado com a técnica de suavização de Kernel (*Kernel smoothing*). Justificaram a escolha dessa metodologia devido ao filtro de partículas SIR ter mostrado ser computacionalmente eficiente, e o método *Kernel smoothing* reduzir a dispersão dos parâmetros. Aplicaram a um modelo distribuído de grande escala e alcançaram resultados satisfatórios representando a dinâmica do transporte de poluentes na bacia hidrográfica do rio Rhine, localizado próximo à cidade de Bauk, Alemanha.

Outro problema importante, e pouco referenciado na literatura, é a assimilação de dados para predição de fontes de poluição difusas (lançamentos instantâneos), modeladas como um experimento por meio de uma perturbação em pulso. Normalmente, as pesquisas realizadas apenas avaliam um evento particular para a calibração de parâmetros, mas não exploram sua utilização para monitoramento em tempo real (ANI *et al.*, 2009; ABDERREZZAK *et al.*, 2015; ZHAO *et al.*, 2014; CHEN

*et al.*, 2012). Dessa forma, não se consegue superar a principal desvantagem da modelagem de fontes difusas, a imprevisibilidade de ocorrência e a impossibilidade de se obter medidas do coeficiente de dispersão longitudinal diretamente no rio, além da computação do “tempo morto” entre a fonte poluidora e o ponto de coleta (JI, 2017).

A maneira de superar estas desvantagens e aplicar o monitoramento em tempo real concentra-se na utilização da técnica de assimilação de dados conjuntamente com análise de problemas inversos, combinando a calibração dos parâmetros com a modelagem em tempo real. Neste contexto, testes para justificar a utilização dessa abordagem podem ser realizados na modelagem de experimentos com traçador em pulso comparando-se às técnicas de estimação *off-line* mais utilizadas na literatura, como, por exemplo, o método heurístico *Simulated Annealing*, utilizado com frequência na identificação de fontes poluidoras que afetam a qualidade de águas subterrâneas (*groundwater resources*) (NEWMAN *et al.*, 2005; YEH *et al.*, 2007; JHA; DATTA, 2013; DATTA *et al.*, 2014). Depois de feito isso, e conforme as potencialidades de cada método, pode-se empregar a assimilação de dados como um sinalizador para tomada de decisões, alarme, controle e inferência virtual (HUTTON *et al.*, 2010; SOCOLOFSKY; JIRKA, 2005).

### 3.3 Fluidodinâmica Computacional

A Mecânica dos Fluidos é uma disciplina importante em diversas áreas do conhecimento relacionadas ao movimento de fluidos. No que diz respeito à avaliação da dispersão de poluentes em rios, tornou-se importante nos últimos anos, como consequência da necessidade de estudar cientificamente o fluxo de fluidos que afetam a qualidade ambiental da água. O principal parâmetro do fluido lançado a ser avaliado é a forma como ele se dispersará, a qual é influenciada principalmente pela turbulência, que favorece a sua rápida dispersão e diluição (BLOCKEN; GUALTIERI, 2012).

Apesar de uma técnica muito útil, pois permite avaliar a forma como o movimento dos fluidos influencia na dispersão dos poluentes, a geometria pode influenciar de maneira drástica os resultados por se tratar da análise de um sistema físico real. Em função disso, alguns autores têm utilizado pacotes computacionais de simulação

tridimensional, pelo fato de que permitem simular qualquer geometria, independentemente das suas condições de contorno, além de apresentarem técnicas de solução numéricas eficientes para as equações que descrevem o movimento dos fluidos (ARDEJANI *et al.*, 2011; KALDI *et al.*, 2015; SCHATZMANN e LEITL, 2011).

A maioria das pesquisas na área ambiental, baseadas em fluidodinâmica computacional, está relacionada (1) à predição de parâmetros da qualidade da água, como o coeficiente de dispersão, e (2) ao processo de decisão, visto que os dados preditos podem fornecer informações úteis, uma vez que podem ser simulados diversos Cenários diferentes (BLOCKEN; GUALTIERI, 2012).

Em relação à predição, para estimar um determinado parâmetro, faz-se necessário conhecer a dinâmica do processo real. Isso é possível através da aquisição de dados experimentais. A solução direta por um pacote computacional também deve ser realizada, pois, como suas equações se baseiam em fenômenos (modelagem fenomenológica), existem parâmetros inerentes do modelo a serem determinados pela comparação com os dados experimentais. O modelo ajustado permite fazer previsões para todos os pontos do domínio (BLOCKEN; GUALTIERI, 2012). Os trabalhos de Ardejani *et al.* (2011) e Hamza *et al.* (2015) utilizam tal abordagem.

A decisão é também outro tópico a se destacar, o qual, apesar de ser interpretado como uma consequência da predição, pode ser analisado separadamente. Isso é explicado pelo fato de que se podem simular diversos Cenários hipotéticos importantes, que auxiliam no gerenciamento dos recursos hídricos, sem a necessidade de se realizar experimentos reais, como visto nos trabalhos de Kaldi *et al.* (2015) e Pinheiro *et al.* (2013).

Apesar de todas as vantagens da Mecânica dos fluidos ambiental, a precisão e a confiabilidade dos *softwares* são preocupantes e os estudos de verificação e validação são imperativos. Como resultado, experimentos de alta qualidade para fornecer dados para estudos de validação também são indispensáveis (BLOCKEN; GUALTIERI, 2012).

### 3.4 Comentários finais

A análise da dispersão de poluentes em rios pode ser realizada de diferentes maneiras, destacando-se a assimilação de dados como uma técnica importante para o monitoramento em tempo real. A assimilação de dados sequencial baseada em métodos Bayesianos, apesar de ser um campo de pesquisa promissor em modelagem hidrológica, sua utilização combinada com problemas inversos ainda é pouco explorada. Por este motivo, o próximo capítulo fundamenta as técnicas de assimilação de dados mais conhecidas na literatura, as quais são aplicadas posteriormente em um problema inverso de dispersão de poluentes em rios.

## 4 ESTIMAÇÃO VIA APROXIMAÇÃO BAYESIANA

O foco deste capítulo concentra-se na teoria do método Bayesiano para assimilação de dados para sistemas dinâmicos não lineares, conhecido na literatura como amostragem de Monte Carlo Sequencial (*Particle Filtering*). Utilizam-se como referências os trabalhos de Arulampalam *et al.* (2002), Doucet *et al.* (2001), Gustafsson (2010) e Liu e West (2001), os quais apresentam uma fundamentação teórica detalhada sobre o método.

O objetivo deste capítulo é servir como uma fundamentação teórica para aplicação da técnica mencionada acima em um problema de dispersão de poluentes com fonte difusa e pontual, nos quais os dados referentes ao sistema são analisados e atualizados à mesma frequência em que são recebidos.

### 4.1 Amostragem de Monte Carlo Sequencial (*Particle Filtering - PF*)

O PF é uma técnica de assimilação de dados sequencial baseada em métodos Bayesianos usada para atualizar os estados de um modelo dinâmico  $X = f(U, x_0, \theta, B)$  quando novas observações tornam-se disponíveis ( $Y$ ). Isto significa dizer que a dinâmica dos estados e das observações deve ser mapeada para aplicação em tempo real. As Equações 8 e 9 são utilizadas para esse fim e representam um espaço de estados genérico de um modelo estocástico (GUSTAFSSON, 2010; HUTTON *et al.*, 2010).

$$x_t = f(x_{t-1}, \theta, u_t) + \eta_t \quad \eta_t \sim N \left[ 0, R_t^\eta \right] \quad (8)$$

$$y_t = h(x_t^*) + \varepsilon_t \quad \varepsilon_t \sim N \left[ 0, R_t^\varepsilon \right] \quad (9)$$

Nestas equações,  $x_t$  e  $x_{t-1}$  representam vetores dos estados do sistema nos tempos  $t$  e  $t-1$ ;  $x_t^*$  denota o estado verdadeiro do modelo;  $f$  representa a estrutura do modelo propagando o sistema de  $t-1$  a  $t$  em resposta ao vetor de entrada  $u_t$ , e  $\theta$  é

um vetor de parâmetros do modelo, invariante no tempo. A observação  $y_t$  está relacionada aos estados do modelo através de um operador de mapeamento  $h_t$ ;  $\eta_t$  representa o erro do modelo e  $\varepsilon_t$  representa o erro de medida, ambos seguindo uma distribuição normal com média zero e variância iguais a, respectivamente,  $R_t^\eta$  e  $R_t^\varepsilon$ .

Após definidas as formulações para o mapeamento dos estados e observações, a principal vantagem do PF reside na modelagem de todos os componentes como variáveis aleatórias, de forma que as incertezas nos parâmetros dos modelos, estrutura, entradas e dados de calibração são levados em conta. Como consequência disso, o cômputo do estado atual  $x_t$ , condicionado a observação  $y_t$ , é uma função de distribuição de probabilidade (PDF) a *posteriori*  $\pi(x_t | y_t)$ , e a sua evolução com o tempo é descrita de acordo com o teorema de Bayes (Equação 10), com as seguintes hipóteses: problema de filtragem (estimação em tempo real); e modelo de Markov de primeira ordem, isto é, depende apenas da informação imediatamente anterior para aproximação da PDF a *posteriori* (DOUCET *et al.*, 2001; ARULAMPALAM *et al.*, 2002).

$$\pi(x_t | y_t) = \frac{\pi(y_t | x_t)\pi(x_t | x_{t-1})}{\pi(y_t)} \quad (10)$$

Nesta equação, a distribuição de probabilidade a *priori*  $\pi(x_t | x_{t-1})$  é modelada com a informação disponível sobre o que se deseja avaliar. O grau de confiança (ou importância) das hipóteses é fornecido pela função de verossimilhança  $\pi(y_t | x_t)$  que corrige e incorpora mais informações à distribuição a *priori*. E, a função  $\pi(y_t)$  é a distribuição marginal das medidas, que atua como uma constante de normalização.

Após definido o problema de estimação (modelagem de estados em tempo real), visto que o modelo é estocástico e atualizado no tempo conforme o teorema de Bayes, ainda é necessário encontrar uma alternativa para amostragem da PDF a *posteriori* de maneira sequencial. A ideia chave é representar a PDF a *posteriori*  $\pi(x_t | y_t)$  por um conjunto de amostras randômicas  $\{x_t^i, w_t^i\}$ , onde  $\{x_t^i, i = 1, \dots, N\}$  são as partículas (N) com peso  $\{w_t^i, i = 1, \dots, N\}$ . Quando o número de partículas torna-se grande, uma representação equivalente para a PDF a *posteriori* é definida de acordo

com a Equação 11. Nesta equação,  $\delta$  é a função delta de Dirac; e os pesos são normalizados para que  $\sum_{i=1}^N w_t^i = 1$  (ARULAMPALAM *et al.*, 2002).

$$\pi(\mathbf{x}_t | \mathbf{y}_t) \approx \sum_{i=1}^N w_t^i \delta(\mathbf{x}_t - \mathbf{x}_t^i) \quad (11)$$

Entretanto, em virtude da dificuldade de amostrar-se diretamente da PDF a *posteriori*, as amostras são extraídas de uma densidade de importância conhecida  $q(\mathbf{x}_t^i | \mathbf{y}_t)$  e os pesos são definidos conforme a Equação 12.

$$w_t^i = \frac{\pi(\mathbf{x}_t | \mathbf{y}_t)}{q(\mathbf{x}_t^i | \mathbf{y}_t)} \quad (12)$$

Usando essa relação pode ser mostrado que a atualização recursiva dos pesos em um caso sequencial pode ser reescrita conforme a Equação 13, baseada na teoria do teorema de Bayes (DOUCET *et al.*, 2001).

$$w_t^i = w_{t-1}^i \frac{\pi(\mathbf{y}_t | \mathbf{x}_t) \pi(\mathbf{x}_t | \mathbf{x}_{t-1})}{q(\mathbf{x}_t^i | \mathbf{x}_{t-1}^i, \mathbf{y}_t)} \quad (13)$$

Esta equação ainda pode ser simplificada, e de acordo com Gustafsson (2010), a escolha mais simples para ela é utilizar a PDF a *priori* como aproximação para  $q(\mathbf{x}_t^i | \mathbf{x}_{t-1}^i, \mathbf{y}_t)$ , resultando na Equação 14.

$$w_t^i = w_{t-1}^i \pi(\mathbf{y}_t | \mathbf{x}_t) \quad (14)$$

Assumindo que os erros de medida são variáveis aleatórias que seguem uma distribuição Gaussiana, com média zero e matriz de covariância conhecida  $L$ , a função de verossimilhança pode ser expressa pela Equação 15 (KAPIO; SOMERSALO, 2004). Nesta função,  $\hat{\mathbf{y}}_t = h(\mathbf{x}_t^*)$  é o vetor contendo a solução do problema direto e  $ND$  é o número de dados medidos disponíveis.

$$\pi(y_t | x_t) = (2\pi)^{-ND/2} |L|^{-1/2} \exp\left\{-\frac{1}{2}[y_t - \hat{y}_t]^T L^{-1}[y_t - \hat{y}_t]\right\} \quad (15)$$

Dessa forma, o PF ocorre recursivamente através da aplicação da Equação 8, e atualizando os pesos das partículas usando as Equações 9, 14 e 15. Este procedimento é original para estimativas de estados, sendo proposto inicialmente por Hammersley e Hanscomb (1964), conhecido como filtro amostragem por importância sequencial (*Sequential Importance Sampling - SIS*). O pseudocódigo para o PF é ilustrado de acordo com a Figura 5, a qual esboça de uma maneira sequencial as etapas de atualização empregando as equações mencionadas acima.

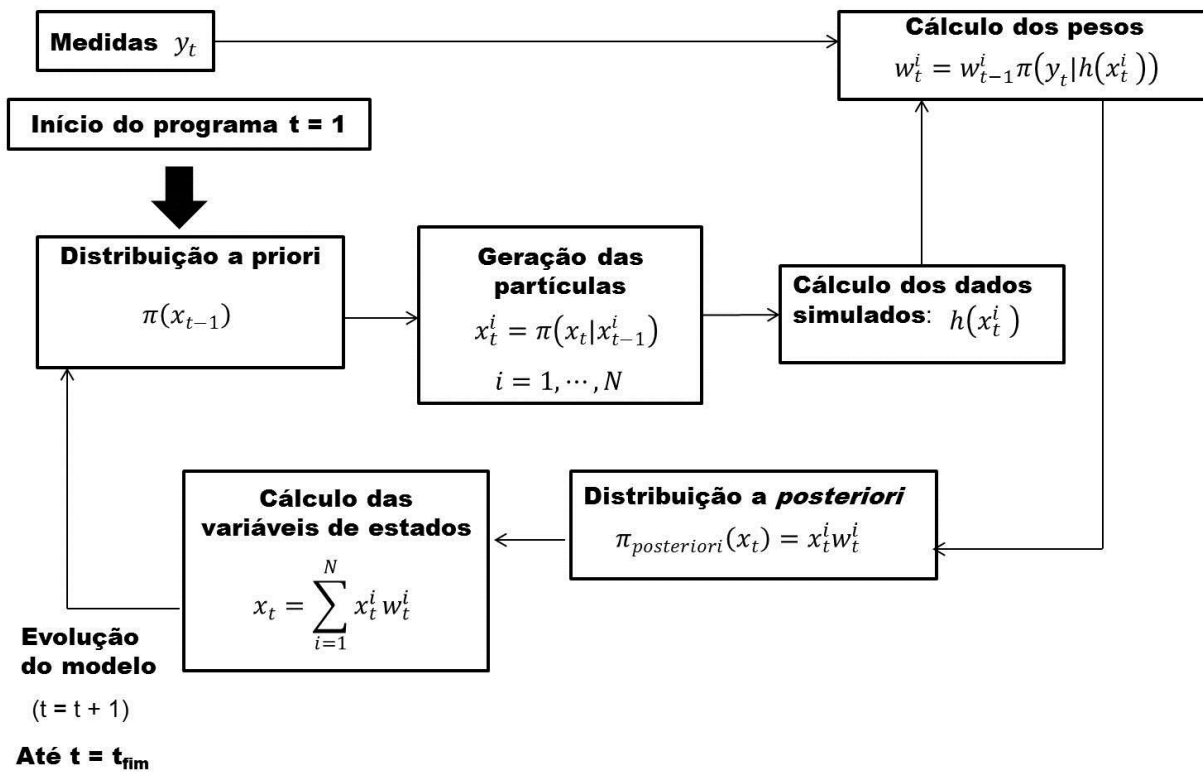


Figura 5 – Pseudocódigo do PF, original para estimativas de estados. Fonte: DA SILVA (2012), modificado pelo autor.

Apesar do PF ser um método para estimativas de estados em tempo real, Gordon *et al.* (1993) constataram um problema na etapa de aproximação da PDF a *posteriori*. Após poucos passos de atualização, mesmo em sistemas de baixa dimensão, apenas poucas partículas recebem pesos significativos, enquanto o restante recebe pesos próximos a zero. Isto significa que a informação estatística para estimação é

baixa e que grande esforço computacional é dedicado para atualizar partículas que contribuem pouco para a estimativa da PDF *a posteriori*.

Este problema é denominado como degeneração das partículas, sendo indicado para reduzir o seu efeito a aplicação de uma etapa adicional no PF original, conhecida como reamostragem (GORDON *et al.*, 1993). O Quadro 3, adaptado de Ripley (1988), sumariza o procedimento através de um algoritmo, cujo objetivo consiste em duplicar partículas tendo alto peso normalizado e descartar partículas com baixo peso, enquanto o número total de partículas permanece inalterado.

Quadro 3 – Método de Ripley (1988) para reamostragem

<p><b>1. Reamostragem</b></p> <p>1) <i>Cálculo da soma dos pesos cumulativos (CSW):</i>  <math>c_i = c_{i-1} + w_t^i</math> para <math>i = 1, \dots, N</math> e <math>c_0 = 0</math></p> <p>2) Tome <math>i=1</math> e gere <math>\xi_1</math> de uma distribuição uniforme <math>U(0, N^{-1})</math></p> <p>3) Para <math>j=1, \dots, N</math>, faça</p> <ul style="list-style-type: none"> <li>• Calcule <math>\xi_j = \xi_1 + N^{-1}(j-1)</math></li> <li>• Enquanto <math>\xi_j \geq c_i</math> faça <math>i = i + 1</math></li> <li>• Designe as partículas <math>x_t^j = x_t^i</math></li> <li>• Designe os pesos para <math>w_t^j = N^{-1}</math></li> </ul>
---

Este procedimento adicional, no pseudocódigo do filtro de partículas SIS, é esquematizado de acordo com a Figura 6 (filtro de partículas SIR), cuja etapa de reamostragem depende de um parâmetro conhecido como tamanho efetivo da amostra  $\hat{N}_{\text{eff}}$ . Este parâmetro foi introduzido por Kong *et al.* (2000), e é uma medida da degeneração das partículas, cujo valor deve ser inferior a  $\hat{N}_{\text{eff}} = 2N/3$  para ativar a reamostragem (GUSTAFSSON, 2010). Adicionalmente, o seu cálculo é executado conforme a Equação 16.

$$\hat{N}_{\text{eff}} = \frac{1}{\sum_{i=1}^N (w_t^{(i)})^2} \quad (16)$$

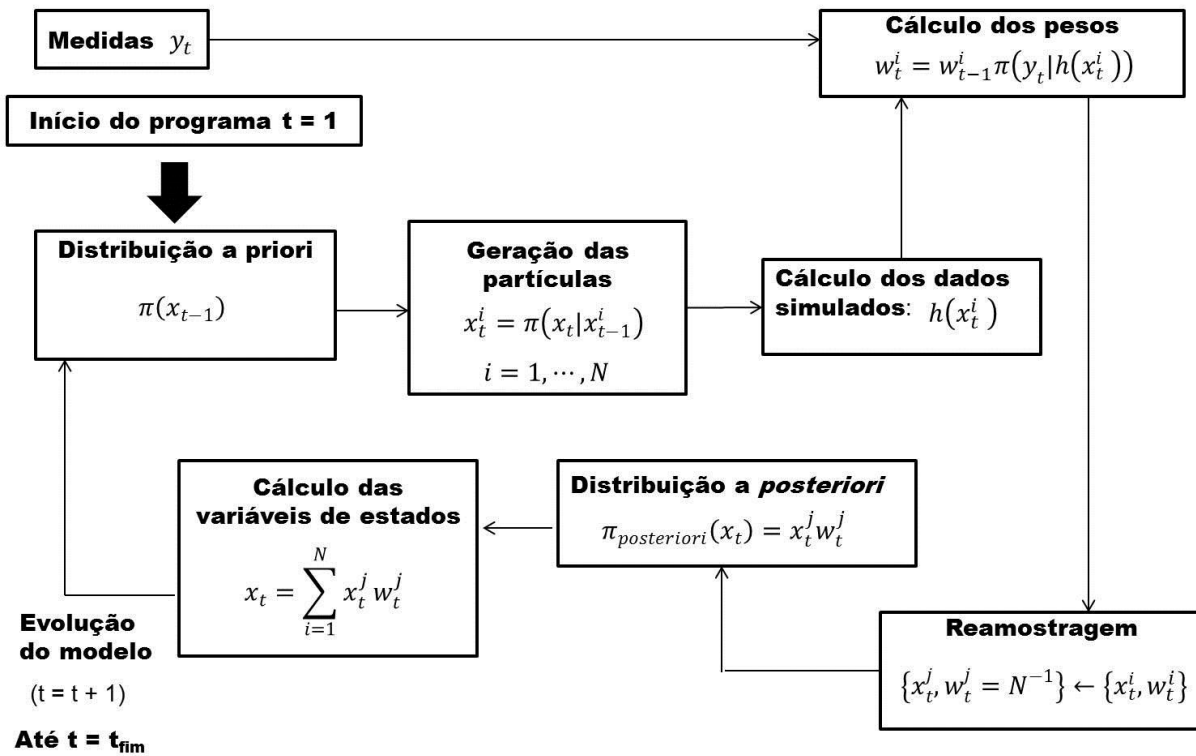


Figura 6 – Pseudocódigo do filtro de Partículas SIR. Fonte: DA SILVA (2012), modificado pelo autor.

