v7.n11.2026

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

PIPELINE DE DADOS MULTIORIGEM PARA MONITORAMENTO AGRÍCOLA DO POLO PETROLINA-PE/JUAZEIRO-BA

MULTI-SOURCE DATA PIPELINE FOR AGRICULTURAL MONITORING OF THE PETROLINA-PE/JUAZEIRO-BA HUB

PIPELINE DE DATOS MULTIORIGEN PARA EL MONITOREO AGRÍCOLA DEL POLO PETROLINA-PE/JUAZEIRO-BA

Wagner Lucas Araujo Miranda[1], Mateus Amorim Silva[2]

e7119474

https://doi.org/10.47820/recima21.v7i11.9474

PUBLICADO: 11/2026

RESUMO

O polo Petrolina-PE/Juazeiro-BA é o principal centro de fruticultura irrigada do semiárido brasileiro, mas os dados públicos sobre produção, clima, água e preços estão dispersos em portais com formato, periodicidade e política de acesso próprios, o que impede cruzá-los de forma reproduzível. Com previsão de El Niño muito forte para 2026, medir a resposta da safra local a eventos passados ajuda a antecipar o próximo. O objetivo é construir e verificar uma base integrada que permita testar essa relação. O método compreendeu seis etapas: levantamento de sete fontes públicas e de suas vias de acesso; extração para uma camada bruta com registro de procedência; padronização com regras de tratamento declaradas; consolidação em modelo dimensional com grão definido por tabela de fato; construção de um indicador de anomalia acumulada de precipitação; e testes de correlação de Pearson com controle de tendência. A base resultante reúne a produção municipal de 1974 a 2024 e é consultada por um painel analítico. Apenas a uva apresentou associação significativa com a anomalia acumulada de precipitação ($r = -0,24$; $p = 0,018$; $n = 97$), com produtividade menor em períodos de chuva acima da média; para a manga, o resultado foi inconclusivo ($p = 0,056$). O índice ENSO defasado em um ano não antecipou a safra seguinte, e a associação entre cotações do agronegócio e valor da produção da manga desaparece sob variação anual. Conclui-se que o efeito climático precisa ser avaliado por cultura e que a base integrada torna essa verificação reproduzível.

PALAVRAS-CHAVE: Engenharia de dados. Modelagem dimensional. Fruticultura irrigada. Variabilidade climática. Vale do São Francisco.

ABSTRACT

The Petrolina-PE/Juazeiro-BA hub is the main irrigated fruit-growing center of the Brazilian semiarid region, but public data on production, climate, water and prices are scattered across portals with their own format, periodicity and access policy, which prevents crossing them reproducibly. With a very strong El Niño forecast for 2026, measuring how the local harvest responded to past events helps anticipate the next one. This article aims to build and verify an integrated database that allows this relationship to be tested. The method comprised six stages: survey of seven public sources and their access routes; extraction into a raw layer with provenance records; standardization under declared treatment rules; consolidation into a dimensional model with a declared grain per fact table; construction of an accumulated precipitation anomaly indicator; and Pearson correlation tests with trend control.

[1] Bacharel em Ciência da Computação pela Faculdade de Petrolina (FACAPE), com atuação em engenharia de dados.
[2] Mestre em Engenharia da Computação e professor do curso de Ciência da Computação da Faculdade de Petrolina (FACAPE).

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

The resulting database holds municipal production from 1974 to 2024 and is queried through an analytical dashboard. Only grape showed a significant association with the accumulated precipitation anomaly (r = −0.24; p = 0.018; n = 97), with lower yield in periods of above-average rainfall; for mango, the result was inconclusive (p = 0.056). The ENSO index lagged by one year did not anticipate the following harvest, and the association between agribusiness stock prices and the value of mango production disappears under year-over-year change. The climate effect must therefore be assessed per crop, and the integrated database makes this verification reproducible.

KEYWORDS: *Data engineering. Dimensional modeling. Irrigated fruit farming. Climate variability. São Francisco Valley.*

RESUMEN

El polo Petrolina-PE/Juazeiro-BA es el principal centro de fruticultura de regadío del semiárido brasileño, pero los datos públicos sobre producción, clima, agua y precios están dispersos en portales con formato, periodicidad y política de acceso propios. Con un pronóstico de El Niño muy fuerte para 2026, medir la respuesta de la cosecha local a eventos pasados ayuda a anticipar el próximo. El objetivo es construir y verificar una base integrada que permita probar esa relación. Se siguieron seis etapas: relevamiento de siete fuentes públicas y de sus vías de acceso; extracción hacia una capa bruta con registro de procedencia; estandarización con reglas de tratamiento declaradas; consolidación en un modelo dimensional con granularidad definida por tabla de hechos; construcción de un indicador de anomalía acumulada de precipitación; y pruebas de correlación de Pearson con control de tendencia. La base resultante reúne la producción municipal de 1974 a 2024 y alimenta un panel analítico. Solo la uva presentó asociación significativa con la anomalía acumulada de precipitación (r = −0,24; p = 0,018; n = 97), con menor productividad en períodos de lluvia por encima del promedio; para el mango, el resultado fue no concluyente (p = 0,056). El índice ENSO desfasado un año no anticipó la cosecha siguiente, y la asociación entre cotizaciones del agronegocio y valor de la producción del mango desaparece bajo variación anual. Se concluye que el efecto climático debe evaluarse por cultivo y que la base integrada hace reproducible esa verificación.

PALABRAS CLAVE: *Ingeniería de datos. Modelado dimensional. Fruticultura de regadío. Variabilidad climática. Valle del São Francisco.*

1. INTRODUÇÃO

O polo formado por Petrolina, em Pernambuco, e Juazeiro, na Bahia, é o principal centro de fruticultura irrigada do semiárido brasileiro: concentra a produção de manga e uva de mesa do submédio São Francisco e responde por parte relevante do emprego agrícola e da pauta de exportação de frutas do país (OLIVEIRA, 2021). O Nordeste produziu 81% da manga nacional em 2023, com Bahia, Pernambuco e o Vale do São Francisco à frente desse total (VIDAL, 2024), e a Produção Agrícola Municipal (PAM) do IBGE cobre esses municípios desde 1974, o que oferece cinco décadas de área, quantidade e valor da produção por município (IBGE, 2024).

Essa produção depende da irrigação a partir do rio São Francisco, cuja vazão varia com um regime de chuvas irregular. Entre 2010 e 2016, o semiárido atravessou a seca mais longa

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

das últimas décadas (MARENGO *et al.*, 2018), com perdas que motivaram discussão sobre adaptação e segurança alimentar (MARENGO *et al.*, 2022); a vazão da bacia caiu de forma prolongada entre 1991 e 2020 (BARBOSA; BURITI, 2024), e os efeitos desses períodos sobre os usos da água são acompanhados anualmente (ANA, 2025). As previsões para 2026 renovam a preocupação: o boletim de julho de 2026 do *Climate Prediction Center* atribui mais de 80% de probabilidade a que o El Niño alcance categoria muito forte no fim do ano (NOAA CLIMATE PREDICTION CENTER, 2026b), aumento já registrado pelo INMET em abril (INMET, 2026). Estimar o que esperar da próxima seca exige, portanto, medir como as safras de manga e uva responderam às anteriores, o que justifica este trabalho.

