

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

**Melhorando a tomada de decisões na construção:
Modelagem não paramétrica de atrasos induzidos pelo clima**

Mateus Borges Comito

Dissertação de Mestrado do Programa Interinstitucional de
Pós-Graduação em Estatística (PIPGEs)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Mateus Borges Comito

Melhorando a tomada de decisões na construção:
Modelagem não paramétrica de atrasos induzidos pelo
clima

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP e ao Departamento de Estatística – DEs-UFSCar, como parte dos requisitos para obtenção do título de Mestre em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística. *VERSÃO REVISADA*

Área de Concentração: Estatística

Orientador: Prof. Dr. Rafael Izbicki

Coorientador: Prof. Dr. Paulo do Canto Hubert Junior

USP – São Carlos
Junho de 2024

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

B532m Borges Comito, Mateus
 Melhorando a tomada de decisões na construção:
Modelagem não paramétrica de atrasos induzidos pelo
clima / Mateus Borges Comito; orientador Rafael
Izbicki; coorientador Paulo do Canto Hubert Junior.
-- São Carlos, 2024.
 82 p.

 Dissertação (Mestrado - Programa
Interinstitucional de Pós-graduação em Estatística) --
Instituto de Ciências Matemáticas e de Computação,
Universidade de São Paulo, 2024.

 1. Estimação de distribuição. 2. Seleção de
modelos. 3. Monte Carlo. 4. Processo Estocástico.
5. Clima. I. Izbicki, Rafael, orient. II. do Canto
Hubert Junior, Paulo, coorient. III. Título.

Mateus Borges Comito

Improving Decision-Making in Construction: Nonparametric
Modeling of Weather-Induced Delays

Dissertation submitted to the Institute of Mathematics and Computer Science – ICMC-USP and to the Department of Statistics – DEs-UFSCar – in accordance with the requirements of the Statistics Interagency Graduate Program, for the degree of Master in Statistics. *FINAL VERSION*

Concentration Area: Statistics

Advisor: Prof. Dr. Rafael Izbicki

Co-advisor: Prof. Dr. Paulo do Canto Hubert Junior

USP – São Carlos
June 2024

Dedico este trabalho a minha família e amigos.

AGRADECIMENTOS

Gostaria de expressar minha profunda gratidão à minha família, pelo apoio incondicional em todos os momentos. Expresso minha gratidão à minha namorada, Marielle Cristina, por todo seu apoio e compreensão durante este período.

Às pessoas que se tornaram meus amigos desde os tempos de graduação, meu agradecimento por estarem ao meu lado.

Agradeço especialmente meu orientador, Rafael. Sou imensamente grato pela oportunidade, pela orientação excepcional e por toda a assistência prestada. Agradeço também meu co-orientador, Paulo do Canto, pelas contribuições e valiosas dicas.

Aos professores do PIPGES, o conhecimento e o crescimento que adquiri foram fundamentais para minha jornada acadêmica e pessoal.

RESUMO

BORGES, M. C. **Melhorando a tomada de decisões na construção: Modelagem não paramétrica de atrasos induzidos pelo clima.** 2024. 82 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2024.

A gestão eficaz de projetos de construção enfrenta desafios significativos devido a frequentes atrasos, muitos dos quais são influenciados por variáveis climáticas. Antecipar esses atrasos é crucial, e embora existam diversos métodos baseados em geradores estocásticos, modelos de impacto na produtividade ou aprendizado de máquina, ainda há uma carência notável de abordagens diretas que modelem a produtividade utilizando exclusivamente dados meteorológicos históricos, os quais são facilmente acessíveis. Além disso, não é suficiente ter apenas uma estimativa pontual do atraso de um projeto específico. É mais útil estimar a incerteza total associada a essa estimativa. Esta dissertação propõe um modelo não paramétrico flexível que visa preencher essas lacunas, estimando a distribuição de probabilidade do tempo de execução de um projeto utilizando apenas informações meteorológicas. Partimos da premissa de que cada tarefa tem uma probabilidade diária de execução. Esse processo envolve um processo estocástico não estacionário, descrito por uma Cadeia de Markov não estacionária em tempo discreto. Utilizamos exclusivamente os dados climáticos para calcular os parâmetros necessários, e a distribuição é estimada por meio de simulação de Monte Carlo. Os resultados destacam a utilidade do modelo na previsão de datas de início ideais, na estimativa precisa da conclusão do projeto, no estabelecimento de limites contratuais para atrasos esperados devido às condições climáticas e na análise de caminhos críticos dentro do projeto. Além disso, apresentamos uma ferramenta matematicamente rigorosa para comparação de modelos, permitindo a otimização de hiperparâmetros e a seleção do modelo de predição mais adequado. Esta investigação contribui para melhorar a tomada de decisões, minimizando os impactos negativos da incerteza na produtividade e nos prazos de construção, resultando numa melhoria global na eficiência do projeto.

