

UNIVERSIDADE ESTADUAL DE MARINGÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

DISSERTAÇÃO DE MESTRADO

HEITOR DELIVIO SANCHES SALINAS

Modelagem Probabilística para Rastreabilidade em Cadeias Produtivas

Maringá
2026

HEITOR DELIVIO SANCHES SALINAS

Modelagem Probabilística para Rastreabilidade em Cadeias Produtivas

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção do Departamento de Engenharia de Produção, Centro de Tecnologia da Universidade Estadual de Maringá, como requisito parcial para obtenção do título de Mestre em Engenharia de Produção.
Área de concentração: Engenharia de Produção

Orientador: Prof. Dr. George Lucas Moraes Pezzott

Coorientadora: Profa. Dra. Francielle Cristina Fenerich

Maringá
2026

FOLHA DE APROVAÇÃO

HEITOR DELIVIO SANCHES SALINAS

Modelagem Probabilística para Rastreabilidade em Cadeias Produtivas

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção do Centro de Tecnologia da Universidade Estadual de Maringá, como requisito parcial para obtenção do título de Mestre em Engenharia de Produção pela Banca Examinadora composta pelos membros:

BANCA EXAMINADORA


Prof. Dr. George Lucas Moraes Pezzott
Presidente Orientador
Universidade Estadual de Maringá – PGP/UEM


Profa. Dra. Gislaine Camila Lapasini Leal
Membro examinador(a) interno
Universidade Estadual de Maringá – PGP/UEM

Documento assinado digitalmente
 **DANILO ALVARES DA SILVA**
Data: 30/04/2026 07:49:44-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Danilo Alvares da Silva
Membro examinador(a) externo
MRC Biostatistics Unit/University of Cambridge


Profa. Dra. Francielle Cristina Fenerich
Coorientadora
Universidade Estadual de Maringá – PGP/UEM

Aprovada em: 29 de abril de 2026.

Local da defesa: Sala 02, Bloco 19, *campus* da Universidade Estadual de Maringá.

Dados Internacionais de Catalogação-na-Publicação (CIP)
(Biblioteca Central Maria Grazia Zolet - UEM, Maringá - PR, Brasil)

S165m

Salinas, Heitor Delivio Sanches

Modelagem probabilística para rastreabilidade em cadeias produtivas / Heitor Delivio Sanches Salinas. -- Maringá, PR, 2026.
80 f. : il. color., figs., tabs.

Orientador: Prof. Dr. George Lucas Moraes Pezzott.

Coorientadora: Profa. Dra. Francielle Cristina Fenerich.

Dissertação (mestrado) - Universidade Estadual de Maringá, Centro de Tecnologia, Departamento de Engenharia de Produção, Programa de Pós-Graduação em Engenharia de Produção, 2026.

1. Rastreabilidade probabilística. 2. Cadeias produtivas complexas. 3. Teorema de Bayes. 4. Cadeias de Markov. 5. Segurança alimentar. I. Pezzott, George Lucas Moraes, orient. II. Fenerich, Francielle Cristina, coorient. III. Universidade Estadual de Maringá. Centro de Tecnologia. Departamento de Engenharia de Produção. Programa de Pós-Graduação em Engenharia de Produção. IV. Título.

CDD 23.ed. 658.7

Modelagem Probabilística para Rastreabilidade em Cadeias Produtivas

RESUMO

Este estudo investiga o uso de modelos probabilísticos para aprimorar a rastreabilidade em cadeias produtivas complexas, com foco na modelagem de incertezas e no suporte à tomada de decisão em cenários de eventos adversos, como contaminações. Inicialmente, é proposta e aplicada uma abordagem baseada no Teorema de Bayes e em Cadeias de Markov para realizar rastreabilidade direta e reversa em uma cadeia produtiva alimentícia real, permitindo a reconstrução probabilística dos fluxos de materiais e a identificação de fontes potenciais de contaminação. Os resultados demonstram que a abordagem probabilística é capaz de representar adequadamente a dinâmica de mistura e recombinação de lotes, fornecendo não apenas a identificação de fontes prováveis, mas também a quantificação explícita da incerteza associada à inferência. Em seguida, o estudo avança ao propor uma extensão metodológica para cenários com múltiplas fontes simultâneas de contaminação, incorporando a modelagem probabilística da propagação por meio de funções dose-resposta não lineares, simulação Monte Carlo e algoritmos de inferência baseados na detecção de padrões estruturais e similaridade entre cenários observados e simulados. Os resultados indicam elevada capacidade de detecção de contaminação, boa performance na distinção entre cenários de fonte única e duas fontes e alta eficiência na identificação das origens, reduzindo significativamente o número de hipóteses a serem investigadas na prática. Do ponto de vista teórico, o trabalho contribui ao integrar probabilidade condicional, cadeias de Markov e modelagem de mistura multiperíodo em um único arcabouço, posicionando o problema de rastreabilidade como um problema inverso estocástico em redes produtivas. Sob a ótica prática, a abordagem proposta oferece suporte à tomada de decisão em contextos industriais, permitindo priorização de investigações, redução de custos e maior eficiência em processos de rastreabilidade e recall.

Palavras-chave: Rastreabilidade probabilística; Cadeias produtivas complexas; Teorema de Bayes; Cadeias de Markov; Múltiplas fontes de contaminação.

Probabilistic Modeling for Supply Chain Traceability

ABSTRACT

This study investigates the use of probabilistic models to enhance traceability in complex production chains, with a focus on uncertainty modeling and decision support in adverse event scenarios such as contamination. Initially, a methodology based on Bayes' Theorem and Markov Chains is proposed and applied to perform forward and backward traceability in a real food production chain, enabling the probabilistic reconstruction of material flows and the identification of potential contamination sources. The results show that the probabilistic approach effectively captures the dynamics of mixing and recombination of batches, providing not only the identification of likely sources but also an explicit quantification of the uncertainty associated with the inference. Subsequently, the study advances by proposing a methodological extension for scenarios involving multiple simultaneous contamination sources, incorporating probabilistic propagation modeling through nonlinear dose-response functions, Monte Carlo simulation, and inference algorithms based on structural pattern detection and similarity between observed and simulated scenarios. The results indicate high contamination detection capability, good performance in distinguishing between single-source and multiple-source scenarios, and high efficiency in identifying contamination origins, significantly reducing the number of hypotheses to be investigated in practice. From a theoretical perspective, the work contributes by integrating Bayesian inference, Markov Chains, and multi-period mixing modeling into a unified framework, positioning traceability as an inverse stochastic problem in production networks. From a practical standpoint, the proposed approach supports industrial decision-making by enabling prioritization of investigations, cost reduction, and improved efficiency in traceability and recall processes.

Keywords: *Probabilistic traceability; Complex production chains; Bayesian Theorem; Markov chains; Multiple contamination sources.*

SUMÁRIO

INTRODUÇÃO	9
1.1 CONTEXTUALIZAÇÃO	9
1.2 PROBLEMÁTICA	11
1.3 JUSTIFICATIVA E CONTRIBUIÇÕES	12
1.4 OBJETIVOS DE PESQUISA.....	13
1.4.1 Objetivos específicos	13
1.5 DESENHO DA PESQUISA.....	14
ARTIGO 1.....	16
2.1 INTRODUÇÃO	17
2.2 FUNDAMENTAÇÃO TEÓRICA.....	18
2.2.1 Rastreabilidade em produtos alimentícios	18
2.2.2 Probabilidade bayesiana e cadeias de markov.....	19
2.3 METODOLOGIA.....	20
2.4 ANÁLISE E RESULTADOS.....	23
2.5 CONCLUSÕES.....	27
2.6 REFERÊNCIAS BIBLIOGRÁFICAS	28
ARTIGO 2.....	30
3.1 INTRODUÇÃO	31
3.2 ENQUADRAMENTO DO PROBLEMA E POSICIONAMENTO TEÓRICO ..	35
3.2.1 Características estruturais do problema	35
3.2.2 Trabalhos relacionados	37
3.3 MODELAGEM PROBABILÍSTICA DA PROPAGAÇÃO DE	
CONTAMINAÇÃO.....	39
3.3.1 Interpretação dos limiares e do parâmetro de inclinação da função dose–resposta	
42	
3.3.2 Algoritmo 1 - Método de detecção de múltiplas fontes baseado em “vales”	42
3.3.3 Algoritmo 2 – Identificação das fontes de contaminação por similaridade entre	
vetor observado e mapa de respostas prováveis	47
3.4 SIMULAÇÕES COM DADOS DE UM CENÁRIO REAL	53
3.4.1 Descrição da cadeia produtiva e do estudo de caso.....	53

3.4.2	Definição dos cenários de contaminação.....	54
3.4.3	Implementação computacional	55
3.5	RESULTADOS E DISCUSSÕES	56
3.5.1	Desempenho do algoritmo 1 (detecção de múltiplas fontes).....	56
3.5.2	Desempenho do algoritmo 2 (detecção da origem da contaminação	62
3.5.3	Implicações práticas para a gestão industrial.....	66
3.6	CONCLUSÃO.....	67
3.7	REFERÊNCIAS	68
	CONSIDERAÇÕES FINAIS.....	72
	REFERÊNCIAS	75

INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

A rastreabilidade pode ser definida como a capacidade de rastrear o histórico, a aplicação ou a localização de um produto ao longo da cadeia produtiva, permitindo acompanhar matérias-primas, lotes intermediários e produtos finais ao longo das etapas de produção e distribuição (ISO 9000:2015; OLSEN; BORIT, 2013). Em cadeias produtivas complexas, essa capacidade torna-se ainda mais relevante, pois os fluxos de materiais podem envolver múltiplos processos, diferentes fornecedores, mistura de insumos e redistribuição de lotes ao longo do tempo (REGATTIERI; GAMBERI; MANZINI, 2007; CHRISTOPHER, 2016).

Cadeias produtivas complexas são entendidas como sistemas compostos por processos interdependentes, nos quais matérias-primas e lotes intermediários podem ser transformados, misturados, fracionados e recombinados até a obtenção do produto final. Essa configuração amplia a dificuldade de rastrear a origem de um evento adverso, uma vez que a composição final de um produto pode resultar da contribuição simultânea de diferentes lotes e etapas produtivas (IVANOV; DOLGUI; SOKOLOV, 2017; HOSSEINI; IVANOV, 2020).

Em setores como o alimentício, farmacêutico e biotecnológico, a rastreabilidade é essencial para garantir segurança, qualidade, conformidade regulatória e capacidade de resposta a eventos críticos, como contaminações e recalls (BADIA-MELIS; MISHRA; RUIZ-GARCÍA, 2015; RINGSBERG, 2014). Nesses contextos, a identificação rápida da origem de uma contaminação permite reduzir impactos à saúde dos consumidores, perdas econômicas, danos reputacionais e

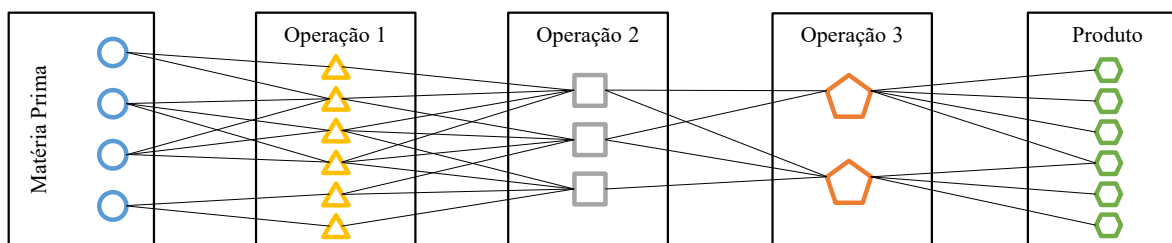
custos associados à investigação de lotes potencialmente afetados (REGATTIERI; GAMBERI; MANZINI, 2007; HORN; FRIEDRICH, 2019).

Apesar de sua importância, a rastreabilidade em cadeias produtivas com mistura de materiais apresenta limitações quando tratada apenas por abordagens determinísticas. Em muitos casos, não há uma relação direta e única entre uma matéria-prima e um produto final, pois diferentes caminhos de propagação podem conectar uma mesma origem a diversos produtos, assim como diferentes origens podem contribuir para um mesmo produto final (DABBENE; GAY; TORTIA, 2014; LU *et al.*, 2019).

Esse desafio torna-se ainda mais relevante quando se considera a possibilidade de múltiplas fontes simultâneas de contaminação. Em cenários de fonte única, a investigação busca identificar o lote ou matéria-prima mais provável de ter originado o evento; entretanto, quando duas ou mais fontes podem estar envolvidas, o espaço de hipóteses cresce de forma combinatória, dificultando a identificação da origem e aumentando a incerteza associada à tomada de decisão (ALTARELLI *et al.*, 2014; CHAI; WANG; ZHU, 2021).

A Figura 1 apresenta uma representação esquemática de uma cadeia produtiva rastreável, na qual diferentes matérias-primas e lotes intermediários percorrem sucessivas etapas de transformação até a obtenção dos produtos finais. As conexões entre os elementos indicam que um mesmo insumo pode contribuir para diferentes lotes ao longo do processo, assim como um produto final pode ser resultado da combinação de múltiplas origens. Essa estrutura ilustra a complexidade da rastreabilidade em sistemas produtivos com mistura, fracionamento e recombinação de materiais, nos quais a identificação da origem de um evento adverso depende do registro das relações entre lotes e da modelagem das contribuições ao longo da cadeia.

Figura 1 – Representação esquemática de uma cadeia produtiva com múltiplas etapas, fluxos de materiais e relações de rastreabilidade entre lotes.



Fonte: (Autor, 2026)

Nesse contexto, métodos probabilísticos oferecem uma alternativa para representar a incerteza inerente à propagação de materiais e à identificação de fontes em cadeias produtivas

complexas. Ao invés de fornecer apenas vínculos determinísticos entre lotes, esses métodos permitem estimar probabilidades, simular cenários e apoiar a priorização de hipóteses de investigação em situações de contaminação observada nos produtos finais (JAYNES; BRETTTHORST, 2003; HOSSEINI; IVANOV, 2020).

1.2 PROBLEMÁTICA

A identificação da origem de uma contaminação em cadeias produtivas complexas constitui um problema desafiador, especialmente quando os produtos finais resultam da mistura de diferentes matérias-primas e lotes intermediários. Nessas condições, a observação de contaminação no produto final não permite, por si só, determinar de forma direta qual matéria-prima originou o evento, exigindo métodos capazes de lidar com incerteza, múltiplos caminhos de propagação e sobreposição de contribuições (DABBENE; GAY; TORTIA, 2014; LU *et al.*, 2019).

O problema torna-se ainda mais complexo quando se admite a possibilidade de duas fontes simultâneas de contaminação. Nesse caso, diferentes combinações de matérias-primas podem gerar padrões semelhantes nos produtos finais, fazendo com que a investigação deixe de ser apenas a busca por uma única origem e passe a envolver a distinção entre cenários de fonte única e fonte dupla, além da identificação das combinações mais prováveis de origem (ALTARELLI *et al.*, 2014; CHAI; WANG; ZHU, 2021).

Assim, o problema de pesquisa desta dissertação pode ser formulado da seguinte maneira: **como desenvolver uma abordagem probabilística capaz de identificar se uma contaminação observada em produtos finais é compatível com uma única fonte ou com duas fontes simultâneas e, a partir disso, indicar quais matérias-primas devem ser priorizadas como origens mais prováveis do evento?** Essa formulação está alinhada à compreensão da rastreabilidade como um problema inverso estocástico em redes produtivas, no qual diferentes hipóteses causais podem explicar um mesmo padrão observado (DE SANCTIS *et al.*, 2008; CAI; YE, 2021).

Para responder a esse problema, a dissertação utiliza inicialmente a modelagem probabilística baseada no Teorema de Bayes e em Cadeias de Markov para representar as relações entre matérias-primas, processos intermediários e produtos finais. Em seguida, propõe uma extensão metodológica para cenários com múltiplas fontes potenciais de contaminação, incorporando função dose-resposta, simulação Monte Carlo e algoritmos voltados à classificação do número de fontes e à identificação das origens mais prováveis (JAYNES; BRETTTHORST, 2003; BROEMELING, 2018; SALINAS *et al.*, 2024).

1.3 JUSTIFICATIVA E CONTRIBUIÇÕES

A identificação rápida e confiável da origem de uma contaminação é fundamental para apoiar decisões industriais relacionadas ao bloqueio de lotes, investigação de matérias-primas, comunicação com fornecedores, definição de ações corretivas e realização de recalls. Quanto maior a incerteza sobre a origem do evento, maior tende a ser o número de lotes investigados ou retidos, o que pode ampliar custos operacionais e impactos econômicos para a organização (BADIA-MELIS; MISHRA; RUIZ-GARCÍA, 2015; HORN; FRIEDRICH, 2019).

Em cadeias produtivas caracterizadas por mistura, fracionamento e recombinação de lotes, a rastreabilidade precisa considerar não apenas a existência de vínculos entre matérias-primas e produtos finais, mas também a intensidade dessas contribuições e a incerteza associada à propagação da contaminação. Por esse motivo, abordagens probabilísticas são justificadas por permitirem transformar registros de produção e proporções de mistura em informações quantitativas para apoio à tomada de decisão (REGATTIERI; GAMBERI; MANZINI, 2007; HOSSEINI; IVANOV, 2020).

A principal contribuição desta dissertação está na proposição de uma abordagem probabilística voltada à identificação de fontes potenciais de contaminação em cadeias produtivas complexas. Diferentemente de abordagens centradas apenas em rastrear vínculos determinísticos ou em assumir uma única origem causal, o trabalho considera explicitamente a possibilidade de duas fontes simultâneas, ampliando a aplicabilidade da rastreabilidade para situações mais próximas da complexidade observada em ambientes industriais reais (ALTARELLI *et al.*, 2014; CHAI; WANG; ZHU, 2021).

Do ponto de vista acadêmico, a pesquisa contribui ao integrar modelagem probabilística, Cadeias de Markov, Teorema de Bayes, função dose-resposta e simulação Monte Carlo em um arcabouço aplicado à rastreabilidade de cadeias produtivas com mistura multiperíodo. Essa integração permite tratar a rastreabilidade como um problema inverso estocástico, no qual a origem mais provável de um evento é inferida a partir de padrões observados nos produtos finais (JAYNES; BRETTHORST, 2003; BROEMELING, 2018; DE SANCTIS *et al.*, 2008).

Do ponto de vista metodológico, o trabalho propõe dois algoritmos complementares no Capítulo 3. O primeiro busca identificar se o padrão de contaminação observado é compatível com uma ou duas fontes, com base na estrutura do vetor de resposta dos produtos finais. O segundo utiliza um mapa de respostas prováveis, construído por simulação Monte Carlo, para comparar cenários observados e simulados, permitindo priorizar matérias-primas ou combinações de

matérias-primas mais prováveis de terem originado o evento (HORN; FRIEDRICH, 2019; LU *et al.*, 2019).

Do ponto de vista prático, a abordagem proposta oferece suporte à tomada de decisão em ambientes industriais, pois permite reduzir o número de hipóteses investigadas, priorizar lotes ou matérias-primas mais prováveis e orientar ações de investigação, bloqueio ou *recall* com maior eficiência. Assim, o método contribui para tornar a rastreabilidade mais operacional e quantitativa, especialmente em contextos nos quais múltiplas fontes podem estar envolvidas e a decisão precisa ser tomada sob incerteza (BADIA-MELIS; MISHRA; RUIZ-GARCÍA, 2015; HOSSEINI; IVANOV, 2020).

1.4 OBJETIVOS DE PESQUISA

O objetivo geral desta pesquisa é propor uma abordagem probabilística para rastreabilidade em cadeias produtivas complexas, com foco na identificação do número de fontes e das origens mais prováveis de contaminação em sistemas caracterizados por mistura, transformação e recombinação de lotes ao longo do processo produtivo.

1.4.1 Objetivos específicos

Para alcançar o objetivo geral, definem-se os seguintes objetivos específicos:

- a) Revisar a literatura sobre rastreabilidade em cadeias produtivas complexas, com ênfase nos desafios associados à incerteza, à mistura de lotes e à identificação de fontes de contaminação.
- b) Aplicar métodos probabilísticos, como o Teorema de Bayes e Cadeias de Markov, para modelar a rastreabilidade direta e reversa em uma cadeia produtiva alimentícia real, conforme desenvolvido no Capítulo 2.
- c) Propor uma metodologia probabilística para avaliar se um padrão de contaminação observado em produtos finais é compatível com uma única fonte ou com duas fontes simultâneas de contaminação.
- d) Desenvolver um algoritmo para identificar, por similaridade entre cenários observados e simulados, as matérias-primas ou combinações de matérias-primas mais prováveis de terem originado a contaminação.

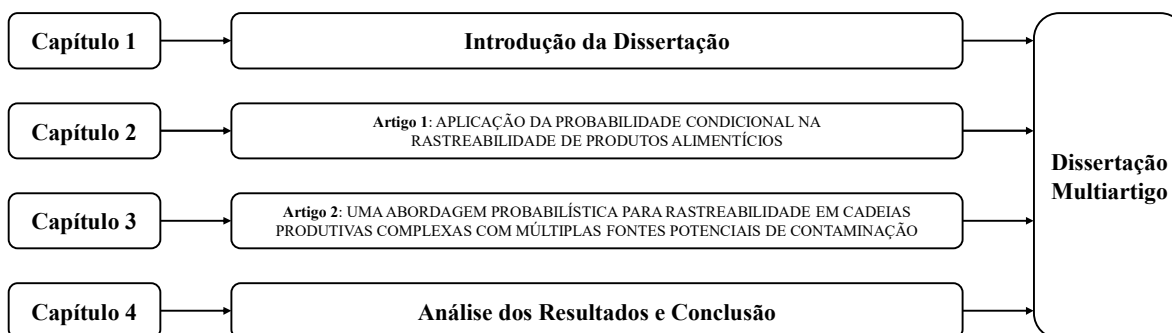
- e) Discutir as contribuições teóricas, metodológicas e práticas da abordagem proposta para apoiar a tomada de decisão em processos de rastreabilidade, investigação de contaminação e priorização de ações corretivas.

1.5 DESENHO DA PESQUISA

Esta dissertação está estruturada no formato multiartigo, integrando uma aplicação metodológica inicial e uma extensão voltada à identificação de múltiplas fontes potenciais de contaminação. O trabalho busca demonstrar como métodos probabilísticos podem apoiar a rastreabilidade em cadeias produtivas complexas, especialmente em contextos nos quais há mistura de lotes, propagação incerta da contaminação e necessidade de tomada de decisão sob incerteza.

A dissertação é composta por quatro capítulos, organizados de forma lógica e sequencial, conforme descrito a seguir e na Figura 2.

Figura 2 – Desenho da Pesquisa



Fonte: (Autor, 2026)

Capítulo 1 – Introdução:

Este capítulo apresenta o contexto da pesquisa, abordando a relevância da rastreabilidade em cadeias produtivas complexas, os desafios relacionados à integração de dados e à gestão de incertezas, bem como a motivação para o uso de métodos probabilísticos. São definidos o problema de pesquisa, o objetivo geral, os objetivos específicos e as contribuições acadêmicas e práticas do estudo, estabelecendo o fio condutor que orienta os capítulos subsequentes.

Capítulo 2 – Artigo 1:

O segundo capítulo corresponde ao primeiro artigo científico, no qual é apresentada a aplicação de métodos probabilísticos, com ênfase no teorema de Bayes e em cadeias de Markov,

para a modelagem da rastreabilidade em uma cadeia produtiva do setor alimentício. O capítulo descreve o desenvolvimento do modelo, a abordagem metodológica adotada e a análise dos resultados obtidos, evidenciando o potencial da inferência probabilística para apoiar a rastreabilidade direta e reversa ao longo do processo produtivo.

Capítulo 3 – Artigo 2:

O terceiro capítulo apresenta o segundo artigo, no qual é proposta uma abordagem probabilística para identificação de múltiplas fontes potenciais de contaminação. São desenvolvidos dois algoritmos: um voltado à distinção entre cenários de fonte única e dupla, e outro voltado à identificação das fontes mais prováveis por similaridade entre o vetor observado e um mapa de respostas simuladas.

Capítulo 4 – Análise integrada dos resultados e conclusões:

O último capítulo apresenta uma análise integrada dos resultados obtidos no Artigo 1 e das contribuições metodológicas desenvolvidas no Artigo 2. São discutidas as implicações teóricas e práticas da abordagem proposta, bem como suas limitações. O capítulo encerra a dissertação com as conclusões gerais do estudo e a indicação de possibilidades de aplicação e desdobramentos futuros da pesquisa.

ARTIGO 1

APLICAÇÃO DA PROBABILIDADE CONDICIONAL NA RASTREABILIDADE DE PRODUTOS ALIMENTÍCIOS

Resumo

A rastreabilidade desempenha um papel importante na gestão da segurança alimentar, especialmente face a eventos adversos. Por outro lado, a teoria das probabilidades oferece uma medida única de incerteza, integrando dados empíricos com conhecimento especializado para estimar a probabilidade de eventos futuros e, com a modelagem via cadeias de Markov, fornece informações valiosas sobre a evolução temporal das sequências de lotes na cadeia produtiva, facilitando a compreensão das transições entre estados e contribuindo para decisões informadas em situações críticas. Neste contexto, este artigo propõe a aplicação da teoria de probabilidades para analisar a rastreabilidade na indústria alimentícia. O estudo atinge seus objetivos utilizando cálculos de probabilidade e desenvolvendo uma árvore de rastreabilidade. A metodologia é aplicada utilizando o software RStudio em um conjunto de dados reais de uma indústria alimentícia e, como resultado, apresenta-se graficamente a cadeia de lotes, juntamente com as matrizes de transição para cada processo, a qual oferece uma visualização clara da rastreabilidade direta e reversa ao longo do tempo/processos.