Os dados necessários para essa medição existem e são públicos, mas estão separados: produção, preços de atacado, clima diário, volume do reservatório, índice oceânico, cotações do agronegócio e câmbio são publicados por sete órgãos diferentes, com vias de acesso desiguais e periodicidade que vai de diária a anual. Nenhuma fonte responde sozinha a uma pergunta que envolva clima, produção e mercado ao mesmo tempo, e cruzá-las manualmente a cada consulta não é reproduzível nem auditável. O problema tratado é, portanto, a ausência de uma base integrada e atualizável sobre a região, e ele se sintetiza na pergunta: os dados públicos disponíveis sobre o polo permitem identificar sinal de que eventos climáticos regionais afetam a produção agrícola e a disponibilidade hídrica, e com que antecedência esse sinal aparece?

O objetivo geral é responder a essa pergunta construindo e verificando a base que permite respondê-la. Os objetivos específicos são: (i) desenvolver um *pipeline* automatizado que extrai, padroniza, valida e carrega as sete fontes, documentando as restrições de acesso de cada uma; (ii) consolidar a série de 1974 a 2024 em um modelo dimensional único e consultável, com grão declarado por tabela de fato; e (iii) medir por correlação estatística a associação entre clima, produção e mercado, reportando o tamanho de amostra de cada teste. Parte da resposta obtida é negativa e é registrada como tal, porque os resultados nulos delimitam o que a base permite concluir.

Para isso, as sete fontes foram integradas por um *pipeline* em camadas, consolidadas em um modelo dimensional e analisadas por correlação de Pearson sobre um indicador de anomalia acumulada de precipitação construído para o estudo. O ineditismo está em reunir, sobre o polo Petrolina/Juazeiro, *pipeline* automatizado, modelo dimensional persistente e análise conjunta de clima, produção e mercado, combinação que nenhum dos trabalhos revisados apresenta (Seção 2.2).

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

2. REFERENCIAL TEÓRICO

2.1. Clima e produção irrigada no polo

A fruticultura do submédio São Francisco recebe água de um reservatório de vazão regularizada, o de Sobradinho (BARRETO; XAVIER JÚNIOR; STOSIC, 2019), e não depende diretamente da chuva que cai sobre a lavoura. O choque climático pode, portanto, chegar a ela por dois caminhos: pela disponibilidade do manancial ou pelo excesso ou déficit de chuva local acumulado ao longo de vários anos, escala em que a seca é caracterizada pela anomalia do acumulado de precipitação, e não por um valor absoluto (MCKEE; DOESKEN; KLEIST, 1993). A fase quente do El Niño-Oscilação Sul (ENSO), monitorada pelo *Oceanic Niño Index* (ONI) a partir da definição de Trenberth (1997), está associada a menos chuva sobre o Nordeste (MARENGO *et al.*, 2018; NOAA CLIMATE PREDICTION CENTER, 2026a), mas com exceções: o gradiente de temperatura do Atlântico tropical responde por parte dos anos secos e chuvosos mesmo sem episódio ENSO ativo (LUCENA; GOMES FILHO; SERVAIN, 2011). Daí a hipótese, testada aqui, de que nas lavouras irrigadas o choque climático chegue primeiro pelo manancial e só depois pela chuva local.

Do lado dos dados, integrar fontes independentes que mudam de formato sem aviso é um problema clássico de engenharia de dados, tratado pela separação em camadas: uma preserva o dado como veio da origem e as seguintes o padronizam e modelam (REIS; HOUSLEY, 2022). Para a consulta analítica, o esquema estrela organiza medidas em tabelas de fato com grão declarado, ligadas a dimensões compartilhadas (KIMBALL; ROSS, 2013). São esses os dois conceitos que estruturam a solução descrita na Seção 3.

2.2. Trabalhos relacionados

Os trabalhos revisados distribuem-se em três frentes. Na primeira, sobre produção agrícola da região, Ferreira (2022) analisa a eficiência produtiva da manga no Distrito de Irrigação Nilo Coelho com dados primários de 75 produtores entre 2014 e 2018, e Araujo (2012) estima por regressão em painel o efeito do ENSO sobre a produtividade de grãos de sequeiro no Nordeste e no Sul entre 1970 e 2002. Nenhum dos dois constrói base própria de longo prazo sobre o polo; aqui a unidade é o município, o dado é público e a série vai de 1974 a 2024.

Na segunda frente, a de integração de dados públicos, Sales (2026) apresenta o Carcará, que integra dados meteorológicos e hidrológicos do semiárido, mas não cruza o clima com produção, preços e reservatório nem declara um modelo dimensional; o pacote *agrobr* (AGROBR, 2026) padroniza fontes do agronegócio sob demanda, sem armazenamento próprio; e Bolfe *et al.* (2019) e Barros *et al.* (2025) publicam dados públicos integrados em escala

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

institucional, para a agropecuária nacional e para indicadores educacionais, respectivamente. Assis e Gomide (2021) documentam o problema de obter dados quando a fonte não oferece via programática, obstáculo que reaparece neste trabalho. Na terceira frente, Lima *et al.* (2023) desenvolvem um aplicativo que recomenda cultivos ao agricultor a partir de dados climáticos, atuando na camada que consome bases como a construída aqui. Nenhum dos trabalhos reúne, sobre o polo Petrolina/Juazeiro, *pipeline* automatizado, modelo dimensional persistente e análise de correlação entre clima, produção e mercado, lacuna que este trabalho ocupa.