Palavras-chave: Clima; Seleção de modelo; Estimação de distribuição; Monte Carlo.

ABSTRACT

BORGES, M. C. Improving Decision-Making in Construction: Nonparametric Modeling of Weather-Induced Delays. 2024. 82 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2024.

Effective construction project management faces significant challenges due to frequent delays, many of which are influenced by climatic variables. Anticipating these delays is crucial, and although various methods based on stochastic generators, productivity impact models, or machine learning exist, there is still a notable lack of direct approaches that model productivity using exclusively historical weather data, which is easily accessible. Moreover, it is not sufficient to have only a point estimate of the delay for a specific project. It is more useful to estimate the total uncertainty associated with this estimate. This dissertation proposes a flexible non-parametric model aimed at filling these gaps by estimating the probability distribution of a project's execution time using only weather information. We start from the premise that each task has a daily probability of execution. This process involves a non-stationary stochastic process, described by a non-stationary discrete-time Markov Chain. We use exclusively climatic data to calculate the necessary parameters, and the distribution is estimated through Monte Carlo simulation. The results highlight the utility of the model in predicting optimal start dates, accurately estimating project completion, establishing contractual limits for expected delays due to weather conditions, and analyzing critical paths within the project. Additionally, we present a mathematically rigorous tool for model comparison, allowing for the optimization of hyperparameters and the selection of the most suitable prediction model. This investigation contributes to improving decision-making, minimizing the negative impacts of uncertainty on productivity and construction timelines, resulting in an overall improvement in project efficiency.

Keywords: Weather; Model selection; Distribution estimation; Monte Carlo.

LISTA DE ILUSTRAÇÕES

- Figura 1 – O diagrama apresenta a sequência de etapas para estimar o tempo de execução de um projeto de construção e avaliar a qualidade preditiva do modelo, juntamente com as seções que introduzem os conceitos necessários. 32
- Figura 2 – Cronograma de construção com a sequência de tarefas distribuídas ao longo de quatro semanas. O comprimento das barras horizontais é a duração planejada de cada tarefa. As dependências entre as tarefas são indicadas por setas, com as abreviações "TC"(Terminar-para-Começar), "CC"(Começar-para-Começar) e "TT"(Terminar-para-Terminar). 34
- Figura 3 – Comparação das funções densidade de probabilidade de diferentes modelos para (a) distribuições com mesma média e variâncias diferentes, (b) distribuições simétricas em relação ao valor observado, (c) distribuições com mesma variância e médias mais distantes do valor observado e (d) distribuições com probabilidade zero para o valor observado. 49
- Figura 4 – Probabilidade de geração ao longo do tempo para cada experimento. Cada gráfico está dividido em duas partes: a fase de treino (linha azul) e a fase de teste (linha laranja). 55
- Figura 5 – Resultados do Experimento 1. (A) Porcentagem de dias trabalhados ao longo dos 12 meses, com cada linha representando um ano no histórico de dados. (B) Hiperparâmetros: (B1) Parâmetros Estimados: Probabilidade ao longo do ano com diferentes configurações de hiperparâmetros. (B2) Desempenho de Hiperparâmetro: Risco associado às combinações de hiperparâmetros, medido pela perda pinball da cauda direita. 56
- Figura 6 – Resultados do Experimento 2. (A) Porcentagem de dias trabalhados ao longo dos 12 meses, com cada linha representando um ano no histórico de dados. (B) Hiperparâmetros: (B1) Parâmetros Estimados: Probabilidade ao longo do ano com diferentes configurações de hiperparâmetros. (B2) Desempenho de Hiperparâmetro: Risco associado às combinações de hiperparâmetros, medido pela perda pinball da cauda direita. 57