Palavras-chave: Rastreabilidade, Teorema de Bayes, Cadeias de Markov.

2.1 INTRODUÇÃO

A rastreabilidade dos alimentos tem o potencial de melhorar a proteção do consumidor, concentrando-se no recolhimento de produtos, na remoção de produtos não consumíveis e na investigação da segurança alimentar. Isto é essencial para garantir a segurança alimentar, manter a qualidade, proteger contra ameaças e satisfazer as necessidades da cadeia de abastecimento (BADIA-MELIS *et al.*, 2015). Um sistema eficiente que forneça informações precisas e oportunas em toda a cadeia de abastecimento pode reduzir significativamente os custos operacionais e aumentar a produtividade (REGATTIERI *et al.*, 2007).

Para estimar a contribuição de diferentes lotes ao longo de uma cadeia produtiva, é necessária uma representação matemática capaz de descrever as relações entre etapas sucessivas do processo e quantificar a incerteza associada à rastreabilidade. Nesse contexto, a probabilidade condicional fornece uma estrutura adequada para representar a dependência entre lotes de entrada, lotes intermediários e produtos finais, permitindo estimar a probabilidade de um determinado lote ter contribuído para a composição de outro. O Teorema de Bayes, por sua vez, possibilita realizar a rastreabilidade reversa, isto é, inferir as origens mais prováveis de um produto final a partir das evidências observadas ao longo da cadeia produtiva (JAYNES; BRETTHORST, 2003; ACZEL; SOUNDERPANDIAN, 2009).

Complementarmente, as Cadeias de Markov permitem representar a evolução sequencial dos lotes entre processos produtivos, assumindo que a transição para uma etapa posterior depende do estado imediatamente anterior. Essa propriedade é adequada para modelar cadeias produtivas compostas por etapas sucessivas, nas quais as proporções de uso dos lotes em cada processo podem ser organizadas em matrizes de transição. Assim, a combinação entre probabilidade condicional, Teorema de Bayes e Cadeias de Markov oferece uma base metodológica para modelar a rastreabilidade direta e reversa em sistemas produtivos com mistura e transformação de lotes (DEGROOT; SCHERVISH, 2012; BROEMELING, 2018).

Este estudo tem como objetivo utilizar métodos probabilísticos, como o teorema de Bayes e cadeias de Markov, para realizar a rastreabilidade na indústria alimentícia. O objetivo é agilizar a tomada de decisão em eventos críticos que exigem coleta e rastreamento de lotes durante o processo produtivo, como contaminações identificadas, variações inesperadas de qualidade e outros incidentes relevantes.

O presente artigo está estruturado em 5 seções. Após a introdução, na seção 2.1, é apresentada a fundamentação teórica, abordando inicialmente o conceito de rastreabilidade em produtos alimentícios e sua importância na segurança alimentar. Em seguida, são discutidos o

Teorema de Bayes e as Cadeias de Markov como ferramentas probabilísticas fundamentais para análise de rastreabilidade na indústria alimentícia, na seção 2.2. A seção 2.3 detalha a metodologia adotada, incluindo a descrição dos dados utilizados, os cálculos de probabilidades condicionais e a construção da árvore de rastreabilidade genérica. Os resultados obtidos e sua análise são apresentados na seção 2.4, destacando as visualizações gráficas da rastreabilidade direta e reversa, bem como as conclusões derivadas desses resultados. Por fim, na seção 2.5, são apresentadas as conclusões finais do estudo, incluindo uma discussão sobre as contribuições teóricas e práticas do trabalho, bem como sugestões para pesquisas futuras na área.

2.2 FUNDAMENTAÇÃO TEÓRICA

2.2.1 Rastreabilidade em produtos alimentícios

A segurança alimentar é um dos principais problemas do mundo atual (HAN *et al.*, 2015). Quando ocorrem escândalos de segurança alimentar, como contaminação de alimentos por bactérias, vírus, parasitas, produtos químicos e/ou toxinas, a gestão da rastreabilidade na cadeia de abastecimento alimentar torna-se mais relevante, pois esses eventos impactam diretamente as empresas e a saúde humana (MEUWISSEN *et al.*, 2003).

A ISO 9000/2015 define rastreabilidade como a capacidade de rastrear o histórico, aplicação ou localização de um objeto. Ao considerar um produto ou serviço, a rastreabilidade pode referir-se a: origem dos materiais e peças, histórico de processamento, distribuição e localização do produto ou serviço após a entrega. Segundo Ragattieri *et al.* (p. 347, 2007), “rastreabilidade” é um conceito relativo a todos os produtos e todos os tipos de cadeia de abastecimento. Num sistema econômico em que as empresas competem entre si num ambiente largamente baseado na satisfação do cliente, a rastreabilidade é um instrumento indispensável na obtenção do consenso do mercado. Segundo Ringsberg (2014), para mitigar os impactos das perturbações na segurança alimentar, as empresas e autoridades da cadeia de suprimento alimentar estão conscientes da necessidade de garantir a rastreabilidade dos alimentos, e a sua gestão tem ganhado cada vez mais importância nas regulamentações. Segundo Alfaro e Rábade (2008), a rastreabilidade é um requisito na indústria alimentar.

Assim sendo, é de suma importância expandir tecnologias que garantam a segurança alimentar de toda a cadeia de abastecimento alimentar, e um sistema de rastreabilidade alimentar que seja capaz de monitorar toda a vida do ciclo alimentar, incluindo produção, processamento, transporte, armazenamento e vendas de alimentos (LIN *et al.*, 2019). Nos últimos anos, muitas

soluções foram propostas com diferentes tecnologias e técnicas emergentes para melhorar a rastreabilidade de animais, plantas e produtos alimentares (SOUPRAYEN *et al.*, 2021). Uma delas é utilizar os princípios Bayesianos e a Cadeia de Markov.

2.2.2 Probabilidade bayesiana e cadeias de markov

A abordagem bayesiana para mensuração de probabilidade, conforme discutida por Jaynes e Bretthorst (2003), destaca-se por ser uma medida da incerteza relacionada a um evento, considerando as evidências disponíveis. Representa a probabilidade de ocorrência de um evento, dado que determinadas circunstâncias já ocorreram. Essa perspectiva oferece aos estatísticos a capacidade única de integrar dados empíricos com conhecimento especializado, utilizando o teorema de Bayes para calcular a probabilidade de uma hipótese específica diante de determinadas evidências (ACZEL e SOUNDERPANDIAN, 2009).

A manipulação de probabilidades na abordagem bayesiana, destacada por Aragón *et al.* (2016), envolve tratá-las como funções de variáveis, seguindo as regras do cálculo de probabilidades. O teorema da probabilidade condicional é a regra fundamental da estatística Bayesiana e fornece a base para o cálculo das probabilidades de eventos, amplamente conhecido como Teorema de Bayes.

Sejam os eventos A_i , $i = 1, 2, \dots$, uma partição do espaço amostral Ω , e B um evento de probabilidade não nula. Dada a ocorrência do evento B, a probabilidade do evento A_i também ter ocorrido é calculado pela fórmula de Bayes como:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)} \quad \text{Eq. 1}$$

Ainda segundo Paulino *et al.* (2019), o Teorema de Bayes continua a ser considerado por muitos como um dos poucos resultados matemáticos que busca descrever o processo de aprendizagem através da experiência. Segundo o autor, esse processo refere-se à mudança na postura inicial em relação a estados, causas ou hipóteses após a obtenção de informações iniciais sobre a ocorrência de determinados eventos (com base em dados resultantes da experiência ou observação). Por outro lado, a cadeia de Markov é um processo estocástico discreto que modela uma sequência de variáveis aleatórias (BROEMELING, 2018). Sua dependência se restringe à observação mais recente, distinguindo-a de cadeias de eventos convencionais, onde a probabilidade está condicionada à sequência completa a partir do valor inicial. Ao integrar a cadeia de Markov no contexto da probabilidade Bayesiana, torna-se interessante considerar como a evolução da cadeia

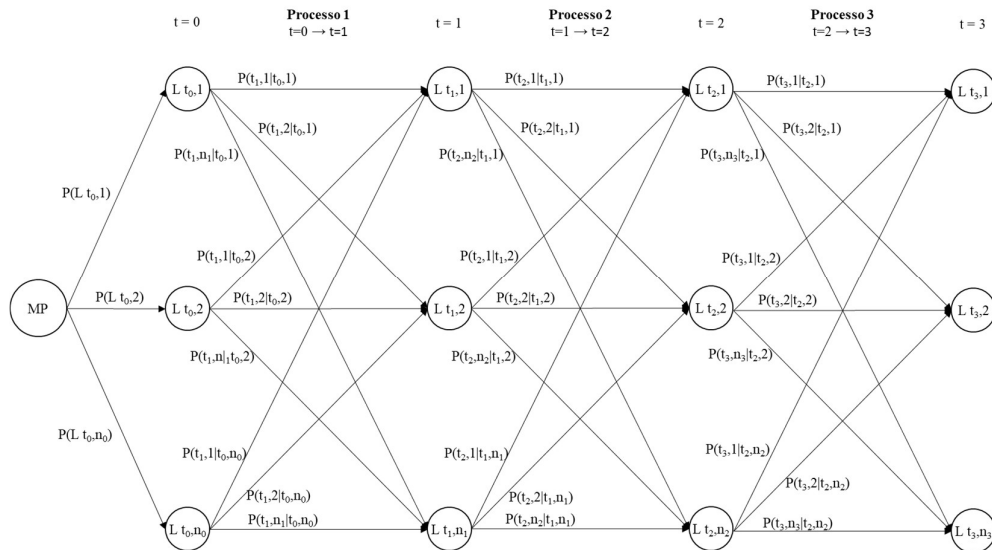
ao longo do tempo pode influenciar crenças e inferências sobre os estados futuros da sequência.

2.3 METODOLOGIA

Este estudo utiliza uma base de dados de uma indústria alimentícia, a qual recebe uma determinada matéria-prima MP e a processa em três diferentes etapas, gerando diferentes lotes ao longo do processamento. A partir da base de dados disponível e utilizando um conceito genérico de rastreabilidade condicional, seguindo o teorema de Bayes, construirão-se os vetores de probabilidade para as matérias primas utilizadas no processamento e as matrizes de transição das cadeias de Markov, a fim de determinar uma árvore de rastreabilidade genérica para uma cadeia produtiva de alimentos utilizando conceitos de probabilidade condicional e aplicar a árvore de probabilidade nos dados da indústria de alimentos.

Considere que tenhamos n_0 matérias primas iniciais e cada processo t produz n_t lotes utilizando lotes produzidos pelo processo anterior ($t - 1$). A identificação dos lotes é dada por Lote t,i , onde t é a classificação do processo ($t = 1,2,3$) e i é a classificação do Lote dentro do processo t ($i = 1,2, \dots, n_t$). A Figura 3 mostra como se dá o desenvolvimento do processo desde a chegada da matéria-prima até a distribuição de lotes.

Figura 3 - Representação gráfica da cadeia de lotes



Fonte: Autores (2024)

Aplicando conceitos de probabilidade, podemos traçar um paralelo entre probabilidade condicional e rastreabilidade. Primeiramente, definimos as proporções iniciais (π_i^0) da matéria prima i :

$$\pi_i^0 = \frac{\text{Total da matéria prima inicial } i}{\text{Total das } n_0 \text{ matérias primas}}, \quad i = 1, 2, \dots, n_0. \quad \text{Eq. 2}$$

Seja $\pi_{i \rightarrow j}^{t \rightarrow t+1}$ a proporção do lote tipo i do processo t utilizado para produzir os lotes tipo j no processo $t + 1$, definida por:

$$\pi_{i \rightarrow j}^{t \rightarrow t+1} = \frac{\text{total do lote } i \text{ do processo } t \text{ utilizado para produzir o lote } j \text{ no processo } t + 1}{\text{total do lote } i \text{ do processo } t} \quad \text{Eq. 3}$$

para todo $i = 1, 2, \dots, n_t, j = 1, 2, \dots, n_{t+1}$ e $t = 1, 2$. De fato, é possível registrar as proporções acima somente em processos consecutivos. Em outras palavras, não é possível quantificar, por exemplo, a quantidade utilizada de cada lote de matéria prima inicial em cada lote produzido pelo processo 2 ou 3, pois existe o processo 1 entre as etapas iniciais e finais (como mostrado na Figura 3), implicando neste caso na impossibilidade de rastreabilidade direta e inversa. Desta forma, torna-se necessário mensurar as quantidades utilizadas de cada lote de matéria-prima nos diferentes lotes gerados em cada processo para tornar a cadeia produtiva descrita acima rastreável.

Seja $L_{0,i}$ a escolha da matéria prima inicial i ($i = 1, 2, \dots, n_0$), com $P(L_{0,i}) = \pi_i^0$ e $P(L_{t_2,j} | L_{t_1,i}) = \pi_{i \rightarrow j}^{t_1 \rightarrow t_2}$ sendo a probabilidade condicional de produzir o lote j no processo t_2 utilizando o lote i do processo t_1 , com $t_1 < t_2$.

A matriz de transição de primeira ordem pode ser definida por:

$$\bigwedge_{t \rightarrow t+1} = [\pi_{i \rightarrow j}^{t \rightarrow t+1}]_{n_t \times n_{t+1}} \quad \text{Eq. 4}$$

para $t = 0, 1, 2$, com:

$$\sum_{i=1}^{n_t} \sum_{j=1}^{n_{t+1}} \pi_{i \rightarrow j}^{t \rightarrow t+1} = n_t \quad \text{Eq. 5}$$

Observe que, utilizando a dependência condicional de primeira ordem na cadeia de Markov nas probabilidades condicionais acima, podemos calcular as probabilidades de transição em processos não sequenciais (de ordem superior). Primeiramente, vejamos que:

$$\begin{aligned} \pi_{i \rightarrow j}^{t \rightarrow t+2} &= P(L_{t+2,j} | L_{t,i}) = \sum_{l=1}^{n_{t+1}} P(L_{t+2,j} \cap L_{t+1,l} | L_{t,i}) \\ &= \sum_{l=1}^{n_{t+1}} P(L_{t+2,j} | L_{t+1,l} \cap L_{t,i}) P(L_{t+1,l} | L_{t,i}) \end{aligned} \quad \text{Eq. 6}$$

$$\begin{aligned}
&= \sum_{l=1}^{n_{t+1}} P(L_{t+2,j}|L_{t+1,l})P(L_{t+1,l}|L_{t,i}) \\
&= \sum_{l=1}^{n_{t+1}} \pi_{l \rightarrow j}^{t+1 \rightarrow t+2} \times \pi_{i \rightarrow l}^{t \rightarrow t+1}
\end{aligned}$$

Logo, a matriz de transição $t \rightarrow t + 2$ é:

$$\bigwedge_{t \rightarrow t+2} = \bigwedge_{t \rightarrow t+1} \left(\bigwedge_{t+1 \rightarrow t+2} \right)^{t'} \quad \text{Eq. 7}$$

o que resulta na rastreabilidade direta entre os processos $0 \rightarrow 2$ e $1 \rightarrow 3$. Agora, para $t \rightarrow t + 3$ temos:

$$\begin{aligned}
\pi_{i \rightarrow j}^{t \rightarrow t+3} &= P(L_{t+3,j} | L_{t,i}) = \sum_{l=1}^{n_{t+2}} P(L_{t+3,j} \cap L_{t+2,l} | L_{t,i}) \\
&= \sum_{l=1}^{n_{t+1}} P(L_{t+3,j} | L_{t+2,l}) P(L_{t+2,l} | L_{t,i}) \\
&= \sum_{l=1}^{n_{t+2}} \pi_{i \rightarrow l}^{t \rightarrow t+2} \times \pi_{l \rightarrow j}^{t+2 \rightarrow t+3}
\end{aligned} \quad \text{Eq. 8}$$

onde $\pi_{i \rightarrow l}^{t \rightarrow t+2}$ é obtido anteriormente. Na forma matricial, segue a matriz de transição:

$$\bigwedge_{t \rightarrow t+3} = \bigwedge_{t \rightarrow t+2} \left(\bigwedge_{t+2 \rightarrow t+3} \right)^{t'} \quad \text{Eq. 9}$$

o que resulta na rastreabilidade direta entre os Processos $0 \rightarrow 3$. De maneira geral, e utilizando o princípio da recorrência, podemos calcular para um número maior de processos utilizando a fórmula:

$$\pi_{i \rightarrow j}^{t \rightarrow t+k} = \sum_{l=1}^{n_{t+k}} \pi_{i \rightarrow l}^{t \rightarrow t+k-1} \pi_{l \rightarrow j}^{t+k-1 \rightarrow t+k} \quad \text{Eq. 10}$$

ou, reescrito como forma matricial, temos:

$$\bigwedge_{t \rightarrow t+k} = \bigwedge_{t \rightarrow t+k-1} \left(\bigwedge_{t+k-1 \rightarrow t+k} \right)^{t'} \quad \text{Eq. 11}$$

para $k = 2, 3, 4 \dots$

Definidas as matrizes de transição de probabilidades condicionais e rastreabilidade direta,

podemos também calcular a rastreabilidade reversa em termos de probabilidades condicionais.

Dado que selecionou-se um lote j no processo $t + k$, seja $\lambda_{i \leftarrow j}^{t \leftarrow t+k}$ a probabilidade condicional de ter usado o lote i do processo t na sua fabricação. Podemos calcular essa probabilidade utilizando novamente o Teorema de Bayes como se segue:

$$\lambda_{i \leftarrow j}^{t \leftarrow t+k} = P(L_{t,i} | L_{t+k,j}) = \frac{P(L_{t+k,j} | L_{t,i}) P(L_{t,i})}{P(L_{t+k,j})} = \frac{\pi_{j \rightarrow i}^{t \rightarrow t+k} \times \pi_i^t}{\pi_j^{t+k}} \quad \text{Eq. 12}$$

onde $\pi_i^t = P(L_{t,i})$, $i = 1, 2, \dots$ são as probabilidades marginais em cada processo t . Para $t = 0$ as probabilidades marginais (material prima inicial) foram previamente definidas pelo vetor:

$$\boldsymbol{\pi}_0 = [\pi_1^0, \pi_2^0, \pi_3^0, \dots, \pi_{n_0}^0] \quad \text{Eq. 13}$$

Para $t > 0$, de forma recorrente a partir do processo inicial $t = 0$, obtemos as demais marginais facilmente por:

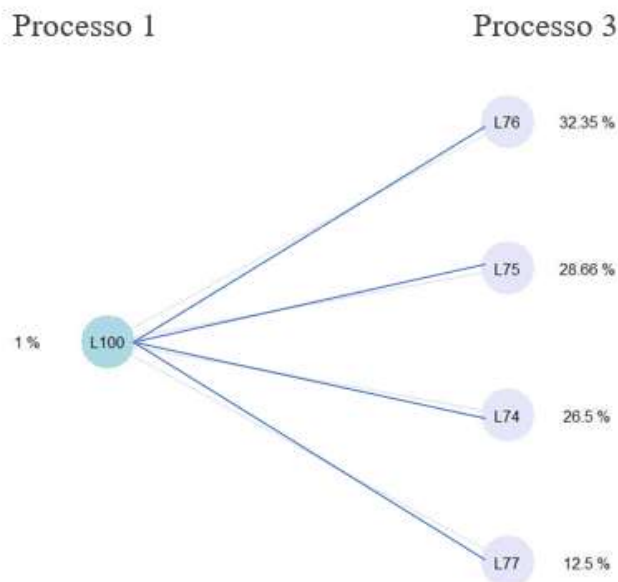
$$\pi_i^t = P(L_{t,i}) = \sum_{j=1}^{n_{t-1}} P(L_{t,i} | L_{j,t-1}) P(L_{j,t-1}) = \sum_{j=1}^{n_0} \pi_{j \rightarrow i}^{t-1 \rightarrow t} \times \pi_j^{t-1} \quad \text{Eq. 14}$$

para todo $t = 1, 2, \dots$

2.4 ANÁLISE E RESULTADOS

Aplicando os cálculos de probabilidades condicionais da seção 2.3 é possível realizar as rastreabilidades diretas e reversas, e assim obter a probabilidade em uma árvore, conforme exemplificado nas Figuras 4 e 5 a seguir. As figuras foram geradas após implementação do método no *software* RStudio.

Figura 4 - Rastreabilidade direta do lote L100

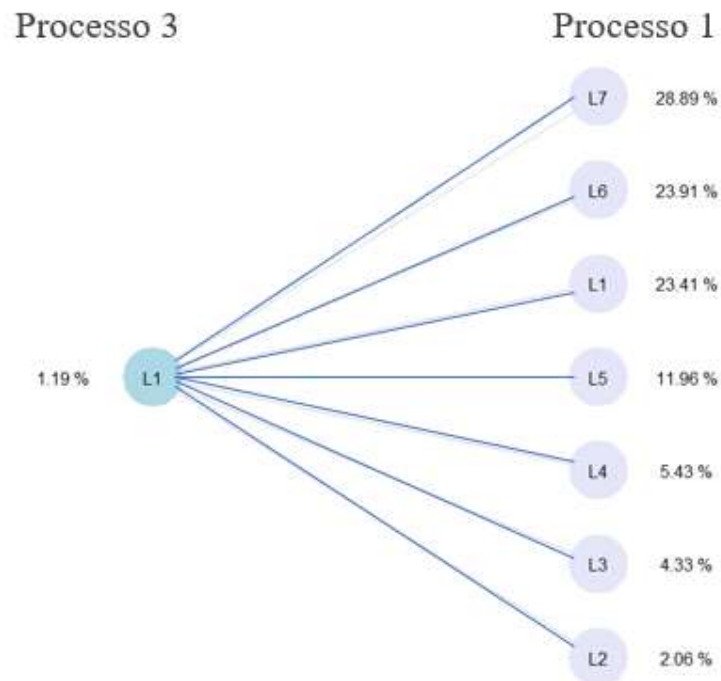


Fonte: Autores (2024)

A Figura 4 informa que, após o Processo 1, temos que o Lote 100 (L100) possui 1% de todos os produtos produzidos por este processo. Do lado direito da Figura, temos a rastreabilidade direta: 32,35% do Lote 100 do processo 1 foi utilizado para produzir o Lote 76 no Processo 3; 28,66% do Lote 100 do processo 1 foi utilizado para produzir o Lote 75 no Processo 3; 26,25% do Lote 100 do processo 1 foi utilizado para produzir o Lote 74 no Processo 3; e 12,50% do Lote 100 do processo 1 foi utilizado para produzir o Lote 77 no Processo 3.

A Figura 5 a seguir apresenta a rastreabilidade reversa. Após o Processo 3, temos que o Lote 1 (L1) possui 1,19% de todos os produtos produzidos por este processo. Do lado direito da Figura, temos a rastreabilidade inversa: 28,89% do Lote 1 do Processo 3 foi produzido utilizando o Lote 7 no Processo 1; 23,91% do Lote 1 do Processo 3 foi produzido utilizando o Lote 6 no Processo 1; 23,41% do Lote 1 do Processo 3 foi produzido utilizando o Lote 6 no Processo 1; e assim sucessivamente.

Figura 5 - Rastreabilidade reversa do lote L1



Fonte: Autores (2024)

Vale ressaltar a diferença das probabilidades na rastreabilidade direta e reversa utilizando como exemplo as figuras acima. Na probabilidade direta, fixa-se o Lote 100 no Processo 1 e calculamos a distribuição desse lote ao longo dos processos de produção, até se chegar no Processo 3. Já na rastreabilidade reversa, fixa-se o Lote 1 do Processo 3, e para este lote, calcula-se a proporção utilizada de cada lote do processo 1 na sua produção.

Os resultados obtidos a partir do modelo de rastreabilidade probabilística proposto evidenciam que a abordagem baseada em probabilidade condicional é capaz de reconstruir, de forma consistente, a contribuição relativa dos lotes ao longo da cadeia produtiva, mesmo em cenários caracterizados por mistura e recombinação de fluxos. A utilização de probabilidades condicionais associadas às matrizes de transição permite capturar a estrutura do processo produtivo e traduzir a propagação de materiais em distribuições probabilísticas interpretáveis, superando abordagens determinísticas tradicionais.

A análise das distribuições *a posteriori* associadas aos lotes de entrada indica que, mesmo na presença de múltiplos caminhos possíveis de propagação, a probabilidade tende a se concentrar em subconjuntos específicos de lotes. Esse comportamento sugere que a estrutura da cadeia produtiva impõe restrições relevantes à propagação de eventos adversos, reduzindo o espaço

efetivo de fontes plausíveis. Do ponto de vista prático, isso implica que a identificação da origem de um evento não depende exclusivamente da observação direta da contaminação, mas também da topologia do processo e das proporções de mistura entre lotes.