Depois de fundamentado o funcionamento do filtro de partículas SIR, uma possível dificuldade com a reamostragem é o empobrecimento da amostra. Isto é, algumas partículas serão idênticas após a etapa de reamostragem e a diversidade entre elas pode ser perdida quando o erro do sistema é pequeno. Esta desvantagem pode ser amenizada ao perturbar as partículas, parâmetros do modelo e variáveis de estado após o passo de atualização. As soluções indicadas para esse propósito são: (1) a utilização de importância auxiliar (filtro de partículas auxiliar ou amostragem e reamostragem por importância auxiliar, conhecido como *Auxiliary Sampling Importance Resampling* - ASIR); e (2) a estimação conjunta de parâmetros e estados, utilizando *Kernel smoothing* (SALAMON; FEYEN, 2009; GORDON *et al.*, 1993).

## 4.2 Estimação conjunta de parâmetros e estados

A estimação de estados em tempo real conforme alguns trabalhos consultados na literatura sobre modelagem hidrológica assumem que um parâmetro ótimo já é conhecido. Isto pode ser uma desvantagem, mesmo em sistemas em que a dinâmica dos parâmetros seja inferior a do estado, por não levar-se em consideração a incerteza paramétrica (HUTTON *et al.*, 2010; SALAMON; FEYEN, 2009).

Segundo Hutton *et al.* (2010), esta metodologia pode ser classificada em aproximações que aplicam assimilação de dados sobre uma série de tempo usada para calibração (parâmetros estáticos), e aproximações que considerem parâmetros variantes no tempo utilizando a técnica *Kernel smoothing*.

O foco, nesta seção, concentra-se na estimação conjunta de parâmetros e estados empregando a amostragem de Monte Carlo sequencial (PF) e *Kernel smoothing*, contribuindo para uma melhor estimativa do filtro de partículas SIR (LIU; WEST, 2001).

A técnica é bem similar à estimação de estados com parâmetros fixos, ocorrendo também recursivamente através da aplicação da Equação 8 para atualização dos estados, e atualizando os pesos das partículas usando as Equações 9, 14 e 15. Entretanto, a amostragem da PDF a *posteriori*  $\pi(x_t, \theta_t | y_t)$  é representada por um conjunto de amostras combinadas  $\{x_t^i, \theta_t^i, w_t^i\}$ , onde  $\{\theta_t^i, i = 1, \dots, N\}$  são as partículas representando a evolução artificial dos parâmetros (LIU; WEST, 2001).

A dificuldade de interpretação dessa técnica está na aproximação da PDF a *posteriori*  $\pi(\theta_t | y_t)$  e amostragem dos parâmetros  $\pi(\theta_t | \theta_{t-1})$ . A Figura 7 ilustra o procedimento adicional no filtro de partículas SIR, que, após cada atualização sucessiva, os parâmetros reamostrados são perturbados adicionando um ruído randômico  $\mu_t^\dagger$ , independente, com média zero e variância igual a  $W$ . Infelizmente, em virtude deste ruído ser adicionado na PDF a *posteriori* dos parâmetros  $\pi(\theta_t | y_t)$  que tem média  $\bar{\theta}_t$  e variância  $V_t$ , acaba resultando em uma “perda de informação” da amostra, que artificialmente têm a variância a *posteriori* aumentada em  $W_t + V_t$ . Esta perda de informação pode ser amenizada usando a função densidade de Kernel relativa à PDF a *posteriori*, que preserva a média do parâmetro original  $\bar{\theta}_t$  e

corrige a variância  $V_t$ . A Equação 17 contabiliza essa perda na etapa de amostragem por meio do fator de desconto  $\omega \in (0,1)$  e Kernel Locations  $m_t^i$  (LIU; WEST, 2001; SALAMON; FEYEN, 2009).

$$m_t^i = \frac{3\omega - 1}{2\omega} \theta_t^i + \left(1 - \frac{3\omega - 1}{2\omega}\right) \cdot \bar{\theta}_t \tag{17}$$

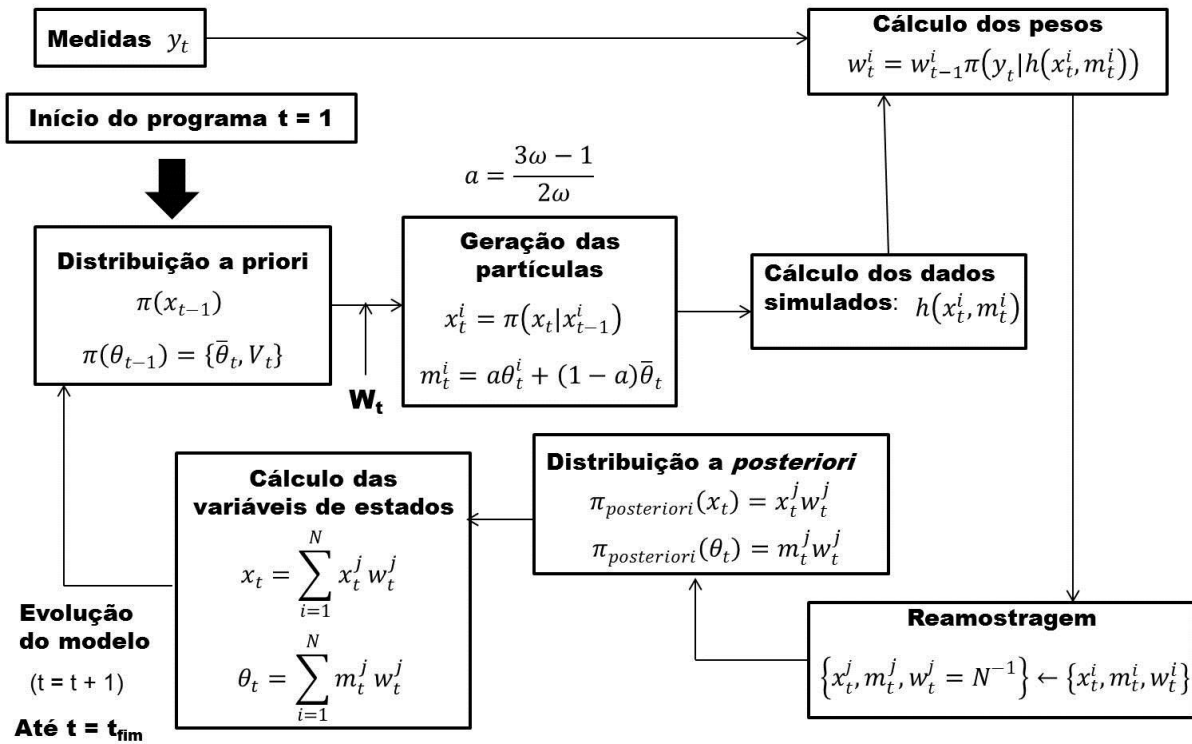


Figura 7 - Pseudocódigo do filtro de Partículas SIR, com Kernel *smoothing*. Fonte: DA SILVA (2012), modificado pelo autor.

### 4.3 Amostragem de Monte Carlo Sequencial utilizando importância auxiliar – ASIR

O filtro ASIR é uma alternativa para contribuir para uma melhor estimativa do filtro de partículas SIR. Foi introduzido por Pitt e Shephard (1999), cujo objetivo consiste em reduzir o empobrecimento da amostra após a etapa de reamostragem. A ideia chave é representar a PDF *a posteriori*  $\pi(x_t, \beta^j | y_t)$  por um conjunto de amostras randômicas auxiliares  $\{u_t^i, w_t^i\}$ , antes da etapa de reamostragem, onde  $\{u_t^i, i = 1, \dots, N\}$  são as partículas no tempo  $t - 1$  ( $\beta^j$ ). Estas partículas auxiliares após

reamostradas serão utilizadas para representar um novo conjunto de amostras randômicas  $\{x_t^i, w_t^i\}$ , onde  $x_t^i = \pi(x_t | x_{t-1}^{ij})$  são as novas partículas e  $w_t^i = \frac{\pi(y_t | x_t^i)}{\pi(y_t | u_t^{\beta^j})}$  os novos pesos (ARULAMPALAM *et al.*, 2002; RISTIC *et al.*, 2004). A Figura 8 ilustra o pseudocódigo do filtro de partículas ASIR, demonstrando-se apenas a etapa de estimação de estados após a reamostragem, uma vez que o procedimento inicial (antes da etapa de reamostragem) é similar ao pseudocódigo do filtro de partículas SIR.

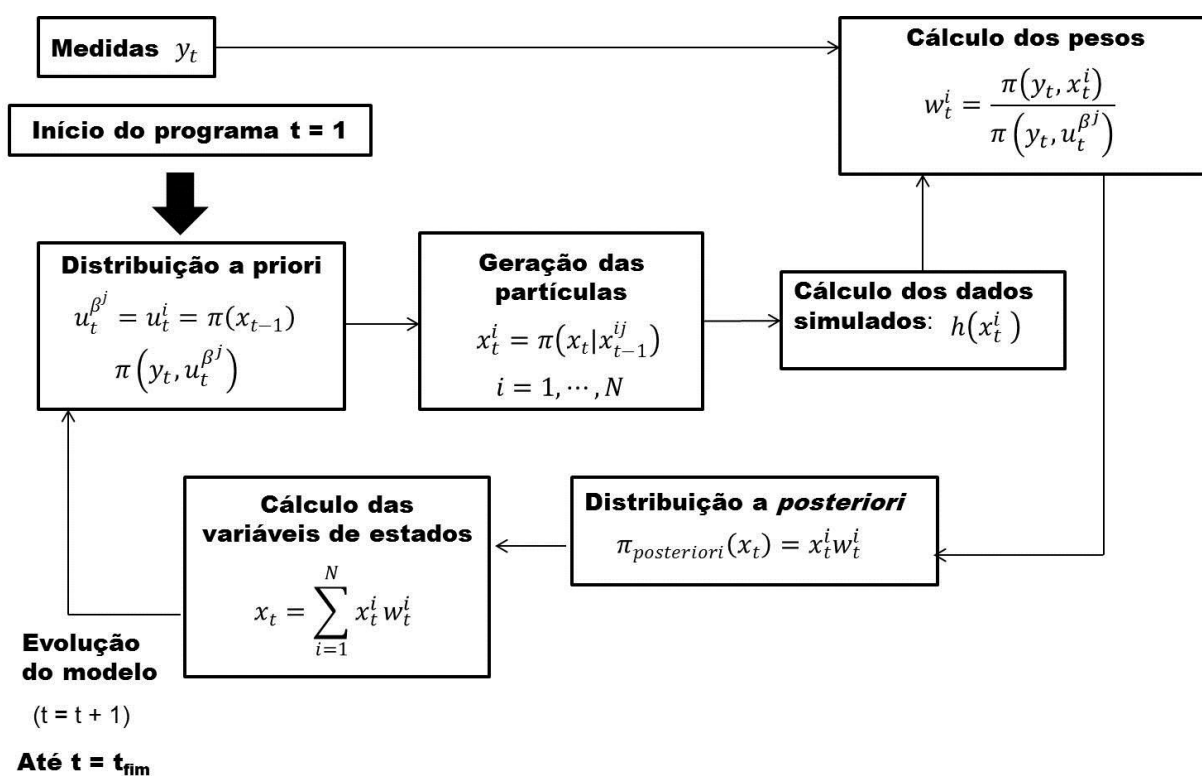


Figura 8 - Pseudocódigo do filtro de Partículas ASIR. Fonte: DA SILVA (2012), modificado pelo autor.

O filtro de partículas ASIR também pode ser utilizado para estimação conjunta de estados e parâmetros com Kernel *smoothing*. O procedimento iterativo emprega os  $m_t^i$  (*Kernel locations*) como variável auxiliar, seguindo até a etapa de reamostragem de maneira idêntica ao algoritmo do filtro de partículas ASIR. Após a reamostragem, uma nova contribuição é dada para o cálculo das partículas conforme a função de densidade de probabilidade normal  $N(\theta | m_k^{(j)}, h^2 V_k)$  (LIU; WEST, 2001).

#### 4.4 Comentários Finais

Chegando ao fim do Capítulo 4, viu-se que o filtro de partículas SIS, SIR e ASIR são destinados exclusivamente à estimação de estados, desta forma não incluem o efeito dos parâmetros na aproximação da *posteriori*. Isto pode ser um problema quando o efeito dos parâmetros desconhecidos tem que ser contabilizado sequencialmente. Por esses motivos, o método Kernel *smoothing* surge como uma alternativa para a descrição de cargas pontuais e difusas em rios, de modo a determinar sequencialmente, além da concentração (estado), os demais parâmetros envolvidos, como dispersão longitudinal, velocidade, distância da fonte de poluição e carga de poluente lançado.

## 5 METODOLOGIA

Este capítulo apresenta a metodologia proposta nesta Dissertação de Mestrado. Considerando-se a principal equação para análise da dispersão de poluentes em rios (EGTPR) e a técnica de amostragem de Monte Carlo Sequencial (*Particle Filtering*), a metodologia também foi baseada em estudos de caso que considerem a imprevisibilidade de ocorrência de fontes de poluição lançadas em um rio, sejam pontuais ou difusas. Neste trabalho, as fontes de poluição difusas a se examinar representam cargas de poluente conservativo lançado de maneira instantânea, enquanto as contínuas caracterizam lançamentos com uma duração de tempo finita ou infinita.

A metodologia será formatada com base nesses dois tipos de lançamentos de poluentes, que compõem os dois estudos de casos investigados. Ainda, são apresentadas as propostas de análises de desempenho para verificar a aplicabilidade de estimação *off-line* e em tempo real para os casos mencionados acima. Ademais, esta metodologia tem como escopo descrever as técnicas aplicadas para alcançar os objetivos desta Dissertação de Mestrado, que se resumem a análise da dispersão de poluentes em rios para diferentes Cenários postos como problemas inversos.

### 5.1 Problema de Lançamento Instantâneo

#### 5.1.1 Descrição do Estudo de Caso – medidas experimentais

A área de estudo localiza-se na região serrana do Rio de Janeiro dentro da bacia do Rio Macaé. O rio estudado é chamado de São Pedro, caracterizado por apresentar o relevo acidentado e grande biodiversidade em remanescentes da Mata Atlântica. Atualmente, vêm sofrendo, entre outros impactos, poluição por esgotos, intensificação de processos erosivos provocados por atividades agropecuárias, perda de solo e assoreamento.

A realidade do rio é preocupante, então diante destas circunstâncias o presente trabalho utilizou como referência um estudo promovido pela Universidade do Estado

do Rio de Janeiro, no instituto politécnico de Nova Friburgo, de modo a contribuir para o entendimento da dinâmica do transporte de poluentes neste rio (SOUZA, 2009). O estudo, primeiramente, fez descrição dos parâmetros físicos e hidrodinâmicos do rio, conseguindo através de um levantamento batimétrico e altimétrico, profundidade média e largura média do rio. Também se obteve a declividade através do levantamento altimétrico realizado. Outro destaque da pesquisa é a determinação experimental da vazão pelo método das seções e método químico. O Método das seções estima a vazão em cursos d'água de maneira indireta, através da medição da velocidade da água em diversos pontos de uma seção transversal do rio. Enquanto, o método químico estima a vazão baseado no princípio de que a vazão do rio não varia dentro do intervalo de tempo de determinação. Todas essas informações são apresentadas na Tabela 1 e confirmam que o trecho do rio estudado geometricamente é uniforme e apresenta largura, profundidade, velocidade e vazão relativamente baixas.

Tabela 1 – Características gerais do trecho do rio estudado

Vazão	Área	Largura	Profundidade	Velocidade	Declividade
0.430 m <sup>3</sup> /s	1.05 m <sup>2</sup>	2.4 m	0.43 m	0.41 m/s	0.03 m/m

As informações completas do trecho do rio estudado permitiram um planejamento do experimento com traçador conservativo injetado em pulso. Foi usado NaCl e liberado um total de 2000 g na linha central de escoamento, sendo registradas as medidas de concentração 50 e 100 m a jusante do ponto de lançamento. Tais medidas foram utilizadas neste presente trabalho, porém considera-se apenas o conhecimento do perfil de concentração inicialmente, sendo necessário determinar a carga de poluente lançado e distância da fonte poluidora para confrontar com os dados de campo.

### 5.1.2 *Formulação Matemática e solução analítica do problema físico*

O problema físico mencionado na seção anterior pode resumidamente ser representado pela Figura 9, em que entre os pontos de injeção e detecção é

necessário conhecer a dinâmica do poluente através de um modelo (EGTPR). Esta equação não é utilizada na sua forma completa (tridimensional) devido ao grande esforço computacional requerido para seu cálculo e ausência de dados para sua validação. Em virtude disso, o modelo unidimensional representado pela Equação 18 é o recomendado na literatura para testes com traçador (LEIBUNDGUT *et al.*, 2009). Nesta equação,  $C = \rho\alpha$  é a concentração,  $U$  é a velocidade média na direção de escoamento do rio e  $E_x$  o coeficiente de dispersão longitudinal.

$$\frac{\partial C}{\partial t} + U \frac{\partial C}{\partial x} = E_x \frac{\partial^2 C}{\partial x^2} \quad (18)$$

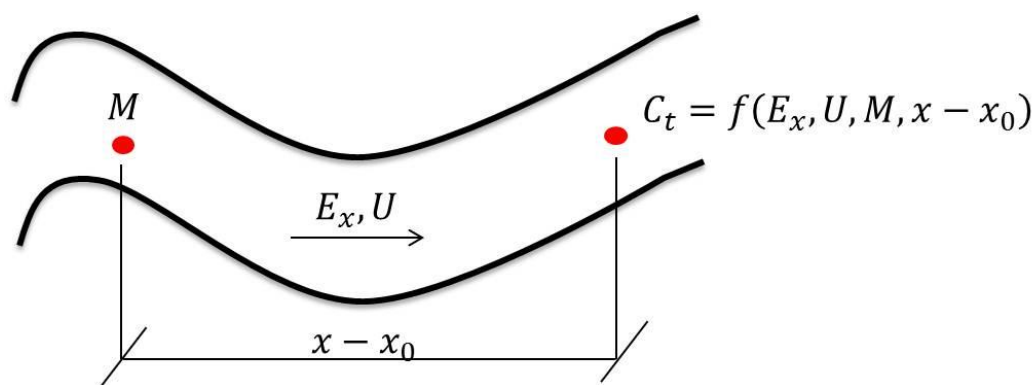


Figura 9 – Representação esquemática do estudo de caso. Fonte: autor.