3. METODOLOGIA

3.1. Caracterização do estudo e recorte

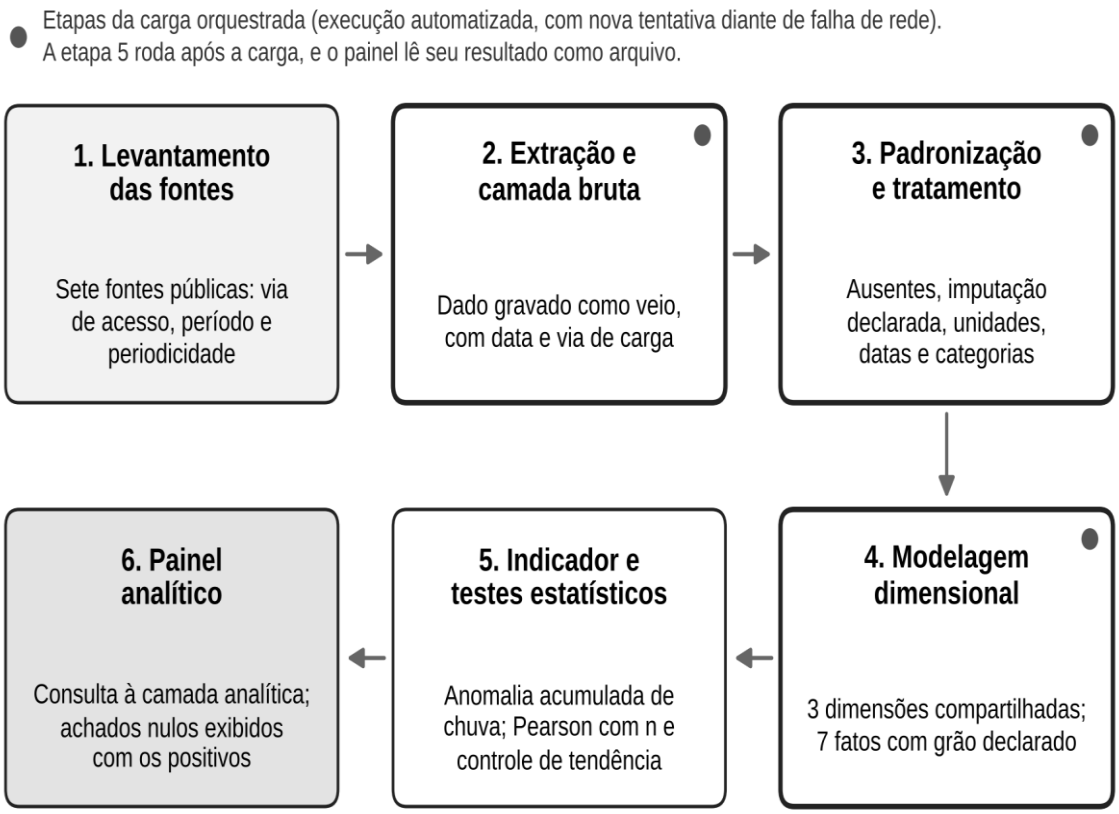
Trata-se de estudo exploratório e correlacional sobre dados secundários de acesso público, sem intervenção sobre as unidades observadas e sem coleta primária. O recorte espacial compreende os municípios de Petrolina e Juazeiro, e o de cultura, a manga e a uva de mesa, as duas lavouras que definem o perfil do polo. O intervalo analisado vai de 1974, primeiro ano em que a PAM publica os dois municípios, a 2024, último ano encerrado na data da coleta; fontes que começam depois, como o preço de atacado (2009) e o reservatório (2000), restringem por conta própria os testes de que participam. A unidade de observação é o par ano–município, desdobrado por cultura, e o tamanho de amostra de cada teste é reportado junto ao coeficiente. Os resultados correspondem à extração concluída em agosto de 2026.

3.2. Etapas de construção da solução

A solução foi construída em seis etapas, resumidas na Figura 1. As etapas de extração, padronização e modelagem formam a carga orquestrada, executada em sequência determinística por um ponto de entrada único e com nova tentativa automática diante de falha transitória de rede. As subseções seguintes descrevem cada etapa e as decisões tomadas nelas.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Figura 1. Etapas de construção da solução, do levantamento das fontes ao painel analítico



Fonte: elaborado pelos autores.

3.2.1. Levantamento das fontes e de suas vias de acesso

A primeira etapa identificou, para cada dimensão do problema, a fonte pública com maior cobertura temporal sobre os dois municípios, e registrou de cada uma a via de acesso, o período disponível e a periodicidade. O resultado está na Tabela 1, que também declara o grão adotado para cada fonte no modelo final. Três fontes expõem interface programática pública, três publicam arquivo estático e uma é lida por biblioteca de dados de mercado. A produção anual do IBGE é a série mais longa e a de menor frequência, com uma observação por ano, e por isso limita o número de pares em todo teste que cruze clima e safra.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Tabela 1. Fontes públicas integradas, via de acesso, período carregado, grão adotado no modelo dimensional e número de registros na camada analítica

Fonte	Órgão	Via de acesso	Período	Grão da tabela de fato	Registros
Produção agrícola	IBGE (SIDRA)	API pública	1974–2024	Ano × município × cultura	189
Preços de atacado	CONAB (PROHORT)	Arquivo estático	2009–2020	Mês × localidade × cultura	220
Ações do agronegócio	B3	Biblioteca de mercado	2009–2026	Dia × ativo	8.748
Clima diário	Open-Meteo (ERA5)	API pública	1974–2026	Dia × localidade	38.418
Reservatório de Sobradinho	ONS	Arquivo estático	2000–2026	Dia	9.720
Índice ENSO (ONI)	NOAA/CPC	Arquivo estático	1974–2026	Trimestre móvel	630
Câmbio USD/BRL	Banco Central (SGS)	API pública	1994–2026	Dia	8.060

Fonte: elaborado pelos autores.

As cotações de AGRO3 (BrasilAgro) e SLCE3 (SLC Agrícola) entram na base como *proxy* do sentimento geral do agronegócio de grande escala, e não como medida do mercado de frutas do polo: as duas empresas produzem predominantemente grãos e fibras no Cerrado, de modo que não se esperava delas resposta a eventos hídricos do submédio São Francisco. O teste com essas séries serve para verificar se a associação com o valor da produção local resiste ao controle de tendência ou se reflete apenas crescimento paralelo, e por isso a expectativa era de resultado nulo na variação anual. A mesma lógica vale para o preço de atacado da manga na praça de Juazeiro, que, por ser da própria cultura e da praça vizinha, funciona como controle mais exigente.

O levantamento também revelou restrições que condicionam a análise. Os pontos de coleta de clima de Petrolina e Juazeiro distam 3,6 km, abaixo da resolução da reanálise ERA5 (HERSBACH *et al.*, 2020), servida pelo Open-Meteo (ZIPPENFENIG, 2023), que fica entre 9 e 25 km, de modo que os dois municípios recebem na prática a mesma série climática; a coincidência foi verificada numericamente e mantida, sem alteração do dado. O preço de atacado cobre manga e uva com lacunas: 2016 está ausente e 2009, 2017 e 2018 são parciais, totalizando 110 observações mensais por cultura. A série de câmbio começa em julho de 1994, para não reunir valores em moedas anteriores ao Plano Real.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

3.2.2. Extração e camada bruta

Cada fonte tem uma rotina de extração própria, que grava o registro como veio da origem em uma camada bruta, com uma tabela por fonte. Cada linha dessa camada carrega duas informações de procedência, o momento da carga e a via de acesso utilizada, o que permite responder, para qualquer registro, quando ele entrou e por onde. A camada bruta é recarregada integralmente a cada execução, de modo que uma mudança de formato na origem se resolve ali, sem alcançar o modelo consultado. A consequência dessa escolha é que a camada guarda sempre o estado atual da fonte, e não seu histórico: um registro que deixa de ser publicado desaparece na execução seguinte.