Um resultado relevante é que o modelo não fornece apenas um ranqueamento ordinal das fontes mais prováveis, mas uma distribuição probabilística completa, preservando informação sobre a dispersão das probabilidades. Essa característica é particularmente importante em contextos industriais, nos quais decisões precisam ser tomadas sob incerteza. Diferentemente de abordagens baseadas em identificação binária da fonte, a distribuição probabilística permite avaliar o grau de confiança associado à inferência, oferecendo suporte mais robusto à tomada de decisão.

Sob a ótica gerencial, os resultados sugerem que a rastreabilidade probabilística pode apoiar estratégias mais eficientes de resposta a eventos críticos. Em vez de adotar medidas amplas e indiferenciadas, como recalls totais, o gestor passa a dispor de informações quantitativas para priorizar lotes, processos ou fornecedores que concentram maior probabilidade de envolvimento no evento adverso. Essa priorização permite reduzir custos operacionais e impactos logísticos, sem comprometer os requisitos de segurança e qualidade.

Além disso, os resultados indicam que a rastreabilidade não deve ser interpretada como uma propriedade absoluta, mas como uma característica dependente da estrutura da cadeia produtiva. Cadeias mais lineares e com menor grau de mistura tendem a apresentar menor incerteza na identificação da fonte, enquanto cadeias altamente interconectadas exibem maior dispersão probabilística. Essa constatação reforça a importância de considerar aspectos estruturais do processo produtivo no desenho e na avaliação de sistemas de rastreabilidade.

De forma geral, os resultados deste artigo demonstram que a abordagem probabilística proposta é adequada para capturar a dinâmica de propagação de materiais em cadeias produtivas complexas, oferecendo não apenas um mecanismo de identificação de fontes, mas também uma métrica quantitativa de incerteza. Esses achados estabelecem uma base sólida para extensões metodológicas mais realistas, como a consideração explícita de múltiplas fontes de contaminação, tema explorado no Artigo 2.

Do ponto de vista de aplicação prática, o modelo probabilístico proposto neste artigo é particularmente adequado para ser utilizado por gestores de qualidade, segurança e operações, bem como por equipes de engenharia de processos e *supply chain*, em indústrias de médio e grande porte cujos processos produtivos envolvem mistura, fracionamento e recombinação de lotes. Setores como alimentos e bebidas, biotecnologia, fermentação industrial, ingredientes especiais e químicos de base apresentam características operacionais compatíveis com o escopo do modelo, uma vez que operam com múltiplas etapas de transformação e dependem fortemente de registros

de consumo de insumos. Em termos de escala, a abordagem mostra-se especialmente apropriada para plantas industriais com rastreabilidade em nível de lote, nas quais dados históricos de produção e proporções de mistura estão disponíveis, permitindo a construção confiável das matrizes de transição. Nessas condições, o modelo oferece suporte quantitativo à priorização de investigações e ações corretivas, sendo aplicável tanto em análises pós-evento quanto como ferramenta de apoio à avaliação preventiva da rastreabilidade do processo produtivo.

Apesar das contribuições do modelo probabilístico apresentado neste artigo, algumas limitações devem ser reconhecidas. A abordagem assume, de forma implícita, que os eventos adversos observados podem ser explicados por uma única fonte dominante de contaminação, o que simplifica a inferência e facilita a interpretação dos resultados, mas pode não refletir adequadamente a complexidade de cadeias produtivas reais. Em sistemas caracterizados por múltiplos pontos de mistura, operações paralelas e variabilidade operacional, é plausível que mais de uma fonte contribua simultaneamente para os produtos finais afetados. Nessas situações, a análise individual de cada produto contaminado pode gerar inferências fragmentadas ou excessivamente dispersas, limitando o suporte à tomada de decisão gerencial. Além disso, o modelo não explora explicitamente mecanismos de agregação de evidências provenientes de múltiplas observações, o que restringe sua capacidade de capturar padrões recorrentes ao longo da cadeia produtiva. Essas limitações motivam a extensão metodológica proposta no Artigo 2, que avança ao incorporar a inferência probabilística de múltiplas fontes potenciais de contaminação, por meio da consolidação iterativa de evidências e da quantificação da incerteza associada à rastreabilidade da rede.

2.5 CONCLUSÕES

Este trabalho realizou a aplicação de métodos probabilísticos, como o teorema de Bayes e as cadeias de Markov, para aprimorar a rastreabilidade na indústria alimentícia. A rastreabilidade desempenha um papel importante na gestão da segurança alimentar, especialmente diante de eventos adversos, como contaminações identificadas, variações inesperadas na qualidade e outros incidentes relevantes.

A integração das cadeias de Markov ao contexto da probabilidade bayesiana oferece insights valiosos sobre a evolução temporal das sequências de lotes na cadeia produtiva. Isso não apenas facilita a compreensão das transições entre os estados, mas também contribui para a tomada de decisões informadas em situações críticas, como a necessidade de rastreamento de lotes em eventos inesperados.

A representação gráfica da cadeia de lotes, juntamente com as matrizes de transição para cada processo, proporciona uma visualização clara e compreensível da rastreabilidade ao longo do tempo. Isso não apenas simplifica a interpretação dos resultados, mas também fornece uma base sólida para análises futuras e melhorias nos processos de rastreabilidade.

Os objetivos do estudo foram atingidos com a determinação de uma árvore de rastreabilidade genérica e a aplicação prática dessa abordagem a dados reais de uma indústria alimentícia. As equações foram implementadas no software RStudio, demonstrando a viabilidade e eficácia da metodologia proposta.

Por fim, foi possível concluir que a aplicação de métodos probabilísticos, como a combinação de princípios bayesianos e cadeias de Markov, apresenta-se como uma ferramenta para aprimorar a rastreabilidade na indústria alimentícia. Esta pesquisa contribui não apenas para a compreensão teórica desses métodos, mas também para sua aplicação prática em ambientes industriais, visando aprimorar a segurança alimentar e a eficiência operacional. Essa abordagem tem o potencial de desempenhar um papel significativo na gestão de eventos críticos ao longo da cadeia produtiva de alimentos, promovendo uma indústria mais segura e confiável.

2.6 REFERÊNCIAS BIBLIOGRÁFICAS

ACZEL, A. D.; SOUNDERPANDIAN, J. *Business Statistics*. 7. ed. McGraw-Hill, 2009. p. 692–788.

ALFARO, J. A.; RÁBADE, L. A. **Traceability as a strategic tool to improve inventory management: A case study in the food industry**. *International Journal of Production Economics*, v. 118, n. 1, p. 104–110, 2008.

ARAGÓN, G. D.; ARANGO, F. O.; ARANDA, F. C. **Calculation of operational value at risk using Bayesian networks for a financial company**. *Contaduría y Administración*, v. 61, n. 1, p. 176–201, 2016. <https://doi.org/10.1016/j.cya.2015.09.009>

BADIA-MELIS, R.; MISHRA, P.; RUIZ-GARCÍA, L. **Food traceability: New trends and recent advances – A review**. *Food Control*, v. 57, p. 393–401, 2015. <https://doi.org/10.1016/j.foodcont.2015.05.005>

BERGONSE, R. V.; BIDARRA, J. M. **Bayesian probability and logistic regression in the assessment of susceptibility to large-scale fire occurrence**. *Finisterra*, v. XLV, n. 89, p. 79–104, 2010.

BROEMELING, L. D. *Bayesian Inference for Stochastic Processes*. Boca Raton: CRC Press, 2018.

DEGROOT, M. H.; SCHERVISH, M. J. *Probability and Statistics*. 4. ed. New York: Addison-Wesley, 2012.

EL-AWADY, A.; PONNAMBALAM, K. **Integration of simulation and Markov chains to support Bayesian networks for probabilistic failure analysis of complex systems.** *Reliability Engineering & System Safety*, v. 211, p. 107511, 2021. <https://doi.org/10.1016/j.ress.2021.107511>

HAN, W.; GU, Y.; WANG, W.; ZHANG, Y.; YIN, Y.; WANG, J.; ZHENG, L. R. **The design of an electronic pedigree system for food safety.** *Information Systems Frontiers*, v. 17, n. 2, p. 275–287, 2015.

ISO 9000:2015. *Quality Management Systems — Fundamentals and Vocabulary*. ABNT, 2015.
JAYNES, E. T.; BRETTTHORST, G. L. *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press, 2003.

LIN, X.; LI, J.; WU, J.; LIANG, H.; YANG, W. **Making knowledge tradable in edge-AI-enabled IoT: A consortium blockchain-based efficient and incentive approach.** *IEEE Transactions on Industrial Informatics*, v. 15, n. 12, p. 6367–6378, 2019.

MEUWISSEN, M. P. M.; VELTHUIS, A. G. J.; HOGVEEN, H.; HUIRNE, R. B. M. **Certification in meat supply chains.** *Journal of Agribusiness*, v. 21, n. 2, p. 167–181, 2003.

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

PAULINO, C. D.; TURKMAN, M. A.; MURTEIRA, B.; SILVA, G. L. *Estatística Bayesiana*. v. 2. Lisboa: Fundação Calouste Gulbenkian, 2019.

PENG, M.; ZHANG, L. M. **Analysis of human risks due to dam-break floods – Part 1: A new model based on Bayesian networks.** *Natural Hazards*, v. 64, p. 903–933, 2012. <https://doi.org/10.1007/s11069-012-0275-5>

REGATTIERI, A.; GAMBERI, M.; MANZINI, R. **Traceability of food products: General framework and experimental evidence.** *Journal of Food Engineering*, v. 81, n. 2, p. 347–356, 2007.

RINGSBERG, H. **Perspectives on food traceability: A systematic literature review.** *Supply Chain Management*, v. 19, n. 5/6, p. 558–576, 2014.

SOUPRAYEN, B.; AYYANAR, A.; K, S. J. **Improvement of C5.0 algorithm using Internet of Things with Bayesian principles for food traceability systems.** *Modern Supply Chain Research and Applications*, v. 3, n. 1, p. 2–23, 2021.

ARTIGO 2

UMA ABORDAGEM PROBABILÍSTICA PARA RASTREABILIDADE EM CADEIAS PRODUTIVAS COMPLEXAS COM MÚLTIPLAS FONTES POTENCIAIS DE CONTAMINAÇÃO

Resumo

Este artigo propõe uma abordagem probabilística para a rastreabilidade em cadeias produtivas complexas, com foco na identificação de fontes potenciais de contaminação em sistemas caracterizados por mistura e transformação de fluxos ao longo do tempo. Diferentemente de modelos tradicionais centrados em uma única origem, o estudo considera explicitamente a possibilidade de dupla fonte, ampliando o escopo da inferência em redes produtivas. A metodologia combina funções dose-resposta não lineares, simulação Monte Carlo e dois algoritmos complementares: o primeiro detecta padrões estruturais denominados “vales” para distinguir cenários de fonte única e dupla; o segundo utiliza um mapa de respostas prováveis e distância euclidiana para identificar as matérias-primas mais compatíveis com o padrão observado. A interpretação dos resultados considera a lógica FIFO (*First In, First Out*) como premissa operacional, uma vez que desvios dessa lógica podem gerar padrões não monotônicos e aumentar a ambiguidade entre cenários de fonte única e dupla. Os resultados, obtidos a partir de dados reais de uma cadeia produtiva alimentar, indicam elevada capacidade de detecção de contaminação, bom desempenho na distinção entre fontes únicas e duplas e eficiência na priorização das origens mais prováveis, reduzindo o número de hipóteses a serem investigadas. Assim, o trabalho contribui para a modelagem probabilística da rastreabilidade e oferece suporte à tomada de decisão em processos industriais de investigação, bloqueio de lotes e *recall*.

Palavras-chave: Rastreabilidade probabilística, Múltiplas fontes de contaminação, Teorema de Bayes, Cadeias produtivas complexas, Simulação Monte Carlo.

3.1 INTRODUÇÃO

A identificação da origem de eventos adversos em redes complexas constitui um problema central em diferentes domínios, incluindo epidemiologia, redes de suprimento e sistemas de infraestrutura. A literatura recente tem demonstrado que a capacidade de localizar fontes em redes depende não apenas da observação de estados finais, mas também da modelagem adequada da dinâmica de propagação e da estrutura topológica subjacente (ALTARELLI *et al.*, 2014; GOMEZ-RODRIGUEZ; LESKOVEC; KRAUSE, 2012). Em redes direcionadas e ponderadas, nas quais múltiplos caminhos conectam origem e observação, abordagens que consideram apenas trajetórias dominantes tendem a subestimar a complexidade do problema (HORN; FRIEDRICH, 2019).

No contexto industrial, essa problemática assume contornos ainda mais desafiadores, uma vez que cadeias produtivas envolvem transformação de insumos, mistura de fluxos e redistribuição multiperíodo. Embora avanços metodológicos tenham sido observados na inferência probabilística em redes físicas e epidemiológicas (DE SANCTIS *et al.*, 2008; SHAGHAGHIAN; COATES, 2016), permanece limitada a aplicação sistemática desses arcabouços ao ambiente de cadeias produtivas industriais com múltiplas origens potenciais de contaminação.

A rastreabilidade em cadeias produtivas industriais é um mecanismo estratégico para mitigar riscos sanitários, econômicos e reputacionais. Em cadeias alimentares e agroindustriais, por exemplo, a capacidade de identificar rapidamente a origem de um lote contaminado reduz a extensão de *recall* e os impactos sistêmicos do evento (HORN; FRIEDRICH, 2019). Estudos demonstram ainda que a própria estrutura da rede influencia a incerteza associada à identificação da fonte, sendo redes mais densas e com maior sobreposição de conexões estruturalmente mais ambíguas do ponto de vista inferencial (LU *et al.*, 2019).

Entretanto, modelos tradicionais de rastreabilidade são, em grande medida, determinísticos, baseando-se na reconstrução de vínculos diretos entre lotes ou na identificação de caminhos específicos na rede. Essa abordagem ignora a multiplicidade de trajetórias plausíveis e a incerteza inerente à mistura de fluxos, podendo conduzir a soluções excessivamente restritivas ou inconsistentes quando há sobreposição estrutural (ALTARELLI *et al.*, 2014; DE SANCTIS *et al.*, 2008). A literatura em redes complexas evidencia que problemas de identificação de fonte são, por natureza, problemas inversos estocásticos sujeitos à não unicidade de solução (CAI; YE, 2021).

O estudo de Salinas *et al.* (2024) contribuiu ao propor uma representação probabilística dos fluxos de materiais em cadeias produtivas, utilizando o Teorema de Bayes e Cadeias de Markov para calcular probabilidades condicionais de contribuição entre matérias-primas, lotes intermediários e produtos finais. Contudo, esse modelo não teve como objetivo estimar a

probabilidade de uma fonte de contaminação, mas sim estruturar a rastreabilidade direta e reversa dos lotes ao longo do processo produtivo. O presente artigo parte dessa matriz de contribuição como base estrutural, mas formula um problema distinto: identificar se o padrão de contaminação observado nos produtos finais é compatível com uma ou duas fontes potenciais e, em seguida, priorizar as matérias-primas ou combinações de matérias-primas mais prováveis de terem originado o evento. Essa abordagem é necessária porque, em cadeias com mistura, fracionamento e recombinação de lotes, diferentes origens ou combinações de origens podem gerar padrões semelhantes de contaminação, ampliando o espaço de hipóteses e a incerteza associada à investigação (ALTARELLI *et al.*, 2014; CHAI; WANG; ZHU, 2021; HORN; FRIEDRICH, 2019).

A consideração de múltiplas fontes altera substancialmente a natureza do problema de identificação de origem. Em cenários de fonte única, embora já exista incerteza associada à estrutura da rede e à observação parcial dos estados, o espaço de hipóteses permanece restrito a um conjunto finito de candidatos individuais. Quando múltiplas origens passam a ser consideradas, o problema deixa de ser apenas uma busca por um nó gerador e passa a envolver combinações de nós potencialmente ativos, modificando a própria estrutura do problema inverso (CHAI; WANG; ZHU, 2021; ALTARELLI *et al.*, 2014).

Primeiramente, surge a não unicidade estrutural ampliada. Em redes com múltiplos caminhos e transformação de fluxos, diferentes combinações de fontes podem gerar distribuições observadas indistinguíveis nos nós finais. Esse fenômeno já é reconhecido em problemas de contaminação hidráulica como característica intrínseca de sistemas com múltiplos caminhos de propagação (DE SANCTIS *et al.*, 2008; CAI; YE, 2021), mas torna-se ainda mais crítico em cadeias produtivas industriais, nas quais há transformação e recombinação de lotes ao longo do tempo. A presença de múltiplas origens não apenas multiplica as soluções possíveis, mas pode gerar configurações estruturalmente equivalentes sob o ponto de vista observacional.

Em segundo lugar, a sobreposição de fluxos introduz um nível adicional de complexidade. Diferentemente de modelos epidemiológicos clássicos do tipo SI ou SIR, nos quais a transmissão ocorre por contato individual (ALTARELLI *et al.*, 2014; BRAUNSTEIN; INGROSSO, 2016), cadeias produtivas envolvem mistura física de materiais, agregação e divisão de lotes. Assim, os efeitos de diferentes fontes não se propagam isoladamente, mas podem se combinar e interagir ao longo das etapas produtivas. Essa mistura estrutural reduz a separabilidade das trajetórias causais e amplia a ambiguidade na inferência posterior.

Como consequência direta, emerge a ambiguidade probabilística. Em cenários de múltiplas fontes, a distribuição posterior deixa de ser concentrada em hipóteses individuais e passa a se distribuir sobre subconjuntos de combinações possíveis, muitas vezes com probabilidades

próximas entre si. Trabalhos que tratam múltiplas sementes em redes temporais já indicam que a incerteza cresce significativamente à medida que o número de fontes aumenta (CHAI; WANG; ZHU, 2021; ZHU *et al.*, 2022). Contudo, tais estudos geralmente assumem independência entre fontes ou ausência de transformação estrutural dos fluxos, o que não se verifica em sistemas produtivos com mistura multiperíodo.

Por fim, há o escalonamento combinatório do espaço de hipóteses. Se uma rede possui n possíveis nós candidatos, o problema de fonte única envolve, no pior caso, n hipóteses. Já no caso de até k fontes simultâneas, o número de combinações cresce segundo coeficientes binomiais, tornando-se rapidamente explosivo mesmo para valores modestos de n e k . Esse crescimento combinatório impõe desafios computacionais significativos e exige estratégias inferenciais mais sofisticadas do que aquelas empregadas em modelos de fonte única, como já observado em contextos de inferência estrutural em redes latentes (GOMEZ-RODRIGUEZ; LESKOVEC; KRAUSE, 2012).

Portanto, múltiplas fontes não representam apenas uma extensão quantitativa do problema original, mas uma mudança qualitativa em sua estrutura inferencial, exigindo arcabouços probabilísticos capazes de lidar simultaneamente com não unicidade, mistura estrutural e explosão combinatória de hipóteses, elementos centrais para o avanço metodológico proposto no presente estudo.

Um aspecto operacional relevante para a interpretação dos padrões de contaminação é a lógica de consumo dos lotes ao longo do processo produtivo. Neste estudo, assume-se que a propagação dos lotes segue predominantemente a lógica FIFO (*First In, First Out*), na qual os materiais que entram primeiro no sistema tendem a ser consumidos primeiro nas etapas subsequentes. Essa premissa é importante porque preserva a coerência temporal entre a entrada das matérias-primas e sua contribuição nos produtos finais, favorecendo a formação de padrões estruturais interpretáveis no vetor de contaminação. Quando essa lógica não é observada, um mesmo lote pode impactar produtos finais em diferentes janelas temporais, gerando padrões não monotônicos que podem dificultar a distinção entre cenários de fonte única e dupla. Assim, embora práticas FIFO sejam recorrentes em sistemas de produção e rastreabilidade, elas não substituem o registro detalhado por lote, mas constituem uma condição operacional relevante para a interpretação dos resultados do método proposto (IFT, 2009; COMBA *et al.*, 2013).

Diante das limitações identificadas nos modelos de fonte única e das lacunas associadas à modelagem de múltiplas origens simultâneas, este artigo propõe uma ampliação estrutural do arcabouço probabilístico desenvolvido anteriormente por SALINAS *et al.* (2024). A primeira contribuição consiste na extensão do modelo markoviano para múltiplas origens simultâneas,

permitindo que o processo produtivo seja representado como resultado da interação conjunta de diferentes nós geradores potenciais. Diferentemente da formulação anterior, na qual a dinâmica era condicionada a uma única hipótese causal, o presente trabalho estrutura o sistema como combinação probabilística de múltiplos vetores iniciais propagados por matrizes de transição multiperíodo.

A segunda contribuição reside na formulação de um modelo para avaliação conjunta de fontes de contaminação, no qual combinações de fontes são tratadas como hipóteses estruturais concorrentes. Em vez de avaliar candidatos individualmente, o método considera explicitamente subconjuntos do espaço de nós, permitindo comparar o grau de aderência entre os padrões simulados e as observações do sistema. Embora inspirado no uso do Teorema de Bayes como mecanismo de atualização de probabilidades condicionais, o procedimento adotado se aproxima de uma abordagem de otimização, na qual se busca a combinação de fontes que melhor explica os dados observados. Essa estratégia responde diretamente ao problema do escalonamento combinatório e da ambiguidade probabilística discutidos anteriormente, aproximando-se conceitualmente de arcabouços de inferência estrutural em redes complexas (ALTARELLI *et al.*, 2014; GOMEZ-RODRIGUEZ; LESKOVEC; KRAUSE, 2012), porém adaptada à lógica de transformação e conservação de fluxos materiais.

Adicionalmente, o artigo introduz o tratamento explícito da mistura multiperíodo, reconhecendo que, em cadeias produtivas industriais, fluxos não apenas se propagam, mas se combinam e se transformam ao longo de etapas sucessivas. Diferentemente de modelos epidemiológicos clássicos ou de transporte sem replicação (HORN; FRIEDRICH, 2019), a proposta considera que lotes intermediários podem resultar da agregação proporcional de diferentes origens, preservando coerência com balanços de massa e matrizes de transição.

Por fim, o estudo incorpora uma avaliação estrutural da identificabilidade, examinando sob quais configurações topológicas e parâmetros de mistura o problema admite discriminação adequada entre hipóteses concorrentes. Essa análise dialoga com o conceito de traceability entropy proposto por LU *et al.* (2019), mas o expande para cenários com múltiplas origens e transformação multiperíodo, contribuindo para a compreensão das condições sob as quais a inferência probabilística é estruturalmente viável.

3.2 ENQUADRAMENTO DO PROBLEMA E POSICIONAMENTO TEÓRICO

A identificação de múltiplas origens de contaminação em cadeias produtivas industriais pode ser formalmente compreendida como um problema inverso estocástico em redes direcionadas com transformação de fluxos. Diferentemente de processos clássicos de difusão epidemiológica ou propagação de informação, nos quais a transmissão ocorre predominantemente por contato individual e replicação de estado (ALTARELLI *et al.*, 2014; BRAUNSTEIN; INGROSSO, 2016), cadeias produtivas envolvem transporte direcionado, transformação física e mistura proporcional de materiais ao longo de etapas sucessivas.

Embora a literatura sobre identificação de fontes em redes tenha avançado significativamente nas últimas duas décadas (GOMEZ-RODRIGUEZ; LESKOVEC; KRAUSE, 2012; FATIMA; RIZVI, 2025), grande parte dessas formulações permanece ancorada em modelos de transmissão singular, com hipóteses estruturais que não capturam plenamente as características de sistemas produtivos industriais. Assim, torna-se necessário posicionar teoricamente o problema tratado neste artigo, explicitando suas particularidades estruturais e inferenciais.

3.2.1 Características estruturais do problema

O problema de identificação de múltiplas fontes em cadeias produtivas industriais apresenta um conjunto de características estruturais que o diferenciam substancialmente dos modelos epidemiológicos e de difusão clássicos.

(1) Rede dirigida e acíclica por períodos, mas com mistura estrutural.

A rede produtiva pode ser representada como direcionada e temporalmente ordenada, uma vez que fluxos seguem etapas produtivas sucessivas. Contudo, diferentemente de grafos aproximadamente arbóreos frequentemente assumidos em modelos epidemiológicos (ALTARELLI *et al.*, 2014), há convergência de múltiplos fluxos em nós intermediários, gerando mistura estrutural e sobreposição de trajetórias causais.

(2) Transformação de lotes ao longo do tempo.

Nos sistemas produtivos, insumos não apenas se propagam, mas são transformados física ou quimicamente, originando novos lotes com composição derivada. Essa característica distingue o problema de modelos de *random walk* ou difusão sem replicação, como em HORN; FRIEDRICH (2019), pois aqui há transformação multiperíodo associada a matrizes de transição que representam proporções de utilização e recombinação.

(3) Observações parciais e tardias.

As evidências de contaminação normalmente são observadas apenas em estágios finais da cadeia, frequentemente após múltiplas etapas de processamento. Tal configuração é análoga a cenários de observação parcial em redes epidemiológicas (BU *et al.*, 2020), porém agravada pelo fato de que a transformação intermediária altera a estrutura probabilística da inferência.

(4) Múltiplas fontes independentes simultâneas.

Diferentemente da maioria dos modelos de identificação de fonte única (CHANG *et al.*, 2018; LIU *et al.*, 2021), o problema considerado admite a possibilidade de múltiplas origens independentes ativas no mesmo horizonte temporal. Essa condição altera qualitativamente o espaço de hipóteses, que passa de linear para combinatório, ampliando a complexidade inferencial.