Para a especificação completa do problema físico, representado pela Equação 18, assume-se: domínio infinito, significando dizer que nos contornos do rio em uma posição suficientemente grande, independentemente do tempo, a concentração é igual a zero; injeção instantânea de massa  $M$  em um ponto do domínio  $x_0$  no instante de tempo  $t = 0$ ; valores de dispersão constantes, independentemente da posição no domínio; campo de velocidade estacionário  $v = [U, 0, 0]$ ; área da seção transversal do rio constante ( $A$ ); e concentração medida inicial de NaCl ( $C_0$ ) igual a 15.5 mg/L, antes da realização do teste com traçador.

Considerando-se as condições acima, segundo Socolofsky e Jirka (2005), há algumas metodologias que podem ser usadas para a obtenção da solução analítica do problema físico, sendo o método da similaridade uma alternativa interessante. Esta metodologia foi proposta por Fischer *et al.* (1967) e tem como principal diferencial a aplicação de análise dimensional para obter a relação entre os

parâmetros e variáveis do problema. Como vantagem em relação aos demais métodos, ao invés de se solucionar um problema de valor de contorno, resolve-se apenas um problema de valor inicial, resultando na Equação 19.

$$C(x,t) = C_0 + \frac{M}{A\sqrt{4\pi E_x t}} \exp\left(-\frac{((x-x_0)-Ut)^2}{4E_x t}\right) \quad (19)$$

Finalizando, a solução analítica (ou exata) é uma alternativa interessante quando o problema a ser avaliado tem uma geometria simples e dimensão baixa, o que permitirá avaliar casos limites e ter uma visão excelente do comportamento do sistema (SOCOLOFSKY; JIRKA, 2005).

### 5.1.3 Problema de Estimação Off-line

O problema de estimação *off-line* conforme os trabalhos da literatura consistem em determinar o coeficiente de dispersão longitudinal  $E_x = f^{-1}(A, x - x_0, U, M, C_0, Y_i)$  para um experimento com perturbação em pulso, conhecido todo o planejamento do ensaio. Isto é insuficiente para aplicações reais devido ao desconhecimento da carga de poluente lançado ( $M$ ) e posição em que se encontra ( $x - x_0$ ). Por esses motivos, propõe-se estimar de maneira conjunta tais parâmetros, incluindo a velocidade média e os valores filtrados de concentração ( $C_i$ ), de modo que  $[E_x, U, x - x_0, M, C_i] = f^{-1}(A, Y_i, C_0)$ , por meio do algoritmo Simulated Annealing (ou Recozimento Simulado) (KIRPATRICK *et al.*, 1983). Para tanto, é utilizada a informação da área transversal ( $A$ ) do rio, concentração inicial ( $C_0$ ) e as observações de concentração  $Y_i$  ao longo do tempo.

Baseando-se em uma analogia com a termodinâmica, o algoritmo emprega uma sequência de temperaturas decrescentes para gerar soluções de um problema de otimização, começando com um valor de temperatura elevado e, a cada nova temperatura (resfriamento), geram-se soluções até que o equilíbrio àquela temperatura seja alcançado. Em seguida, a temperatura é rebaixada e o processo prossegue até o congelamento, ou seja, quando não se obtém mais melhoras. Sua

representação de uma maneira estruturada e sequencial é ilustrada de acordo com o Quadro 4, adaptado de Kirkpatrick *et al.* (1983).

Quadro 4 – Pseudocódigo do algoritmo Simulated Annealing.

```

Dado ( $\alpha$ ,  $SA_{max}$ ,  $T_0$ ,  $T_c$  e  $S$ ) Faça
 $S^* \leftarrow S$ ;                                {melhor solução obtida até então}
iterT  $\leftarrow$  0;                            {número de iterações na temperatura T}
 $T \leftarrow T_0$ ;                             {Temperatura corrente}
Enquanto ( $T > T_c$ ) Faça
    Enquanto (iterT  $>$   $SA_{max}$ ) Faça
        Gerar (um vizinho  $S' \in N(S)$ )
         $\Delta \leftarrow f(S') - f(S)$ ;
        Se ( $\Delta < 0$ )
             $S \leftarrow S'$ 
            Se ( $f(S') < f(S)$ )
                 $S^* \leftarrow S'$ ;
            Fim-se;
        Senão
            Tomar ( $x \in [0,1]$ )
            Se  $x < e^{-\Delta/T}$ 
                 $S \leftarrow S'$ ;
            Fim-se;
        Fim-se;
    Fim-Enquanto;
 $T \leftarrow \alpha * T$ ; iterT  $\leftarrow$  iterT + 1;
Fim-Enquanto;
Retornar ( $S^*$ );

```

Para a implementação do *Simulated Annealing*, definiram-se seus parâmetros atentando-se às seguintes condições: a temperatura inicial ( $T_0 = 1000$ ) deve ser alta o bastante para permitir movimentos livres entre soluções vizinhas; o valor utilizado para  $\alpha = 0,975$  permite simular a condição de resfriamento lento; a temperatura de congelamento foi definida como  $T_c = 0,01$ , embora, em teoria, devesse ser igual à zero; O número de iterações para atingir o equilíbrio térmico ( $SA_{max}$ ) foi escolhido de acordo com a dimensão do problema, sendo 9 uma aproximação razoável para gerar boas soluções vizinhas; e a solução inicial (S) foi definida como o erro quadrático médio (*Root Mean Square Error - RMS*) mostrado na Equação 20. Nesta expressão,  $P = [E_x, U, x - x_0, M]$  é o vetor de parâmetros,  $C_i(P)$  a concentração estimada no tempo  $t_i$  (Equação 19),  $Y_i$  é a concentração medida no tempo  $t_i$ , e  $N$  o número total de medidas. Logo, a partir de uma estimativa inicial para a função objetivo, a minimização do RMS é almejada.

$$RMS(P) = \left( \sum_{i=1}^N [Y_i - C_i(P)]^2 / N \right)^{0,5} \quad (20)$$

Além disso, a cada iteração, uma solução vizinha é gerada aleatoriamente no espaço de busca com novos parâmetros do modelo ( $P_{next}$ ), conforme a Equação 21, até obter-se uma melhora da aproximação dos dados (isto é, redução do RMS). Nesta expressão,  $F$  é uma variável randômica com distribuição normal, tal que  $F \sim N(0,1)$ ; e  $\sigma$  é o desvio padrão dos erros de medidas.

$$P_{next} = P(1 + F\sigma) \quad (21)$$

O desempenho do algoritmo será avaliado, considerado o desconhecimento inicial sobre o problema (informação *a priori* ou parâmetros que compõe a solução inicial), em duas condições distintas:

- 1) a primeira utiliza como aproximações os dados de Souza (2009), entretanto, modificando-se aleatoriamente com uma incerteza de 50% somente os valores iniciais de massa do poluente e da posição em que se encontra a fonte poluidora.

- 2) e a condição seguinte, também baseada no trabalho de Souza (2009), considera que os valores são gerados aleatoriamente pelo algoritmo para todos os componentes de P com 50% de incerteza.

Adicionalmente, a convergência do procedimento meta-heurístico de estimação, é verificada por meio da obtenção de n amostras independentes e identicamente distribuídas de  $M_1, M_2, \dots, M_n$ ,  $(x - x_0)_1, (x - x_0)_2, \dots, (x - x_0)_n$ ,  $(E_x)_1, (E_x)_2, \dots, (E_x)_n$  e  $U_1, U_2, \dots, U_n$ . Depois de obtido a amostra randômica para cada uma das variáveis, realizam-se três análises para inferir sobre o nível de confiança das estimativas.

Identificação da distribuição de probabilidade das estimativas. Isto pode ser feito por meio da função *probplot* do Matlab (R2009b), que possui como parâmetros de entrada, na sua forma mais simples, apenas as n amostras e a distribuição especificada. A saída da função é um gráfico que compara a distribuição das n amostras com a distribuição definida pelo usuário. E, por meio de uma linha de referência que passa através dos quartis inferior e superior das n amostras, é possível julgar se os dados seguem ou não a distribuição de interesse.

Intervalo de Confiança para  $\mu$ , variância desconhecida. Depois de identificada a distribuição de probabilidade de cada uma das amostras, procede-se apenas a definição do intervalo de confiança para amostras que seguem uma distribuição normal. Isto é feito usando a estatística t de Student expressada pela Equação 22. Nesta equação,  $\bar{X}$  é a média amostral, S o erro padrão da média amostrada,  $\mu$  é a média verdadeira,  $\alpha$  é o nível desejado para o intervalo de confiança, t o valor do teste nos quartis  $\alpha$  e  $1 - \alpha$  e  $v = n - 1$  os graus de liberdade.

$$p\left(\bar{X} - t_{\frac{\alpha}{2}, n-1} \cdot \frac{S}{\sqrt{N}} < \mu < \bar{X} + t_{1-\frac{\alpha}{2}, n-1} \cdot \frac{S}{\sqrt{N}}\right) = 1 - \alpha \quad (22)$$

Ajuste da distribuição dos dados e análise Box Plot. Essas análises são realizadas, respectivamente, por meio das funções *histfit* e *boxplot* presentes no Matlab (R2009b), as quais, na sua forma mais simples, dependem apenas da entrada dos dados. A saída de *histfit* é um histograma dos dados ajustados com a distribuição normal. Enquanto, a saída de *boxplot* mostra um gráfico na forma de caixa dos múltiplos dados. Em cada caixa, a marca central é a mediana, as fronteiras das

caixas são os quartis, e as linhas que saem da fronteira da caixa até a sua extensão final são os pontos mais extremos que o algoritmo considera não ser um *outlier* (isto é, uma observação atípica com um grande afastamento em relação aos dados).

#### 5.1.4 Problema de Estimação em Tempo Real

A estimação em tempo real aplicando a técnica de assimilação de dados sequencial baseada no método Bayesiano dos Filtros de Partículas (PF) será usada para atualizar as estimativas dos parâmetros  $[E_x, U, x - x_0, M, C_i] = f^{-1}(A, Y_i, C_0)$ , quando novas observações de concentração estiverem disponíveis. A solução deste problema inclui uma etapa de problema direto (simulação) e uma etapa de problema inverso (estimação) detalhadas abaixo.

Problema Direto. Esta etapa da estimação consiste na simulação das partículas por meio do modelo matemático do sistema físico. Isso é feito para obter uma aproximação da variável observada considerando os valores de  $E_x, U, x - x_0, M$  dados pelas  $i = 1, \dots, N_{part}$  partículas. A simulação numérica foi obtida com a Equação 19 (solução analítica do modelo) considerada a incerteza  $\sigma_{mod}$  para o modelo, de modo que o valor simulado para cada partícula é dado por  $\hat{C}_t^i = C_{t,mod}^i + \eta_t \cdot \sigma_{mod}$ , em que  $C_{t,mod}^i$  é a resposta do modelo e  $\eta_t$  é uma variável randômica tal que  $\eta_t \sim N(0,1)$ .

Problema Inverso. O problema inverso concentra-se na atualização das etapas do PF conforme descrito no Capítulo 4. Inicialmente, nas abordagens de filtro SIR e filtro SIR com Kernel *Smoothing*, é feita a evolução temporal de cada um dos elementos de  $P = [E_x, U, x - x_0, M]$  seguindo um processo Markoviano por meio da amostragem de Monte Carlo  $P_t = P_{t-1} + (v_t)_P \sigma_P$ , em que  $\sigma_P$  é a incerteza associada e  $(v_t)_P \sim N(0, \sigma_P^2)$  é uma variável aleatória.

Estas etapas são então aplicadas aos dados do experimento com perturbação em pulso simulando o lançamento de carga de poluente conservativo em um rio em tempo real. São utilizadas diferentes abordagens do PF para avaliar o desempenho

das estimativas, incluindo análise sobre o efeito da escolha da informação *a priori*  $P_{t-1}$  e como sua evolução é afetada variando-se o número de partículas e incerteza associada. Nesta situação, são testadas configurações dos filtros de partículas SIR variando-se o número de partículas arbitrariamente conforme o conjunto  $N_{\text{part}} = \{100, 500, 1000\}$ , bem como a incerteza associada ao que se deseja estimar com  $\sigma_p = \{25\%, 40\%, 60\%\}$  em relação à estimativa inicial.

Diante dessas informações, de acordo com o Quadro 5, apresentam-se os elementos necessários para a evolução artificial dos parâmetros para os dois Cenários de problema inverso detalhados abaixo.

**Cenário 1.** Desconhecimento sobre a carga de poluente conservativo (M) e a posição em que se encontra ( $x - x_0$ ). Para gerar a condição inicial  $P_0$ , a partir da melhor estimativa de Souza (2009), acrescentou-se uma incerteza de 50% a tais parâmetros. Além disso, foram fixados valores de dispersão longitudinal ( $E_x$ ) e a velocidade (U).

**Cenário 2.** Semelhante ao Cenário anterior, porém determinaram-se conjuntamente os valores de  $E_x$  e U, utilizando-se Kernel *smoothing*.

Quadro 5 – Evolução artificial dos parâmetros para os diferentes Cenários de solução inversa. Fonte: autor.

<b>Cenário 1</b>	
$M_t = M_{t-1} + (v_t)_M \sigma_M$	$(x - x_0)_t = (x - x_0)_{t-1} + (v_t)_{x-x_0} \sigma_{x-x_0}$
$(E_x)_t = (E_x)_{t-1}$	$U_t = U_{t-1}$
<b>Cenário 2</b>	
$M_t = M_{t-1} + (v_t)_M \sigma_M$	$(x - x_0)_t = (x - x_0)_{t-1} + (v_t)_{x-x_0} \sigma_{x-x_0}$
$(E_x)_t^i = \frac{3\omega - 1}{2\omega} (E_x)_t^j + \left(1 - \frac{3\omega - 1}{2\omega}\right) \cdot (\bar{E}_x)_t$	$U_t^i = \frac{3\omega - 1}{2\omega} U_t^j + \left(1 - \frac{3\omega - 1}{2\omega}\right) \cdot \bar{U}_t$

O desempenho de estimação para os Cenários mencionados anteriormente será quantificado por meio de métricas baseadas no tempo computacional entre as iterações, no resíduo da estimação  $\varepsilon_t = \hat{C}_t - Y_t$ , em que  $\hat{C}_t$  é o valor filtrado das observações  $Y_i$  disponíveis, e no intervalo de confiança das variáveis e parâmetros estimados, por meio dos seguintes critérios.

Erro quadrático médio (RMS). Este critério dado pela Equação 23 é uma medida da dispersão dos resíduos, em que  $\hat{y}_t$  e  $y_t$  são respectivamente os valores medidos e os estimados e  $N$  o total de medidas obtidas. Para caracterizar um bom desempenho da ferramenta de estimação, os valores de RMS devem ser próximos a zero.

$$\text{RMS} = \left( \frac{1}{N} \sum_{t=1}^N [\hat{y}_t - y_t]^2 \right)^{\frac{1}{2}} \quad (23)$$

Coefficiente de determinação ( $R^2$ ). O bem conhecido  $R^2$  expressa o quão associado dois conjuntos de dados são entre si. Este coeficiente foi usado para analisar o grau de aproximação de valores medidos pelas soluções dos filtros SIR por meio da função *corr* do matlab.

Coefficiente de autocorrelação do resíduo (SC). A autocorrelação pode ser definida como a correlação entre integrantes de séries de observações ordenadas no tempo (Box *et al.*, 1994). A partir deste conceito aplicado à série dos resíduos da estimação, este trabalho emprega a seguinte expressão  $SC = \max_{\tau_{lag} > 0} |\text{autocorr}(\varepsilon(\tau) - \varepsilon(\tau - \tau_{lag}))|$  para caracterizar o desempenho do filtro (Faria *et al.*, 2018). Este critério é o máximo valor absoluto de autocorrelação entre a série do resíduo  $\varepsilon(\tau)$  e a série atrasada no tempo  $\varepsilon(\tau - \tau_{lag})$ . Neste caso, para classificar um resíduo como ruído branco, o que é desejável, os valores de SC devem ser próximos de zero ou serem menores que os limites (*bounds*), que definem o intervalo de significância estatística ao nível de 95% de confiança. Para a obtenção dos valores de autocorrelação e dos limites, foi utilizada a função *autocorr* do Matlab.

Intervalo de Confiança, função prctile. Esta função presente no Matlab (R2009b) retorna percentis do vetor de dados dado como entrada. Esta aproximação é aplicada na etapa de estimação da PDF a *posteriori*, de forma que as partículas utilizadas como vetor de entrada retornam um intervalo de 95% de confiança. Isto supre necessidade de utilização da relação  $\hat{y}_t \pm 1,96\sigma_t^y$ , em que  $\hat{y}_t$  é a estimativa e  $\sigma_t^y$  é a variância das partículas ambas calculadas no tempo t e de acordo com o número de partículas.

Coeficiente de reamostragem (REA). A ideia básica deste coeficiente é, por meio da Equação 24, fornecer o tamanho efetivo da amostra para todas as observações sequenciais. Nesta equação, REA é um valor percentual e  $N_{part}$  é o número inicial de partículas antes da etapa de reamostragem.

$$REA(\%) = \frac{\hat{N}_{eff}}{N_{part}} \cdot 100 \quad (24)$$

## 5.2 Problema de Lançamento Contínuo

### 5.2.1 Descrição do Estudo de Caso – medidas sintéticas

Este estudo de caso também se baseia no trabalho de Souza (2009) quanto às informações relatadas na seção 5.1.1, exceto pela forma diferente de lançamento do poluente. Considerou-se que foi injetado continuamente e invariavelmente a uma taxa de  $\dot{M} = 2000 \text{ g/s}$  durante um intervalo de tempo  $\tau = 220 \text{ s}$ . A ideia desta nova forma de análise é criar um Cenário teórico em que se objetiva identificar em tempo real a carga de poluente conservativo lançado, a posição em que se encontra e o intervalo de duração, conhecidas as observações nas posições de detecção.

### 5.2.2 Sistema Físico e Formulação Matemática

O sistema físico é formulado matematicamente em conformidade com as características do rio estudado por Souza (2009) e modeladas com o mesmo enfoque dado na seção 5.1.2 quanto à equação diferencial parcial utilizada. Contudo, difere quanto à especificação da condição inicial  $C(x, 0) = C_0$  e condições de contorno, que se caracterizam como  $C(x_0, t) = \dot{M}/(A \cdot U)$  durante o intervalo de tempo  $0 \leq t < \tau$ ; e como  $C(x_0, t) = 0$  após o instante de tempo  $\tau$ . Ademais, ainda é válida a definição de domínio infinito na direção positiva do contorno  $x$  ( $C(\infty, t) = 0$ ).