3.2.3. Adaptação às restrições de acesso

Duas fontes exigiram investigação antes da implementação. O plano inicial previa obter os preços de atacado por consulta ao portal da CONAB, mas os serviços internos usados pela própria interface recusam chamadas feitas fora de sessão autenticada do navegador. A coleta foi então redirecionada para os relatórios históricos do PROHORT/CONAB, publicados como planilha, mas contendo tabelas em HTML, que foram lidas e normalizadas; a via testada e bloqueada ficou registrada no código. O episódio repete o padrão descrito por Assis e Gomide (2021) para fontes públicas sem contrato de acesso. Na API do IBGE, a adaptação foi de outra natureza: o serviço rejeita os delimitadores de consulta quando codificados no formato padrão dos clientes HTTP, e a requisição passou a ser montada com o endereço literal. Uma oitava fonte prevista, a de exportações, teve o acesso automatizado bloqueado em todas as tentativas e foi descartada.

3.2.4. Padronização e regras de tratamento

A padronização converte cada fonte para unidades, datas e categorias comuns, segundo regras declaradas. Os marcadores de ausência do IBGE são convertidos em valor nulo. A área plantada falta sistematicamente entre 1974 e 1987 e foi imputada a partir da área colhida, opção sustentada pela correlação de 0,99 entre as duas nos anos em que ambas existem; os registros imputados recebem uma marcação própria e ficam fora das análises que dependem da razão entre as áreas, na qual a imputação daria razão igual a um por construção. Datas mensais são ancoradas no primeiro dia do mês, valores no formato numérico brasileiro são convertidos para o padrão decimal, e a uva de mesa, que o IBGE registra como "Uva" e o PROHORT como "Uva Itália", é conciliada na dimensão de cultura. O ONI é ancorado no mês central de cada trimestre e classificado pelo limiar de ±0,5 °C. Valores monetários anteriores a 1994 são excluídos de

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

qualquer agregação, porque a série nominal é dominada por trocas de moeda e hiperinflação, com quedas de 97% em 1993 e de 98,5% em 1994 mesmo com produção física crescente.

3.2.5. Modelagem dimensional

A camada analítica segue o esquema estrela, com três dimensões compartilhadas e sete tabelas de fato, uma por fonte. A dimensão de tempo é um calendário diário de 1974 a 2030, a de localidade contém os dois municípios e a de cultura contém manga e uva. O grão de cada tabela de fato está declarado na Tabela 1, e cada uma tem restrição de unicidade sobre as colunas do seu grão, o que torna a carga idempotente: executá-la duas vezes sobre a mesma extração não gera duplicatas, e a promoção para a camada analítica preserva o histórico já carregado, atualizando apenas os registros alterados. Seis das sete tabelas de fato referenciam a dimensão de tempo por chave estrangeira, de modo que o próprio gerenciador de banco garante a integridade referencial.

A tabela de produção é a exceção deliberada. Por ter periodicidade anual, ligá-la a um calendário diário exigiria eleger um dia arbitrário para representar a safra, data que não corresponde a nenhum fato observado, pois a PAM informa apenas o total do ano. Optou-se por preservar o grão real da fonte, guardando ano, município e cultura como atributos com unicidade composta. O custo dessa decisão é que a integridade dessas três colunas não é verificada pelo gerenciador, ressalva relevante por ser essa a tabela que sustenta as estatísticas de produção.

3.2.6. Verificação da solução

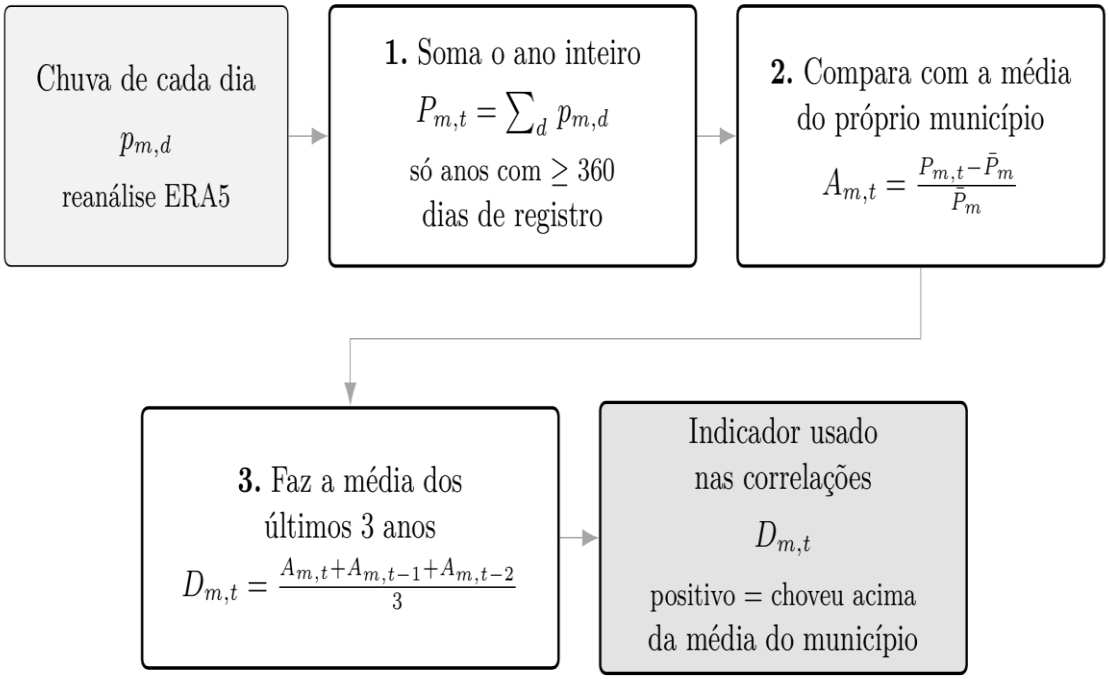
A solução é verificada por 98 testes automatizados em três níveis: testes de unidade sobre a leitura de cada fonte e sobre as funções de cálculo do painel, testes de integração sobre os fluxos de carga e um teste de ponta a ponta sobre a renderização do painel. Os testes de unidade usam trechos reais de cada fonte, incluindo valores ausentes e formatos inconsistentes, e rodam sem rede e sem banco, de modo que a indisponibilidade de uma origem não quebra a verificação. A cobertura de linhas é de 62% sobre o conjunto do código e de 81% desconsiderando as rotinas de análise, que não têm testes. O ambiente é declarado como código e pode ser recriado em outra máquina sem configuração manual; o código-fonte, a configuração do ambiente experimental e as instruções de execução estão em repositório público, disponível em https://github.com/Wagner2525/tcc-pipeline-dados-agricolas.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

3.3. Indicador de anomalia acumulada de precipitação

O indicador que relaciona clima e produção é construído em três passos sobre a série diária de chuva, resumidos na Figura 2: a precipitação é somada por ano e município, apenas nos anos com pelo menos 360 dias de registro, para que um ano incompleto não seja lido como ano seco; cada ano é comparado, em termos relativos, com a média histórica do próprio município; e o resultado é acumulado em janela retroativa de três anos, critério que capta déficit ou excesso acumulado em manancial e solo profundo, relevante para lavoura irrigada (MCKEE; DOESKEN; KLEIST, 1993). Valores positivos do indicador correspondem a chuva acima da média histórica do município na janela, e uma correlação negativa com a produtividade significa produtividade menor nos períodos de chuva acima da média.