(5) Conservação e mistura de massa.

Ao contrário de modelos epidêmicos baseados em transmissão binária de estado (SI, SIR), cadeias produtivas obedecem a princípios de conservação de massa. Lotes resultantes podem conter proporções de múltiplas origens, exigindo tratamento explícito de mistura proporcional, conceito ausente em grande parte dos modelos de difusão baseados em contato (ALTARELLI *et al.*, 2014; BRAUNSTEIN; INGROSSO, 2016).

(6) Não unicidade de solução.

Assim como observado em problemas de contaminação hidráulica (DE SANCTIS *et al.*, 2008; CAI; YE, 2021), diferentes combinações de fontes podem produzir padrões observacionais indistinguíveis. Contudo, em cadeias produtivas com mistura multiperíodo, essa não unicidade é amplificada pela agregação estrutural ao longo do tempo.

(7) Dependência estrutural entre caminhos.

Em redes de difusão clássicas, muitas formulações assumem independência condicional entre transmissões ao longo das arestas (GOMEZ-RODRIGUEZ; LESKOVEC; KRAUSE, 2012). No contexto produtivo, entretanto, caminhos que convergem em um mesmo nó tornam-se estruturalmente dependentes devido à mistura física de fluxos, o que inviabiliza suposições simplificadas de independência entre trajetórias.

Em conjunto, essas características evidenciam que métodos epidemiológicos clássicos, ainda que sofisticados, não são suficientes para capturar plenamente a dinâmica de múltiplas origens em cadeias produtivas com transformação e mistura. O problema exige um arcabouço probabilístico capaz de integrar dinâmica markoviana multiperíodo, conservação de massa e inferência conjunta de combinações causais, fundamentando o posicionamento teórico e metodológico desenvolvido nas seções seguintes.

3.2.2 Trabalhos relacionados

A literatura sobre identificação de fontes em redes evoluiu significativamente nas últimas décadas, especialmente no contexto de epidemias, redes temporais, infraestrutura física e cadeias de suprimento. Entretanto, embora esses avanços ofereçam importantes fundamentos conceituais e metodológicos, a maioria das abordagens permanece ancorada em hipóteses estruturais que limitam sua aplicação direta a cadeias produtivas industriais com múltiplas fontes simultâneas e mistura multiperíodo. A seguir, os principais trabalhos são organizados em blocos temáticos, destacando convergências e lacunas.

(i) Modelos de fonte única em redes epidêmicas

Grande parte da literatura concentra-se na identificação de uma única origem em processos de difusão modelados sob dinâmicas do tipo SI ou SIR. Trabalhos como ALTARELLI *et al.* (2014), BRAUNSTEIN; INGROSSO (2016), CHANG *et al.* (2018) e LIU *et al.* (2021) demonstram que a inferência bayesiana, *Belief Propagation* ou estimação por *Maximum a Posteriori* melhoram substancialmente a identificação da fonte sob observação parcial. Estudos como MUNTONI *et al.* (2024) e BU *et al.* (2020) avançam ao tratar incerteza temporal e observações incompletas.

Apesar da sofisticação inferencial, esses modelos compartilham uma característica central: assumem uma única origem ativa por evento ou uma estrutura causal baseada em “pai único” por infecção. Além disso, operam em contextos de transmissão por contato, sem transformação física ou mistura estrutural de fluxos. Assim, embora relevantes do ponto de vista probabilístico, não capturam a complexidade de cadeias produtivas com múltiplas origens simultâneas.

(ii) Inferência em redes temporais e redes latentes

Outro conjunto relevante de trabalhos trata da reconstrução de redes e cadeias causais quando a estrutura não é completamente observável. GOMEZ-RODRIGUEZ; LESKOVEC; KRAUSE (2012) demonstram ser possível inferir redes latentes de difusão a partir de múltiplas cascatas temporais. SHAGHAGHIAN; COATES (2016) e SHAGHAGHI; COATES (2017) propõem inferência bayesiana conjunta da estrutura e dos tempos latentes, inclusive em regime *online*. XU *et al.* (2025) e XU (2025) integram dados temporais, genômicos e de rede social para reconstruir cadeias de transmissão com maior precisão.

Esses trabalhos ampliam a capacidade de inferência sob incerteza estrutural e temporal. Contudo, mesmo quando tratam múltiplas sementes, permanecem baseados em processos de transmissão discreta e causalidade singular por evento, sem considerar explicitamente conservação de massa, transformação multiperíodo ou mistura física de fluxos.

(iii) Problemas inversos em redes físicas (água e esgoto)

A literatura de redes hidráulicas oferece paralelos diretos ao problema de rastreabilidade industrial. DE SANCTIS *et al.* (2008), CAI; YE (2021) e YANG; HUANG; LIU (2024) formulam a identificação de fontes como problema inverso estocástico, incorporando incerteza sensorial e modelagem físico-hidráulica. A dissertação de YANG (2013) avança ao propor inferência probabilística em tempo quase real e permitir múltiplas injeções simultâneas.

Esses estudos são particularmente relevantes por tratarem explicitamente não unicidade de solução e mistura de fluxos. Entretanto, permanecem ancorados em redes contínuas de transporte hidráulico, nas quais não há transformação discreta de lotes nem recombinação multiperíodo típica de cadeias produtivas industriais.

(iv) Difusão em redes de suprimento

A aproximação mais direta com cadeias produtivas aparece em LU *et al.* (2019) e HORN; FRIEDRICH (2019). LU *et al.* (2019) introduzem o conceito de *traceability entropy* para avaliar a capacidade estrutural de uma rede em permitir identificação da fonte. HORN; FRIEDRICH (2019) modelam surtos alimentares como *random walk* em rede direcionada e ponderada, considerando todos os caminhos possíveis entre origem e observação.

Embora esses trabalhos avancem ao tratar redes de suprimento explicitamente, ambos assumem fonte única e não incorporam mistura estrutural multiperíodo com múltiplas origens simultâneas.

(v) Modelos com múltiplas fontes

Alguns estudos abordam explicitamente múltiplas origens. CHAI; WANG; ZHU (2021) tratam múltiplas fontes em redes temporais por meio de particionamento inspirado em Voronoi. ZHU *et al.* (2022) propõem identificação baseada em caminhos sob baixa taxa de infecção. STERCHI *et al.* (2023) exploram *active querying* para reduzir incerteza sob observação restrita.

Essas contribuições demonstram que múltiplas fontes alteram substancialmente o espaço inferencial. Contudo, mantêm dinâmicas de difusão sem transformação física ou mistura proporcional de fluxos. Mesmo revisões amplas, como FATIMA; RIZVI (2025), evidenciam que a maior parte da literatura permanece concentrada em contextos epidemiológicos ou informacionais, com hipóteses estruturais simplificadas. Da mesma forma, modelos variacionais e baseados em aprendizado profundo, como ZHOU *et al.* (2020), avançam na representação de incerteza, mas continuam orientados a cascatas com origem singular.

Em síntese, os trabalhos analisados consolidam três grandes avanços: (i) tratamento explícito da incerteza via inferência bayesiana ou máxima verossimilhança, (ii) consideração de observações parciais e redes temporais, (iii) reconhecimento da influência estrutural da topologia na identificabilidade.

Entretanto, a maioria das formulações assume fonte única, causalidade singular ou ausência de mistura estrutural multiperíodo. Mesmo quando múltiplas fontes são consideradas, raramente há integração simultânea entre:

- conservação e mistura proporcional de massa,
- transformação multiperíodo de lotes,
- inferência conjunta de combinações causais em redes produtivas discretas.

É precisamente nessa interseção que se localiza a lacuna específica que este artigo busca preencher: desenvolver um arcabouço probabilístico capaz de tratar múltiplas fontes simultâneas em cadeias produtivas industriais com transformação e mistura estrutural, expandindo a base estabelecida por SALINAS *et al.* (2024) para um cenário de maior complexidade estrutural e inferencial.

3.3 MODELAGEM PROBABILÍSTICA DA PROPAGAÇÃO DE CONTAMINAÇÃO

A propagação de contaminação em redes produtivas complexas raramente ocorre de forma determinística. Em processos industriais que envolvem mistura de insumos, transformações intermediárias e redistribuição de fluxos, a presença de uma matéria-prima contaminada não implica necessariamente que todos os produtos derivados apresentem contaminação detectável. Assim, em vez de assumir uma relação determinística entre entrada contaminada e saída contaminada, este trabalho modela a ocorrência de contaminação no produto final como um evento probabilístico dependente da fração de insumo contaminado presente na formulação.

Considere uma rede produtiva composta por um conjunto de matérias-primas $i \in I$ e um conjunto de produtos finais $j \in J$. Seja a_{ij} a fração da matéria-prima i utilizada na composição do produto j , que será determinada com base na formulação probabilística detalhada no Capítulo 2, de forma que:

$$\sum_{i \in I} a_{ij} = 1, \forall j \in J \quad \text{Eq. 15}$$

Considere ainda uma hipótese H que define o conjunto $S(H) \subseteq I$ de matérias-primas contaminadas. A fração total de material contaminado incorporado ao produto j sob a hipótese H é dada por:

$$x_j(H) = \sum_{i \in S(H)} a_{ij} \quad \text{Eq. 16}$$

onde $x_j(H)$ representa a proporção de insumos contaminados presentes na formulação do produto j .

A ocorrência de contaminação observável no produto final é então modelada como uma variável aleatória binária:

$$Y_j \sim \text{Bernoulli}(p_j) \quad \text{Eq. 17}$$

em que $Y_j = 1$ indica que o produto j foi observado contaminado e $Y_j = 0$ indica ausência de contaminação detectável.

A probabilidade de contaminação p_j é definida como uma função não linear da fração contaminada $x_j(H)$:

$$p_j = g(x_j(H)) \quad \text{Eq. 18}$$

Neste estudo, adota-se uma função dose–resposta sigmoideal do tipo logística, amplamente utilizada para modelar relações probabilísticas entre intensidade de exposição e ocorrência de um evento. Essa formulação permite representar uma transição suave entre regimes de baixa e alta probabilidade de contaminação à medida que aumenta a fração de matéria-prima contaminada incorporada ao produto:

$$g(x) = \begin{cases} 0, & x \leq LI \\ \left(\frac{1}{1 + e^{-k(x - \frac{LI}{2})}} \right), & LI < x < LS \\ 1, & x \geq LS \end{cases} \quad \text{Eq. 19}$$

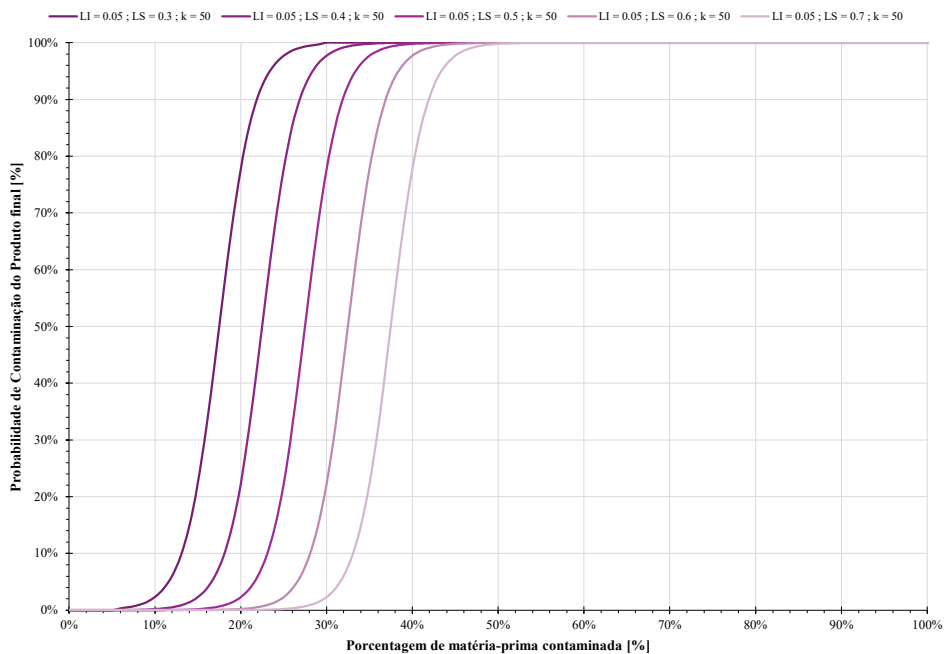
onde:

- LI representa o limiar inferior de contaminação,
- LS representa o limiar superior de contaminação,
- k controla a inclinação da curva e, consequentemente, a rapidez com que a probabilidade de contaminação aumenta com a fração contaminada.

O parâmetro de inclinação k é calibrado de forma que a função logística produza probabilidades próximas de 0,05 e 0,95 nos pontos LI e LS , respectivamente. Essa parametrização permite representar uma região intermediária na qual a ocorrência de contaminação é probabilística, refletindo variabilidades inerentes ao processo produtivo, como heterogeneidade de mistura, degradação do contaminante e limites de detecção analítica.

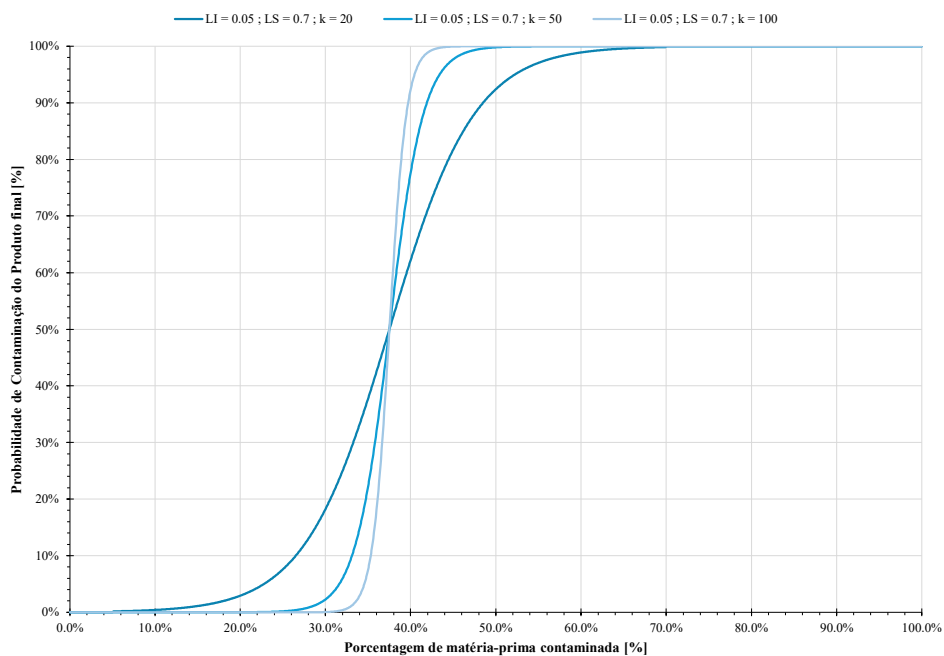
Essa formulação permite capturar de forma realista o comportamento de propagação de contaminação na rede produtiva, preservando ao mesmo tempo propriedades desejáveis de continuidade e monotonicidade da função dose–resposta.

Figura 6 – Exemplo da função Sigmoidal variando o LS.



Fonte: Autor (2026)

Figura 7 – Exemplo da função Sigmoidal variando o k .



Fonte: Autor (2026)

3.3.1 Interpretação dos limiares e do parâmetro de inclinação da função dose–resposta

Os limiares inferior (LI) e superior (LS) delimitam diferentes regimes na relação entre a fração de matéria-prima contaminada e a probabilidade de contaminação do produto final. Abaixo de LI, a contaminação é considerada negligenciável, uma vez que pequenas frações de insumo contaminado tendem a ser diluídas ao longo do processo produtivo, resultando em baixa probabilidade de detecção no produto final. Acima de LS, a contaminação torna-se praticamente certa, refletindo situações em que a proporção de material contaminado é suficientemente elevada para superar efeitos de diluição e variabilidade do processo.

Entre esses dois limiares, observa-se uma região intermediária de transição probabilística, na qual a ocorrência de contaminação depende de fatores estocásticos inerentes ao sistema produtivo, como heterogeneidade da mistura, eficiência das operações unitárias e limitações analíticas. Essa estrutura é consistente com modelos de dose–resposta, nos quais a probabilidade de ocorrência de um evento adverso varia de forma não linear com o nível de exposição.

O parâmetro k determina a sensibilidade da probabilidade de contaminação em relação à fração contaminada: valores elevados produzem transições mais abruptas entre os regimes de baixa e alta probabilidade, aproximando o comportamento de um limiar determinístico, enquanto valores menores resultam em transições mais suaves, ampliando a região probabilística intermediária.

No contexto deste trabalho, o parâmetro k pode ser interpretado como um indicador da intensidade de propagação da contaminação na rede produtiva, refletindo a sensibilidade do sistema às variações na fração de matéria-prima contaminada. Sua calibração é realizada de forma a garantir que a transição da função logística ocorra no intervalo delimitado por LI e LS, preservando a coerência entre a interpretação operacional dos limiares e o comportamento probabilístico do modelo.

3.3.2 Algoritmo 1 - Método de detecção de múltiplas fontes baseado em “vales”

O objetivo deste algoritmo é identificar, a partir do padrão de contaminação observado nos produtos finais, se o sistema é caracterizado por uma única fonte de contaminação ou por múltiplas fontes. Neste Artigo, especificamente duas fontes. Assume-se que os parâmetros da função dose–resposta logística, o limiar inferior LI , o limiar superior LS e o parâmetro de inclinação k , são previamente conhecidos. Esses parâmetros foram definidos com base nas características do

processo produtivo analisado e no comportamento esperado da propagação da contaminação ao longo das operações unitárias, conforme discutido nas seções anteriores.

Como dados de entrada, o algoritmo utiliza a matriz de transição $A = [a_{ij}]$, que representa a contribuição de cada matéria-prima $i \in I$ na composição de cada produto final $j \in J$, além do conjunto de matérias-primas I e do conjunto de produtos finais J . A partir dessa estrutura, são gerados cenários simulados de contaminação por meio de uma abordagem probabilística, utilizando a função dose–resposta previamente parametrizada.

Para cada hipótese de contaminação, considerando uma ou duas matérias-primas como potenciais fontes, constrói-se o vetor de contribuição total para os produtos finais. No caso de uma única fonte i , o vetor de contribuição é dado por $x_j = a_{ij}$. No caso de duas fontes i e k , o vetor resulta da soma das contribuições individuais, ou seja, $x_j = a_{ij} + a_{kj}$, respeitando a estrutura da matriz de transição.

Em seguida, para cada produto final j , calcula-se a probabilidade de contaminação p_j por meio da função dose–resposta logística previamente definida. A partir dessas probabilidades, são gerados cenários simulados de contaminação por meio de amostragem de variáveis aleatórias Bernoulli, resultando em um vetor binário de resposta:

$$y = (y_1, y_2, \dots, y_{|J|}) \quad \text{Eq. 20}$$

onde $y_j = 1$ indica que o produto final j foi classificado como contaminado, e $y_j = 0$ indica ausência de contaminação detectável.

O algoritmo então procede à classificação do padrão observado nesse vetor de resposta com base no método dos “vales”. Inicialmente, verifica-se se há contaminação detectável no sistema. Caso:

$$\sum_{j \in J} y_j = 0 \quad \text{Eq. 21}$$

o cenário é classificado como “**SEM CONTAMINAÇÃO**”, não sendo possível inferir a presença de fontes contaminantes.

Nos casos em que existe ao menos um produto contaminado, o vetor y é analisado quanto à sua estrutura, com o objetivo de identificar padrões que indiquem a presença de múltiplas fontes de contaminação. Para isso, calcula-se o vetor de diferenças sucessivas entre elementos adjacentes:

$$d_j = y_{j+1} - y_j, j = 1, 2, \dots, |J| - 1 \quad \text{Eq. 22}$$

A seguir, obtém-se o vetor de sinais dessas diferenças:

$$s_j = \text{sign}(d_j) \quad \text{Eq. 23}$$

em que a função sinal assume valores em $\{-1,0,1\}$, representando, respectivamente, transições decrescentes, constantes e crescentes no vetor de contaminação.

Para evitar distorções decorrentes de trechos constantes, todos os valores nulos são removidos do vetor de sinais. Adicionalmente, elimina-se a repetição de sinais consecutivos, preservando apenas os pontos de mudança efetiva no comportamento do vetor. O resultado é uma sequência reduzida que descreve a dinâmica essencial das variações do padrão de contaminação.

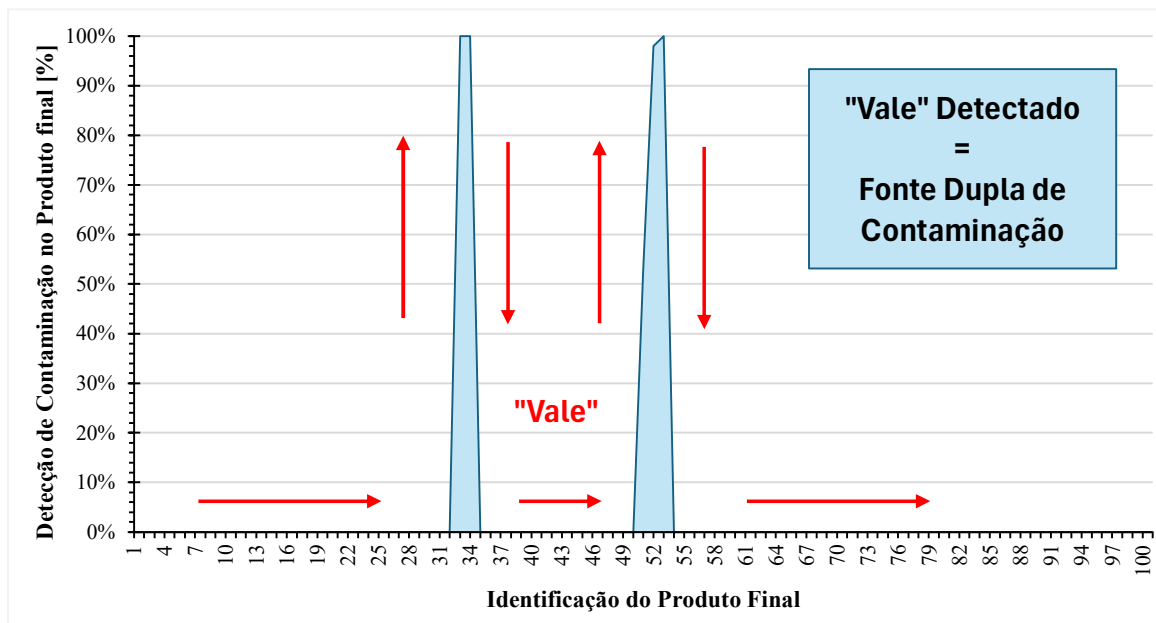
A partir dessa sequência reduzida, busca-se a ocorrência de padrões do tipo:

$$(-1, 1, -1),$$

os quais caracterizam um comportamento de **decréscimo seguido de crescimento e novo decréscimo**, configurando um “vale” no padrão de contaminação.

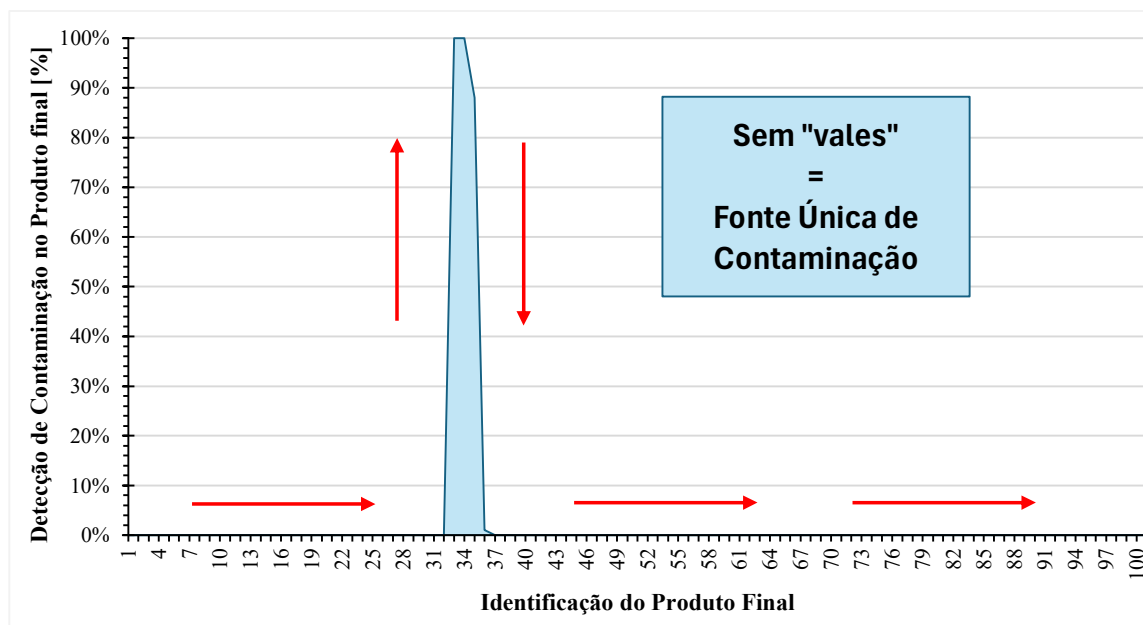
Seguem exemplos gráficos do método de detecção de “vales”:

Figura 8 – Exemplo gráfico da detecção dos “vales” para o caso de fonte de contaminação dupla.