### 5.2.3 Medidas sintéticas

As medidas sintéticas de carga de poluente conservativo foram simuladas em uma posição 100 m a jusante do ponto de injeção. De acordo com Socolofsky e Jirka (2005), a solução para uma fonte pontual contínua é obtida integrando-se no tempo a solução para uma fonte instantânea. A duração da injeção é  $\tau$ , e a forma geral da solução é mostrada pela Equação 25.

$$C(x, t) = \int_0^t \frac{\dot{M}}{A \sqrt{4\pi E_x (t - \tau)}} \cdot \exp\left(-\frac{((x - x_0) - U(t - \tau))^2}{4E_x (t - \tau)}\right) d\tau \quad (25)$$

Esta equação solucionada analiticamente, de acordo com Benedini e Tsakiris (2013) e Socolofsky e Jirka (2005), durante o intervalo de tempo  $0 \leq t < \tau$ , resulta na Equação 26. Após o intervalo de tempo concluído, subtraindo-se da Equação 26 o que deveria ser o efeito iniciando-se em  $t = \tau$ , obtém-se a Equação 27. Nestas equações,  $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$  é a função erro complementar.

$$C(x, t) = \frac{\dot{M}}{2AU} \left\{ \operatorname{erfc}\left(\frac{(x - x_0) - Ut}{2\sqrt{E_x t}}\right) + \exp\left(\frac{U(x - x_0)}{2E_x}\right) \cdot \operatorname{erfc}\left(\frac{(x - x_0) + Ut}{2\sqrt{E_x t}}\right) \right\} \quad (26)$$

$$C(x,t) = \frac{\dot{M}}{2AU} \left\{ \operatorname{erfc} \left[ \frac{(x-x_0)-Ut}{2\sqrt{E_x t}} \right] - \operatorname{erfc} \left[ \frac{(x-x_0)-U(t-\tau)}{2\sqrt{E_x(t-\tau)}} \right] \right. \\ \left. + \exp \left( \frac{U(x-x_0)}{E_x} \right) \left[ \operatorname{erfc} \left[ \frac{(x-x_0)+Ut}{2\sqrt{E_x t}} \right] - \operatorname{erfc} \left[ \frac{(x-x_0)+U(t-\tau)}{2\sqrt{E_x(t-\tau)}} \right] \right] \right\} \quad (27)$$

Depois de obtidas as soluções analíticas para descrição do comportamento do sistema, e conhecendo-se as características hidrológicas do rio e forma de injeção do poluente, pode se obter os resultados ilustrados na Figura 10. Nesta figura, 20 amostras de concentração simuladas (solução analítica) foram obtidas 100 s após o início do lançamento considerando um intervalo de amostragem igual a 20 s; feito isso, acrescentaram aos valores exatos simulados com a solução analítica uma componente de erros randômicos (Equação 28) para gerar as medidas ruidosas (medidas sintéticas) de concentração do poluente;

$$C_t = C_{\text{exata},t} + w_t \cdot \sigma_{\text{med}} \quad (28)$$

Nesta proposta,  $C_t$  são as medidas sintéticas contendo erros randômicos no instante de tempo  $t$ ;  $C_{\text{exata},t}$  são as medidas simuladas exatas;  $\sigma_{\text{med}} = 0,05 C_{\text{max}}$  é o desvio padrão dos erros de medida em relação a concentração máxima possível dada por  $C_{\text{max}} = \dot{M}/(A \cdot U)$ ; e  $w_t$  é uma variável randômica tal que  $w_t \sim N(0,1)$ .

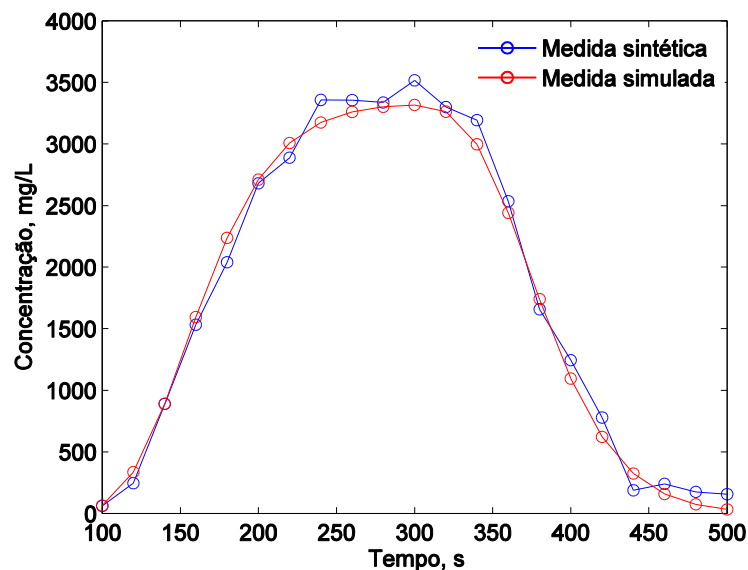


Figura 10 – Concentração de poluente na posição de detecção.

#### 5.2.4 Problema de Estimação em tempo real

Objetiva-se aplicar a técnica de assimilação de dados sequencial baseada em métodos Bayesianos, via filtro de partículas (PF), para atualizar  $[E_x, U, x - x_0, \dot{M}, C_i, \tau] = f^{-1}(A, Y_i)$  quando novas observações de concentração tornarem-se disponíveis. O enfoque é semelhante ao dado na seção 5.1.4 quanto à solução do problema inverso, análise sobre o efeito da escolha da informação a priori  $P_{t-1}$  e como sua evolução é afetada variando-se o número de partículas e incerteza associada. Contudo, difere quanto à solução do problema direto, o qual foi resolvido durante o intervalo de tempo  $0 \leq t < \tau$  integrando-se a Equação 25 por meio do Método de Simpson 1/3 composto, resultando na Equação 29. Nesta equação,  $h = (t - 0)/N$  é a largura de cada subintervalo ( $N$ ) igualmente espaçado, o peso é 4 em todos os pontos  $\tau(i)$  com índice par, e 2 em todos os pontos  $\tau(i)$  com índices ímpares (SOCOLOFSKY E JIRKA, 2005; GILAT E SUBRAMANIAN, 2008).

$$C(x, t) \approx \frac{h}{3} \left( \begin{aligned} & f(E_x, U, x - x_0, \dot{M}, t, \tau(0)) + 4 \sum_{i=2,4,6}^N f(E_x, U, x - x_0, \dot{M}, t, \tau(i)) \\ & + 2 \sum_{i=3,5,7}^{N-1} f(E_x, U, x - x_0, \dot{M}, t, \tau(i)) + f(E_x, U, x - x_0, \dot{M}, t, \tau(t)) \end{aligned} \right) \quad (29)$$

A solução numérica para o intervalo de tempo  $t > \tau$  foi obtida subtraindo-se da Equação 29 o que deveria ser o efeito iniciando-se em  $t = \tau$ , resultando na Equação 30.

$$C(x, t) \approx \frac{h}{3} \left( \begin{aligned} & f(E_x, U, x - x_0, \dot{M}, t, \tau(0)) + 4 \sum_{i=2,4,6}^N f(E_x, U, x - x_0, \dot{M}, t, \tau(i)) \\ & + 2 \sum_{i=3,5,7}^{N-1} f(E_x, U, x - x_0, \dot{M}, t, \tau(i)) + f(E_x, U, x - x_0, \dot{M}, t, \tau(t)) \end{aligned} \right) \quad (30)$$

$$- \frac{h}{3} \left( \begin{aligned} & f(E_x, U, x - x_0, \dot{M}, t - \tau, \tau(0)) + 4 \sum_{i=2,4,6}^N f(E_x, U, x - x_0, \dot{M}, t - \tau, \tau(i)) \\ & + 2 \sum_{i=3,5,7}^{N-1} f(E_x, U, x - x_0, \dot{M}, t - \tau, \tau(i)) + f(E_x, U, x - x_0, \dot{M}, t - \tau, \tau(t)) \end{aligned} \right)$$

Ademais, o valor simulado para cada partícula foi dado por  $\hat{C}_t^i = C_{t,mod}^i + \eta_t \cdot \sigma_{mod}$ , em que  $C_{t,mod}^i$  é a resposta do modelo,  $\sigma_{mod}$  é a incerteza para o modelo e  $\eta_t$  é uma variável randômica tal que  $\eta_t \sim N(0,1)$ .

Completando, esta proposta de utilizar a solução numérica ao invés da solução analítica tem como objetivo evitar o chamado crime inverso (KAPIO; SOMERSALO, 2004). Isto é, empregar a mesma solução do problema direto para obtenção das medidas sintéticas e estimativas em tempo real, de modo que os resultados seriam inconclusivos e não comprovariam a efetividade do PF – SIR em reduzir estatisticamente o ruído da medida.

Dito isso, foram testadas configurações dos filtros de partículas SIR, igualmente ao feito na seção 5.1.4, variando-se o número de partículas arbitrariamente conforme o conjunto  $N_{part} = \{100,500,1000\}$ , bem como a incerteza associada ao que se deseja estimar com  $\sigma_p = \{25\%,40\%,60\%\}$  em relação à estimativa inicial. Ademais, os elementos necessários para a simulação dos dois Cenários de solução inversa foram também similares aos propostos na seção 5.1.4, salvo o acréscimo de incerteza em  $\tau$  e incerteza do modelo igual a  $\sigma_{mod} = \frac{\dot{M}}{AU} \cdot 0,005 \text{ mg/L} \cdot s$  e  $\sigma_{med} = \frac{\dot{M}}{AU} \cdot 0,05 \text{ mg/L} \cdot s$ . E, por fim, todas essas mudanças foram analisadas utilizando os mesmos critérios de desempenho da seção 5.1.4.

### 5.3 Comentários Finais

Este capítulo de metodologia trouxe dois estudos de casos diferentes quanto à forma de lançamento da carga de poluente conservativo em um rio. Os diferenciais da proposta desta Dissertação, em relação aos trabalhos encontrados na literatura consultada, concentram-se na estruturação de uma ferramenta de estimação para análise do problema de dispersão de poluente considerando tanto uma abordagem *off-line* como em tempo real (*on-line*), cujos desempenhos foram verificados por meio de um bom conjunto de métricas e critérios estatísticos.



## 6 RESULTADOS E DISCUSSÃO

### 6.1 Problema de Lançamento Instantâneo

Esta seção é proposta para analisar o desempenho das estimativas em tempo real e *off-line* para o experimento com perturbação em pulso simulando a carga de poluente conservativo lançado instantaneamente em um rio, em diferentes Cenários de busca da solução do problema inverso. Isto foi feito respeitando-se as diferenças entre as duas abordagens e para os Cenários mencionados nas seções 5.1.3 e 5.1.4.

Além disso, os códigos utilizados nas simulações deste trabalho foram implementados no *software* Matlab (R2009b) em um computador com processador Intel Pentium P6000, 1.87GHz e 3 Gb de RAM.

#### 6.1.1 Análise com Simulated Annealing (Estimação Off-line)

A Figura 11 ilustra o desempenho das estimativas obtidas para a primeira condição proposta em 5.1.3, considerando 120 repetições do algoritmo *Simulated Annealing* e um desvio padrão de 0,5 para  $M$  e  $x - x_0$ . Esses resultados caracterizam a distribuição de probabilidade das amostras de  $M$  e  $x - x_0$  em relação à distribuição normal.

Essas análises permitiram concluir que as amostras de  $M$  e  $x - x_0$  detalhados nas Figuras 12 e 13, respectivamente, podem ser avaliados por meio das propriedades da distribuição normal, as quais são sumarizadas de acordo com a Tabela 2, constando medidas de tendência central, medidas de dispersão, medidas de posição e intervalo de confiança.

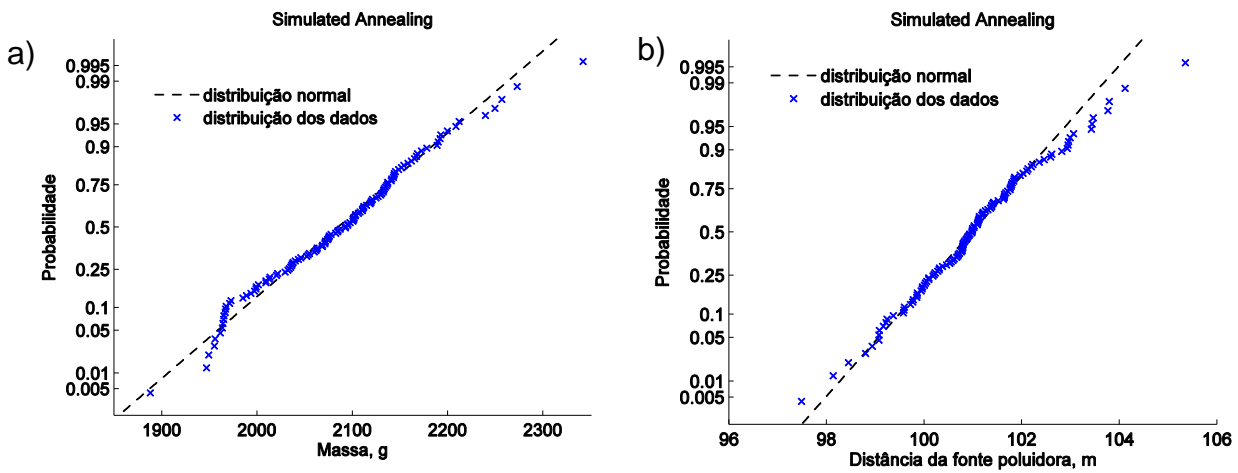


Figura 11 – Caracterização da distribuição de probabilidade dos valores de massa (a) e distância da fonte poluidora (b), considerando 120 repetições do algoritmo *Simulated Annealing*. Fonte: autor.

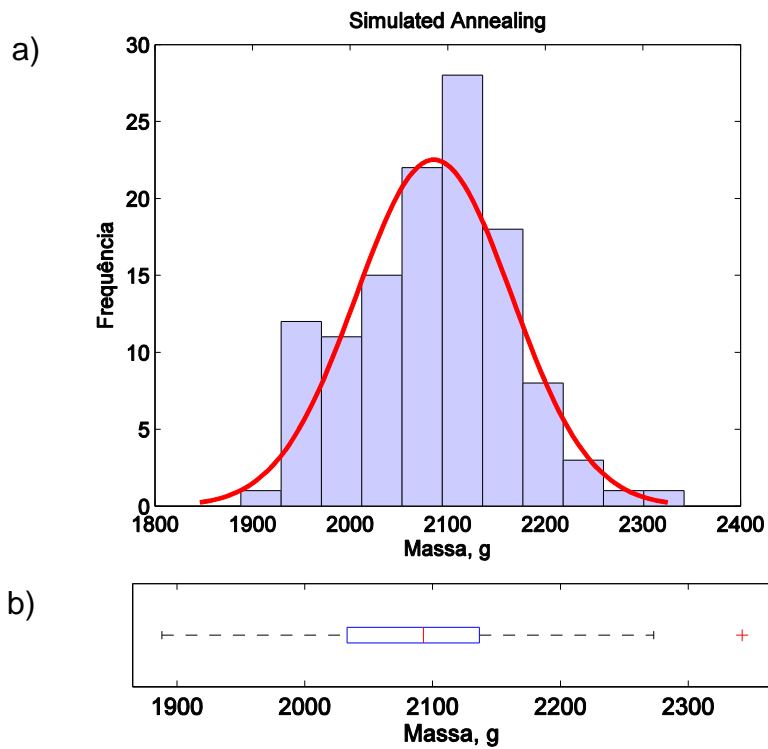


Figura 12 – Análise detalhada das amostras de massa ajustada pelo algoritmo *Simulated Annealing*: (a) Histograma com ajuste pela curva normal; (b) análise *Box Plot*. Fonte: autor.

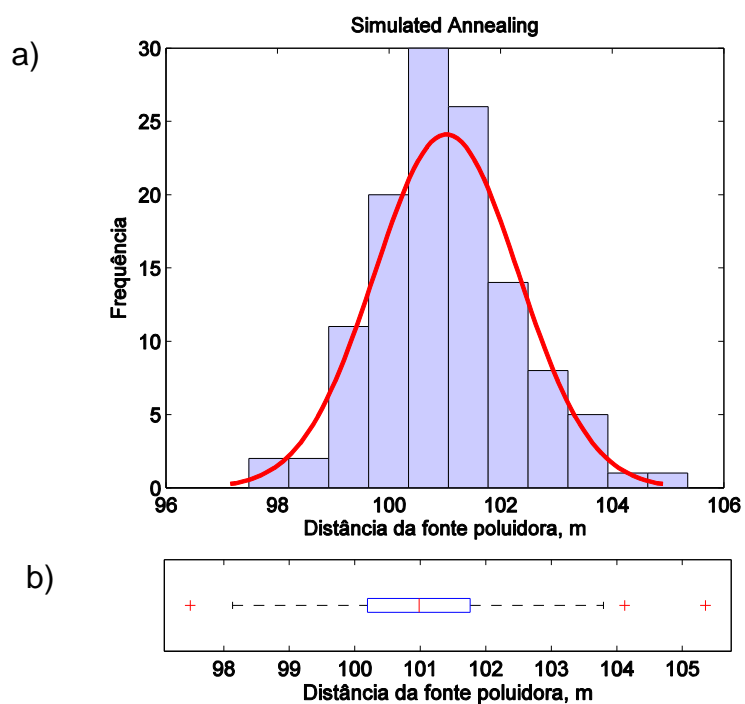


Figura 13 – Análise detalhada das amostras de distância da fonte poluidora ajustada pelo algoritmo *Simulated Annealing*: (a) Histograma com ajuste pela curva normal; (b) análise *Box Plot*. Fonte: autor.

Tabela 2 – Propriedades estatísticas da distribuição normal dos dados de M e  $x - x_0$ . Fonte: autor.

Propriedades	Massa (g)	Distância da fonte poluidora (m)
Média	2085,7	101,03
Variância	6361,1	1,6643
Desvio Padrão	79,7564	1,2901
Mínimo	1888,0	97,48
1º quartil	2033,0	100,20
Mediana	2092,8	100,98
3º quartil	2136,5	101,76
Máximo	2342,2	105,35
Intervalos de Confiança		
Média	2071,3 – 2100,1	100,8 – 101,26
Mediana	2078,3 – 2107,2	100,74 – 101,21

Diante destas informações, as médias amostrais de  $M$  e  $x - x_0$  aproximaram-se dos valores experimentais ( $M = 2000,0$  g,  $x - x_0 = 100,00$  m) e corrigiram os iniciais ( $M = 1178,5$  g,  $x - x_0 = 194,26$  m). Isto, dentro do limite de confiança, estabelecido como 95% e considerando o teste de hipótese  $t$ , válido neste caso, devido as amostras seguirem uma distribuição normal, mas apresentarem variância da população desconhecida. Ademais, também plotou-se os valores de RMS para cada repetição do algoritmo, e conforme a Figura 14 verificou-se uma melhora no ajuste inicial do  $RMS = 18,68$  em todas as ocasiões.

Todas essas informações mostraram o caráter heurístico do método *Simulated Annealing*, que apesar de relativa facilidade de implementação e produzir boas soluções rapidamente, não garante afirmar otimalidade da solução obtida quando apenas uma repetição é realizada (KIRPATRICK *et al.*, 1983). Além disso, mostrou-se uma alternativa para os métodos ditos exatos, como o método de Levenberg-Marquardt utilizado por Souza (2009) para o ajuste da dispersão longitudinal  $E_x$ .

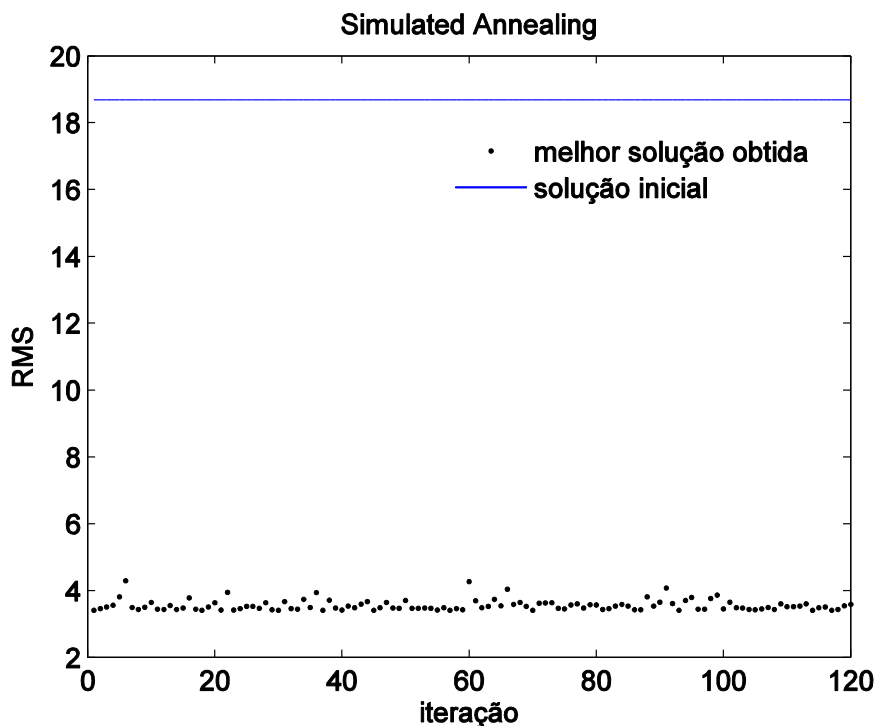


Figura 14 – Valores de RMS obtidos considerando 120 repetições do algoritmo *Simulated Annealing* em relação à condição inicial. Fonte: autor.