Figura 7 – Resultados do Experimento 3. (A) Porcentagem de dias trabalhados ao longo dos 12 meses, com cada linha representando um ano no histórico de dados. (B) Hiperparâmetros: (B1) Parâmetros Estimados: Probabilidade ao longo do ano com diferentes configurações de hiperparâmetros. (B2) Desempenho de Hiperparâmetro: Risco associado às combinações de hiperparâmetros, medido pela perda pinball da cauda direita.	58
Figura 8 – Resultados do Experimento 4. (A) Porcentagem de dias trabalhados ao longo dos 12 meses, com cada linha representando um ano no histórico de dados. (B) Hiperparâmetros: (B1) Parâmetros Estimados: Probabilidade ao longo do ano com diferentes configurações de hiperparâmetros. (B2) Desempenho de Hiperparâmetro: Risco associado às combinações de hiperparâmetros, medido pela perda pinball da cauda direita.	59
Figura 9 – Cronograma de construção com a sequência de tarefas distribuídas ao longo de 160 dias, com suas respectivas durações indicadas no topo do cronograma. As tarefas têm dependências "TC"(Terminar-para-Começar), indicadas pelas setas.	61
Figura 10 – Contagem de estações para cada combinação de hiperparâmetros selecionada, utilizando a perda pinball ponderada pela cauda direita. Os gráficos de barras na diagonal apresentam a contagem de estações para cada hiperparâmetro individualmente, enquanto os gráficos de calor mostram a contagem para pares de hiperparâmetros.	63
Figura 11 – Comparação do desempenho do modelo proposto com o gerador climático usando três métricas de perda diferentes: (a) Pinball da cauda direita, (b) Pinball e (c) Densidade. Cada ponto representa o risco e seu intervalo de incerteza para uma estação climática. Os pontos laranjas indicam que o modelo proposto teve um desempenho melhor, os pontos azuis indicam que o gerador climático foi melhor, e os pontos cinza mostram diferenças não significativas. A linha tracejada representa a linha de identidade, onde ambos os modelos teriam o mesmo desempenho.	64
Figura 12 – Probabilidade de execução de cada tarefa ao longo dos dias do ano, para a estação 113. Cada linha representa uma tarefa específica.	66
Figura 13 – Distribuição estimada da duração do projeto, segmentada por mês de início, para a estação 113. As distribuições são calculadas usando um estimador da densidade por kernel, para aproximar por uma distribuição contínua e usando um padrão mais suave para facilitar a visualização. Na parte inferior, as linhas observadas representam a duração real do projeto.	67

Figura 14 – Série temporal do percentual de dias trabalhados para cada mês na estação 9. Cada gráfico corresponde a uma tarefa específica do cronograma de construção. As linhas dentro de cada gráfico representam os dados de diferentes anos.	70
Figura 15 – Comparação do quantil esperado com o quantil real para cada ano.	78