Fonte: Autor (2026)

Figura 9 – Exemplo gráfico da falta de detecção de “vales” para o caso de fonte de contaminação única.



Fonte: Autor (2026)

A presença de pelo menos um vale indica um comportamento bimodal no vetor de resposta, sugerindo a existência de múltiplas fontes de contaminação atuando no sistema. Nesse caso, o cenário é classificado como **“2 FONTES”**. Por outro lado, na ausência desse padrão, o vetor é interpretado como monomodal, sendo classificado como **“1 FONTE”**.

Esse procedimento permite identificar, de forma robusta, a presença de múltiplas fontes de contaminação com base exclusivamente na estrutura do vetor de resposta, sem a necessidade de ajuste adicional de parâmetros. Essa interpretação está condicionada à hipótese de escoamento em regime *FIFO* (*First In, First Out*), na qual a propagação da contaminação ao longo da cadeia produtiva preserva a ordem temporal de entrada das matérias-primas, resultando em padrões de contribuição monotônicos nos produtos finais.

Nesse contexto, a utilização do método dos “vales” torna-se particularmente adequada, uma vez que múltiplas fontes tendem a gerar padrões não monotônicos no vetor de resposta devido à superposição de contribuições ao longo da cadeia produtiva. Por outro lado, cenários de fonte única, sob a hipótese de *FIFO*, tendem a produzir distribuições mais regulares e unimodais, permitindo sua distinção com base na estrutura do sinal observado.

Dessa forma, o algoritmo proposto consolida em uma única etapa a geração probabilística dos cenários de contaminação e a classificação estrutural dos padrões observados, constituindo uma abordagem eficiente e coerente com a dinâmica do sistema produtivo analisado.

Para cenários em que a hipótese verdadeira é de fonte única, define-se, para cada matéria-prima $i \in I$, o Índice de Bimodalidade:

$$IB_i = \frac{\text{Número de Simulações Classificadas como "FONTE ÚNICA"}}{N - \text{Número de Simulações Classificadas como "SEM CONTAMINAÇÃO"}} \quad \text{Eq. 24}$$

Esse índice mede a capacidade do método em preservar o comportamento monomodal esperado.

O Índice de Detecção de Contaminação é dado por:

$$IDC = \frac{\text{Número de Matérias-primas com Detecção de Contaminação}}{|I|} \quad \text{Eq. 25}$$

Posteriormente, calcula-se o valor médio do índice de bimodalidade entre todas as matérias-primas avaliadas:

$$IB_{\text{médio}} = \frac{1}{|I|} \sum_{i \in I} IB_i \quad \text{Eq. 26}$$

Com base nesses dois indicadores, define-se o **Índice Composto de Contaminação (ICC)**, calculado como a média aritmética entre $IB_{\text{médio}}$ e IDC :

$$ICC = \frac{IB_{\text{médio}} + IDC}{2} \quad \text{Eq. 27}$$

Para cenários em que a hipótese verdadeira é de fonte dupla, define-se, para cada combinação $(i, \ell) \in \mathcal{P}$, o Índice de Bimodalidade:

$$IB_{i,\ell} = \frac{\text{Número de Simulações Classificadas como "FONTE DUPLA"}}{N - \text{Número de Simulações Classificadas como "SEM CONTAMINAÇÃO"}} \quad \text{Eq. 28}$$

Esse índice mede a capacidade do método em identificar corretamente padrões bimodais.

O Índice de Detecção de Contaminação é dado por:

$$IDC = \frac{\text{Número de Combinações com Detecção de Contaminação}}{|\mathcal{P}|} \quad \text{Eq. 29}$$

Posteriormente, constrói-se uma matriz triangular superior contendo os valores de $IB_{i,\ell}$ para cada combinação (i, ℓ) , de forma a registrar o desempenho de cada hipótese de contaminação dupla para cada conjunto de parâmetros testado.

A partir dessa matriz, calcula-se o **Índice Médio de Bimodalidade**, dado por:

$$IB_{\text{médio}} = \frac{1}{|\mathcal{P}|} \sum_{(i,\ell) \in \mathcal{P}} IB_{i,\ell} \quad \text{Eq. 30}$$

Com base em $IB_{\text{médio}}$ e IDC , define-se o **Índice Composto de Contaminação (ICC)** para cenários de fonte dupla como:

$$ICC = \frac{IB_{\text{médio}} + IDC}{2} \quad \text{Eq. 31}$$

Esse índice sintetiza, em uma única medida, a capacidade do modelo de detectar contaminação e, simultaneamente, de classificar corretamente o padrão observado como sendo de fonte dupla.

Por fim, define-se o **Índice Composto de Contaminação Global** como:

$$ICC_{\text{Global}} = \frac{ICC_{\text{fonte única}} + ICC_{\text{fonte dupla}}}{2} \quad \text{Eq. 32}$$

Esse índice sintetiza, de forma integrada, a capacidade do método em detectar contaminação e classificar corretamente cenários de fonte única e de fonte dupla.

Dessa forma, o algoritmo proposto consolida em uma única estrutura metodológica: a geração probabilística de cenários de contaminação, a classificação estrutural dos padrões por meio do método dos vales e a avaliação quantitativa do desempenho para diferentes configurações de contaminação.

Essa abordagem permite verificar se o método é capaz de reproduzir adequadamente tanto padrões monomodais quanto bimodais, garantindo consistência com o comportamento esperado do sistema produtivo e fornecendo uma métrica objetiva para validação dos resultados.

3.3.3 Algoritmo 2 – Identificação das fontes de contaminação por similaridade entre vetor observado e mapa de respostas prováveis

O objetivo deste algoritmo é identificar, a partir de um padrão observado de contaminação nos produtos finais, qual matéria-prima, ou combinação de duas matérias-primas, apresenta maior probabilidade de ter originado o evento de contaminação. Para isso, utiliza-se um procedimento baseado na construção prévia de um **mapa de respostas prováveis**, seguido da comparação entre o vetor de observação e os vetores armazenados nesse mapa por meio da distância euclidiana.

Assim como no Algoritmo 1, assume-se que os parâmetros da função dose-resposta logística, o limiar inferior LI , o limiar superior LS e o parâmetro de inclinação k , são previamente conhecidos e compatíveis com o processo produtivo analisado. O ponto de partida do algoritmo é

novamente a matriz de transição $A = [a_{ij}]$, obtida a partir do Capítulo 2, a qual representa a contribuição de cada matéria-prima $i \in I$ na composição de cada produto final $j \in J$.

O algoritmo é composto por duas etapas principais: (i) a construção do mapa de respostas prováveis e (ii) a busca, nesse mapa, da combinação de matérias-primas que apresenta maior similaridade com o vetor de contaminação observado.

Etapa 1 – Construção do mapa de respostas prováveis

Inicialmente, constrói-se uma matriz triangular superior, doravante denominada **mapa**, em que cada posição armazena um vetor representativo da resposta esperada dos produtos finais para uma hipótese específica de contaminação. Cada hipótese pode corresponder tanto à contaminação de uma única matéria-prima quanto à contaminação simultânea de duas matérias-primas.

No caso de contaminação por fonte única, para cada matéria-prima $i \in I$, extrai-se da matriz de transição o vetor de contribuições dessa matéria-prima nos produtos finais, dado por:

$$x_j = a_{ij}, \forall j \in J \quad \text{Eq. 33}$$

No caso de contaminação por fonte dupla, para cada combinação (i, ℓ) , com $i < \ell$, considera-se o vetor de contribuição conjunta:

$$x_j = a_{ij} + a_{\ell j}, \forall j \in J \quad \text{Eq. 34}$$

Em seguida, aplica-se a função dose–resposta logística a cada elemento do vetor x , utilizando os parâmetros k , LI e LS previamente definidos, obtendo-se um vetor de probabilidades de contaminação detectável nos produtos finais. A partir dessas probabilidades, realizam-se $N = 1000$ simulações Monte Carlo, nas quais o estado de contaminação de cada produto final é gerado por variáveis aleatórias Bernoulli.

Cada simulação produz um vetor binário de resposta. Ao final das N repetições, calcula-se a média elemento a elemento desses vetores simulados, gerando um **vetor médio de resposta provável** associado à hipótese de contaminação considerada. Esse vetor representa, para cada produto final, a frequência média de contaminação detectável esperada sob aquela hipótese.

O procedimento é repetido para todas as matérias-primas individualmente e para todas as combinações duplas possíveis. Como a combinação (i, ℓ) produz o mesmo resultado que a combinação (ℓ, i) , armazena-se apenas uma das possibilidades, de modo que o mapa assume a forma de uma **matriz triangular superior**. Assim, cada posição dessa matriz contém o vetor médio de resposta provável correspondente a uma hipótese específica de contaminação de fonte única ou dupla.

Etapa 2 – Geração do vetor observado e busca no mapa

Após a construção do mapa, considera-se um vetor de observação de contaminação nos produtos finais, denotado por $\mathbf{y}^{obs}$. Esse vetor representa um cenário observado de contaminação e é gerado de forma probabilística a partir da mesma função dose–resposta logística, considerando-se uma hipótese verdadeira de contaminação por uma ou duas matérias-primas. Diferentemente da etapa anterior, porém, utiliza-se apenas **uma única simulação**, de modo que o vetor observado seja binário, composto apenas por zeros e uns.

Em seguida, mede-se a distância euclidiana entre o vetor observado $\mathbf{y}^{obs}$ e cada um dos vetores médios armazenados no mapa. Seja $\mathbf{m}_{i\ell}$ o vetor médio de resposta provável associado à hipótese (i, ℓ) . A distância euclidiana entre o vetor observado e essa hipótese é dada por:

$$d(i, \ell) = \sqrt{\sum_{j \in J} (y_j^{obs} - m_{i\ell,j})^2} \quad \text{Eq. 35}$$

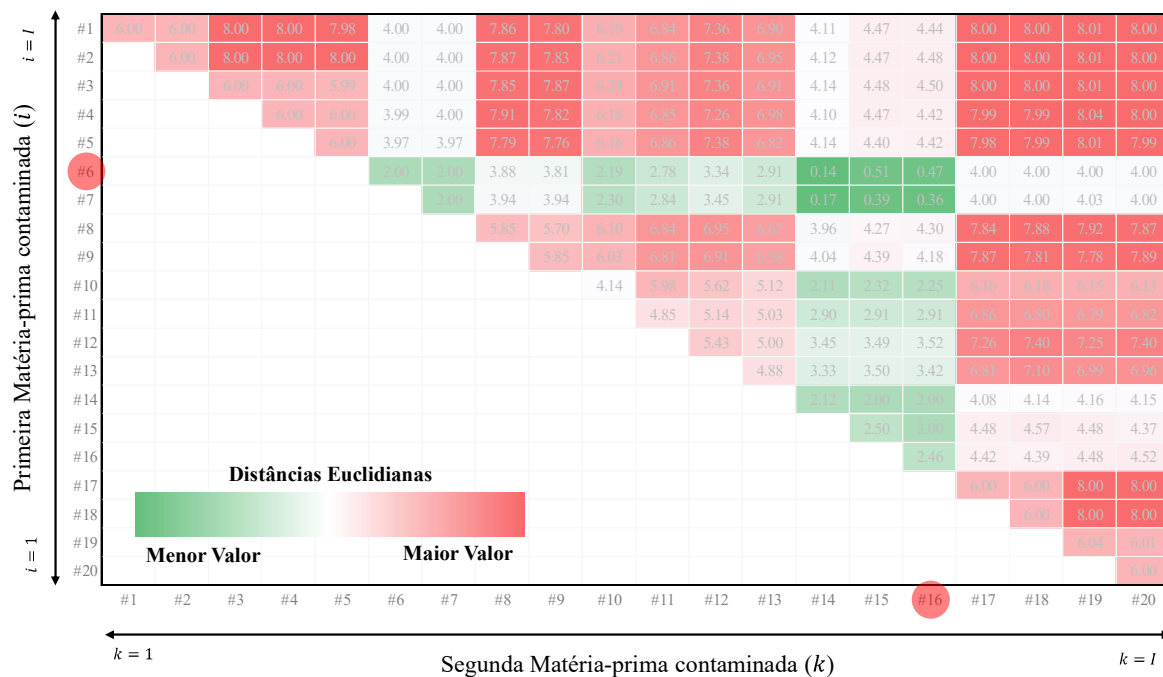
em que $m_{i\ell,j}$ representa o valor do produto final j no vetor médio armazenado no mapa para a combinação (i, ℓ) .

Os valores de distância calculados são então organizados em uma nova matriz triangular superior, com a mesma estrutura do mapa original, denominada **matriz de distâncias**. Cada elemento dessa matriz expressa o grau de dissimilaridade entre o padrão observado e a hipótese correspondente de contaminação. Quanto menor for o valor da distância euclidiana, maior é a similaridade entre o vetor observado e o vetor armazenado no mapa, indicando que aquela hipótese é mais compatível com o evento de contaminação observado.

Dessa forma, a primeira estimativa para a origem da contaminação é obtida pela busca do menor valor da matriz de distâncias. A posição desse menor valor identifica a matéria-prima, ou a combinação de matérias-primas, mais próxima do padrão observado. Na prática, essa hipótese indica quais lotes devem ser priorizados para análise laboratorial ou rastreamento detalhado.

Como exemplo para um melhor entendimento, segue na Figura 10 um corte da matriz triangular superior que contempla apenas 20 lotes. Cada valor dentro da matriz corresponde a distância euclidiana entre o vetor de observação de contaminação e os valores médios armazenados no mapa. Para o exemplo, a origem das contaminações então nos lotes #6 e #16.

Figura 10 – Exemplo gráfico do corte (Lote #1 até #20) do mapa de distâncias euclidianas quando a origem da contaminação está nos lotes #6 e #16.



Fonte: Autor (2026)

Etapa 3 – Busca iterativa até identificação das fontes

O menor valor da matriz de distâncias nem sempre corresponde exatamente à combinação verdadeira de matérias-primas contaminadas. Assim, o algoritmo adota um procedimento iterativo de busca por similaridade.

Se a hipótese associada ao menor valor de distância não contiver nenhuma das matérias-primas efetivamente contaminadas, esse valor é desconsiderado e o algoritmo passa a buscar o próximo menor valor da matriz.

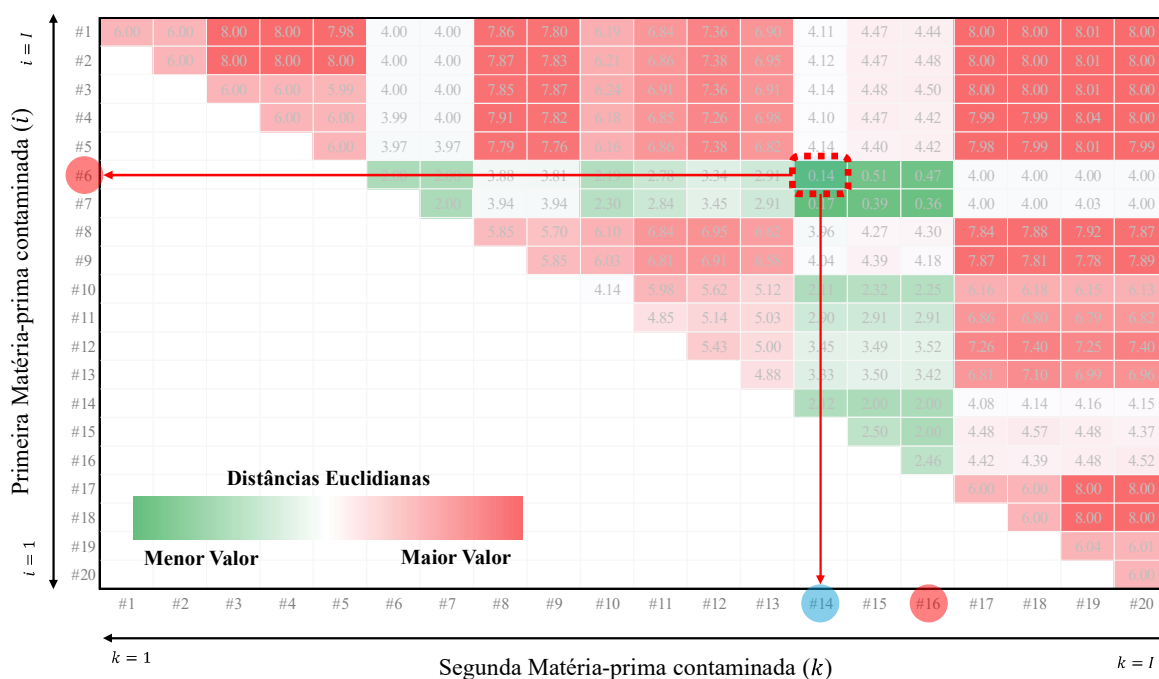
Se a hipótese associada ao menor valor contiver **uma** das matérias-primas corretas, considera-se que houve um acerto parcial. Nesse caso, restringe-se a busca às combinações que contenham a matéria-prima já identificada, reduzindo o espaço de busca e aumentando a probabilidade de localizar corretamente a segunda fonte de contaminação, quando houver. Em termos geométricos, na matriz triangular superior isso significa que a busca passa a se concentrar na linha ou coluna associada à matéria-prima já encontrada, em vez de considerar todas as combinações possíveis.

Se a hipótese associada ao menor valor contiver **as duas** matérias-primas corretas, a busca é encerrada, pois a origem da contaminação foi corretamente identificada.

Esse processo iterativo é repetido até que sejam encontradas ambas as fontes de contaminação, no caso de cenários de fonte dupla, ou a única fonte verdadeira, no caso de cenários de fonte única. O número de hipóteses examinadas até o acerto final constitui uma medida prática da eficiência do algoritmo, pois representa quantas matérias-primas ou combinações precisariam ser efetivamente investigadas na prática até localizar a origem do problema.

Novamente como exemplo, traz-se na Figura 11 o corte da matriz triangular superior que contempla apenas 20 lotes. Cada valor dentro da matriz corresponde a distância euclidiana entre o vetor de observação de contaminação e os valores médios armazenados no mapa. No exemplo gráfico, o método indicaria como primeira opção de testagem (menor distância euclidiana) as matérias-primas #6 e #14. Na prática, apenas a matéria-prima #6 teria a contaminação confirmada e a hipótese de contaminação pela matéria-prima #14 seria descartada.

Figura 11 – Exemplo gráfico da primeira iteração do corte (Lote #1 até #20) do mapa de distâncias euclidianas quando a origem da contaminação está nos lotes #6 e #16.

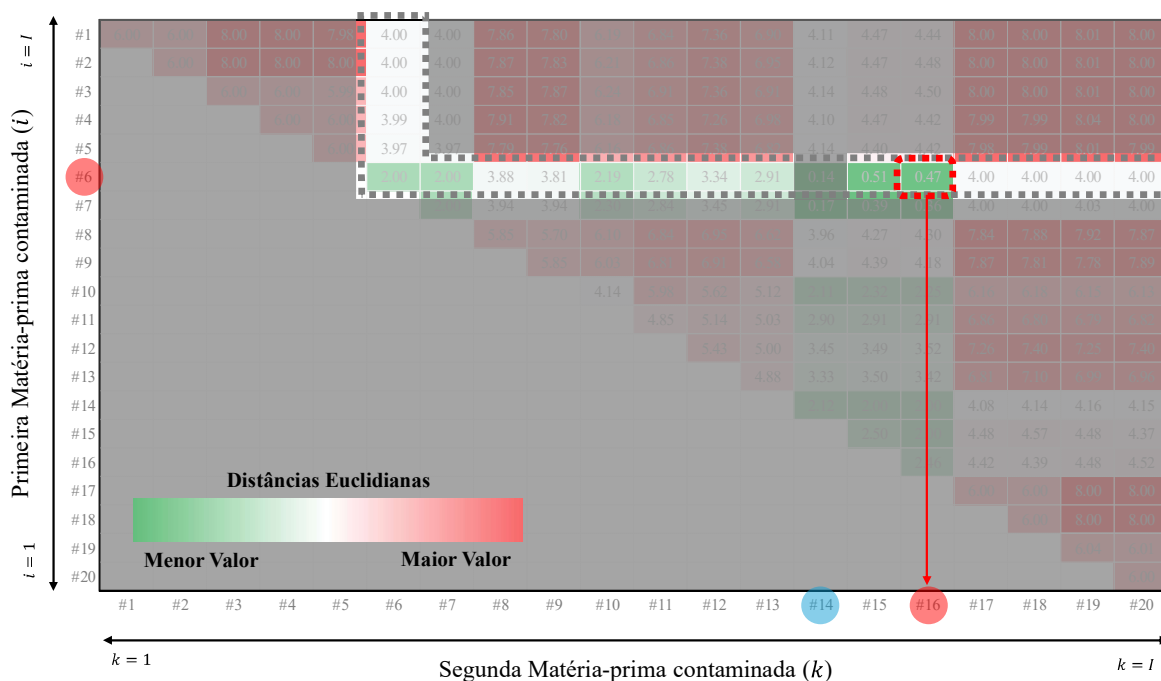


Fonte: Autor (2026)

Como passo seguinte na iteração para identificação da origem de contaminação, o método restringiria a busca às combinações que contenham a matéria-prima já identificada. Graficamente exemplificado na Figura 12, a busca passa a se concentrar na linha ou coluna associada à matéria-prima já encontrada. No exemplo, o método teria como próxima menor distância euclidiana a

combinação entre os lotes #6 e #16 e, na prática, o lote #16 seria testado e teria a contaminação confirmada encerrando a iteração para identificar as origens da contaminação.

Figura 12 – Exemplo gráfico da segunda iteração do corte (Lote #1 até #20) do mapa de distâncias euclidianas quando a origem da contaminação está nos lotes #6 e #16.



Etapa 4 – Avaliação do desempenho do algoritmo

Para avaliar o desempenho do método proposto, simulam-se múltiplos cenários de contaminação observada. Neste trabalho, foram geradas 1000 observações de contaminação por fonte única e 1000 observações de contaminação por fonte dupla. Para cada observação, aplica-se o procedimento de busca no mapa e registra-se a quantidade de matérias-primas ou combinações que precisaram ser analisadas até que a origem real da contaminação fosse corretamente encontrada.

Essa avaliação permite medir a eficiência do algoritmo em termos de priorização investigativa. Em vez de exigir a análise exaustiva de todas as matérias-primas ou de todas as combinações possíveis, o método propõe uma ordenação baseada em similaridade, permitindo direcionar inicialmente os esforços para as hipóteses mais prováveis. Dessa forma, o algoritmo contribui para reduzir o número de análises necessárias e, consequentemente, o custo e o tempo associados à investigação de eventos de contaminação em redes produtivas complexas.

Em síntese, o Algoritmo 2 complementa o Algoritmo 1 ao transformar os padrões simulados de contaminação em um mecanismo operacional de inferência. Enquanto o primeiro algoritmo permite distinguir entre cenários de uma ou duas fontes por meio do método dos vales, o segundo algoritmo utiliza essa estrutura probabilística para construir um mapa de referência e localizar, por similaridade, as matérias-primas mais prováveis de terem originado a contaminação observada. Assim, o método proposto oferece não apenas uma ferramenta de classificação, mas também um procedimento de apoio à decisão para direcionar a investigação prática das fontes contaminantes no sistema produtivo estudado.

3.4 SIMULAÇÕES COM DADOS DE UM CENÁRIO REAL

3.4.1 Descrição da cadeia produtiva e do estudo de caso

O estudo de caso considerado neste trabalho baseia-se em uma cadeia produtiva alimentar representativa, estruturada de forma sequencial e composta por múltiplas etapas de processamento nas quais diferentes lotes intermediários são gerados a partir da combinação de lotes do estágio anterior. Essa estrutura é consistente com cadeias produtivas reais do setor alimentício, caracterizadas por processos de mistura, fracionamento e recombinação de insumos ao longo do fluxo produtivo (REGATTIERI; GAMBERI; MANZINI, 2007; RINGSBERG, 2014).

A cadeia produtiva analisada é composta por três estágios principais de processamento, além da etapa inicial de recebimento de matérias-primas. No estágio inicial ($t = 0$), o sistema recebe um conjunto de n_0 matérias-primas distintas, cada uma representando um lote de entrada potencialmente rastreável. Nos estágios subsequentes ($t = 1, 2, 3$), essas matérias-primas são processadas e combinadas para gerar novos lotes intermediários e finais, conforme proporções registradas nos dados operacionais.

Em cada estágio t , os lotes são identificados como $L_{t,i}$, com $i = 1, 2, \dots, n_t$, onde n_t representa o número de lotes gerados naquele estágio. As relações de consumo entre lotes consecutivos são descritas por matrizes de transição probabilísticas, construídas a partir das proporções reais de uso dos lotes do processo t na produção dos lotes do processo $t + 1$. Essas matrizes são as mesmas definidas e validadas no Capítulo 2 (SALINAS *et al.*, 2024), garantindo consistência metodológica entre os estudos.

O conjunto de dados utilizado corresponde a registros históricos de produção de uma indústria alimentícia real, previamente anonimizada por confidencialidade. Os dados incluem

volumes produzidos, proporções de mistura entre lotes e sequenciamento temporal dos processos, permitindo reconstruir as matrizes de transição e as distribuições marginais em cada etapa.