Finalizando a análise da primeira condição, a Figura 15 ilustra o ajuste obtido da curva de concentração, dadas as médias de  $M$  e  $x - x_0$ , em relação aos dados

experimentais e condição inicial de busca (RMS = 18,68). Também é visível a melhora no ajuste dos dados, levando a concluir que o método *Simulated Annealing* pode ser utilizado para estimação *off-line* de carga de poluente lançado instantaneamente em um rio.

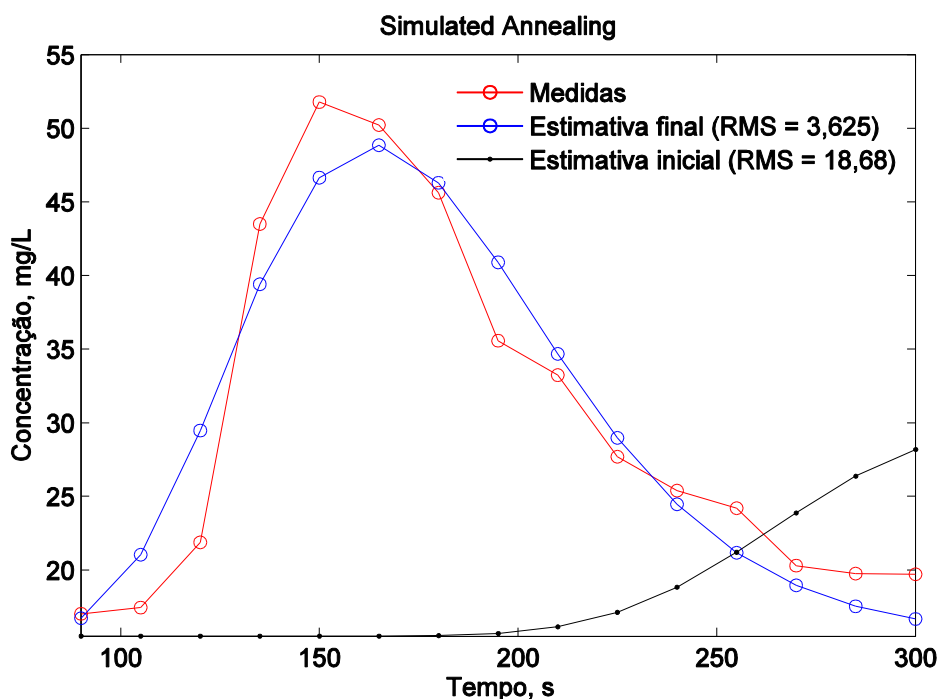


Figura 15 – Ajuste da concentração de poluente, estimando a massa do poluente e a posição de lançamento pelo algoritmo *Simulated Annealing*. Fonte: autor.

Outro Cenário de solução proposto na seção 5.1.3 inclui a estimação conjunta de  $E_x$  e  $U$  considerando o mesmo número de repetições e incertezas da condição anterior. Esta nova proposta tem como objetivo verificar se o acréscimo de incerteza nos parâmetros  $E_x$  e  $U$  levam a uma melhor estimativa de  $M$  e  $x - x_0$ .

Da mesma forma que no caso anterior, por meio da Figura 16, pode se verificar se as amostras de  $M$ ,  $x - x_0$ ,  $E_x$  e  $U$  seguiram um padrão de distribuição normal.

Todas as distribuições não se desviaram significativamente do perfil de distribuição normal. Diante disso, uma análise estatística detalhada para as amostras de  $M$ ,  $x - x_0$ ,  $E_x$  e  $U$  pode ser vista nas Figuras 17 a 20, respectivamente. Ademais, uma análise adicional é possível por meio da Tabela 3.

Como consequência desses resultados, uma ampliação da região de busca leva a uma aproximação mais confiável da média ajustada de  $M$  e  $x - x_0$ . De acordo com

Kirpatrick *et al.* (1983), isto é explicado pela forma diferente de estruturação da vizinhança (mais completa), garantindo dessa forma que qualquer solução seja alcançável a partir da outra.

Todas essas análises justificam a utilização dessa nova abordagem para estimação *off-line* da carga de poluente conservativo lançado instantaneamente em um rio e são confirmadas por meio das Figuras 21 e 22, que ilustram o ajuste obtido da curva de concentração de poluente e redução dos valores de RMS em todas as condições, respectivamente. Apesar de todas essas vantagens, é um método que depende do conhecimento completo da curva de passagem do traçador (300 s), e que em média demandou um tempo igual a 670 s para completar as 120 repetições do algoritmo.

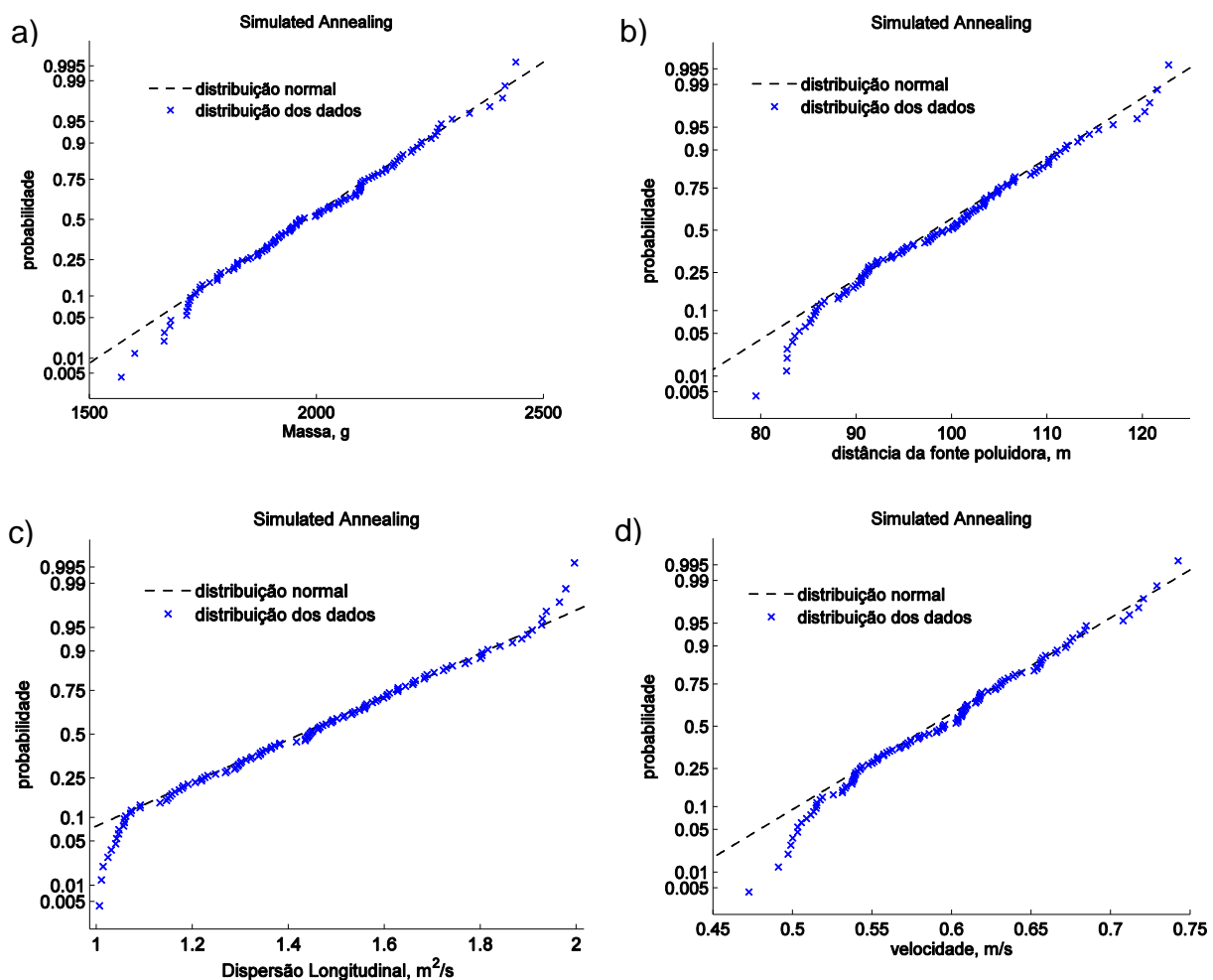


Figura 16 – Caracterização da distribuição de probabilidade dos valores de massa (a), distância da fonte poluidora (b), dispersão longitudinal (c) e velocidade (d), considerando 120 repetições do algoritmo *Simulated Annealing*. Fonte: autor.

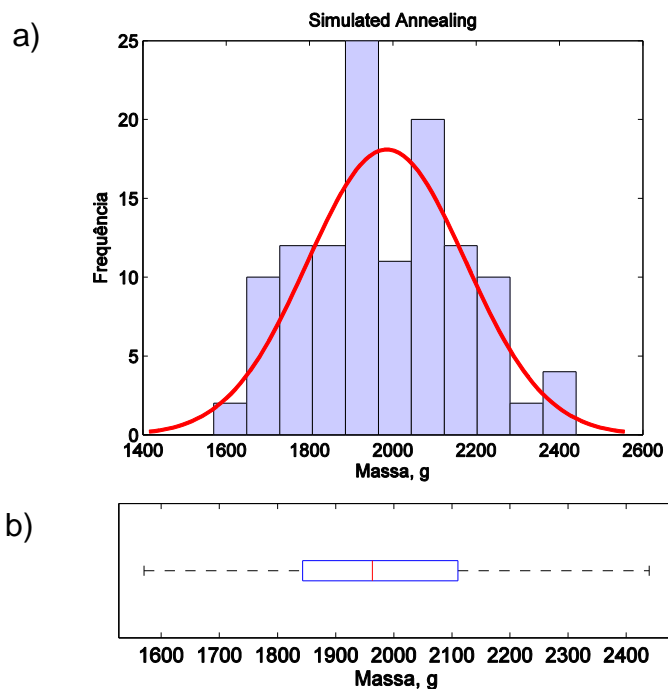


Figura 17 – Análise detalhada das amostras de massa ajustada pelo algoritmo *Simulated Annealing*: (a) Histograma com ajuste pela curva normal; (b) análise *Box Plot*. Fonte: autor.

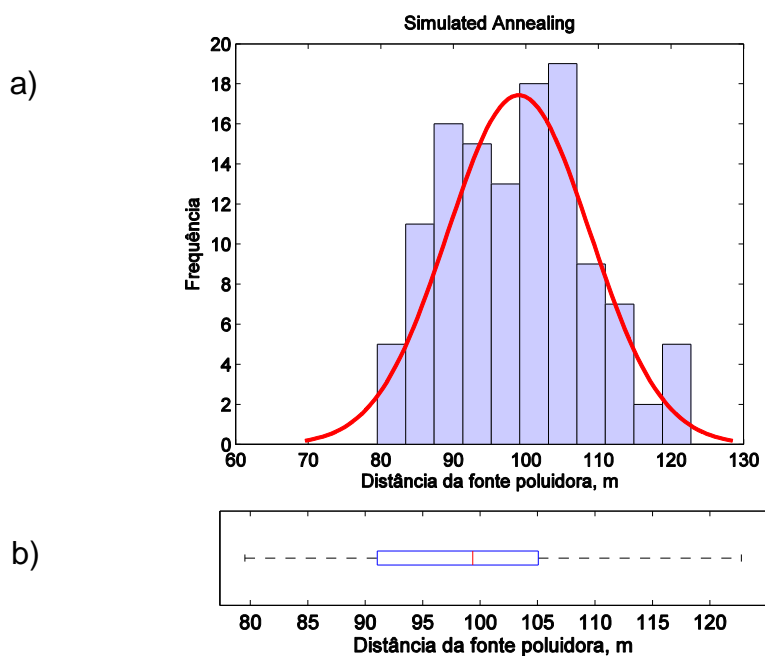


Figura 18 – Análise detalhada das amostras de distância da fonte poluidora ajustada pelo algoritmo *Simulated Annealing*: (a) Histograma com ajuste pela curva normal; (b) análise *Box Plot*. Fonte: autor.

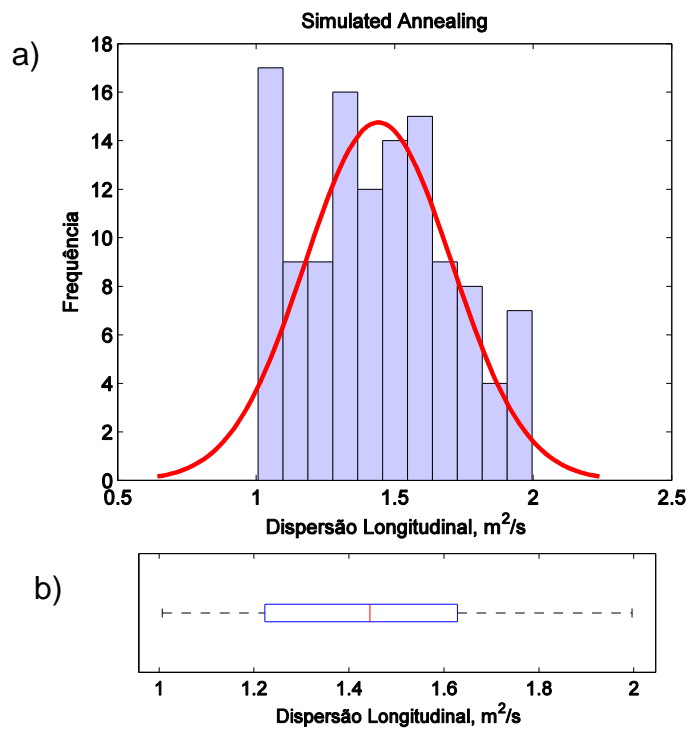


Figura 19 – Análise detalhada das amostras de dispersão longitudinal ajustada pelo algoritmo *Simulated Annealing*: (a) Histograma com ajuste pela curva normal; (b) análise *Box Plot*. Fonte: autor.

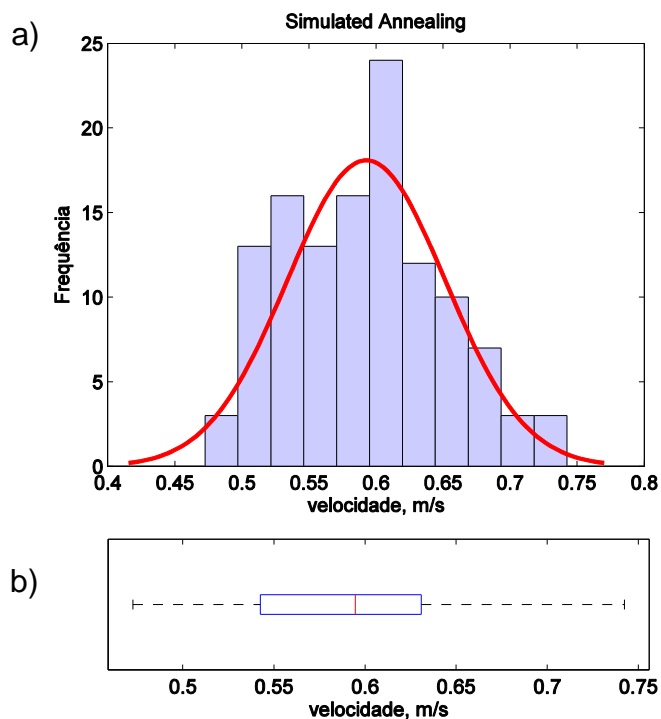


Figura 20 – Análise detalhada das amostras de velocidade ajustada pelo algoritmo *Simulated Annealing*: (a) Histograma com ajuste pela curva normal; (b) análise *Box Plot*. Fonte: autor.

Tabela 3 - Propriedades estatísticas da distribuição normal dos dados de M,  $x - x_0$ ,  $E_x$  e U. Fonte: autor.

Propriedade	Massa (g)	Distância da fonte poluidora (m)	Dispersão Longitudinal ( $m^2/s$ )	Velocidade (m/s)
Média	1985,2	99,01	1,44	0,59
Variância	36092	96,14	0,0704	0,0035
Desvio Padrão	189,98	9,8053	0,2654	0,059
Mínimo	1570,2	79,51	1,0065	0,4726
1º quartil	1843,1	91,05	1,2229	0,5427
Mediana	1963,5	99,36	1,4442	0,5948
3º quartil	2110,5	105,07	1,6284	0,6309
Máximo	2439,5	122,74	1,9964	0,7424

Intervalos de Confiança				
Média	1950,9 – 2019,5	97,24 – 100,78	1,392 – 1,488	0,582 – 0,603
Mediana	1929,1 – 1997,8	97,58 – 101,13	1,396 – 1,492	0,584 – 0,605

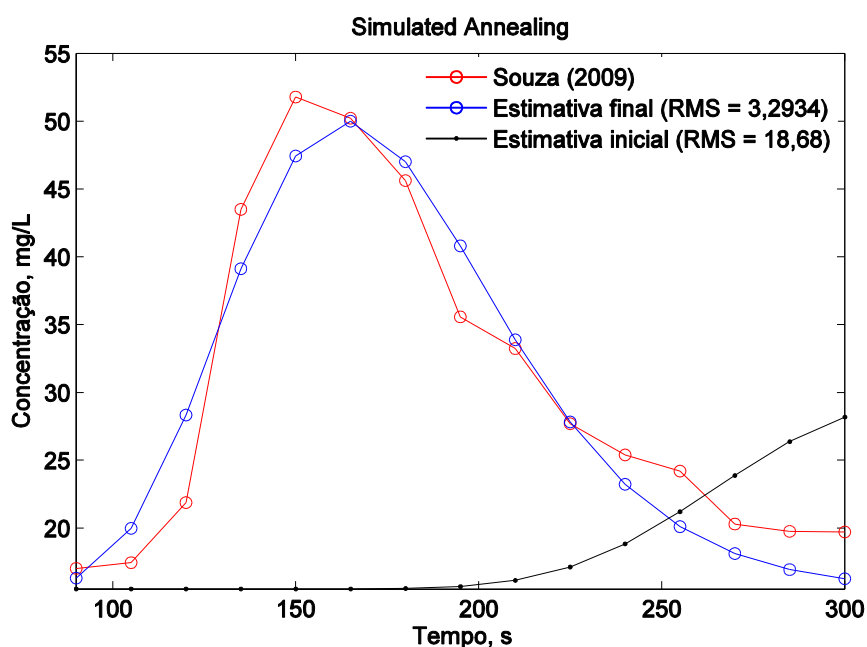


Figura 21 – Ajuste da concentração de poluente, estimando a massa do poluente, a posição de lançamento, dispersão longitudinal e velocidade pelo algoritmo

*Simulated Annealing*. Fonte: autor.