LISTA DE ALGORITMOS

Algoritmo 1 – Estimativa do tempo de execução do projeto	40
--	----

LISTA DE TABELAS

Tabela 1	– Limites climáticos para as tarefas de terraplanagem (τ_1), concretagem (τ_2), cofragem (τ_3), siderurgia (τ_4), pinturas exteriores (τ_5) e pavimentação de asfalto (τ_6), considerando as variáveis climáticas temperatura mínima (T_{min}), temperatura média (T_{mean}), temperatura máxima (T_{max}), precipitação (P), ocorrência de granizo (P_{hail}), ocorrência de neve (P_{snow}), ocorrência de tempestade elétrica (E) e velocidade do vento (V).	35
Tabela 2	– Exemplos de dados diários das condições meteorológicas, incluindo temperatura mínima (T_{min}), temperatura média (T_{mean}), temperatura máxima (T_{max}), precipitação (P), precipitação de granizo (P_{hail}), precipitação de neve (P_{snow}), tempestade elétrica (E) e velocidade do vento (V).	36
Tabela 3	– Exemplos de dados diários indicando se foi possível realizar as tarefas de terraplanagem (τ_1), concretagem (τ_2), cofragem (τ_3), siderurgia (τ_4), pinturas exteriores (τ_5) e pavimentação de asfalto (τ_6) no ano y e dia do ano i	36
Tabela 4	– Exemplos de dados diários da temperatura média (T_{mean}), precipitação (P), realização de tarefas de terraplanagem (τ_1) e o elemento $H^{(t)l,y,j}$ para o ano y , dia do ano j e local l	44
Tabela 5	– Lista de hiperparâmetros utilizados no modelo, com seus respectivos intervalos e descrições detalhadas.	46
Tabela 6	– Valores de perda quadrática, de densidade e de pinball para diferentes grupos e modelos. Cada grupo contém modelos caracterizados por uma distribuição específica $c + BetaBin(100, a, b)$ com parâmetros c , a e b	51
Tabela 7	– Descrição e probabilidade de geração para cada experimento. Cada experimento é descrito em termos do padrão anual (senoidal ou complexo) e da presença ou ausência de tendência. As probabilidades de geração são expressas em função do dia do ano j e ano y	54
Tabela 8	– Desempenho comparativo do modelo proposto e do gerador climático para cada estação climática, utilizando três métricas de perda: Pinball Cauda Direita, Pinball e Densidade. Cada valor é o risco estimado e seu intervalo de incerteza. Os valores com um asterisco (*) indicam diferenças estatisticamente significativas. A coluna modelo escolhido indica qual modelo foi escolhido com base no melhor desempenho.	68

Tabela 9 – Estatísticas da distribuição estimada da duração do projeto, segmentada por mês de início para a estação 113. As estatísticas incluem a média, desvio padrão e vários quantis (5%, 25%, 50%, 75%, 95%) para cada mês do ano.	69
Tabela 10 – Hiperparâmetros selecionados para cada tipo de tarefa na estação 9.	69
Tabela 11 – Hiperparâmetros selecionados para cada estação climática utilizando a perda pinball da cauda direta, perda pinball não ponderada e perda de densidade.	76

LISTA DE SÍMBOLOS

\mathcal{T} — Conjunto de tarefas com p elementos

t — Índice da tarefa, onde $t = 1, \dots, p$

Z — Variável aleatória que descreve o tempo de execução do projeto

$i^{(t)}$ — Dia de início da tarefa t

$D^{(t)}$ — Variável aleatória que descreve o tempo de execução da tarefa t

$d^{(t)}$ — Duração planejada da tarefa, desconsiderando fatores climáticos

$\theta_j^{(t)}$ — Probabilidade da t -ésima tarefa ser executada no dia do ano j

$P_j^{(t)}(s)$ — Probabilidade de transição de $X_{j-1}^{(t)} = s$ para $X_j^{(t)} = 1$

B — Número de amostras de Monte Carlo

$\mathbf{H}^{(t)}$ — Matriz binária tridimensional que indica a execução da tarefa t

\mathcal{L} — Conjunto de estações climáticas disponíveis

\mathcal{L} — Conjunto de anos disponíveis

L — Função de perda

R — Risco do modelo

SUMÁRIO

1	INTRODUÇÃO	25
1.1	Contribuições	26
1.2	Relação com outros trabalhos	27
1.2.1	Modelando $p(z)$	27
1.2.2	Seleção de modelo	28
2	FUNDAMENTAÇÃO TEÓRICA	31
2.1	Cronograma de construção	32
2.2	Variáveis climáticas	33
2.3	Modelagem de Z : Tempo de Execução do Projeto	36
2.4	Modelagem de $X_i^{(t)}$: A tarefa t foi executada no dia i	37
2.4.1	Distribuição de Bernoulli independentes	38
2.4.2	Cadeia de Markov	38
2.5	Simulação Monte Carlo	39
2.6	Comparação de modelos	41
3	METODOLOGIA	43
3.1	Estimando $\theta^{(t)}$, $P^{(t)}(\mathbf{0})$ e $P^{(t)}(\mathbf{1})$	43
3.1.1	Escolhendo $w_{ano}(y_0, y)$ - Decaimento exponencial	44
3.1.2	Escolhendo $w_{dia}(j, j')$ - Média móvel	45
3.1.3	Escolhendo $w_{loc}(l_0, l)$ - Distância euclidiana	45
3.1.4	Escolhendo $f_s^{(t)}$ - Ondas senoidais	45
3.1.5	Cadeia de Markov	46
3.1.6	Hiperparâmetros	46
3.2	Seleção de modelos	47
3.2.1	Perda de densidade	47
3.2.2	Perda pinball ponderada	48
3.2.3	Ilustração	49
4	EXPERIMENTOS	53
4.1	Dados simulados	53
4.2	Dados reais	60
4.2.1	Otimização de hiperparâmetros	61
4.2.2	Comparação de modelos	62