Para fins de análise, assume-se que a contaminação é identificada apenas nos produtos finais ($t = 3$), refletindo cenários típicos em que alertas são disparados a partir de testes laboratoriais, reclamações ou notificações regulatórias após a distribuição do produto (BADIA-MELIS; MISHRA; RUIZ-GARCÍA, 2015; RINGSBERG, 2014). A partir dessas observações, aplica-se a rastreabilidade reversa probabilística para inferir as fontes de contaminação das matérias primas.

Este estudo de caso permite avaliar o algoritmo proposto em cenários com fonte única e múltiplas fontes simultâneas, analisando como a complexidade estrutural impacta a capacidade de inferência e a incerteza associada à identificação das origens. Essa perspectiva é coerente com o argumento central de Lu *et al.* (2019) de que estrutura e mistura aumentam o espaço de fontes plausíveis e, portanto, elevam a incerteza capturada pela entropia.

3.4.2 Definição dos cenários de contaminação

A avaliação da eficácia de métodos de rastreabilidade em cadeias produtivas complexas exige a consideração explícita de diferentes cenários de contaminação, uma vez que a dinâmica de propagação e a capacidade de inferência dependem do número, da localização e da interação entre origens contaminantes na rede produtiva (HOSSEINI; IVANOV, 2020). Em contextos reais de segurança de alimentos, eventos podem ter origem em uma única fonte ou em múltiplas fontes simultâneas decorrentes de falhas independentes em diferentes etapas do processo (BADIA-MELIS; MISHRA; RUIZ-GARCÍA, 2015).

Com o objetivo de avaliar o algoritmo sob diferentes níveis de complexidade, definem-se dois cenários principais de contaminação, ambos baseados na observação de produtos finais contaminados no último estágio ($t = 3$). Em todos os cenários, assume-se que a identificação ocorre após a finalização do processo produtivo, refletindo práticas usuais de monitoramento (RINGSBERG, 2014; REGATTIERI; GAMBERI; MANZINI, 2007).

Cenário A — Contaminação com fonte única.

Considera-se um evento originado em uma única fonte ao longo da cadeia, tipicamente associada a um lote de matéria-prima. Trata-se da hipótese clássica mais comum em formulações de localização de fonte, em que busca-se um único candidato mais provável (SHAH; ZAMAN, 2011; HORN; FRIEDRICH, 2019). A partir do conjunto de produtos finais contaminados, aplica-se a rastreabilidade reversa probabilística para cada produto, obtendo-se distribuições sobre os

possíveis lotes de origem (SALINAS *et al.*, 2024; HOSSEINI; IVANOV, 2020). Esse cenário é utilizado como linha de base para comparação.

Cenário B — Contaminação com duas fontes simultâneas.

Considera-se um evento com duas fontes independentes localizadas em lotes distintos. A presença de múltiplas fontes implica que produtos finais contaminados carregam contribuições combinadas provenientes de diferentes origens, elevando a complexidade da inferência por superposição de caminhos de propagação. Nessa condição, o algoritmo proposto agrega evidências provenientes de múltiplos produtos contaminados para construir uma distribuição global Φ^* , permitindo que múltiplos candidatos se destaquem simultaneamente, sem exigir que se fixe previamente o número ou localização das origens (SALINAS *et al.*, 2024; HOSSEINI; IVANOV, 2020; ZHANG *et al.*, 2019).

3.4.3 Implementação computacional

Toda a aplicação do algoritmo foi implementada em ambiente computacional, utilizando planilhas eletrônicas com suporte a macros no Microsoft Excel para a execução das simulações Monte Carlo e cálculo dos indicadores propostos. As matrizes de transição utilizadas como entrada foram previamente definidas no Capítulo 2 (SALINAS *et al.*, 2024).

As macros desenvolvidas automatizam as etapas de cálculo das probabilidades de contaminação, geração dos cenários simulados, classificação dos padrões de contaminação e cálculo dos indicadores de desempenho do modelo, incluindo o Índice de Bimodalidade (IB), o Índice de Detecção de Contaminação (IDC) e o Índice Composto de Contaminação (ICC).

Essa abordagem permite a execução sistemática do algoritmo para diferentes combinações de parâmetros, garantindo consistência nos resultados e rastreabilidade do processo de simulação. A utilização de ferramentas computacionais para automação de experimentos e análise de cenários é consistente com recomendações da literatura em modelagem probabilística aplicada a cadeias produtivas (ZHANG *et al.*, 2019), contribuindo para a reprodutibilidade e transparência dos resultados obtidos.

3.5 RESULTADOS E DISCUSSÕES

3.5.1 Desempenho do algoritmo 1 (detecção de múltiplas fontes)

O desempenho do Algoritmo 1 foi avaliado com base na sua capacidade de detectar a presença de contaminação e distinguir corretamente entre cenários de fonte única e de múltiplas fontes, utilizando os indicadores Índice de Detecção de Contaminação (IDC), Índice de Bimodalidade (IB) e Índice Composto de Contaminação (ICC). Em resumo, os índices encontrados são apresentados na Tabela 1, abaixo:

Tabela 1 – Desempenho global do algoritmo 1.

Cenário	IDC	IB	ICC
Fonte Única	98,36%	91,34%	94,85%
Fonte Dupla	99,99%	80,46%	90,23%

Fonte: Autor (2026)

Para o cenário de fonte única, considerando os parâmetros $k = 50$, $LI = 5\%$ e $LS = 30\%$, o modelo apresentou elevado desempenho global, com $IDC = 98,4\%$, $IB = 91,3\%$ e $ICC = 94,9\%$. Esses resultados indicam que o método é altamente eficaz tanto na detecção da contaminação quanto na preservação do comportamento monomodal esperado para esse tipo de configuração.

Tabela 2 – Análise individual do algoritmo 1 para diferentes cenários de contaminação única nas matérias-prima (MP) e suas respectivas acurácias de bimodalidade.

Cenário	Acurácia	Cenário	Acurácia	Cenário	Acurácia	Cenário	Acurácia	Cenário	Acurácia
MP #1	100,0%	MP #26	100,0%	MP #51	100,0%	MP #76	100,0%	MP #101	97,4%
MP #2	Sem contaminação	MP #27	100,0%	MP #52	97,7%	MP #77	100,0%	MP #102	100,0%
MP #3	Sem contaminação	MP #28	98,5%	MP #53	100,0%	MP #78	100,0%	MP #103	100,0%
MP #4	100,0%	MP #29	100,0%	MP #54	100,0%	MP #79	100,0%	MP #104	99,7%
MP #5	100,0%	MP #30	99,7%	MP #55	98,8%	MP #80	0,0%	MP #105	100,0%
MP #6	100,0%	MP #31	100,0%	MP #56	100,0%	MP #81	100,0%	MP #106	100,0%
MP #7	100,0%	MP #32	100,0%	MP #57	100,0%	MP #82	0,0%	MP #107	99,4%
MP #8	100,0%	MP #33	100,0%	MP #58	100,0%	MP #83	0,3%	MP #108	12,7%
MP #9	100,0%	MP #34	82,0%	MP #59	95,4%	MP #84	100,0%	MP #109	99,6%
MP #10	100,0%	MP #35	100,0%	MP #60	100,0%	MP #85	0,0%	MP #110	1,1%

MP #11	100,0%	MP #36	99,1%	MP #61	99,4%	MP #86	0,2%	MP #111	100,0%
MP #12	100,0%	MP #37	100,0%	MP #62	100,0%	MP #87	55,2%	MP #112	100,0%
MP #13	100,0%	MP #38	100,0%	MP #63	98,7%	MP #88	6,3%	MP #113	99,5%
MP #14	99,9%	MP #39	100,0%	MP #64	100,0%	MP #89	97,4%	MP #114	100,0%
MP #15	100,0%	MP #40	100,0%	MP #65	100,0%	MP #90	27,7%	MP #115	100,0%
MP #16	100,0%	MP #41	85,5%	MP #66	100,0%	MP #91	100,0%	MP #116	96,2%
MP #17	100,0%	MP #42	100,0%	MP #67	100,0%	MP #92	99,9%	MP #117	100,0%
MP #18	100,0%	MP #43	100,0%	MP #68	100,0%	MP #93	100,0%	MP #118	100,0%
MP #19	98,2%	MP #44	100,0%	MP #69	100,0%	MP #94	99,7%	MP #119	100,0%
MP #20	100,0%	MP #45	100,0%	MP #70	99,6%	MP #95	100,0%	MP #120	100,0%
MP #21	100,0%	MP #46	100,0%	MP #71	100,0%	MP #96	99,6%	MP #121	100,0%
MP #22	99,6%	MP #47	100,0%	MP #72	100,0%	MP #97	100,0%	MP #122	100,0%
MP #23	100,0%	MP #48	99,6%	MP #73	99,2%	MP #98	18,7%		
MP #24	100,0%	MP #49	100,0%	MP #74	100,0%	MP #99	100,0%		
MP #25	100,0%	MP #50	100,0%	MP #75	100,0%	MP #100	99,4%		

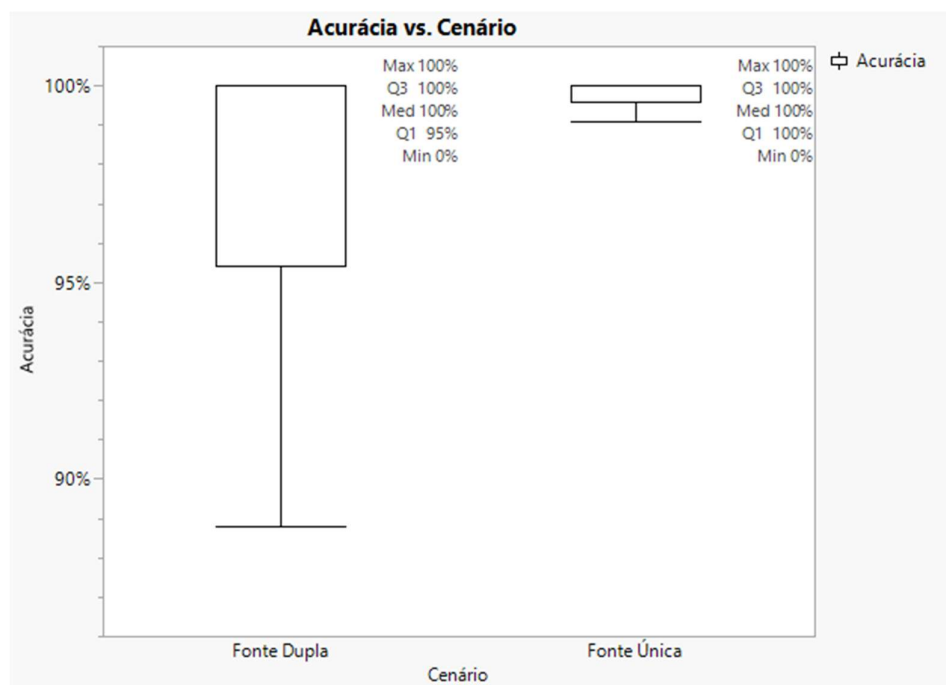
Fonte: Autor (2026)

A análise individual por matéria-prima, conforme apresentado na Tabela 2, evidencia que a acurácia do método é majoritariamente elevada, com grande parte das matérias-primas apresentando taxas próximas a 100%. No entanto, observa-se um comportamento heterogêneo para matérias-primas com baixa participação na composição dos produtos finais (matéria-prima #2 e matéria-prima #3). Nesses casos, quando a fração de contribuição encontra-se abaixo ou próxima ao limiar inferior (LI), o modelo frequentemente não detecta contaminação no produto final. Esse resultado é consistente com a formulação proposta, uma vez que pequenas contribuições tendem a permanecer na região de diluição da função dose–resposta, resultando em baixa probabilidade de detecção.

Adicionalmente, identificou-se uma limitação estrutural associada a matérias-primas que não seguem o sistema FIFO (*First In, First Out*). Nessas situações, um mesmo lote pode impactar diferentes produtos finais em períodos distintos, gerando padrões não monotônicos no vetor de resposta. Como consequência, o algoritmo pode detectar corretamente a presença de contaminação, mas apresentar ambiguidade na classificação entre fonte única e múltiplas fontes. Esse comportamento decorre do fato de que o modelo assume implicitamente uma propagação temporal contínua, não capturando explicitamente efeitos de recirculação ou defasagem temporal dos lotes. Esta condição está associada aos lotes de matéria-prima: 80, 82, 83, 85, 86, 87, 88, 89, 90, 91, 93, 98, 107, 108 e 110.

Para o cenário com duas fontes simultâneas, o modelo apresentou desempenho ainda mais elevado na detecção de contaminação, com $IDC \approx 99,99\%$, evidenciando alta sensibilidade à presença de múltiplas origens. O índice de bimodalidade foi de $IB \approx 80,46\%$, resultando em um $ICC \approx 90,23\%$, o que indica boa capacidade de identificação estrutural, embora inferior ao cenário de fonte única.

Figura 13 – *Boxplot* dos resultados de acurácia para o algoritmo 1 para ambos os cenários de contaminação única e dupla.

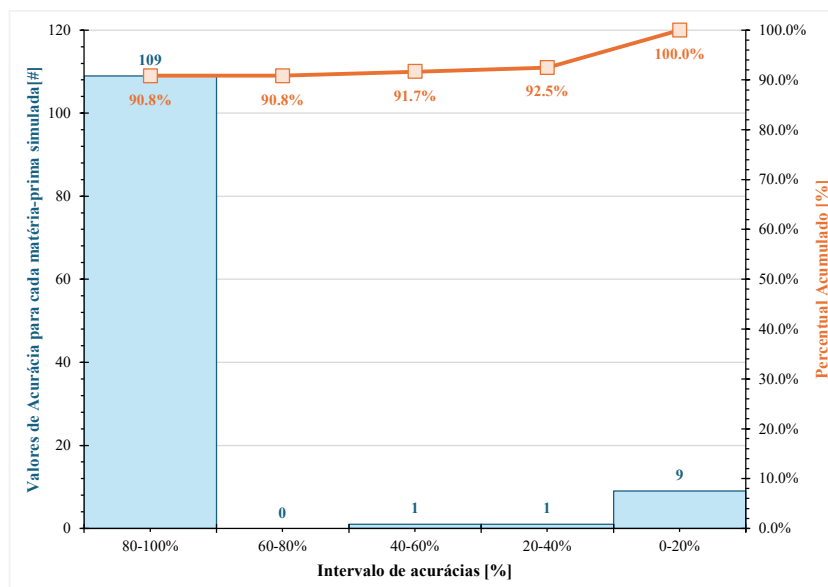


Fonte: Autor (2026)

A distribuição da acurácia do Algoritmo 1 para o cenário de fonte única, apresentada na Figura 14, evidencia forte concentração de valores próximos a 100%, com mediana igual ou muito próxima de 1. Esse comportamento indica elevada robustez do método para a maioria das matérias-primas.

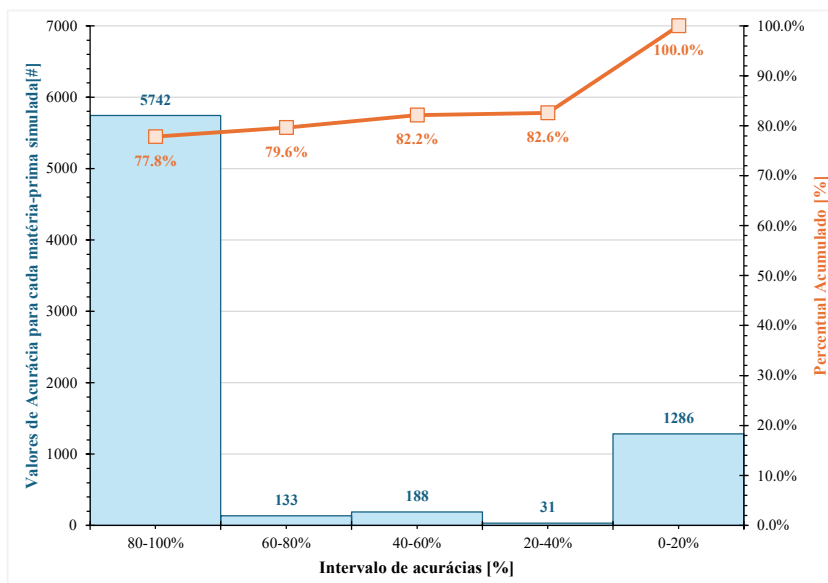
Observam-se, entretanto, alguns *outliers* com baixa acurácia, associados a matérias-primas de baixa participação na composição dos produtos finais. Esses casos corroboram a limitação do modelo em detectar contaminações cuja contribuição se encontra abaixo ou próxima ao limiar inferior da função dose–resposta.

Figura 14 – Histograma dos resultados de acurácia para o algoritmo 1 para o cenário de contaminação única.



Fonte: Autor (2026)

Figura 15 – Histograma dos resultados de acurácia para o algoritmo 1 para o cenário de contaminação dupla.



Fonte: Autor (2026)

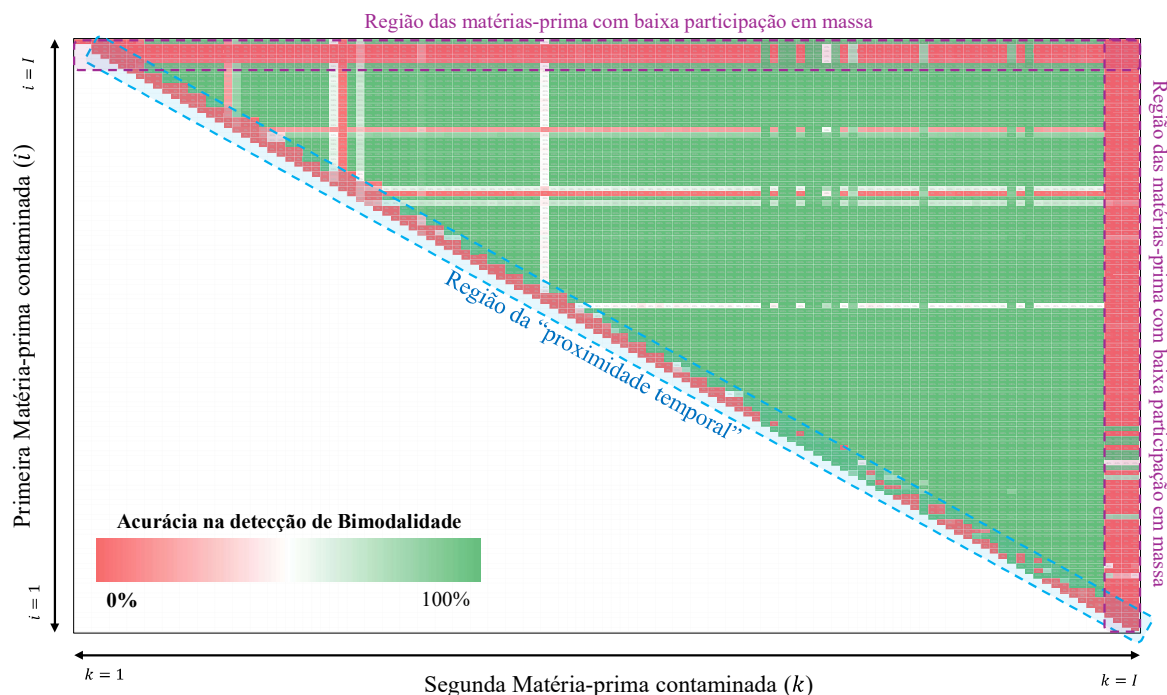
O histograma da acurácia reforça esse comportamento, indicando forte concentração de resultados na faixa superior (acima de 90%), com uma cauda inferior composta por poucos casos de baixo desempenho. Essa assimetria reflete a dependência do modelo em relação à contribuição de massa das matérias-primas no produto final.

A análise da matriz triangular superior de acurácia, representada através da Figura 16, revela que a maioria das combinações de matérias-primas apresenta desempenho elevado, frequentemente superior a 90%. Esse resultado confirma que, de forma geral, a superposição de duas fontes de contaminação aumenta significativamente a probabilidade de detecção no produto final, uma vez que a soma das contribuições tende a deslocar o sistema para regiões acima do limiar superior (LS) da função dose–resposta.

Entretanto, alguns padrões específicos de falha foram identificados. Para combinações envolvendo matérias-primas com baixa participação em massa, o modelo frequentemente detecta contaminação, mas a classifica como sendo de fonte única. Esse comportamento ocorre porque a contribuição combinada das duas fontes pode ainda permanecer próxima à região de transição da função sigmoidal, não gerando padrões suficientemente distintos para caracterizar bimodalidade.

Além disso, em situações em que as duas fontes seguem o sistema FIFO e apresentam proximidade temporal, ou seja, $i \approx k$, observou-se dificuldade na distinção entre uma e duas fontes. Nesses casos, a propagação das contribuições ocorre de forma sobreposta ao longo do tempo, resultando em padrões suavizados que reduzem a ocorrência de “vales” no vetor de resposta.

Figura 16 – Matriz de triângulo superior com os resultados de acurácia para o cálculo do índice de bimodalidade do Algoritmo 1.



Fonte: Autor (2026)

Esses resultados evidenciam que, embora o método seja altamente eficaz na detecção de contaminação, a capacidade de distinção estrutural entre uma e duas fontes depende de características específicas da rede produtiva, como a distribuição de massa das matérias-primas e a dinâmica temporal dos fluxos.

Do ponto de vista prático, os resultados sugerem que, em aplicações industriais, a investigação da origem da contaminação deve considerar não apenas a hipótese mais provável identificada pelo modelo, mas também um conjunto de candidatos próximos. Em particular, recomenda-se a análise de um intervalo de um a dois lotes antes e após a origem estimada, estratégia que aumenta a robustez da identificação frente às limitações associadas à dinâmica temporal e à mistura estrutural do sistema.

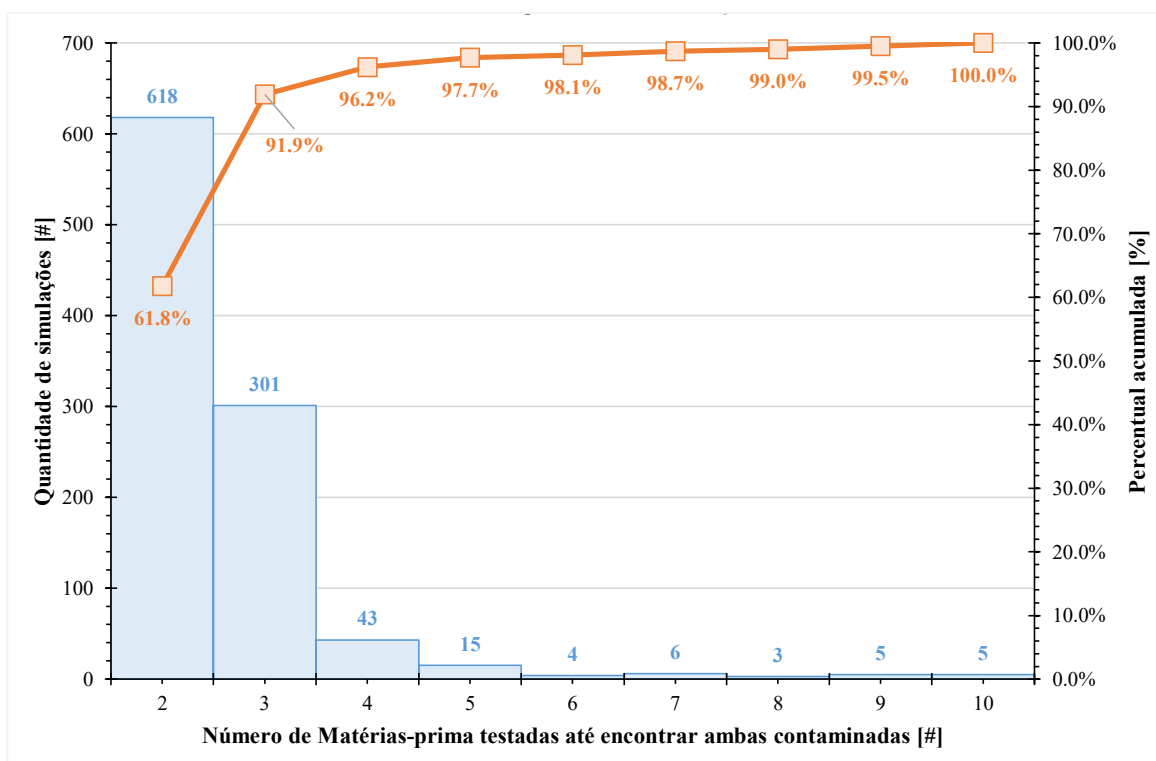
Em síntese, o Algoritmo 1 demonstrou elevada capacidade de detecção de contaminação e boa performance na distinção entre cenários de fonte única e múltiplas fontes, ao mesmo tempo em que revelou limitações importantes associadas à baixa contribuição de massa e à não observância de estruturas temporais do tipo FIFO. Essas observações são fundamentais para a interpretação dos resultados e para a aplicação prática do método em cadeias produtivas reais.