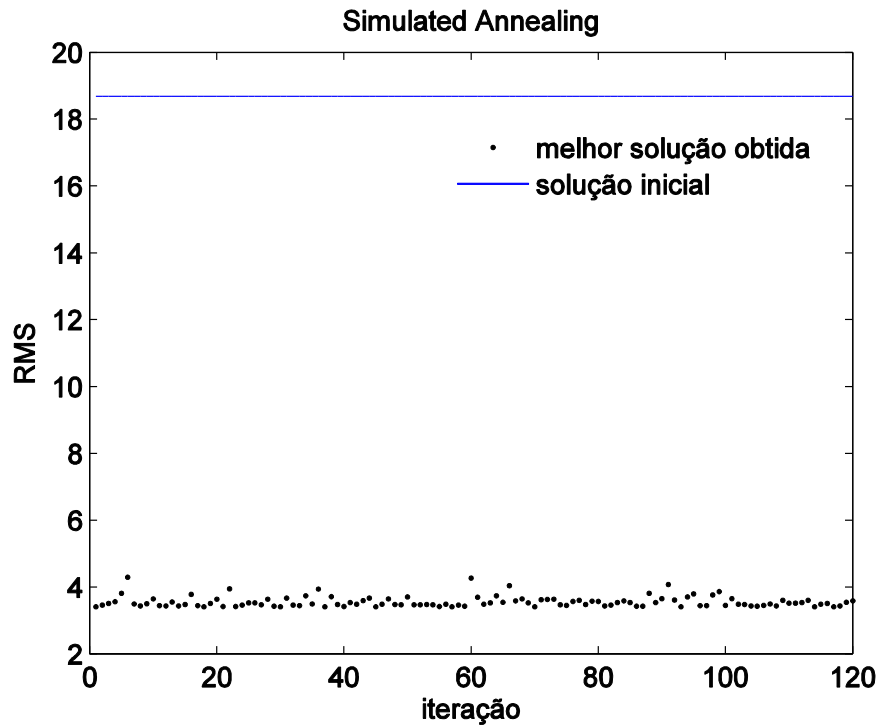


Figura 22 – Valores de RMS obtidos considerando 120 repetições do algoritmo *Simulated Annealing* em relação à condição inicial. Fonte: autor.

### 6.1.2 Análise com Filtro de Partículas (Estimação On-line)

Os resultados foram simulados considerando uma incerteza  $\sigma_{med} = 1,55 \text{ mg/L}$  e  $\sigma_{mod} = 0,78 \text{ mg/L}$  e a informação *a priori* como sendo a mesma utilizada na resolução pelo método *Simulated Annealing*.

#### 6.1.2.1 Cenário 1: apenas $M$ e $x - x_0$ são desconhecidos

Os resultados sumarizados na Tabela 4 trazem os valores encontrados para os critérios de desempenho: erro *RMS*, coeficiente de determinação  $R^2$ , valor mínimo do coeficiente de reamostragem ( $REA_{min}$ ) e autocorrelação *SC*. Este índice foi avaliado com atraso no tempo de 0 até 20 intervalos (*lags*), cujos limites de significância foram iguais a  $\pm 0,5164$  ao nível de 95% de confiança. De modo geral, pode-se notar que o aumento da incerteza  $\sigma_p$  levou a uma melhora do desempenho

do filtro, pois, comparando-se os resultados com o mesmo número de partículas, houve redução dos critérios RMS e SC e aumento do coeficiente  $R^2$  e  $REA_{\min}$ .

Tabela 4 – Análise de desempenho do filtro de partículas SIR.

$\sigma_P/P_0$ (%)	$N_{\text{part}}$	RMS	$R^2$	SC*	$REA_{\min}$ (%)
	<b>100</b>	<b>20,3640</b>	<b>0,9094</b>	<b>0,7011</b>	<b>1,1612</b>
25	500	13,2815	0,9751	0,5643	0,2335
	1000	7,3444	0,9919	0,3830	0,2621
40	100	10,3461	0,9828	0,3989	2,3384
	500	8,0924	0,9924	0,2976	1,1031
	1000	5,8900	0,9938	0,1889	1,0196
60	100	8,0980	0,9915	0,2383	2,1664
	500	6,1528	0,9927	0,2330	2,7146
	<b>1000</b>	<b>5,0695</b>	<b>0,9945</b>	<b>0,3291</b>	<b>3,1732</b>

\* Limites de significância iguais a  $\pm 0,5164$  ao nível de 95% de confiança.

Para melhor analisar o resultado acima, construíram-se os gráficos, conforme as Figuras 23 e 24, respectivamente, para os ajustes obtidos para a pior (2ª linha da Tabela 4) e a melhor condição examinadas (última linha da Tabela 4). Observou-se que, na pior condição avaliada, não se atendeu satisfatoriamente a nenhum dos critérios de desempenho nem foi possível alcançar os valores experimentais de  $M = 2000$  g e  $x - x_0 = 100$  m em relação à informação *a priori* ( $M = 1178,5$  g e  $x - x_0 = 194,26$  m). Em contrapartida, os resultados para a melhor condição examinada atenderam a todos os critérios de desempenho, não resultaram em colapso da região de credibilidade e aproximaram-se dos valores de Souza (2009).

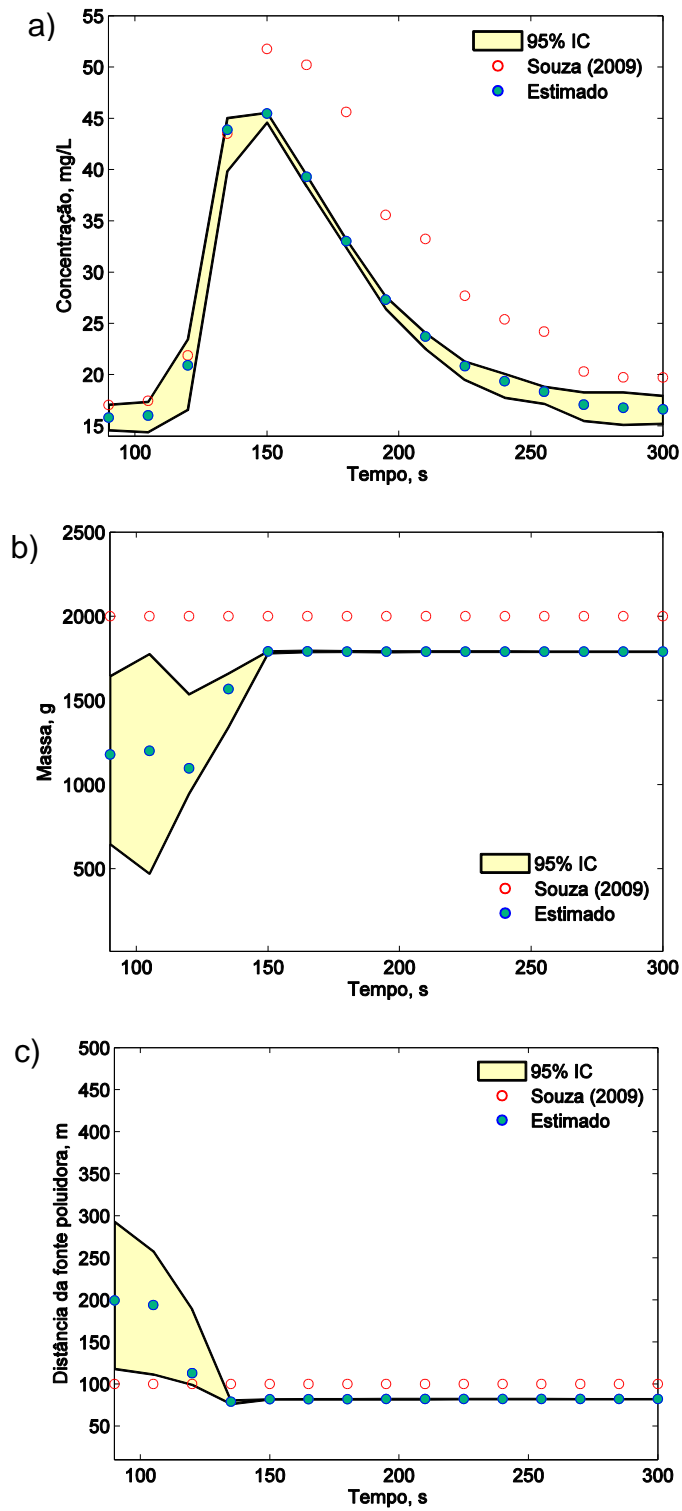


Figura 23 – Estimativas da concentração em tempo real (a), carga de poluente conservativo (b) e distância da fonte poluidora (c), considerando a pior condição examinada da Tabela 4:  $N_{\text{part}} = 100$  e  $\sigma_p / P_0 = 25\%$ . Fonte: autor.

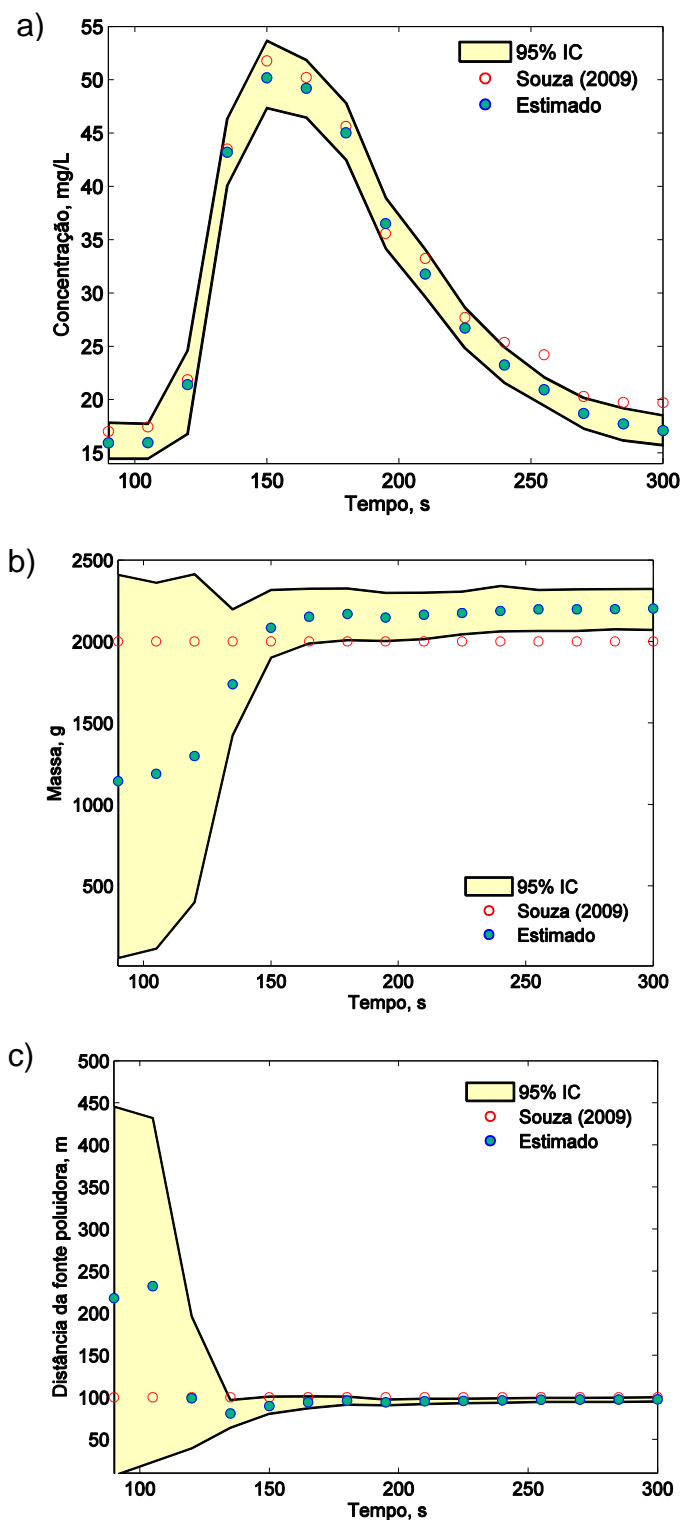


Figura 24 – Estimativas da concentração em tempo real (a), carga de poluente conservativo (b) e distância da fonte poluidora (c), considerando a melhor condição examinada da Tabela 4:  $N_{part} = 1000$  e  $\sigma_p / P_0 = 60\%$ . Fonte: autor.

Esses resultados são ressaltados e explicados através das Figuras 25 e 26, respectivamente, que fazem um comparativo da análise da autocorrelação do

resíduo e número efetivo de partículas para as condições mencionadas anteriormente. Observou-se que, na pior condição examinada ( $REA_{\min} = 1,1612\%$  ou  $N_{\text{eff}} \sim 1$ ), entre 120 e 255 s, o número de partículas foi insuficiente (Figura 25a) para levar a estimativas satisfatórias, o que acabou ocasionando o colapso da região de credibilidade das curvas de concentração, carga de poluente e distância da fonte poluidora (Figura 23). Ademais, também se notou que a autocorrelação do resíduo superou o limite da região de credibilidade (Figura 26a), indicando que o PF – SIR não foi capaz de filtrar as incertezas de medida da concentração, o que poderia ser um indicativo da necessidade de aumentar o número de partículas ou incerteza dos parâmetros. Feito isso, na melhor condição examinada, em virtude do maior espaço de busca e maior número de partículas -  $REA_{\min} = 3,1732\%$  ( $N_{\text{eff}} \sim 32$ ), realmente houve efetividade da técnica de reamostragem (Figura 25b) e redução da autocorrelação do resíduo de estimação (Figura 26b). Outro questionamento ao avaliar a Figura 25 está no fato de que entre 120 e 150 s, mesmo na melhor condição, há uma redução significativa do REA. Isto é explicado devido às estimativas nessa faixa serem mais sensíveis a erros de medidas, isto é, grandes variações de  $E_x$ ,  $U$ ,  $M$  e  $x - x_0$  correspondem a pequenas variações de concentração, o que torna mais difícil o processo de cálculo dos pesos e, conseqüentemente, a reamostragem.

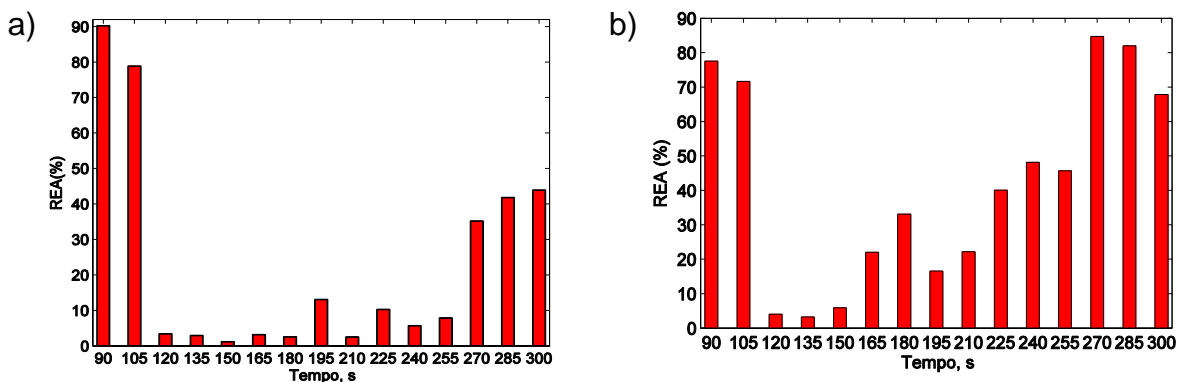


Figura 25 – Análise do REA para  $N_{\text{part}} = 100$  e  $\sigma_p / P_0 = 25\%$  (a) e  $N_{\text{part}} = 1000$  e  $\sigma_p / P_0 = 60\%$  (b). Fonte: autor.

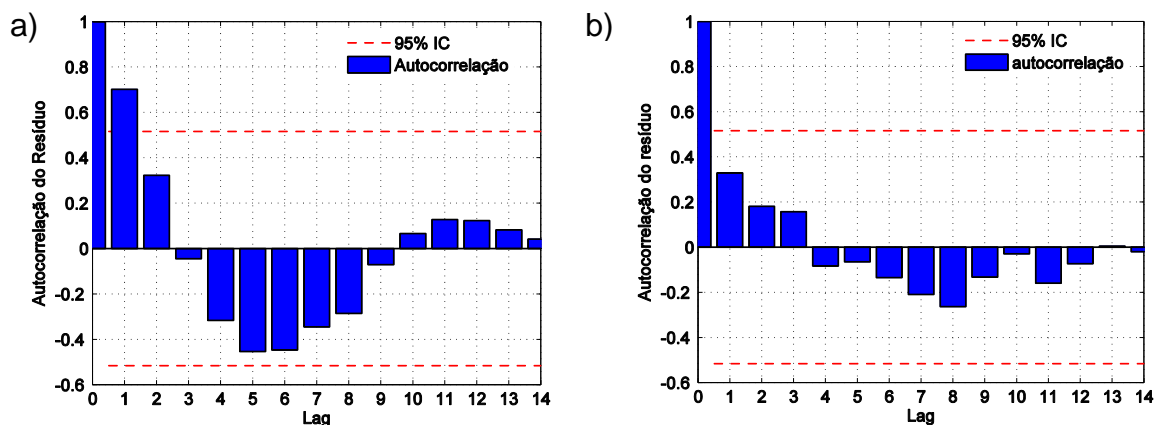


Figura 26 – Análise da autocorrelação do resíduo de estimação para  $N_{\text{part}} = 100$  e  $\sigma_p / P_0 = 25\%$  (a) e  $N_{\text{part}} = 1000$  e  $\sigma_p / P_0 = 60\%$  (b). Fonte: autor.

#### 6.1.2.2 Cenário 2: Método Kernel smoothing para $E_x$ e $U$

Os resultados obtidos estão presentes na Tabela 5. O índice SC foi avaliado com atraso no tempo de 0 até 20 intervalos (*lags*), cujos limites de significância foram iguais a  $\pm 0,5164$  ao nível de 95% de confiança. Ademais, houve uma significativa redução do RMS na melhor condição examinada em comparação ao Cenário anterior, cerca de 2 vezes, levando a concluir que o acréscimo de incerteza nos parâmetros  $E_x$  e  $U$  permite uma melhor estimativa da distribuição de concentração do experimento com perturbação em pulso.

Uma análise adicional desse resultado é ilustrada nas Figuras 27 e 28, incluindo as estimativas obtidas em tempo real da dispersão longitudinal e velocidade de escoamento do rio. Diante disso, a pior condição examinada (2ª linha da Tabela 5) apresentou os mesmos resultados incoerentes como no Cenário 1. Já, em relação a melhor condição examinada (última linha da Tabela 5), atenderam-se todos os critérios de desempenho e houve uma melhor aproximação de  $M = 2000$  g e  $x - x_0 = 100$  m a partir de 150 s. Isto evidencia a utilidade do incremento de incerteza paramétrica nos resultados comparado a melhor condição do Cenário 1. Foi possível aproximar melhor a distância da fonte poluidora (Figura 28c) e o valor da carga de poluente (Figura 28b) com desvio pequeno em relação ao valor exato. Ademais, nota-se a evolução artificial dos parâmetros em torno de estimativas diferentes das originais (Figuras 28d e 28e).

As justificativas estatísticas dos resultados acima são expostas conforme as Figuras 29 e 30. Observou-se similaridade com os resultados ilustrados pelas Figuras 25 e 26, destacando-se a análise do REA, o qual na melhor condição (Figura 29b) é superior em magnitude em um número de intervalos maior (165 – 300 s) comparado a mesma condição do Cenário 1. Isto corrobora com as descrições feitas no Capítulo 4, conforme a literatura consultada, validando dessa forma a utilização do método Kernel *smoothing* no aumento do número efetivo de partículas, ou seja, evitando o empobrecimento da amostra. Ademais, quanto à autocorrelação de resíduo, observa-se conforme a Figura 30b, real redução do resíduo de estimação na melhor condição examinada.

Tabela 5 – Análise de desempenho do filtro de partículas SIR com Kernel *smoothing*.