4.2.3	<i>Data de inicio do projeto</i>	65
4.2.4	<i>Escolha de hiperparâmetros por tarefa</i>	65
5	CONCLUSÃO	71
APÊNDICE A	OTIMIZAÇÃO DE HIPERPARÂMETRO	75
APÊNDICE B	CONJUNTO DE DADOS	77
REFERÊNCIAS	79

INTRODUÇÃO

Quando um projeto de construção se prolonga além das datas estipuladas no contrato, caracteriza-se como atraso na construção. A extensa literatura sobre o tema investiga detalhadamente essa complexa questão, identificando meticulosamente suas causas primárias em diversos países, regiões e tipos de projetos. Os atrasos na construção podem ser causados por diversos fatores, incluindo a incompetência das partes envolvidas, falta de disponibilidade de materiais e equipamentos, e condições climáticas desfavoráveis (DURDYEV; HOSSEINI, 2020; SANNI-ANIBIRE; ZIN; OLATUNJI, 2022; SEPASGOZAR *et al.*, 2019). Esses atrasos podem ter consequências negativas tanto para os proprietários quanto para os empreiteiros, resultando em penalidades contratuais e limitando a capacidade produtiva dos recursos.

Os eventos climáticos representam uma fonte significativa de incerteza nos projetos de construção, uma vez que afetam diretamente a produtividade (ASSAF; AL-HEJJI, 2006; SCHULDT *et al.*, 2021). A modelagem do atraso causado por esses eventos é um desafio, pois eles variam não só no tempo e espaço, mas também em função das diferentes tarefas de construção, que exigem condições climáticas específicas para serem executadas. Portanto, uma gestão de projeto eficaz deve considerar com precisão essas condições climáticas particulares de cada tarefa, a fim de minimizar os potenciais impactos negativos na produtividade e no cumprimento dos prazos.

A engenharia e gestão da construção é um campo interdisciplinar que abrange gestão de recursos humanos, planejamento de projetos, ciências sociais e métodos de previsão (LIU *et al.*, 2014). Este trabalho tem como foco estimar a duração de um projeto devido a atrasos climáticos, que afetam significativamente a gestão e planejamento de projetos de construção. Além disso, geralmente não se está interessado em ter uma estimativa pontual do tempo de execução de um determinado projeto. Em vez disso, é mais útil estimar a incerteza total sobre tal quantidade. Para resolver esta questão, propomos um modelo para estimar a incerteza geral sobre o tempo de conclusão de um determinado projeto, estimando a distribuição de probabilidade de Z , a duração

do projeto em dias. Este modelo permite uma melhor tomada de decisão nas diferentes etapas do projeto de construção, como selecionar a data de início mais adequada e obter uma estimativa da data de término do projeto através do cálculo dos quantis de $p(z)$. Além disso, o modelo também pode ser atualizado durante a execução do projeto para fornecer estimativas atualizadas para a data de término do projeto e verificar se as condições climáticas foram atípicas durante a execução.

Ademais, oferecemos uma ferramenta matematicamente rigorosa que pode ser usada para otimizar hiperparâmetros associados ao nosso modelo, bem como hiperparâmetros usados em qualquer outro modelo para $p(z)$. Esta ferramenta também pode ser usada para comparar o desempenho de vários modelos quanto ao tempo de atraso.

1.1 Contribuições

A incerteza em torno de Z é crucial para a tomada de decisões assertivas; no entanto, a maioria dos estudos foca na obtenção de previsões pontuais. Portanto, a principal contribuição deste trabalho é apresentar um novo modelo não paramétrico para estimar a distribuição de probabilidade de Z , capaz de incorporar informações climáticas para modelar habilmente os atrasos.