3.5.2 Desempenho do algoritmo 2 (detecção da origem da contaminação)

Os resultados do Algoritmo 2 foram avaliados a partir de simulações de Monte Carlo ($N = 1000$ para cada cenário), considerando cenários com uma única fonte de contaminação, dupla origem e uma análise combinada de ambos os casos. O objetivo foi avaliar não apenas a acurácia do método, mas também o esforço necessário para identificação das matérias-primas contaminadas e os principais fatores associados aos erros de classificação.

Inicialmente, para o cenário de **dupla fonte de contaminação**, observa-se que o algoritmo apresenta elevado desempenho na identificação das matérias-primas contaminadas. Conforme ilustrado na Figura 17, aproximadamente 61,8% das simulações identificam ambas as fontes após a análise de apenas duas matérias-primas, enquanto cerca de 91,9% dos casos são resolvidos com até três verificações. A curva de percentual acumulado indica que mais de 99% das simulações convergem para a solução correta com até cinco testes, evidenciando a eficiência do método mesmo em um cenário de maior complexidade combinatória.

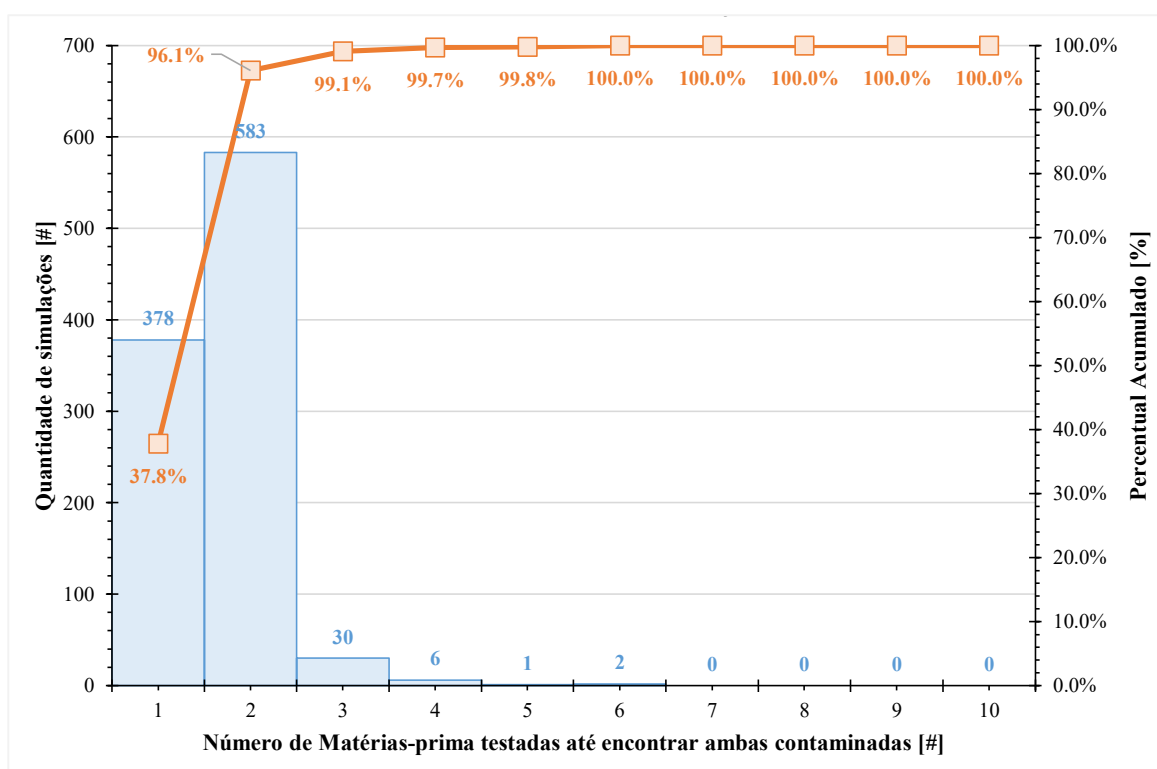
Figura 17 – Distribuição do número de matérias-primas testadas até a identificação correta de duas fontes de contaminação, com curva de percentual acumulado ($N = 1000$).



Fonte: Autor (2026)

Para o caso de **fonte única de contaminação**, os resultados indicam um comportamento ainda mais eficiente. A Figura 18 demonstra que cerca de 96,1% das simulações identificam corretamente a fonte com apenas duas verificações, atingindo mais de 99% de acerto com até três análises. Esse resultado reforça que, em cenários menos complexos, o algoritmo mantém elevada precisão com baixo esforço de rastreabilidade.

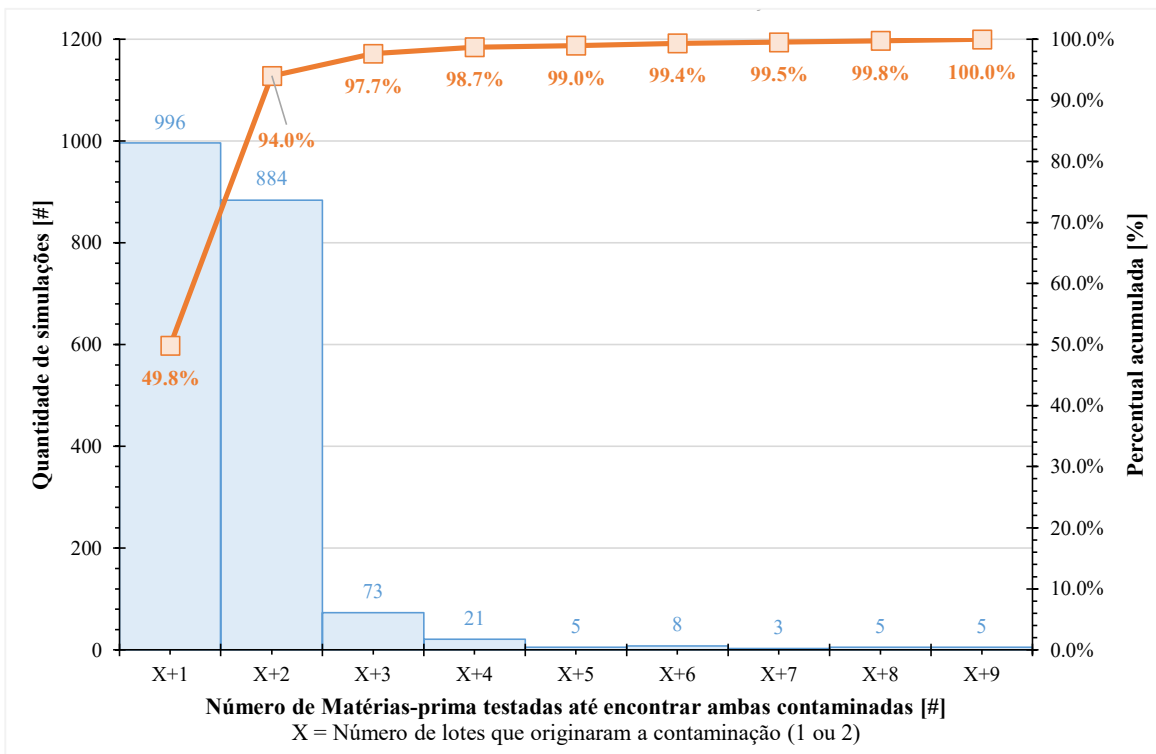
Figura 18 – Distribuição do número de matérias-primas testadas até a identificação da fonte única de contaminação, com curva de percentual acumulado (N = 1000).



Fonte: Autor (2026)

A análise conjunta dos cenários, apresentada na Figura 19, permite comparar diretamente o impacto da complexidade do problema no desempenho do algoritmo. Observa-se que, mesmo ao considerar simultaneamente casos com uma e duas fontes de contaminação, o método mantém alta eficiência: aproximadamente 94% das simulações são resolvidas com até duas verificações adicionais em relação ao número real de fontes ($X+2$), enquanto mais de 99% convergem com até quatro verificações. Esse resultado evidencia a robustez do algoritmo frente à incerteza quanto ao número de fontes de contaminação.

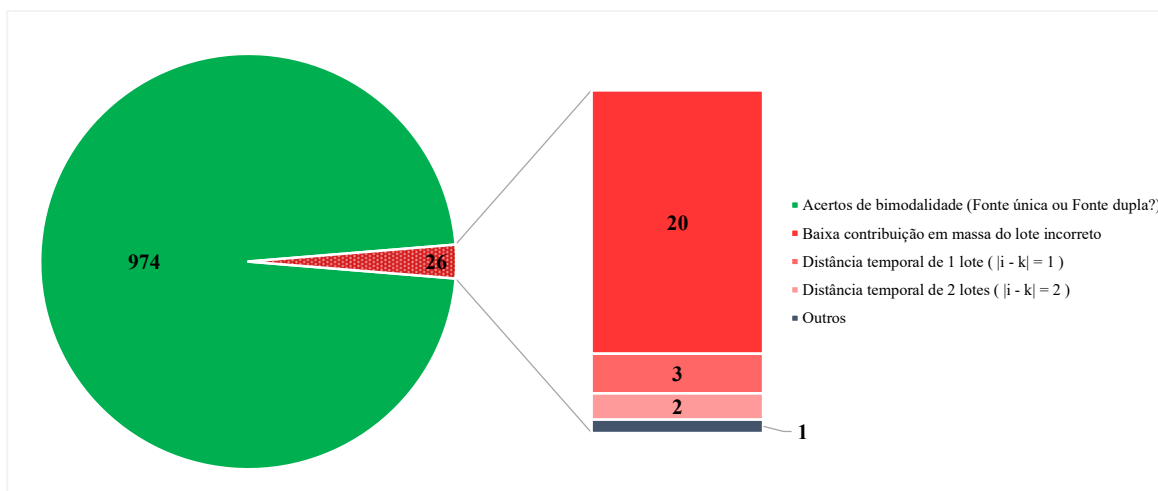
Figura 19 – Distribuição do número de matérias-primas testadas até a identificação das fontes de contaminação em cenários combinados (fonte única e dupla), com curva de percentual acumulado (N = 1000).



Fonte: Autor (2026)

Além da análise de desempenho, foi realizada uma **estratificação dos erros de classificação**, permitindo compreender as limitações do modelo. Para o cenário de dupla fonte, a taxa de acerto na determinação da bimodalidade é elevada (97,4%), sendo os erros residuais predominantemente associados à baixa contribuição em massa das matérias-primas contaminadas. Esse comportamento indica que, quando a participação relativa da fonte contaminada no produto final é reduzida, o sinal probabilístico torna-se menos discriminante, dificultando sua identificação.

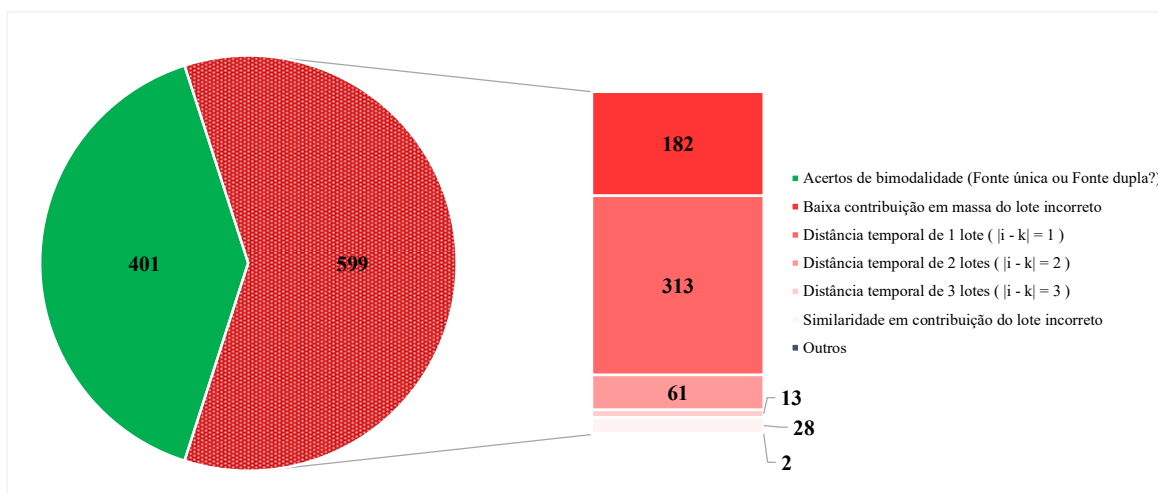
Figura 20 – Estratificação dos erros na determinação da bimodalidade para cenários com dupla fonte de contaminação (N = 1000).



Fonte: Autor (2026)

Por outro lado, no cenário de fonte única, observa-se uma redução significativa na taxa de acerto da determinação da bimodalidade (40,1%), com predominância de classificações incorretas como dupla fonte. A análise detalhada dos erros revela que esse comportamento está principalmente associado à similaridade entre perfis de contribuição e à proximidade temporal entre lotes, fatores que geram padrões de resposta semelhantes aos observados em cenários de contaminação múltipla. Esse resultado evidencia uma limitação importante do modelo, especialmente em sistemas nos quais há sobreposição significativa entre fluxos de produção.

Figura 21 – Estratificação dos erros na determinação da bimodalidade para cenários com fonte única de contaminação (N = 1000).



Fonte: Autor (2026)

Esse resultado também pode ser explicado pela própria estrutura do espaço de busca. No mapa de cenários gerado, apenas 122 das 7503 combinações possíveis correspondem a casos de contaminação por uma única fonte, evidenciando um forte desbalanceamento em favor de cenários com múltiplas fontes. Essa assimetria faz com que padrões associados à contaminação dupla sejam estatisticamente mais frequentes, aumentando a probabilidade de associação indevida nesses casos. Assim, a maior dificuldade na identificação de fontes únicas está diretamente relacionada à baixa representatividade desses cenários no conjunto de soluções possíveis.

De forma geral, os resultados demonstram que o Algoritmo 2 é altamente eficaz na identificação de matérias-primas contaminadas, apresentando elevada acurácia e baixo esforço médio de busca, especialmente quando o número de fontes é conhecido ou quando a contribuição em massa das fontes contaminadas é significativa. No entanto, a distinção entre uma e duas fontes de contaminação mostra-se sensível a características estruturais do sistema, como distribuição de massas e proximidade temporal entre lotes, o que deve ser considerado em aplicações práticas.

3.5.3 Implicações práticas para a gestão industrial

Do ponto de vista gerencial, os algoritmos propostos neste estudo oferecem uma estrutura operacional clara para suporte à tomada de decisão em eventos de contaminação em cadeias produtivas complexas. Diferentemente de abordagens tradicionais, baseadas em rastreabilidade determinística e frequentemente associadas a ações amplas e conservadoras, como recalls generalizados, o método desenvolvido permite a priorização probabilística de hipóteses de origem, reduzindo significativamente o esforço investigativo.

Na prática, a aplicação do método pode ser estruturada em três etapas principais. Inicialmente, a partir da identificação de produtos finais contaminados, o gestor pode utilizar o **Algoritmo 1** para inferir se o evento está associado a uma única fonte ou a múltiplas fontes de contaminação. Essa distinção é crítica, pois define a estratégia de investigação: enquanto cenários de fonte única permitem buscas direcionadas, cenários com múltiplas fontes exigem abordagens combinatórias e maior abrangência analítica.

Na sequência, o **Algoritmo 2** permite ranquear, por similaridade probabilística, as matérias-primas ou combinações de matérias-primas mais prováveis de terem originado o evento. Esse ranqueamento transforma um problema originalmente combinatório, potencialmente envolvendo milhares de hipóteses, em uma lista ordenada de investigação, na qual os primeiros candidatos concentram maior probabilidade de acerto. Como evidenciado nos resultados, na maioria dos casos

a identificação correta das fontes ocorre após a análise de um número reduzido de hipóteses, frequentemente inferior a três ou quatro combinações.

Essa característica tem implicações diretas na operação industrial. Em vez de mobilizar recursos para análise extensiva de todos os lotes possíveis, o gestor pode direcionar testes laboratoriais, auditorias de processo e inspeções operacionais de forma priorizada, reduzindo custos, tempo de resposta e impacto logístico. Em contextos regulatórios, como segurança alimentar e farmacêutica, essa abordagem também contribui para recalls mais precisos e menos disruptivos.

Além disso, o método pode ser utilizado de forma preventiva. Ao simular cenários de contaminação e analisar a resposta do sistema produtivo, torna-se possível identificar regiões da cadeia com baixa identificabilidade, ou seja, pontos em que diferentes fontes produzem padrões observacionais semelhantes. Essas regiões representam riscos estruturais do ponto de vista de rastreabilidade e podem orientar melhorias no desenho do processo, como redução de mistura excessiva, aumento da segregação de lotes ou aprimoramento da coleta de dados.

Por fim, a abordagem probabilística proposta permite que a tomada de decisão seja realizada de forma explícita sob incerteza, fornecendo não apenas uma hipótese mais provável, mas também a distribuição de plausibilidade entre diferentes cenários. Isso representa um avanço relevante em relação a modelos determinísticos, pois permite ao gestor balancear risco, custo e tempo de resposta de maneira quantitativa e fundamentada.

3.6 CONCLUSÃO

Este artigo teve como objetivo propor uma abordagem probabilística para a identificação de múltiplas fontes de contaminação em cadeias produtivas complexas, ampliando o arcabouço metodológico desenvolvido no Artigo 1 para cenários mais realistas do ponto de vista industrial.

Os resultados obtidos demonstram que a introdução de múltiplas fontes altera qualitativamente o problema de rastreabilidade, tornando-o estruturalmente mais complexo devido à combinação de efeitos como mistura de fluxos, não unicidade de solução e crescimento combinatório do espaço de hipóteses. Nesse contexto, a abordagem proposta mostrou-se adequada para lidar com essa complexidade ao integrar modelagem probabilística da propagação de contaminação, simulação estocástica e mecanismos de inferência baseados em similaridade.

O Algoritmo 1 evidenciou elevada capacidade de detecção de contaminação e bom desempenho na distinção entre cenários de fonte única e dupla, embora com limitações associadas à baixa contribuição de massa e a efeitos estruturais como não conformidade com FIFO. Já o

Algoritmo 2 demonstrou alta eficiência na identificação das fontes de contaminação, sendo capaz de reduzir significativamente o número de hipóteses a serem investigadas na prática, mesmo em cenários com elevada complexidade combinatória.

Do ponto de vista teórico, o trabalho contribui ao propor uma formulação que integra três elementos raramente tratados de forma conjunta na literatura: (i) múltiplas fontes simultâneas, (ii) transformação e mistura multiperíodo de fluxos e (iii) inferência probabilística baseada em evidências agregadas. Essa integração posiciona o problema de rastreabilidade industrial como um caso particular de problema inverso estocástico em redes com conservação de massa, ampliando o escopo das abordagens existentes.

Sob a ótica prática, os resultados indicam que o método pode ser utilizado como ferramenta de apoio à decisão em ambientes industriais, permitindo priorizar investigações, reduzir custos operacionais e aumentar a eficiência na resposta a eventos críticos. A capacidade de transformar um problema combinatório em um processo ordenado de busca probabilística representa um avanço relevante para aplicações em segurança alimentar, biotecnologia e processos químicos.

Como limitações, destaca-se a sensibilidade do modelo a fatores estruturais da rede produtiva, como distribuição de massa e proximidade temporal entre lotes, bem como o desbalanceamento intrínseco entre cenários de fonte única e múltiplas fontes. Esses aspectos indicam oportunidades de aprimoramento metodológico.

Como trabalhos futuros, recomenda-se a extensão do modelo para múltiplas fontes superiores a duas, a incorporação explícita de dinâmica temporal não FIFO, bem como a integração com técnicas de aprendizado de máquina para calibração automática dos parâmetros da função dose-resposta e melhoria da capacidade preditiva.

Em síntese, este artigo avança no estado da arte ao propor um arcabouço probabilístico robusto e aplicável para rastreabilidade em cadeias produtivas complexas, contribuindo tanto para o desenvolvimento teórico quanto para a aplicação prática em contextos industriais reais.

3.7 REFERÊNCIAS

ALTARELLI, F.; BRAUNSTEIN, A.; DALL'ASTA, L.; LAGE-CASTELLANOS, A.; ZECCHINA, R. **Bayesian inference of epidemics on networks via belief propagation**. *Physical Review Letters*, v. 112, n. 11, 118701, 2014. <https://doi.org/10.1103/PhysRevLett.112.118701>

BADIA-MELIS, R.; MISHRA, P.; RUIZ-GARCÍA, L. **Food traceability: New trends and recent advances – A review**. *Food Control*, v. 57, p. 393–401, 2015. <https://doi.org/10.1016/j.foodcont.2015.05.005>.

BOUZEMBRAK, Y.; MARVIN, H. J. P. **Prediction of food fraud type using data from Rapid Alert System for Food and Feed (RASFF) and Bayesian network modelling.** *Food Control*, v. 61, p. 180–187, 2018. <https://doi.org/10.1016/j.foodcont.2015.09.026>.

BOUZEMBRAK, Y.; LIU, N.; MU, W.; GAVAI, A.; MANNING, L.; BUTLER, F.; MARVIN, H. J. P. **Data driven food fraud vulnerability assessment using Bayesian Network: Spices supply chain.** *Food Control*, v. 164, 2024. <https://doi.org/10.1016/j.foodcont.2024.110616>

BRAUNSTEIN, A; INGROSSO, A. **Inference of causality in epidemics on temporal contact networks.** *Scientific Reports*, v. 6, 27538, 2016. DOI: 10.1038/srep27538

BU, F; AIELLO, A. E.; XU, J; VOLFOVSKY, A. **Likelihood-based inference for partially observed epidemics on networks.** *Journal of the American Statistical Association*, v. 117, n. 537, p. 510–526, 2020. <https://doi.org/10.1080/01621459.2020.1790376>

CAI, J.; YE, Z. **Contamination Source Identification: A Bayesian Framework Integrating Physical and Statistical Models.** *IEEE Transactions on Industrial Informatics*, vol. 17, no. 12, pp. 8189-8197, 2021, doi: 10.1109/TII.2021.3062146.

CHAI, Y.; WANG, Y.; ZHU, L. **Information sources estimation in time-varying networks.** *IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY*, v. 16, p. 2621–2636, 2021. DOI: 10.1109/TIFS.2021.3050604

CHANG, B.; Chen, E.; Zhu, F.; Liu, Q.; Xu, T.; Wang, Z. **Maximum a Posteriori Estimation for Information Source Detection.** *IEEE Transactions on Systems, Man and Cybernetics: Systems*. 50, (6), 2242-2256. DOI: 10.1109/TSMC.2018.2811410

COMBA, L.; BELFIORE, F.; DABBENE, F.; GAY, P. Methods for traceability in food production processes involving bulk products. *Biosystems Engineering*, v. 116, n. 1, p. 51–63, 2013.

DE SANCTIS, A. E.; Boccelli, D. L.; SHANG, F. UBER, J. G. **Probabilistic approach to characterize contamination sources with imperfect sensors.** *World Environmental and Water Resources Congress 2008: Ahupua'a*, 2008. [https://doi.org/10.1061/40976\(316\)512](https://doi.org/10.1061/40976(316)512)

FATIMA, N.; RIZVI, S. A. **Identifying infection origins in evolving networks: a comprehensive review.** *Medical Data Mining*, v. 8, issue 2, no 12, 2025. <https://doi.org/10.53388/MDM202508012>

GOMEZ-RODRIGUEZ, M; LESKOVEC, J; KRAUSE, A. **Inferring networks of diffusion and influence.** *ACM Transactions on Knowledge Discovery from Data*, v. 5, n. 4, Arigo 21, 2012. <http://doi.acm.org/10.1145/2086737.2086741>

HORN, A. L.; FRIEDRICH, H. **Locating the source of large-scale outbreaks of foodborne disease.** *Journal of The Royal Society Interface*, v. 16, n. 151, 2019. DOI: 10.1098/rsif.2018.0624.

HOSSEINI, S.; IVANOV, D. **Bayesian networks for supply chain risk, resilience and ripple effect analysis: A literature review.** *Expert Systems with Applications*, v. 161, 2020. <https://doi.org/10.1016/j.eswa.2020.113649>.

IFT – INSTITUTE OF FOOD TECHNOLOGISTS. Traceability (Product Tracing) in Food Systems: An IFT Report Submitted to the FDA. *Comprehensive Reviews in Food Science and Food Safety*, v. 8, n. 1, p. 92–158, 2009.

JAYNES, E. T.; BRETTHORST, G. L. **Probability theory: The logic of science**. Cambridge: Cambridge University Press, 2003.

LIU, H.; BAO, Q.; QIU, H.; ZU, M.; SHI, B. **Source identification of asymptomatic spread on networks**. *IEEE Access*, v. 9, p. 34142 - 34155, 2021. DOI: 10.1109/ACCESS.2021.3061599

LU, X; HORN, A. L.; SU, J.; JIANG, J. **A universal measure for network traceability**. *Omega*, v. 87, p. 191–204, 2019. <https://doi.org/10.1016/j.omega.2018.09.004>

MUNTONI, A. P.; MAZZA, F.; BRAUNSTEIN, A.; CATANIA, G.; DALL’ASTA, L. **Effectiveness of probabilistic contact tracing in epidemic containment: The role of superspreaders and transmission path reconstruction**, *PNAS Nexus*, Volume 3, Issue 9, p. 377, 2024. <https://doi.org/10.1093/pnasnexus/pgae377>

PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B. **Estatística Bayesiana**. 2. ed. Lisboa: Fundação Calouste Gulbenkian, 2019.