$\sigma_P/P_0$ (%)	$N_{part}$	RMS	$R^2$	SC*	REA <sub>min</sub> (%)
	<b>100</b>	<b>32,2039</b>	<b>0,6344</b>	<b>0,7684</b>	<b>1,0119</b>
25	500	5,3493	0,9964	0,7560	1,9252
	1000	3,2031	0,9981	0,4288	0,4288
40	100	5,9416	0,9888	0,3342	2,0444
	500	3,5482	0,9972	0,3680	1,0116
	1000	2,4775	0,9986	0,3909	3,2028
60	100	4,5849	0,9960	0,4078	4,4060
	500	2,5925	0,9979	0,3524	1,6631
	<b>1000</b>	<b>1,9014</b>	<b>0,9985</b>	<b>0,3748</b>	<b>3,2151</b>

\* Limites de significância iguais a  $\pm 0,5164$  ao nível de 95% de confiança.

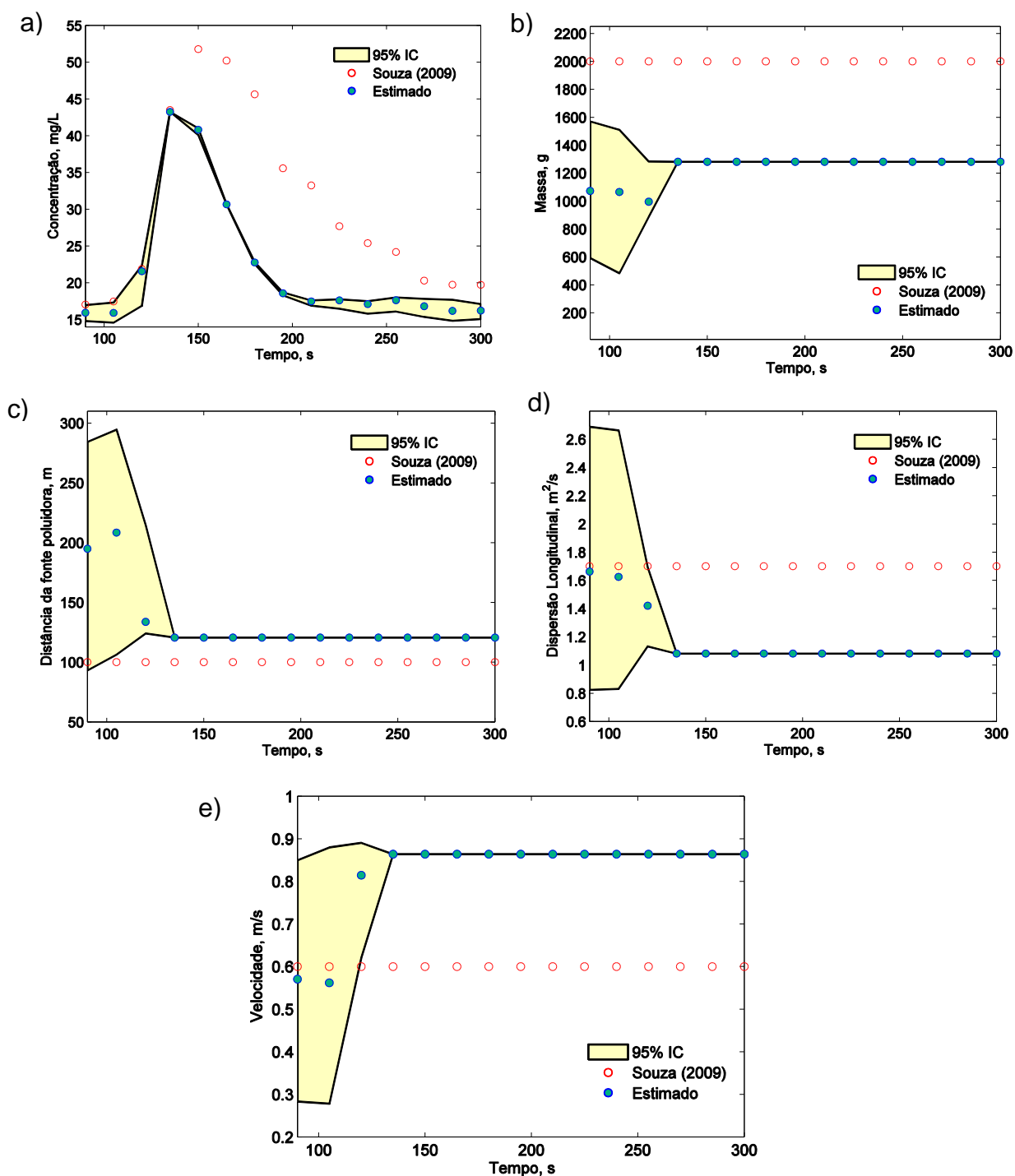


Figura 27 – Estimativas da concentração em tempo real (a), carga de poluente conservativo (b), distância da fonte poluidora (c), dispersão longitudinal (d) e velocidade (e), considerando a pior condição examinada da Tabela 5:  $N_{\text{part}} = 100$  e  $\sigma_p / P_0 = 25\%$ . Fonte: autor.

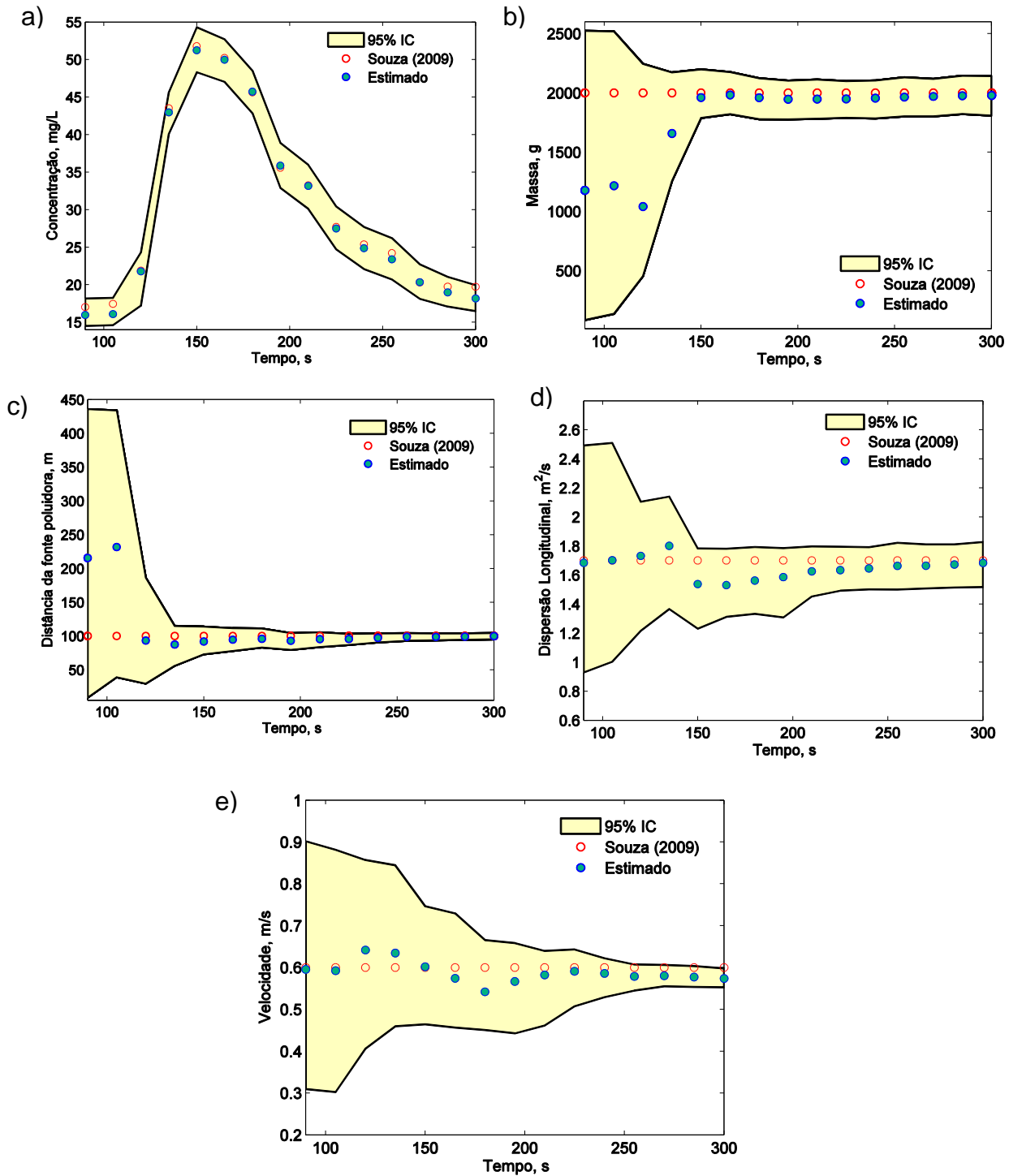


Figura 28 – Estimativas da concentração em tempo real (a), carga de poluente conservativo (b), distância da fonte poluidora (c), dispersão longitudinal (d) e velocidade (e), considerando a melhor condição da Tabela 5:  $N_{\text{part}} = 1000$  e  $\sigma_p / P_0 = 60\%$  Fonte: autor.

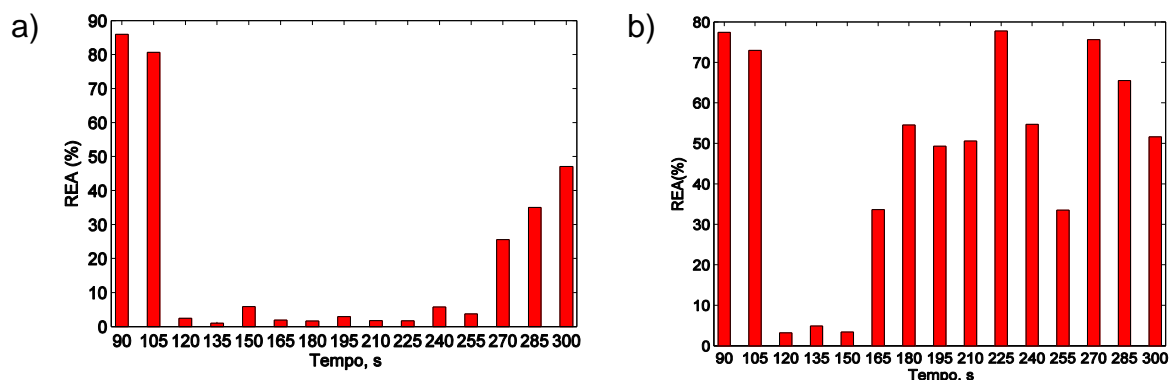


Figura 29 – Análise do REA para  $N_{part} = 100$  e  $\sigma_P / P_0 = 25\%$  (a) e  $N_{part} = 1000$  e  $\sigma_P / P_0 = 60\%$  (b). Fonte: autor.

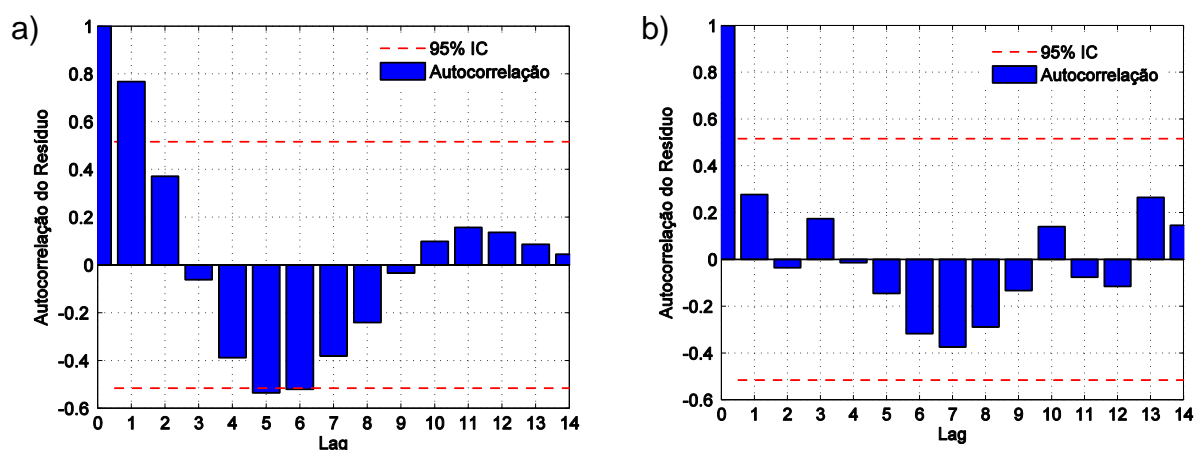


Figura 30 – Análise da autocorrelação do resíduo de estimação para  $N_{part} = 100$  e  $\sigma_P / P_0 = 25\%$  (a) e  $N_{part} = 1000$  e  $\sigma_P / P_0 = 60\%$  (b). Fonte: autor.

## 6.2 Problema de Lançamento Contínuo

Os resultados e discussões vinculados a esta seção são apenas para o filtro de partículas (estimação em tempo real) devido à especificidade do estudo de caso retratado. Este exemplo é representado por medidas sintéticas em que se objetiva comprovar a efetividade do PF em reduzir o ruído de estimação da concentração medida e, conseqüentemente, auxiliar no entendimento da dinâmica do sistema. Vale ressaltar que o método de Simulated Annealing não foi empregado aqui devido ao filtro de partículas SIR no estudo de caso anterior, mesmo na condição de

número de partículas igual a 1000, ser responsável por um ganho de tempo igual a 815 s comparado ao Simulated Annealing, confirmando-se como uma alternativa para possíveis aplicações de controle, inferência virtual e alarme por precisar de aproximadamente 5 s para completar cada simulação, tempo inferior ao de amostragem (15 s).

Além disso, os resultados foram simulados considerando uma incerteza  $\sigma_{med} = 158,73 \text{ mg/L}$  e  $\sigma_{mod} = 15,873 \text{ mg/L}$ , número de subintervalos igual a 100 para integração numérica utilizando o método de Simpson 1/3 e  $\tau = 120,30 \text{ s}$ , resultado esse escolhido de maneira aleatória, diferente dos demais parâmetros que continuam com os mesmos valores descritos por Souza (2009), ou seja,  $\dot{M} = 2000 \text{ g/s}$ ,  $E_x = 1,7 \text{ m}^2/\text{s}$  e  $U = 0,6 \text{ m/s}$ .

### 6.2.1 Análise com Filtro de Partículas (Estimação On-line)

Os resultados sumarizados na Tabela 6 trazem os valores encontrados para os critérios: erro *RMS*, coeficiente de determinação  $R^2$ , valor mínimo da reamostragem  $REA_{min}$  e autocorrelação *SC*. Este índice foi avaliado com atraso no tempo de 0 até 20 intervalos (*lags*), cujos limites de significância foram iguais a  $\pm 0,4364$  ao nível de 95% de confiança. De modo geral, pode-se notar que o aumento da incerteza  $\sigma_p$  levou a uma melhora do desempenho de ambos os filtros, pois, comparando-se os resultados com o mesmo número de partículas, houve redução dos critérios *RMS* e *SC* e aumento do coeficiente  $R^2$  e  $REA_{min}$ .

Contudo, o filtro de partículas SIR com Kernel *Smoothing* demonstrou-se mais vantajoso, pois manteve os valores de *SC* inferiores ao limite de credibilidade em um número maior de ocasiões e reduziu o empobrecimento da amostra ao aumentar os valores de  $REA_{min}$ . Para confirmar tais afirmações, é mostrado na Figura 31 um comparativo do REA entre as duas condições grifadas em negrito na Tabela 6. Nesta figura, principalmente entre o período de transição da concentração do traçador (135 a 255 s), observa-se um aumento no REA proporcionado pelo método Kernel *smoothing* (Figura 31a), o que corrobora com as informações ditas por Liu e West (2001).

Tabela 6 – Análise de Desempenho do filtro de partículas SIR.

$\sigma_P/P_0$ (%)	$N_{part}$	RMS	$R^2$	$SC^*$	$REA_{min}$ (%)
<b>Filtro de Partículas SIR – <math>E_x</math> e <math>U</math>, fixos.</b>					
25	100	4383,2	0,6440	0,8643	2,0781
	500	2509,400	0,9041	0,8255	1,0271
	1000	675,9488	0,9926	0,5350	1,2237
40	100	1258,200	0,9670	0,7362	1,3697
	500	603,3884	0,9919	0,5446	3,5539
	1000	572,2133	0,9922	0,5261	1,9546
60	100	850,6992	0,9869	0,6985	3,1519
	500	457,0051	0,9941	0,4144	1,4445
	<b>1000</b>	<b>431,3439</b>	<b>0,9937</b>	<b>0,4243</b>	<b>1,8706</b>
<b>Filtro de Partículas SIR – técnica de estimação de Kernel para <math>E_x</math> e <math>U</math></b>					
25	100	3188,5	0,8160	0,8556	1,5013
	500	704,0566	0,9943	0,4815	1,5507
	1000	415,8403	0,9931	0,5559	1,6986
40	100	937,2963	0,9934	0,5718	1,3704
	500	275,9881	0,9956	0,3856	2,6342
	1000	258,8805	0,9973	0,2856	3,0751
60	100	786,8896	0,9889	0,6674	8,6564
	500	241,7320	0,9976	0,3067	4,7700
	<b>1000</b>	<b>206,5792</b>	<b>0,9980</b>	<b>0,3183</b>	<b>3,6710</b>

\* Limites de significância iguais a  $\pm 0,4364$  ao nível de 95% de confiança.

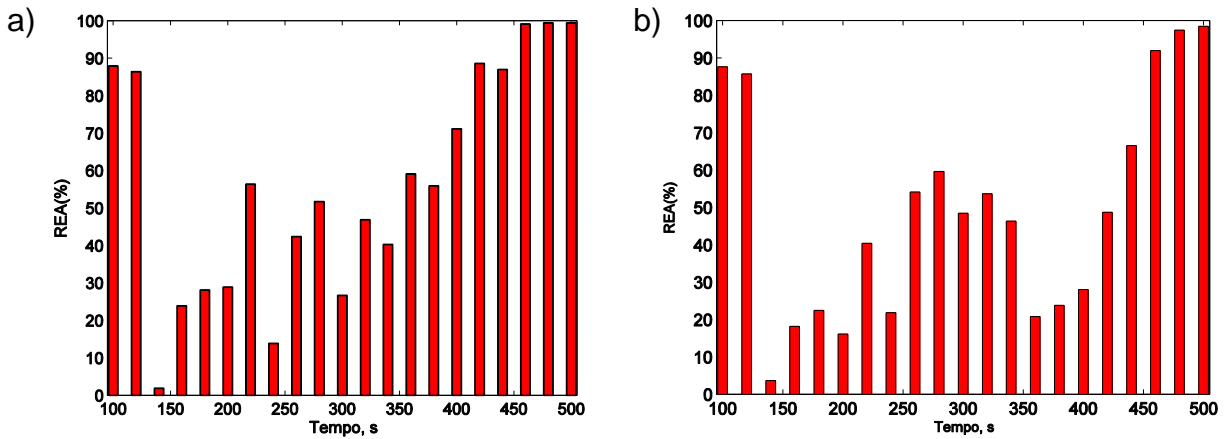


Figura 31 – Análise do REA para o filtro de partículas SIR com Kernel *Smoothing* (a) e filtro de partículas SIR (b), considerando  $N_{\text{part}} = 1000$  e  $\sigma_p / P_0 = 60\%$ . Fonte: autor.

O monitoramento em tempo real da dinâmica do processo que resultou na melhor condição de desempenho é mostrado de acordo com a Figura 32. A Figura 33 mostra o ajuste dos parâmetros  $E_x$ ,  $U$ ,  $M$  e  $x - x_0$  devido à atualização da concentração (estado) com o tempo. Nesta figura, observa-se que aos 250 s do experimento é conhecida toda a dinâmica do processo. Isto evidencia a utilização de assimilação de dados, computada pelo filtro de partículas SIR com Kernel *Smoothing*, responsável por produzir estimativas (condição com maior número de partículas) em um tempo 10 s inferior ao de amostragem (20 s).