A abordagem convencional envolve duas etapas: modelar as condições climáticas e, em seguida, avaliar a viabilidade de realizar as tarefas de acordo com essas previsões. Por outro lado, nossa abordagem é mais direta, modelando diretamente a probabilidade de conclusão de cada tarefa em um determinado dia, considerando todas as tarefas envolvidas no plano de construção. Em seguida, combinamos essas informações para obter uma estimativa mais precisa de $p(z)$. Além disso, nosso modelo também inclui componentes capazes de considerar mudanças climáticas em suas previsões (SCHULDT *et al.*, 2021).

Em outras palavras, nosso objetivo é modelar a dinâmica temporal da realização de uma tarefa específica ao longo do ano, considerando os fatores climáticos. Para isso, assumimos que cada tarefa tem uma probabilidade diária de execução, podendo ser representada como um experimento de Bernoulli. Este processo envolve um processo estocástico não estacionário, descrito por uma Cadeia de Markov não estacionária em tempo discreto. Utilizamos exclusivamente os dados climáticos para calcular os parâmetros necessários, e as amostras de $p(z)$ são obtidas por meio de simulação de Monte Carlo.

Nosso modelo possui hiperparâmetros que controlam o processo de aprendizagem. Como segunda contribuição deste trabalho, apresentamos uma ferramenta matematicamente rigorosa para a escolha dos valores dos hiperparâmetros.

1.2 Relação com outros trabalhos

1.2.1 Modelando $p(z)$

Numerosos estudos investigam os atrasos em projetos de construção utilizando diversas variáveis e conjuntos de dados. Nosso método proposto visa estimar $p(z)$ utilizando exclusivamente variáveis climáticas derivadas de dados meteorológicos diários.

Uma abordagem convencional em duas fases envolve a utilização de modelos de geração climática para simular valores climáticos futuros dia a dia e, posteriormente, projetar seu impacto no cronograma de construção (SCHULDT *et al.*, 2021; APIPATTANAVIS *et al.*, 2007). A primeira etapa pode ser realizada por meio de modelos climáticos estocásticos (JUNG *et al.*, 2016; APIPATTANAVIS *et al.*, 2010; THORPE; KARAN, 2008; SHAHIN; ABOURIZK; MOHAMED, 2011; RICHARDSON, 1981) ou métodos de aprendizado de máquina (HANOON *et al.*, 2021). A segunda etapa avalia como o clima gerado afeta a produtividade, com uma literatura substancial dedicada à compreensão do impacto do clima e outros fatores na produtividade do trabalho (GOLNARAGHI *et al.*, 2019). A abordagem em duas fases pode ser eficaz na estimativa de $p(z)$ se modelos estocásticos confiáveis estiverem disponíveis, especialmente para projetos de curto prazo. No entanto, essa abordagem difere da proposta neste trabalho, que modela diretamente a produtividade. A ferramenta de seleção de modelos proposta nesta dissertação pode ser utilizada para comparar qualquer conjunto de modelos, incluindo os baseados na abordagem em duas fases. Vale destacar que existem estudos focados exclusivamente na modelagem do impacto na produtividade, sem necessariamente prever atrasos (HERAVI; ESLAMDOOST, 2015; SONG; ABOURIZK, 2008; CHINNADURAI *et al.*, 2016; IBBS; SUN, 2017; GOLNARAGHI *et al.*, 2019; LI *et al.*, 2016; DUFFY *et al.*, 2012). Esses estudos oferecem informações valiosas sobre a perda de produtividade e estratégias de mitigação, mas não visam modelar o atraso global do projeto.

Métodos de aprendizado de máquina também podem ser utilizados para prever diretamente a duração total do projeto (GONDIA *et al.*, 2020; YASEEN *et al.*, 2020; EGWIM *et al.*, 2021; WAUTERS; VANHOUCKE, 2014; PEWDUM; RUJIRAYANYONG; SOOKSATRA, 2009). Em geral, esses estudos identificam fatores gerais que impactam os atrasos, medem seu impacto principalmente por meio de pesquisas ou dados históricos de outros projetos, e treinam algoritmos com base nessas informações. O impacto do clima é frequentemente representado por uma variável que captura fatores externos. O conjunto de dados utilizado nesses estudos difere significativamente do nosso, pois utilizamos apenas dados meteorológicos históricos. Embora seja possível fornecer incertezas, os métodos de aprendizado de máquina são normalmente empregados para oferecer previsões pontuais. Recentemente, alguns estudos utilizaram modelos que estimam incertezas para avaliar o custo do projeto (CHAKRABORTY *et al.*, 2020).