REGATTIERI, A.; GAMBERI, M.; MANZINI, R. **Traceability of food products: General framework and experimental evidence**. *Journal of Food Engineering*, v. 81, n. 2, p. 347–356, 2007. <https://doi.org/10.1016/j.jfoodeng.2006.10.032>.

RINGSBERG, H. **Perspectives on food traceability: A systematic literature review**. *Supply Chain Management: An International Journal*, v. 19, n. 5/6, p. 558–576, 2014. <https://doi.org/10.1108/SCM-01-2014-0026>

RUSKEY, B.; ROSENBERG, E. **Minimizing risk in Bayesian supply chain networks**. *Computers & Operations Research*, v. 141, 2022. <https://doi.org/10.1016/j.cie.2022.108134>.

SALINAS, H. D. S. S.; Martins, R. P.; Campos, I. M.; Galdamez, E. V. C.; Pezzott, G. L. M. **Aplicação da probabilidade condicional na Rastreabilidade de produtos alimentícios**. XLIV ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 2024.

SHAGHAGHIAN, S.; COATES, M. Bayesian inference of diffusion networks with unknown infection times. *IEEE Statistical Signal Processing Workshop (SSP)*, 2016. <https://doi.org/10.48550/arXiv.1602.08114>

SHAGHAGHIAN, S.; COATES, Mark. **Online Bayesian inference of diffusion networks**. *IEEE Transactions on Signal Processing*, v. 65, n. 4, p. 992–1005, 2017. <https://doi.org/10.48550/arXiv.1611.01086>

SHAH, D.; ZAMAN, T. **Rumors in a network: Who’s the culprit?** *IEEE Transactions on Information Theory*, v. 57, n. 8, p. 5163–5181, 2011. <https://doi.org/10.1109/TIT.2011.2158885>.

SHANNON, C. E. **A mathematical theory of communication**. *Bell System Technical Journal*, v. 27, p. 379–423, 1948. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.

STERCHI, M.; HILFIKER, L.; GRÜTTER, R.; BERNSTEIN, A. **Active querying approach to epidemic source detection on contact networks**. *Scientific Reports, Nature Portfolio*, 13:11363, 2023. <https://doi.org/10.1038/s41598-023-38282-8>

XU, J. Unravelling transmission dynamics of infectious diseases with Bayesian and Network analysis. 2025. Tese (Doutorado) – University of Georgia, 2015

XU, J., HU, H., ELLISON, G.; YU, L.; WHALEN, C. C.; LIU, L. **Bayesian estimation of transmission networks for infectious diseases**. *J. Math. Biol.* 90, 29 (2025). <https://doi.org/10.1007/s00285-025-02193-1>

XUANYU, Y.; ANNI, L.; QIANG, C.; JUN, L. **Bayesian network modeling applied to food risks: Data from General Administration of Customs of China as an example**, *Applied Food Research*, Volume 4, Issue 2, 2024, 100565, ISSN 2772-5022, <https://doi.org/10.1016/j.afres.2024.100565>.

YANG, L; HUANG, B.; LIU, J. **Identification of illicit discharges in sewer networks by an SWMM-Bayesian coupled approach**. *Water Science & Technology*, v. 90, n 3, 951, 120601, 2024. doi: 10.2166/wst.2024.233

YANG, X. **Real-time probabilistic contamination source identification in water distribution systems**. Tese (Doutorado) – University of Cincinnati, Cincinnati, 2013.

ZHANG, H.; FENG, H.; CUI, Y.; WANG, Y. **A Fuzzy Bayesian Network Model for Quality Control in O2O e-Commerce**. *INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL*, v. 15, 2020. <https://doi.org/10.15837/ijccc.2020.1.3783>.

ZHANG, X.-T.; SU, J.-H.; FANG, M.; JIANG, J.; CHEN, Y.-W.; LU, X. **Modeling and Bayesian Inference of Traceability in Food Supply Networks**. *Scientific Reports*, v. 9, n. 1, 2019.

ZHOU, F.; XU, X.; ZHANG, K.; TRAJCEVSKI, G.; ZHONG, T. **Variational information diffusion for probabilistic cascades prediction**. *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications*, 2020. DOI: 10.1109/INFOCOM41043.2020.9155349

ZHU, P.; CHENG, L.; GAO, C.; WANG, Z.; LI, X. **Locating multi-sources in social networks with a low infection rate**. *IEEE Transactions on Network Science and Engineering*, v. 9, n. 3, p. 1853–1865, 2022. DOI: 10.1109/TNSE.2022.3153968

CONSIDERAÇÕES FINAIS

A presente dissertação teve como objetivo propor e desenvolver uma abordagem probabilística para a rastreabilidade em cadeias produtivas complexas, com ênfase na modelagem de incertezas, na inferência reversa e no suporte à tomada de decisão em cenários de eventos adversos, como contaminações e desvios de qualidade. A pesquisa foi estruturada no formato multiartigo, permitindo a integração entre uma aplicação empírica consolidada e o desenvolvimento de uma extensão metodológica voltada a cenários de maior complexidade.

No âmbito da introdução, foi possível evidenciar que a crescente complexidade das cadeias produtivas modernas, caracterizadas por múltiplos fornecedores, processos interdependentes e mistura de fluxos ao longo do tempo, impõe limitações relevantes aos modelos tradicionais de rastreabilidade, predominantemente determinísticos. Esses modelos, embora úteis para reconstrução de vínculos diretos entre lotes, mostram-se insuficientes para capturar a incerteza inerente aos processos produtivos e a multiplicidade de trajetórias plausíveis associadas à propagação de eventos adversos.

Nesse contexto, o Artigo 1 demonstrou, de forma aplicada, que a integração do Teorema de Bayes com Cadeias de Markov constitui uma abordagem eficaz para modelar a rastreabilidade direta e reversa em cadeias produtivas reais. A construção das matrizes de transição e o uso de probabilidades condicionais permitiram representar a propagação de materiais como distribuições probabilísticas, superando a rigidez das abordagens determinísticas. Como principal contribuição, o modelo possibilitou não apenas a identificação de fontes prováveis de contaminação, mas também a quantificação explícita da incerteza associada a essa inferência, fornecendo suporte mais robusto à tomada de decisão em ambientes industriais.

A partir dessas evidências, o Artigo 2 avançou ao tratar explicitamente cenários com múltiplas fontes simultâneas de contaminação, reconhecendo que, em sistemas produtivos reais, eventos adversos podem ter origens múltiplas e interdependentes. A introdução de múltiplas fontes alterou qualitativamente o problema de rastreabilidade, ampliando o espaço de hipóteses e

intensificando desafios como não unicidade de solução, mistura estrutural e crescimento combinatório. Para lidar com essa complexidade, foi proposto um arcabouço metodológico que integra modelagem probabilística da propagação de contaminação, simulação Monte Carlo e algoritmos de inferência baseados em padrões estruturais e similaridade.

Os resultados obtidos indicam que a abordagem proposta é capaz de detectar com alta precisão a presença de contaminação e distinguir, de forma consistente, entre cenários de fonte única e múltiplas fontes. Adicionalmente, o método demonstrou elevada eficiência na identificação das matérias-primas contaminadas, reduzindo significativamente o número de hipóteses que precisam ser investigadas na prática. Esses resultados reforçam o potencial da abordagem como ferramenta de apoio à decisão em contextos industriais, especialmente em situações críticas que exigem respostas rápidas e fundamentadas, como recalls e investigações de segurança alimentar.

Do ponto de vista teórico, a dissertação contribui ao posicionar o problema de rastreabilidade como um problema inverso estocástico em redes com transformação e mistura de fluxos, integrando conceitos de inferência bayesiana, cadeias de Markov e teoria da informação. A introdução de múltiplas fontes e a consideração explícita da mistura multiperíodo representam avanços relevantes em relação à literatura existente, que, em sua maioria, permanece limitada a cenários de fonte única ou a modelos de difusão sem transformação estrutural.

Sob a ótica prática, os resultados demonstram que a rastreabilidade probabilística pode transformar a forma como organizações lidam com eventos adversos. Em vez de abordagens amplas e conservadoras, o gestor passa a dispor de um mecanismo quantitativo para priorizar investigações, otimizar recursos e reduzir impactos operacionais e financeiros. Além disso, a possibilidade de utilização do modelo em análises preventivas permite identificar fragilidades estruturais da cadeia produtiva, contribuindo para o desenho de sistemas mais robustos e rastreáveis.

Como limitações, destaca-se que o modelo proposto depende da qualidade e disponibilidade de dados operacionais para a construção das matrizes de transição, bem como da correta parametrização da função dose-resposta. Adicionalmente, fatores como não conformidade com estruturas FIFO, baixa contribuição de massa e alta sobreposição de fluxos podem impactar a capacidade de inferência, devendo ser considerados na interpretação dos resultados.

Como perspectivas futuras, recomenda-se a ampliação do modelo para cenários com múltiplas fontes superiores a duas, a incorporação explícita de dinâmicas temporais mais complexas, como recirculação de lotes, e a integração com técnicas de aprendizado de máquina para calibração automática de parâmetros e melhoria da capacidade preditiva. A aplicação do

modelo em diferentes setores industriais também representa uma oportunidade relevante para validação externa e ampliação de sua generalização.

REFERÊNCIAS

- ANAND, G.; GAVAI, A.; KAPTAN, M. **Applying Federated Learning to Combat Food Fraud in Food Supply Chains**. International Journal of Food Science and Technology, v. 58, p. 1073-1087, 2023.
- AUNG, M. M.; CHANG, Y. S. **Traceability in a food supply chain: Safety and quality perspectives**. Food Control, v. 39, p. 172-184, 2014. doi: 10.1016/j.foodcont.2013.11.007.
- AZIZSAFAEI, M. **A systemic approach for risk and resilience assessment of food supply chains**. University of Northampton, 2023. Disponível em: northampton.ac.uk.
- BAI, C.; SARKIS, J. **Flexibility in reverse logistics: A framework and evaluation approach**. Journal of Cleaner Production, v. 47, p. 306-318, 2013. doi: 10.1016/j.jclepro.2012.08.011.
- BASNET, S.; BAHOOTOROODY, A.; MONTEWKA, J.; CHAAL, M. **Selecting cost-effective risk control options for advanced maritime operations**. Ocean Engineering, v. 281, p. 115295, 2023.
- BECHINI, A.; CIMINO, M. G. C. A.; MARCELLONI, F.; TOMASI, A. **Patterns and technologies for enabling supply chain traceability through collaborative e-business**. Information and Software Technology, v. 50, n. 4, p. 342-359, 2008. doi: 10.1016/j.infsof.2007.02.017.
- BECKMAN, C.; BERNANDER, E. **Traceability in legal pharmaceutical supply chains-ensuring safety and quality of prescribed medicinal products**. Chalmers University of Technology, 2015.
- BEHNKE, K.; JANSSEN, M. **Boundary conditions for traceability in food supply chains using blockchain technology**. International Journal of Information Management, v. 52, p. 101-112, 2020.
- BORIT, M.; OLSEN, P. **Seafood traceability systems: Gap analysis of inconsistencies in standards and norms**. FAO Fisheries and Aquaculture Circular, v. 1134, p. 1-12, 2016.
- BOSONA, T.; GEBRESENBET, G. **Food traceability as an integral part of logistics management in food and agricultural supply chain**. Food Control, v. 33, n. 1, p. 32-48, 2013. doi: 10.1016/j.foodcont.2013.02.004
- BOUZEMBRAK, Y.; CAMENZULI, L.; JANSSEN, E.; VAN DER FELS-KLERX, H. J. **Application of Bayesian networks in the development of herbs and spices sampling monitoring system**. Food Control, v. 83, p. 38-44, 2018.
- BOUZEMBRAK, Y.; MARVIN, H. J. P. **Impact of Drivers of Change, Including Climatic Factors, on the Occurrence of Chemical Food Safety Hazards in Fruits and Vegetables: A Bayesian Network Approach**. Food Control, v. 97, p. 67-76, 2019.
- BOUZEMBRAK, Y.; LIU, N.; MU, W.; GAVAI, A.; MANNING, L.; BUTLER, F.; MARVIN, H. J. P. **Data driven food fraud vulnerability assessment using Bayesian network: Spices**

supply chain. Food Control, v. 164, p. 110616, 2024.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento (MAPA). **Instrução Normativa nº 2, de 9 de janeiro de 2018.** Diário Oficial da União. Disponível em: <https://www.in.gov.br>.

BRASIL. Agência Nacional de Vigilância Sanitária (Anvisa). **Resolução da Diretoria Colegiada - RDC nº 24, de 8 de junho de 2015.** Diário Oficial da União. Disponível em: <https://www.in.gov.br>.

CAI, L. **Product safety and security in the global supply chain.** Journal of Operations Management, 2011.

CASINO, F.; KANAKARIS, V.; DASAKLIS, T. K. **Blockchain-based food supply chain traceability: A case study in the dairy sector.** International Journal of Production Research, v. 59, n. 1, p. 141-151, 2021. doi: 10.1080/00207543.2020.1789238.

CHEN, R. **Blockchain application in food supply information security.** Journal of the Knowledge Economy, v. 8, n. 2, p. 530-545, 2017. doi: 10.1007/s13132-017-0482-5.

CHRISTOPHER, M. **Logistics & supply chain management.** Pearson, 2016.

CHRISTOPHER, M.; HOLWEG, M. **Supply chain 2.0 revisited: A framework for managing volatility-induced risk in the supply chain.** International Journal of Physical Distribution & Logistics Management, v. 47, n. 9, p. 7-19, 2017.

DABBENE, F.; GAY, P.; TORTIA, C. **Traceability issues in food supply chains: A review.** Biosystems Engineering, v. 120, p. 65-80, 2014. doi: 10.1016/j.biosystemseng.2013.09.006.

FOSCHI, M. **Tracciabilità geografica di prodotti alimentari tradizionali attraverso profili multi-elementali e spettroscopici combinati alla chemiometria.** University of Bologna, 2021.

GO FAIR. FAIR Principles. Disponível em: <https://www.go-fair.org/fair-principles/>. Acesso em: 10 maio 2026.

HALEEM, A.; KHAN, S.; KHAN, M. I. **Traceability implementation in food supply chain: A grey-DEMATEL approach.** Information Processing in Agriculture, v. 6, n. 4, p. 360-368, 2019. doi: 10.1016/j.inpa.2019.03.001.

HAOUD, N. E. L.; HASNAOUI, M. E. L. **Supply chain and Industry 4.0: Impact and performance analysis: Case of Biomerieux.** Proceedings of the IEEE International Conference on Industrial Engineering and Operations Management, p. 1-6, 2019.

HOSSEINI, S.; IVANOV, D. **Bayesian networks for supply chain risk, resilience and ripple effect analysis: A literature review.** Expert Systems with Applications, v. 130, p. 61-75, 2020. doi: 10.1016/j.eswa.2019.03.005.

HOSSEINI, S.; IVANOV, D. **A new resilience measure for supply networks with the ripple effect considerations: A Bayesian network approach.** Annals of Operations Research, v. 177, p. 29-45, 2022. doi: 10.1007/s10479-019-03350-8.

HULT, G. T. M. **Global supply chain management: Leveraging processes, measurements, and tools for strategic corporate advantage**. McGraw-Hill Education, 2011.

IACOBELLI, M. **An analysis of supplier delivery performance: The FCA case**. Politecnico di Torino, 2018.

ISO. **ISO 9000:2015 – *Quality management systems — Fundamentals and vocabulary***. 2015

ISO. **ISO 22005: Traceability in the feed and food chain – General principles and basic requirements for system design and implementation**. 2011.

JI, G.; HU, L.; TAN, K. H. **A Study on Decision-Making of Food Supply Chain Based on Big Data**. Journal of Systems Science and Systems Engineering, v. 26, no. 2, p. 183-198, 2017.

KAPTAN, M.; BAYAZIT, O. **Fuzzy Bayesian Network Analysis of the Factors Causing Food Losses in Reefer Containers**. Food Control, v. 134, p. 104671, 2023.

KELEPOURIS, T.; PRAMATARI, K.; DOUKIDIS, G. I. **RFID-enabled traceability in the food supply chain**. Industrial Management & Data Systems, v. 107, n. 2, p. 183-200, 2007. doi: 10.1108/02635570710723804.

KUMAR, S.; SCHMITZ, S. **Managing recalls in a consumer product supply chain – Root cause analysis and measures to mitigate risks**. International Journal of Production Research, v. 49, n. 1, p. 235-253, 2011. doi: 10.1080/00207540903436605.

LIU, N.; LIU, C.; DUDAS, T. N.; LOC, M. C.; BAGI, F. F.; VAN DER FELS-KLERX, H. J. **Improved Aflatoxins and Fumonisin Forecasting Models for Maize (PREMA and PREFUM), Using Combined Mechanistic and Bayesian Network Modeling—Serbia as a Case Study**. Frontiers in Microbiology, v. 12, Article ID 643604, 2021.

LIU, N.; VAN DER FELS-KLERX, H. J.; DUDAS, T. N.; BAGI, F. F. **Automated Food Safety Early Warning System in the Dairy Supply Chain Using Machine Learning**. Frontiers in Microbiology, v. 12, Article ID 643604, 2021.

LIU, J.; CAO, X.; ZHOU, H.; LI, L.; ZHAO, P. **A digital twin-driven approach towards traceability and dynamic control for processing quality**. Advanced Engineering Informatics, v. 48, p. 101295, 2021. doi: 10.1016/j.aei.2021.101295.

LOCKAMY III, A.; MCCORMACK, K. **Modeling supplier risks using Bayesian networks**. Industrial Management & Data Systems, v. 112, n. 2, p. 227-246, 2012. doi: 10.1108/02635571211204317.

MA, J.; SHENG, Q. Z.; XIE, D.; CHUAH, J. M.; QIN, Y. **Efficiently managing uncertain data in RFID sensor networks**. World Wide Web, v. 18, n. 1, p. 173-194, 2015. doi: 10.1007/s11280-014-0283-3.

MANNING, L.; SOON, J. M. **Food safety, food fraud, and food defense: A fast evolving literature**. Journal of Food Science, v. 81, n. 4, p. R823-R834, 2016. doi: 10.1111/1750-3841.13256.

MARVIN, H. J. P.; VAN DER FELS-KLERX, H. J.; BOULOGNE, T.; CAMENZULI, L. **Expert-driven Methodology to Assess and Predict the Effects of Drivers of Change on Vulnerabilities in a Food Supply Chain**. Food Research International, v. 134, p. 109423, 2020.

MOE, T. **Perspectives on traceability in food manufacture**. Trends in Food Science & Technology, v. 9, n. 5, p. 211-214, 1998. doi: 10.1016/S0924-2244(98)00037-5.

MUSA, A.; GUNASEKARAN, A.; YUSUF, Y. **Supply chain product visibility: Methods, systems and impacts**. Expert Systems with Applications, v. 41, n. 1, p. 176-194, 2014. doi: 10.1016/j.eswa.2013.07.020.

NAGARAJAN, R.; SCUTARI, M.; LÈBRE, S. **Bayesian Networks in R with Applications in Systems Biology**. 2013. Use R!, Vol. 48, Springer (US)

OLSEN, P.; BORIT, M. **How to define traceability**. Trends in Food Science & Technology, v. 29, n. 2, p. 142-150, 2013. doi: 10.1016/j.tifs.2012.10.003.

OPARA, L. U.; MDITSHWA, A. **Food traceability from field to plate**. Outlook on Agriculture, v. 30, n. 1, p. 47-64, 2001. doi: 10.5367/000000001101293724.

OPARA, L. U.; MDITSHWA, A. **A review on the role of packaging in securing food system: Adding value to food products and reducing losses and waste**. Food Research International, v. 50, n. 1, p. 98-108, 2013. doi: 10.1016/j.foodres.2012.09.052.

PATRO, P. K. **Blockchain-based solutions for product recall and traceability management in supply chain**. 2022.

QAZI, A.; QUIGLEY, J.; DICKSON, A. **Evaluation of control strategies for managing supply chain risks using Bayesian Belief Networks**. IEEE International Conference on Industrial Engineering and Systems Management (IESM), 2015. doi: 10.1109/IESM.2015.7380298.

REBOLLEDO, C.; PACHÉ, G.; MENG, Q.; GRANT, D. B. **Diffusion of sustainable supply chain management: Toward a conceptual framework**. Supply Chain Forum, v. 13, n. 1, p. 24-37, 2012.

REGATTIERI, A.; GAMBERI, M.; MANZINI, R. **Traceability of food products: General framework and experimental evidence**. Journal of Food Engineering, v. 81, n. 2, p. 347-356, 2007. doi: 10.1016/j.jfoodeng.2006.10.032.

RINGSBERG, H. **Implementation of global traceability standards: Incentives and opportunities**. British Food Journal, v. 117, n. 7, p. 1950-1966, 2015.

RUSKEY, B.; ROSENBERG, E. **Minimizing Risk in Bayesian Supply Chain Networks**. International Journal of Production Research, v. 60, no. 14, p. 4323-4340, 2022.

SAKIB, N.; HOSSAIN, N. U. I.; NUR, F.; TALLURI, S.; JARADAT, R.; LAWRENCE, J. M. **An Assessment of Probabilistic Disaster in the Oil and Gas Supply Chain Leveraging Bayesian Belief Network**. International Journal of Production Economics, v. 235, Article ID 108107, 2021.

SARKAR, S. **Digital traceability of pharmaceutical drugs in supply chain**. International Journal, 2022.

SARKAR, S.; RASHIDI, K.; GÖLGEÇI, I.; GLIGOR, D. M. **Exploring pillars of supply chain competitiveness: Insights from leading global supply chains**. Production Planning & Control, v. 35, n. 3, p. 47-64, 2024.

SCUTARI, M.; DENIS, J.-B. **Bayesian Networks with Examples in R**. 2. ed. Texts in Statistical Science, Chapman & Hall/CRC, 2021.

SILVA, R. B.; MATTOS, C. A. **Critical success factors of a drug traceability system for creating value in a pharmaceutical supply chain (PSC)**. International Journal of Environmental Research and Public Health, v. 16, n. 11, p. 1972, 2019. doi: 10.3390/ijerph16111972.

SHARMA, S. K.; ROUTROY, S.; CHANDA, U. **Supply-side risk modelling using Bayesian network approach**. Supply Chain Forum: An International Journal, v. 23, n. 4, p. 298-309, 2022. doi: 10.1080/16258312.2021.1988697.

SOBERANIS, I. E. D. **An extended Bayesian network approach for analyzing supply chain disruptions**. 2010.

SONG, J.; LIU, H.; HUANG, X. **A Dynamic Source Tracing Method for Food Supply Chain Quality and Safety Based on Big Data**. Journal of Food Quality, v. 2022, Article ID 3495678, 2022.

SOON, J. M.; WAHAB, I. R. **A Bayesian Approach to Predict Food Fraud Type and Point of Adulteration**. Foods, v. 11, no. 3, Article ID 328, 2022.

STORØY, J.; THAKUR, M.; OLSEN, P. **The TraceFood Framework – Principles and guidelines for implementing traceability in food value chains**. Journal of Food Engineering, v. 115, n. 3, p. 375-384, 2013.

TANG, C. S.; TOMLIN, B. **The power of flexibility for mitigating supply chain risks**. International Journal of Production Economics, v. 116, n. 1, p. 12-27, 2008. doi: 10.1016/j.ijpe.2008.07.008.

THAKUR, M.; HURBURGH, C. R. **Framework for implementing traceability system in the bulk grain supply chain**. Journal of Food Engineering, v. 95, n. 4, p. 617-626, 2009. doi: 10.1016/j.jfoodeng.2009.07.022.

WHO. **Substandard, spurious, falsely labelled, falsified and counterfeit (SSFFC) medical products**. Organização Mundial da Saúde, 2011. Disponível em: <https://www.who.int>.

WILKINSON, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, v. 3, 160018, 2016.

WOGNUM, P. M.; TRIENEKENS, J. H.; VAN DER VORST, J. G. A. J.; BLOEMHOF, J. M.; VAN DER MEER, D. **Systems for sustainability and transparency of food supply chains – Current status and challenges**. Advanced Engineering Informatics, v. 25, n. 1, p. 65-76, 2011. doi: 10.1016/j.aei.2010.06.001.

YANG, J.; LIU, H. **Research of Vulnerability for Fresh Agricultural-Food Supply Chain Based on Bayesian Network**. International Journal of Agricultural Economics, v. 45, p. 123-135, 2018.

ZENODO. Zenodo: a universal repository for research outputs. OpenAIRE/CERN. Disponível em: <https://www.openaire.eu/zenodo-guide>. Acesso em: 10 maio 2026.

ZHANG, H.; CUI, Y. **A model combining a Bayesian network with a modified genetic algorithm for green supplier selection**. Simulation, v. 95, n. 7, p. 651-662, 2019. doi: 10.1177/0037549719826306.

ZHANG, H.; FENG, H.; CUI, Y.; WANG, Y. **A Fuzzy Bayesian Network Model for Quality Control in O2O e-Commerce**. International Journal of Computers Communications & Control, v. 15, no. 1, Article ID 1003, 2020.

ZHANG, M.; HU, H.; ZHAO, X. **Developing product recall capability through supply chain quality management**. International Journal of Production Economics, v. 223, p. 107529, 2020. doi: 10.1016/j.ijpe.2020.107529.