Figura 2. Construção do indicador de anomalia acumulada de precipitação, da série diária ao valor usado nos testes



Fonte: elaborado pelos autores.

3.4. Método de análise estatística

A associação entre as séries é medida por correlação linear de Pearson, ao nível α = 0,05 (BUSSAB; MORETTIN, 2017), reportando o coeficiente *r*, o valor-*p* bilateral e o número de

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

pares *n*, sem o qual um coeficiente não é interpretável nas séries curtas deste trabalho. Resultados com valor-*p* acima de α são tratados como não significativos, ainda que próximos do limiar. Quando falta dado, o descarte é par a par, e recortes com menos de cinco pares completos são descartados sem cálculo. Os testes cobrem quatro hipóteses: anomalia acumulada de precipitação e produtividade, por cultura e com as culturas agregadas; o ONI como alerta antecipado, no mesmo ano e defasado em um ano; o volume útil de Sobradinho e a produtividade; e a relação entre mercado e valor da produção. Adotou-se o ONI, e não o RONI, que passou a ser o índice oficial de monitoramento, porque a série analisada é retrospectiva, o ONI mantém o mesmo critério em toda a sua extensão e é sobre ele que se apoia a literatura de referência (TRENBERTH, 1997; ARAUJO, 2012).

A série de referência dos testes de mercado é o valor da produção da manga por tonelada, obtido pela razão entre o valor da produção e a quantidade produzida publicados pelo IBGE. Séries que crescem de forma sustentada por décadas produzem coeficiente alto entre si mesmo quando nada as liga; por isso, os testes de mercado foram feitos duas vezes, com as séries em nível e com cada série substituída por sua variação percentual anual, o que remove a tendência comum:

$$\Delta x_t = [(x_t - x_{t-1}) / x_{t-1}] \times 100 \quad (1)$$

Esses testes têm $n = 11$ em nível e $n = 10$ em variação, porque o pareamento é limitado pela série de atacado, que cobre 2009 a 2020 sem 2016; três desses anos têm cobertura mensal parcial, o que os torna menos confiáveis em uma amostra já pequena. Em todos os casos, o coeficiente de Pearson mede associação linear e não estabelece causalidade; as ressalvas decorrentes da dependência temporal das séries, da imputação e da representatividade do indicador climático são discutidas na Seção 4.6.

4. RESULTADOS E DISCUSSÃO

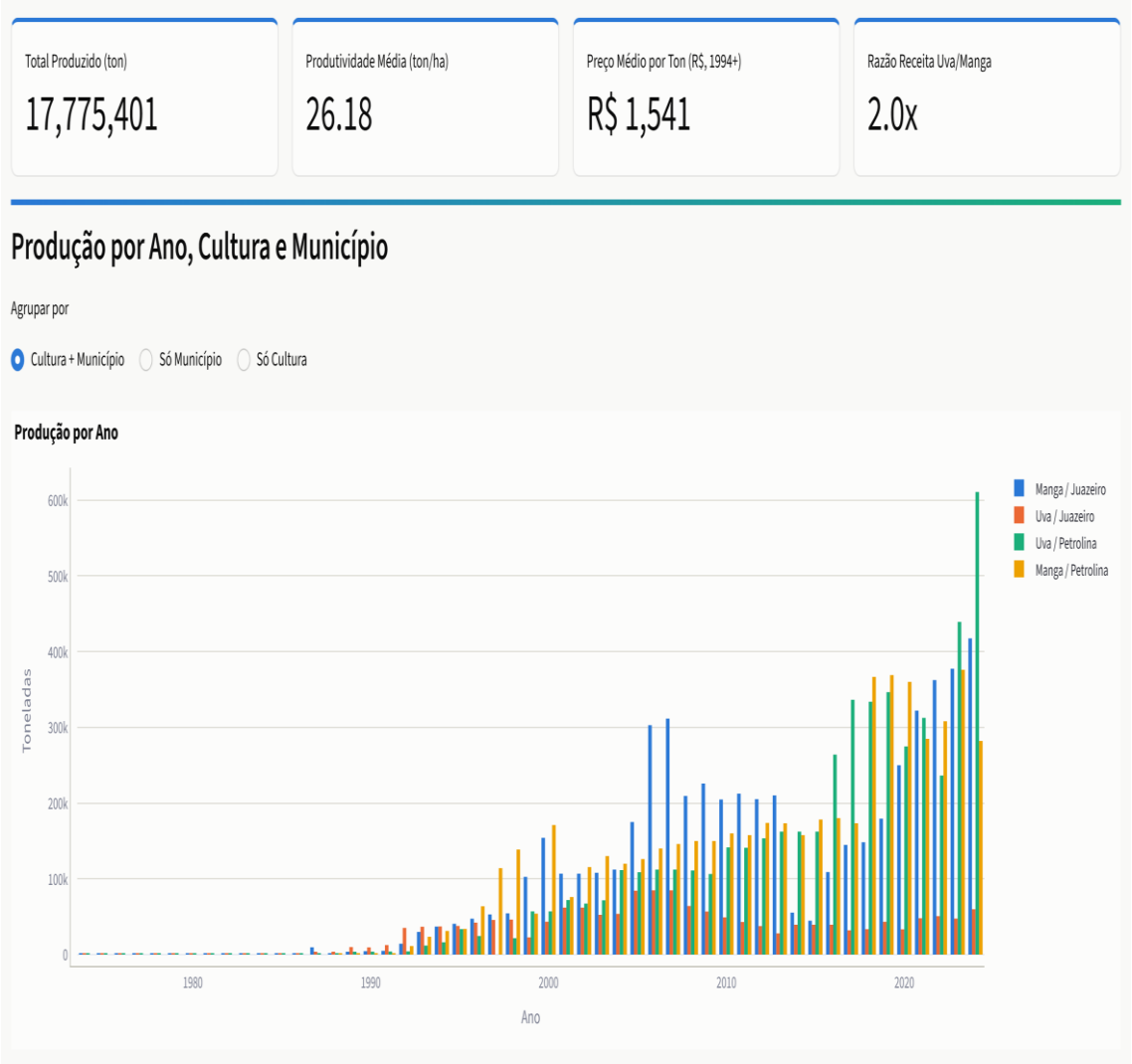
4.1. A base integrada

O *pipeline* carregou as sete fontes da Tabela 1 no modelo dimensional, reunindo cerca de 66 mil registros na camada analítica, com a produção municipal de 1974 a 2024 e as séries diárias atualizadas até agosto de 2026. O painel analítico, ilustrado na Figura 3, consulta diretamente essa camada e organiza em uma única página os indicadores-chave, a produção por ano, a relação entre clima e produtividade, o papel regulador de Sobradinho, a comparação de preços em nível e em variação e as limitações metodológicas, com filtros de período,

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

município e cultura. Os coeficientes entre clima e produtividade chegam ao painel por arquivo gerado pela etapa de análise, executada após a carga.