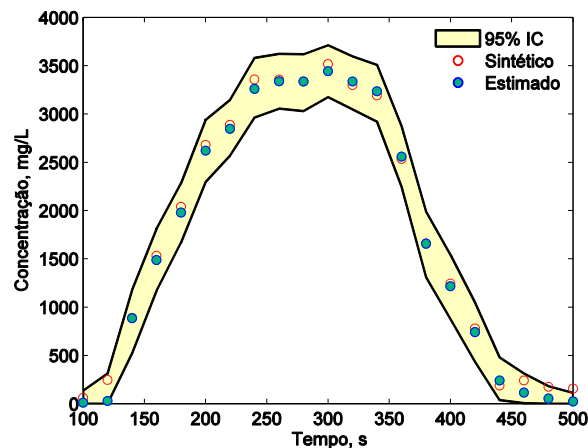


Figura 32 – Concentração estimada em tempo real, considerando método de Kernel *Smoothing* para  $E_x$  e  $U$ ,  $N_{\text{part}} = 1000$  e  $\sigma_p / P_0 = 60\%$ . Fonte: autor.

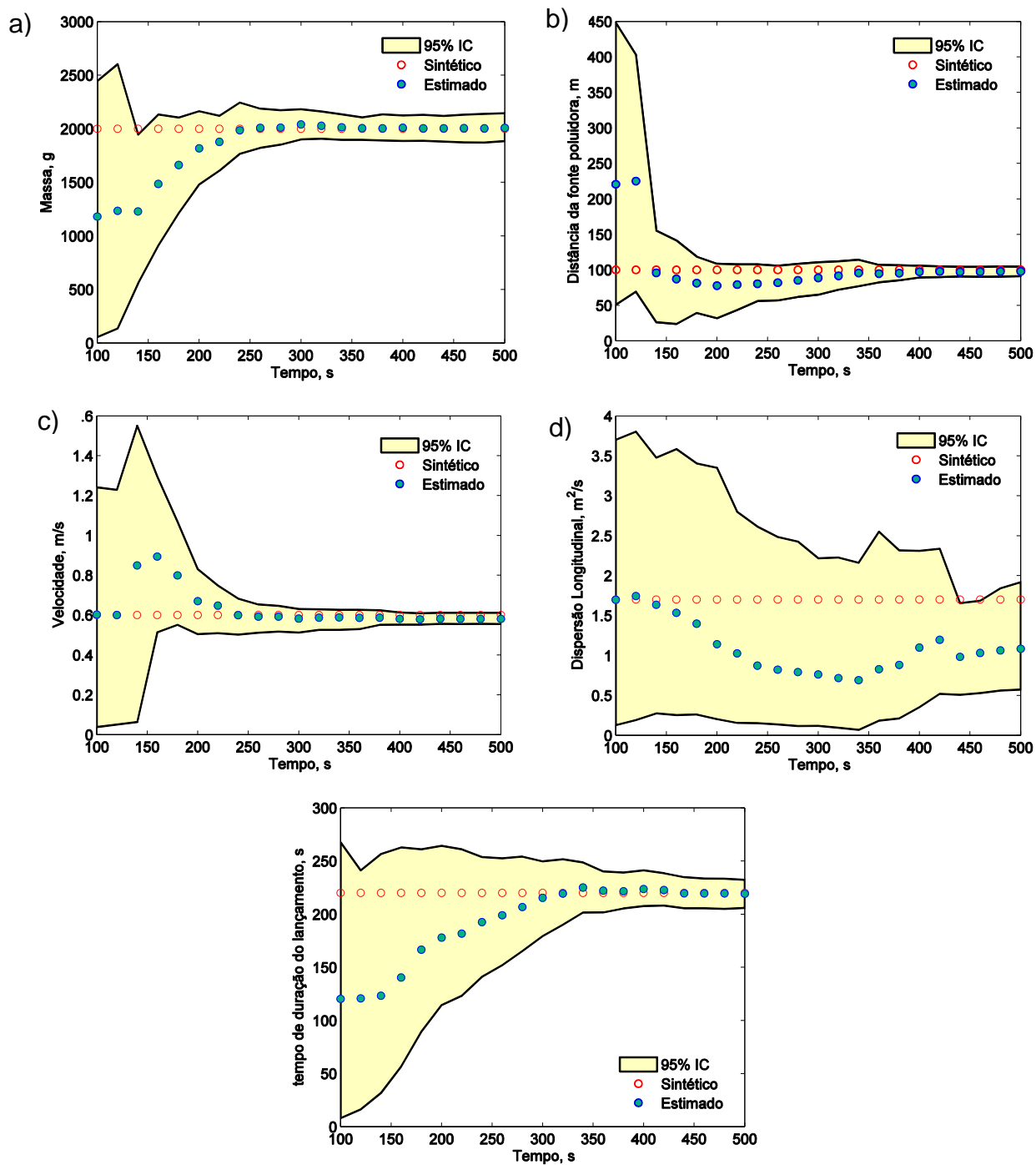


Figura 33 – Ajuste em tempo real da massa (a), distância da fonte de poluição (b), velocidade (c), dispersão longitudinal (d) e tempo de duração do lançamento (e), considerando Kernel Smoothing para  $E_x$  e  $U$ ,  $N_{\text{part}} = 1000$  e  $\sigma_p / P_0 = 60\%$ . Fonte:

autor.



## 7 CONCLUSÕES E SUGESTÕES

A análise da carga de poluentes conservativos liberados em rios é um tema vinculado a muitas pesquisas acadêmicas, distinguindo-se as abordagens de modelagem, controle e análise de processos dependendo da fonte responsável pela poluição (análise da carga de poluentes conservativos lançados instantaneamente ou de maneira contínua em um rio).

O enfoque alternativo desta dissertação de mestrado quanto aos temas relacionados acima se concentrou na análise de diferentes cenários de problemas inversos, baseando-se no algoritmo *Simulated Annealing* (estimação *off-line*) e no Filtro de Partículas SIR (estimação em tempo real). Constatou-se que a técnica de assimilação de dados em tempo real melhora o entendimento destes fenômenos em relação à carga de poluente lançado, posição em que se encontra e, quando contínuo, a duração do intervalo de lançamento do poluente.

Neste contexto, os resultados obtidos tanto para estimação *off-line* quanto para estimação em tempo real foram representativos para análise da dispersão de poluentes em rios frente às abordagens normalmente utilizadas na literatura que não realizam análise completa da dinâmica do sistema.

Especificamente, o método *Simulated Annealing* empregado em uma sequência de repetições demonstrou ser uma técnica de estimação eficiente para aplicações *off-line*. Isto foi confirmado por meio de análises estatísticas, destacando-se a boa adequação dos intervalos da estatística *t* da média da carga de poluente lançado (1950,9 – 2019,5) e posição em que se encontra (97,24 – 100,78) quaisquer que sejam a informação fornecida inicialmente. Apesar de todas essas vantagens, é um método que depende do conhecimento completo da curva de passagem do traçador, e que em média demandou um tempo igual a 670 s para completar as 120 simulações. A questão mais importante quanto ao tempo gasto é que este método pode ser aplicado apenas após o lançamento do poluente. Com isso, pode não ser adequado para utilização como um sensor virtual para monitoramento ambiental.

Neste sentido, a proposta PF – SIR para diferentes cenários mostrou-se eficiente na estimação em tempo real conforme as medidas do poluente estão disponíveis. Além disso, superou a principal desvantagem do método *Simulated Annealing* ao proporcionar a partir do tempo 155 s estimativas conclusivas sobre a carga do poluente lançado e posição em que se encontra o lançamento. Diante dessas

circunstâncias, em média o PF – SIR com Kernel *Smoothing* levou a um ganho de tempo de 815 s em comparação a técnica *off-line*, confirmando-se como uma alternativa para possíveis aplicações de controle, inferência virtual e alarme.

Também se aplicou a estimação em tempo real para o caso envolvendo o lançamento contínuo de duração finita. Comprovou-se a utilidade desta técnica, que permitiu identificar as mudanças de dinâmica do sistema 260 s antes que o experimento se completasse, indicando que a abordagem é válida em identificar os parâmetros em tempo real e reduzir estatisticamente o ruído de medida.

Completando, pontua-se que este trabalho de Dissertação de Mestrado apresenta uma contribuição acadêmica relevante e pode ser utilizado como material de pesquisa quando se planeja realizar um estudo da dinâmica de modelos de dispersão de poluentes em rios, sendo também indicado para trabalhos que avaliam a dispersão de poluentes em grande escala, contexto este que não foi retratado nesta ocasião, e é indicado para trabalhos futuros. Outras sugestões de análise do problema de dispersão de poluentes em rios podem ser dadas: uso de CFD para geometrias de escoamento mais complexas, com interferência de zonas mortas e efeitos de turbulência; análise da dispersão de poluentes que sofrem alterações químicas ao longo do escoamento; e melhorar a eficiência computacional dos métodos de estimação de modo a identificar a presença, a quantidade e a posição de lançamento de múltiplos agentes poluentes em rios.

## REFERÊNCIAS

ABI, T. A.; NWOSU, P. C. The effect of oil spillage on the soil of Eleme in Rivers State of the Niger Delta area of Nigeria. **Res. J. Environ. Sci**, v. 3, n. 3, p. 316-320, 2009.

ABDERREZZAK, Kamal El Kadi; ATA, Riadh; ZAOUI, Fabrice. One-dimensional numerical modelling of solute transport in streams: The role of longitudinal dispersion coefficient. **Journal of Hydrology**, v. 527, p. 978-989, 2015.

ALLISON, Chibuike et al. Dealing with Oil Spill Scenarios in the Niger Delta: Lessons from the Past. In: **The Political Ecology of Oil and Gas Activities in the Nigerian Aquatic Ecosystem**. 2018. p. 351-368.

ANI, Elisabeta-Cristina et al. Development, calibration and evaluation of two mathematical models for pollutant transport in a small river. **Environmental Modelling & Software**, v. 24, n. 10, p. 1139-1152, 2009.

ARDEJANI, Faramarz Doulati et al. Application of Computational Fluid Dynamics (CFD) for simulation of acid mine drainage generation and subsequent pollutants transportation through groundwater flow systems and rivers. In: **Computational Fluid Dynamics Technologies and Applications**. InTech, 2011.

ARULAMPALAM, M. Sanjeev et al. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. **IEEE Transactions on signal processing**, v. 50, n. 2, p. 174-188, 2002.

BAEK, Kyong Oh; SEO, Il Won. On the methods for determining the transverse dispersion coefficient in river mixing. **Advances in Water Resources**, v. 90, p. 1-9, 2016.

BENEDINI, Marcello; TSAKIRIS, George. **Water quality modelling for rivers and streams**. Springer Science & Business Media, 2013.

BIRD, R. B., STEWART, W. E., LIGHTFOOT, E. N., **Fenómenos de transporte**. 2ed. Limusa Wiley, Mexico, 2006.

BLOCKEN, Bert; GUALTIERI, Carlo. Ten iterative steps for model development and evaluation applied to Computational Fluid Dynamics for Environmental Fluid Mechanics. **Environmental Modelling & Software**, v. 33, p. 1-22, 2012.

Box, G. E. P.; Jenkins, G. M.; Reinsel, G. C. (1994), "**Time Series Analysis: Forecasting and Control**". 3rd ed., Prentice Hall, Englewood Cliffs.

CHEN, Dingjiang et al. A Bayesian approach for calculating variable total maximum daily loads and uncertainty assessment. **Science of the total environment**, v. 430, p. 59-67, 2012.

DA SILVA, Wellington Betencurte. **Aplicação de filtros de partículas para a assimilação de dados em problemas de fronteira móvel**. 2012. Tese de Doutorado. Tese de Doutorado, PEM/COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.

DATTA, Bithin et al. Optimal unknown pollution source characterization in a contaminated groundwater aquifer: evaluation of a developed dedicated software tool. **Journal of Geoscience and Environment Protection**, v. 2, p. 41-51, 2014.

DECHANT, Caleb M.; MORADKHANI, Hamid. Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. **Water Resources Research**, v. 48, n. 4, 2012.

DE SOUZA PAROLIN, Radael *et al.* **Análise de sensibilidade e estimação de uma fonte de contaminantes no estuário do Rio Macaé**.

DOUCET, Arnaud; DE FREITAS, Nando; GORDON, Neil. An introduction to sequential Monte Carlo methods. In: **Sequential Monte Carlo methods in practice**. Springer, New York, NY, 2001. p. 3-14.

DUARTE, António AL; BOAVENTURA, Rui A. Rocha. Dispersion modelling in rivers for water sources protection, based on tracer experiments: case studies. **Waste management, water pollution, air pollution, indoor climate (WWAI'08)(ISI Book)**, p. 205-210, 2008.

FARIA, R.R. et al. Ajuste dos parâmetros de dispersão no Rio Macaé com o método Simulated Annealing. **Essentia Editora**. 2018

FARIA, R.R. et al. Análise de desempenho do filtro de partículas SIR usando a autocorrelação do resíduo. **Essentia Editora**. 2018.

FISCHER, Hugo B. The mechanics of dispersion in natural streams. **Journal of the Hydraulics division**, v. 93, n. 6, p. 187-216, 1967.

GUSTAFSSON, Fredrik. Particle filter theory and practice with positioning applications. **IEEE Aerospace and Electronic Systems Magazine**, v. 25, n. 7, p. 53-82, 2010.

HAMZA, Sonia Ben et al. Simulation of pollutant dispersion of a free surface flow in coastal water. **Ocean Engineering**, v. 108, p. 81-97, 2015.

HAMMERSLEY, J. M. e HANSCOMB, D. C., **Monte Carlo Methods**, London: *Chapman & Hall*, 1964.

HUTTON, C. J. et al. **Uncertainty Quantification and Reduction in Urban Water Systems (UWS) Modelling**: Evaluation report. PREPARED Enabling Change. Available from: [www.prepared-fp7.eu/viewer/file.aspx](http://www.prepared-fp7.eu/viewer/file.aspx), 2011.

HUTTON, C. J. et al. Real-Time Modelling and Data Assimilation Techniques for Improving the Accuracy of Model Predictions: Scientific Report. 2010.

JHA, Manish; DATTA, Bithin. Three-dimensional groundwater contamination source identification using adaptive simulated annealing. **Journal of Hydrologic Engineering**, v. 18, n. 3, p. 307-317, 2012.

JI, Zhen-Gang. **Hydrodynamics and water quality: modeling rivers, lakes, and estuaries**. John Wiley & Sons, 2017.

JIRKA, Gerhard H.; WEITBRECHT, Volker. Mixing models for water quality management in rivers: continuous and instantaneous pollutant releases. In: **Water quality hazards and dispersion of pollutants**. Springer, Boston, MA, 2005. p. 1-34.

KAPIO, J. e SOMERSALO, E., 2004, **Statistical and Computational Inverse Problems**, *applied Mathematical Sciences* 160, Springer-Verlag.

Kirkpatrick, S.; Gelatt, C.D.; Vecchi, M.P. (1983), **Optimization by Simulated Annealing**. Science, 220, 671-680.

LAUNAY, M. et al. Calibrating pollutant dispersion in 1-D hydraulic models of river networks. **Journal of Hydro-environment Research**, v. 9, n. 1, p. 120-132, 2015.

LEIBUNDGUT, C., MALOSZEWSKI, P., Külls, C., 2009. **Tracers in Hydrology**. Wiley-Blackwell.

LEISENRING, Marc; MORADKHANI, Hamid. Analyzing the uncertainty of suspended sediment load prediction using sequential data assimilation. **Journal of hydrology**, v. 468, p. 268-282, 2012.

LIU, Jane; WEST, Mike. Combined parameter and state estimation in simulation-based filtering. In: **Sequential Monte Carlo methods in practice**. Springer, New York, NY, 2001. p. 197-223.

LIU, Yong et al. Water quality modeling for load reduction under uncertainty: a Bayesian approach. **Water Research**, v. 42, n. 13, p. 3305-3314, 2008.

LUGON, Jader; SILVA NETO, Antonio J.; RODRIGUES, Pedro Paulo Gomes Watts. Assessment of dispersion mechanisms in rivers by means of an inverse problem approach. **Inverse Problems in Science and Engineering**, v. 16, n. 8, p. 967-979, 2008.

MANSOURI, Majdi; DUMONT, Benjamin; DESTAIN, Marie-France. Modeling and prediction of nonlinear environmental system using Bayesian methods. **Computers and electronics in agriculture**, v. 92, p. 16-31, 2013.

NEWMAN, Mark et al. A hybrid method for inverse characterization of subsurface contaminant flux. **Journal of contaminant hydrology**, v. 81, n. 1-4, p. 34-62, 2005.

ORLANDE, H.; COLAÇO, M.; DULIKRAVICH, G.; VIANNA, F.; DA SILVA, W. B.; FONSECA, H.; FUDYM, O. State estimation problems in heat transfer. *International Journal for Uncertainty Quantification*, v. 2, p. 239-258, 2012.

ÖZİŞİK, M. N. (1993). **Heat conduction**. John Wiley & Sons.

PINHEIRO, LAR et al. Aplicação de Simulação computacional à dispersão de poluentes no baixo rio Amazonas: potenciais riscos à captação de água na orla de Macapá. **Amazônia: Ciência & Desenvolvimento**, v. 4, n. 7, p. 27-44, 2008.

PITT, M. AND SHEPHARD, N., "Filtering via simulation: Auxiliary particle filters," *J. Amer. Statist. Assoc.*, vol. 94, no. 446, pp. 590–599, 1999.

RIPLEY, B . **Stochastic Simulation**. Hoboken, NJ: Wiley, 1988.

RODRIGUES, Pedro Paulo Gomes Watts et al. Evaluation of dispersion parameters for River São Pedro, Brazil, by the simulated annealing method. **Inverse Problems in Science and Engineering**, v. 21, n. 1, p. 34-51, 2013.

RISTIC, B., ARULAMPALAM, S., GORDON, N., 2004. **Beyond the Kalman Filter**, Artech House, Boston.

SCHATZMANN, M., LEITL, B., 2011. Issues with validation of urban flow and dispersion CFD models. **J. Wind Eng. Ind. Aerodyn.** 99 (4), 169e186.

SALAMON, Peter; FEYEN, Luc. Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter. **Journal of Hydrology**, v. 376, n. 3-4, p. 428-442, 2009.

SAMUEL, Jos et al. Assessing model state and forecasts variation in hydrologic data assimilation. **Journal of hydrology**, v. 513, p. 127-141, 2014.

SOCOLOFSKY, Scott A.; JIRKA, Gerhard H. **Special topics in mixing and transport processes in the environment**. Engineering Lectures. 5th Edition. Texas A&M University, 2005.

SOUSA, E. P. **Avaliação de Mecanismos dispersivos em rios através de problemas inversos**. 2009. Dissertação de Mestrado, Nova Friburgo.

TARANTOLA, A. **Inverse Problem Theory**. Amsterdam: Elsevier, 1987.

TELLES, Wagner Rambaldi et al. **Simulação de uma Pluma de Contaminantes no Rio Macaé Utilizando Redes Neurais Artificiais.**

WU, Haichun et al. Deep Tunnel for regulating combined sewer overflow pollution and flood disaster: A case study in Guangzhou City, China. **Water**, v. 8, n. 8, p. 329, 2016.

YEH, Hund-Der; CHANG, Tung-Hua; LIN, Yu-Chung. Groundwater contaminant source identification by a hybrid heuristic approach. **Water Resources Research**, v. 43, n. 9, 2007.

ZENG, Yuhong; HUAI, Wenxin. Estimation of longitudinal dispersion coefficient in rivers. **Journal of Hydro-environment Research**, v. 8, n. 1, p. 2-8, 2014.

ZHAO, Ying et al. A Bayesian method for multi-pollution source water quality model and seasonal water quality management in river segments. **Environmental modelling & software**, v. 57, p. 216-226, 2014

ZHOU, Qianqian et al. Adaptation to urbanization impacts on drainage in the city of Hohhot, China. **Water Science and Technology**, v. 73, n. 1, p. 167-175, 2016.