Figura 3. Painel analítico: indicadores-chave e produção por ano, cultura e município



Fonte: elaborado pelos autores.

Os resultados dos testes estão reunidos na Tabela 2, que concentra os valores de *r*, *p* e *n*; as subseções 4.2 a 4.4 descrevem os achados, e a interpretação é feita na subseção 4.5.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Tabela 2. Hipóteses testadas por correlação de Pearson (α = 0,05). Anomalia positiva indica chuva acima da média histórica do município; nas linhas de ano seguinte, o preditor é do ano t e a resposta, do ano t + 1

Hipótese testada	Variáveis	n	r	p	Conclusão
Anomalia acumulada de precipitação e produtividade					
Culturas agregadas	Anomalia de 3 anos; t/ha	98	−0,1575	0,1215	Não significativo
Manga	Anomalia de 3 anos; t/ha	86	+0,2072	0,0556	Não significativo
Uva	Anomalia de 3 anos; t/ha	97	−0,2405	0,0176	Significativo
Índice ENSO (ONI) como alerta antecipado					
ONI e produtividade, mesmo ano	ONI médio; t/ha	51	+0,1054	0,4619	Não significativo
ONI e produtividade, ano seguinte	ONI médio; t/ha	50	+0,0946	0,5133	Não significativo
ONI e anomalia de chuva, mesmo ano	ONI médio; anomalia de 3 anos	49	−0,1346	0,3565	Não significativo
ONI e anomalia de chuva, ano seguinte	ONI médio; anomalia de 3 anos	49	−0,2022	0,1636	Não significativo
Volume útil de Sobradinho e produtividade					
Volume e safra do mesmo ano	Volume útil médio (%); t/ha	25	−0,0346	0,8696	Não significativo
Volume e safra do ano seguinte	Volume útil médio (%); t/ha	24	−0,1768	0,4085	Não significativo
Mercado e valor da produção da manga por tonelada					
AGRO3, séries em nível	Fechamento médio anual; R$/t	11	+0,8402	0,0012	Significativo
SLCE3, séries em nível	Fechamento médio anual; R$/t	11	+0,8265	0,0017	Significativo
Atacado, séries em nível	Preço médio anual; R$/t	11	+0,7994	0,0031	Significativo
AGRO3, variação anual	Variação anual de ambas (%)	10	+0,2652	0,4590	Não significativo
SLCE3, variação anual	Variação anual de ambas (%)	10	+0,0334	0,9271	Não significativo
Atacado, variação anual	Variação anual de ambas (%)	10	+0,5457	0,1027	Não significativo

Fonte: elaborado pelos autores.

4.2. Clima e produtividade

Entre os testes de clima e produtividade, apenas o da uva foi significativo, com coeficiente negativo: a produtividade da uva tende a ser menor nas janelas de três anos em que a chuva ficou acima da média histórica do município. O resultado da manga não atingiu o nível de significância adotado e é tratado como inconclusivo, assim como o das duas culturas agregadas, que também não mostrou associação. Na série do indicador, o extremo seco é a

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

janela que fecha em 2014 (−0,47, cobrindo 2012 a 2014) e o extremo úmido, a que fecha em 1980 (+0,46). O extremo seco se completa, portanto, antes do El Niño de 2015 e 2016, e as janelas que fecham em 2015, 2016 e 2017 já registram recuperação parcial.

4.3. Antecedência: índice ENSO e reservatório

Nenhum dos testes com o ONI foi significativo, com ou sem defasagem de um ano, seja contra a produtividade, seja contra a anomalia de precipitação. O volume útil de Sobradinho também não se associou à produtividade, no mesmo ano ou no seguinte. Duas hipóteses complementares tampouco se confirmaram: a comparação entre produção anterior e posterior à expansão da irrigação, com corte em 1997, não separou os dois períodos, e a razão entre área colhida e plantada, tomada como indicador de perda, não se associou à anomalia de chuva (r = 0,00; n = 74). Observa-se na série, sem teste que o confirme, que a produção caiu 27,7% em 2014, quando o volume útil ainda estava em torno de 41%, e que o reservatório só desceu à faixa de 9% a 18% entre 2015 e 2017.

4.4. Mercado e valor da produção

Comparadas em nível, as três séries de mercado acompanham o valor da produção da manga por tonelada, todas com associação significativa. Substituindo as séries pela variação anual, nenhum dos três coeficientes permanece significativo, inclusive o do preço de atacado da própria manga.

4.5. Discussão

O achado sustentado pela base é a associação negativa entre chuva acumulada acima da média e produtividade da uva. Uma explicação ligada à sensibilidade da cultura ao excesso hídrico é plausível, mas não é testada aqui, pois a base cobre clima, produção e mercado, e não fisiologia vegetal ou manejo. O resultado da manga não autoriza afirmar efeito em nenhuma direção: o coeficiente positivo ficou próximo do limiar, mas abaixo do nível adotado, e pede amostra maior antes de qualquer conclusão. A ausência de associação com as culturas agregadas é coerente com esse quadro, já que reunir uma cultura com efeito significativo e outra sem efeito definido dilui o sinal; o recorte por cultura é, portanto, condição para que o efeito da uva apareça.

O resultado nulo do ONI não contradiz o de Araujo (2012), porque as duas medições não são equivalentes: aquele estudo cobre culturas de sequeiro em macrorregiões até 2002, e aqui se trata de dois municípios com lavouras irrigadas por água regularizada (BARRETO; XAVIER

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

JÚNIOR; STOSIC, 2019). O coeficiente entre o ONI e a anomalia de chuva do ano seguinte tem o sinal esperado pela literatura regional, mas fica longe do nível adotado, e o papel do Atlântico tropical na chuva do Nordeste (LUCENA; GOMES FILHO; SERVAIN, 2011) reduz a expectativa de que um índice do Pacífico, isolado, antecipe a safra local. A hipótese de que o choque chegaria à lavoura irrigada primeiro pelo manancial não se confirmou: o reservatório não se associou à produtividade, enquanto a chuva local se associou à da uva. A queda de 2014 pode estar ligada a preço, manejo, área colhida, evento fitossanitário ou ao déficit de chuva dos anos anteriores, fatores que as sete fontes não permitem separar; atribuí-la a restrição de irrigação exigiria dados de outorga e de operação dos perímetros irrigados.

No mercado, o resultado confirma o esperado para um *proxy* de escala nacional: a associação em nível vem da tendência de crescimento comum ao período, e não de resposta a um mesmo evento, já que desaparece quando a tendência é removida. O caso do atacado é o mais ilustrativo, porque nem o preço da própria manga, na praça vizinha à produção, sustenta a associação sob controle de tendência.

Para os produtores do polo, a consequência prática é que o efeito climático precisa ser avaliado por cultura: o que a base sustenta hoje é a associação entre períodos prolongados de chuva acima da média e menor produtividade da uva, e um aviso único dirigido ao polo trataria como equivalentes lavouras que, nos dados, não respondem da mesma forma. Para a gestão hídrica, o ganho é de outra natureza: volume do reservatório, clima e produção passaram a estar na mesma base, com grão declarado, e a pergunta sobre o efeito de Sobradinho na safra deixou de depender de coleta manual.

4.6. Limitações

Quatro ressalvas estatísticas delimitam o alcance das correlações. A primeira é que o coeficiente de Pearson mede associação linear e não isola efeitos: um coeficiente significativo é compatível com a influência de terceiras variáveis que mudaram ao longo das cinco décadas, como preço, área colhida, manejo e tecnologia de irrigação, e por isso nenhum resultado deste trabalho é lido como relação de causa e efeito. A segunda é a dependência temporal: as observações anuais não são independentes entre si, porque produtividade e cotações carregam tendência e persistência de um ano para o outro, e o indicador de anomalia, por ser média móvel de três anos, compartilha dois anos de dados entre janelas consecutivas. Essa dependência reduz o número efetivo de observações independentes abaixo do *n* reportado e tende a tornar o valor-*p* otimista, o que reforça a leitura cautelosa do resultado da uva e o tratamento do resultado

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

da manga como inconclusivo; o controle de tendência por variação anual foi aplicado apenas aos testes de mercado.

A terceira ressalva diz respeito à imputação. A área plantada imputada entre 1974 e 1987 não entra nos testes de produtividade, que usam o rendimento médio publicado pelo IBGE, calculado sobre a área colhida, e os registros imputados foram excluídos da análise da razão entre áreas; qualquer uso futuro da área plantada nesse período, porém, carrega a hipótese de que ela acompanhou a área colhida. A quarta é a representatividade do indicador climático: a resolução da reanálise torna os dois municípios indistinguíveis, e a chuva sobre o município não mede diretamente a água aplicada nas lavouras irrigadas, que depende da vazão do rio e da operação dos perímetros. O indicador capta, portanto, o regime climático local, e não a disponibilidade hídrica efetiva das lavouras.

Há ainda limitações que vêm, em sua maioria, das fontes. Nenhuma fonte tem recorte municipal próprio, e a produção chega uma vez por ano, o que mantém a unidade de observação no par ano–município e as amostras em poucas dezenas de pares mesmo sobre cinco décadas. Do lado da solução, a camada analítica não aplica filtros de faixa de valor, a tabela de produção fica fora da integridade referencial verificada pelo gerenciador, o esquema evolui sem mecanismo de migração e a etapa de análise roda fora do fluxo orquestrado. Nenhum teste cruzou clima com preço de atacado, cuja série é curta demais para o indicador de três anos, e as séries de uva no atacado e de câmbio estão carregadas sem análise associada. Uma das variações anuais dos testes de mercado abrange dois anos, por causa da ausência de 2016.

5. CONSIDERAÇÕES FINAIS

Este trabalho investigou se os dados públicos disponíveis sobre o polo Petrolina/Juazeiro permitem identificar sinal de que eventos climáticos afetam a produção agrícola e a disponibilidade hídrica, e com que antecedência. Há sinal na relação entre anomalia acumulada de precipitação e produtividade da uva, que tende a ser menor em períodos de chuva acima da média; para a manga, o resultado é inconclusivo, e com as culturas agregadas não há associação. Sobre antecedência, a resposta é negativa dentro do que foi testado: nem o ONI defasado nem o volume do reservatório anteciparam a safra. No mercado, a associação entre cotações e valor da produção desaparece quando se controla a tendência comum às séries.

A contribuição principal é a base construída e o método que a produziu: sete fontes públicas de acesso desigual, uma delas sem interface programática, integradas em um modelo dimensional com grão declarado, carregadas por um *pipeline* orquestrado e verificado por testes

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

automatizados, e consultadas por um painel que exibe achados positivos e nulos com o mesmo destaque. É essa combinação, e não cada peça isolada, que distingue o trabalho dos revisados, e ela torna reproduzível uma verificação que antes dependia de cruzamento manual.

Como trabalhos futuros, a combinação de anomalia de precipitação, fase ENSO e nível do reservatório já disponível na base serve de entrada para modelos preditivos de produtividade, a ampliação da amostra pode esclarecer o resultado inconclusivo da manga, e o tratamento explícito da dependência temporal, com correção do número efetivo de observações ou modelos de séries temporais, pode confirmar ou refinar o resultado da uva. Do lado da engenharia, recomenda-se implementar a validação de faixas de valor, uma dimensão de ano que permita ligar a produção ao modelo por chave estrangeira, o versionamento das extrações e do esquema e a integração da etapa de análise ao fluxo orquestrado. Do lado dos dados, recomenda-se analisar a série de atacado da uva, investigar o câmbio como fator de confusão nas cotações, buscar fonte climática de maior resolução, como estações de superfície, e solicitar às prefeituras e à CODEVASF, pela Lei de Acesso à Informação, registros de safra e de outorga em periodicidade menor que a anual.

REFERÊNCIAS

AGROBR. **Documentação**. 2026. Disponível em: https://agrobr.dev/docs/. Acesso em: 28 jul. 2026.

ANA. **Conjuntura dos recursos hídricos no Brasil 2025**. Brasília: ANA, 2025. Disponível em: https://conjuntura-2025.webflow.io/. Acesso em: 29 jul. 2026.

ARAUJO, P. H. C. **Eventos climáticos extremos**: os efeitos dos fenômenos El Niño e La Niña sobre a produtividade agrícola das regiões Nordeste e Sul do Brasil. 2012. Dissertação (Mestrado) – Universidade Federal de Viçosa, Viçosa, 2012. Disponível em: http://locus.ufv.br/handle/123456789/68. Acesso em: 25 out. 2026.

ASSIS, W. V. de; GOMIDE, J. V. B. Web scraping em dados públicos: método para extração de dados dos gastos públicos dos vereadores da Câmara Municipal de Belo Horizonte. **Informação & Informação**, Londrina, v. 26, n. 4, p. 319–341, 2021. DOI: https://doi.org/10.5433/1981-8920.2021v26n4p319.

BARBOSA, H. A.; BURITI, C. de O. Assessment of long-term streamflow response to flash drought in the São Francisco River Basin over the last three decades (1991–2020). **Water**, v. 16, n. 16, p. 2271, 2024. DOI: https://doi.org/10.3390/w16162271.

BARRETO, I. D. de C.; XAVIER JÚNIOR, S. F. A.; STOSIC, T. Long-term correlations in São Francisco river flow: the influence of Sobradinho dam. **Revista Brasileira de Meteorologia**, v. 34, n. 2, p. 293–300, 2019. DOI: https://doi.org/10.1590/0102-77863340242.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

BARROS, A. *et al.* Criação de um pipeline de dados abertos conectados para indicadores educacionais do PNE. *In*: LATIN AMERICAN SYMPOSIUM ON DIGITAL GOVERNMENT (LASDIGOV), 12., 2025, Maceió. **Anais** [...]. Porto Alegre: Sociedade Brasileira de Computação, 2025. p. 333–340. DOI: https://doi.org/10.5753/lasdigov.2025.9336.

BOLFE, E. L. *et al.* Base de dados da agropecuária brasileira: planejamento estratégico e desenvolvimento. **Brazilian Journal of Development**, Curitiba, v. 5, n. 1, p. 201–214, 2019. Disponível em: https://www.alice.cnptia.embrapa.br/alice/handle/doc/1109219. Acesso em: 29 jul. 2026.

BUSSAB, W. de O.; MORETTIN, P. A. **Estatística básica**. 9. ed. São Paulo: Saraiva, 2017.

FERREIRA, C. B. **Análise da produtividade agrícola no Vale do São Francisco**: um estudo diante da escassez de recursos hídricos. 2022. Tese (Doutorado) – Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2022. Disponível em: https://teses.usp.br/teses/disponiveis/11/11132/tde-10062022-151607/. Acesso em: 25 out. 2026.

HERSBACH, H. *et al.* The ERA5 global reanalysis. **Quarterly Journal of the Royal Meteorological Society**, v. 146, n. 730, p. 1999–2049, 2020. DOI: https://doi.org/10.1002/qj.3803.

IBGE. **Produção Agrícola Municipal**: culturas temporárias e permanentes. Rio de Janeiro: IBGE, 2024. Disponível em: https://sidra.ibge.gov.br/pesquisa/pam/tabelas. Acesso em: 29 jul. 2026.

INMET. **Previsões recentes apontam novo aumento da probabilidade do El Niño em 2026**. Brasília: INMET, 24 abr. 2026. Disponível em: https://portal.inmet.gov.br/noticias/previs%C3%B5es-recentes-apontam-novo-aumento-da-probabilidade-do-el-ni%C3%B1o-em-2026. Acesso em: 29 jul. 2026.

KIMBALL, R.; ROSS, M. **The data warehouse toolkit**: the definitive guide to dimensional modeling. 3. ed. Indianapolis: Wiley, 2013.

LIMA, I. S. V. de *et al.* Amigo Agricultor: aplicativo móvel de orientação para plantio agrícola. *In*: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 14., 2023, Natal. **Anais** [...]. Porto Alegre: Sociedade Brasileira de Computação, 2023. p. 426–431. DOI: https://doi.org/10.5753/sbiagro.2023.26589.

LUCENA, D. B.; GOMES FILHO, M. F.; SERVAIN, J. Avaliação do impacto de eventos climáticos extremos nos oceanos Pacífico e Atlântico sobre a estação chuvosa no Nordeste do Brasil. **Revista Brasileira de Meteorologia**, v. 26, n. 2, p. 297–312, 2011. DOI: https://doi.org/10.1590/S0102-77862011000200013.

MARENGO, J. A. *et al.* Climatic characteristics of the 2010–2016 drought in the semiarid Northeast Brazil region. **Anais da Academia Brasileira de Ciências**, v. 90, n. 2, supl. 1, p. 1973–1985, 2018. DOI: https://doi.org/10.1590/0001-3765201720170206.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

MARENGO, J. A. *et al.* Drought in Northeast Brazil: a review of agricultural and policy adaptation options for food security. **Climate Resilience and Sustainability**, v. 1, n. 1, p. e17, 2022. DOI: https://doi.org/10.1002/cli2.17.

MCKEE, T. B.; DOESKEN, N. J.; KLEIST, J. The relationship of drought frequency and duration to time scales. *In*: CONFERENCE ON APPLIED CLIMATOLOGY, 8., 1993, Anaheim. **Proceedings** [...]. Boston: American Meteorological Society, 1993. p. 179–184. Disponível em: https://climate.colostate.edu/pdfs/relationshipofdroughtfrequency.pdf. Acesso em: 29 jul. 2026.

NOAA CLIMATE PREDICTION CENTER. **Cold and warm episodes by season**: historical El Niño/La Niña episodes (1950–present). College Park: NOAA, 2026a. Disponível em: https://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php. Acesso em: 29 jul. 2026.

NOAA CLIMATE PREDICTION CENTER. **El Niño/Southern Oscillation (ENSO) diagnostic discussion**. College Park: NOAA, 9 jul. 2026b. Disponível em: https://www.cpc.ncep.noaa.gov/products/analysis_monitoring/enso_advisory/ensodisc.shtml. Acesso em: 29 jul. 2026.

OLIVEIRA, P. D. D. de. Situação econômica da fruticultura irrigada no submédio do São Francisco: avaliação dos últimos anos. **Revista Ibero-Americana de Humanidades, Ciências e Educação**, v. 7, n. 6, p. 823–842, 2021. DOI: https://doi.org/10.51891/rease.v7i6.1435.

REIS, J.; HOUSLEY, M. **Fundamentals of data engineering**: plan and build robust data systems. Sebastopol: O'Reilly Media, 2022.

SALES, M. M. F. **Carcará**: sistema para integração de dados meteorológicos e hidrológicos no semiárido brasileiro. 2026. Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação) – Universidade Federal do Ceará, Crateús, 2026. Disponível em: https://repositorio.ufc.br/handle/riufc/84644. Acesso em: 29 jul. 2026.

TRENBERTH, K. E. The definition of El Niño. **Bulletin of the American Meteorological Society**, v. 78, n. 12, p. 2771–2777, 1997. DOI: https://doi.org/10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2.

VIDAL, M. de F. Fruticultura (manga). **Caderno Setorial ETENE**, Fortaleza, v. 9, n. 371, 2024. Disponível em: https://www.bnb.gov.br/revista/cse/article/view/3135. Acesso em: 29 jul. 2026.

ZIPPENFENIG, P. **Open-Meteo.com Weather API**. Zenodo, 2023. DOI: https://doi.org/10.5281/zenodo.7970649.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.