Outro método bem estabelecido para analisar cronogramas e incertezas em projetos é a Técnica de Avaliação e Revisão de Programas (PERT) (SLYKE, 1963), com uma vasta

literatura propondo extensões e estudos de caso (REZAIE *et al.*, 2007; SADEGHI; FAYEK; PEDRYCZ, 2010; KOKKAEW; WIPULANUSAT, 2014; BALLESTEROS-PÉREZ *et al.*, 2017a; BALLESTEROS-PÉREZ *et al.*, 2017b). A informação básica necessária para o sistema PERT são as distribuições paramétricas das durações das atividades, fornecidas por especialistas ou com base em informações de outros projetos. O cronograma de construção é simulado utilizando valores da distribuição, e a estimativa de $p(z)$ é obtida através de simulações de Monte Carlo. O modelo desenvolvido por Ballesteros-Pérez *et al.* (2017a) pode ser visto como uma extensão onde, ao invés de obter a duração da tarefa através de uma distribuição paramétrica, utiliza-se experimentos de Bernoulli para verificar a possibilidade de realização da tarefa em cada dia do ano até atingir a duração planejada. Diferentemente do método tradicional, a incerteza vem diretamente dos dados meteorológicos. O modelo assume que a produtividade diária segue um comportamento senoidal ao longo do ano, tornando-se paramétrico.

O modelo aqui proposto para $p(z)$ é uma extensão deste modelo senoidal, modelando diretamente os dias úteis usando dados meteorológicos históricos. Embora encapsule o modelo de onda senoidal, permite a escolha de outros métodos, incluindo uma versão não paramétrica mais flexível e o uso de cadeias de Markov para adicionar dependência entre dias. Abordagens não paramétricas tendem a ter melhor desempenho se um conjunto de dados de treinamento maior estiver disponível (HASTIE *et al.*, 2009). Portanto, espera-se que a versão não paramétrica do modelo tenha melhor desempenho com uma grande quantidade de dados meteorológicos históricos ou quando os dados não apresentam o comportamento paramétrico esperado. A ferramenta de seleção de modelo proposta auxilia na escolha da melhor abordagem para os dados disponíveis..

1.2.2 Seleção de modelo

A validação de modelos de previsão de atrasos em projetos é frequentemente realizada por meio de estudos de caso (EL-RAYES; MOSELHI, 2001; THORPE; KARAN, 2008; APIPAT-TANAVIS *et al.*, 2010; SHAHIN; ABOURIZK; MOHAMED, 2011; JUNG *et al.*, 2016; GUO; CHEN; CHIU, 2017; BALLESTEROS-PÉREZ *et al.*, 2017a). Embora esses estudos forneçam informações valiosas sobre os modelos, eles oferecem uma compreensão limitada sobre como diferentes abordagens se comparam entre si. Nossas ferramentas complementam esses estudos de caso ao proporcionar uma metodologia matematicamente rigorosa para a comparação de modelos que estimam $p(z)$.

Uma abordagem alternativa para a comparação de modelos é a validação cruzada ou a divisão de dados para calcular métricas de aprendizado de máquina (ML) que avaliam a precisão das previsões, como utilizamos em nosso trabalho (GONDIA *et al.*, 2020). No entanto, em vez de usar métricas convencionais de ML, como o erro quadrático médio, que são projetadas para comparar previsões pontuais, apresentamos funções de perda adaptadas para medir quão bem uma estimativa de $p(z)$ descreve a incerteza total em torno de Z .

Este trabalho está organizado da seguinte forma: no [Capítulo 2](#) realizamos uma revisão dos conceitos que ajudam a entender os métodos utilizados. No [Capítulo 3](#), descrevemos os métodos propostos e, no [Capítulo 4](#), apresentamos um estudo de simulação e uma aplicação em dados reais.

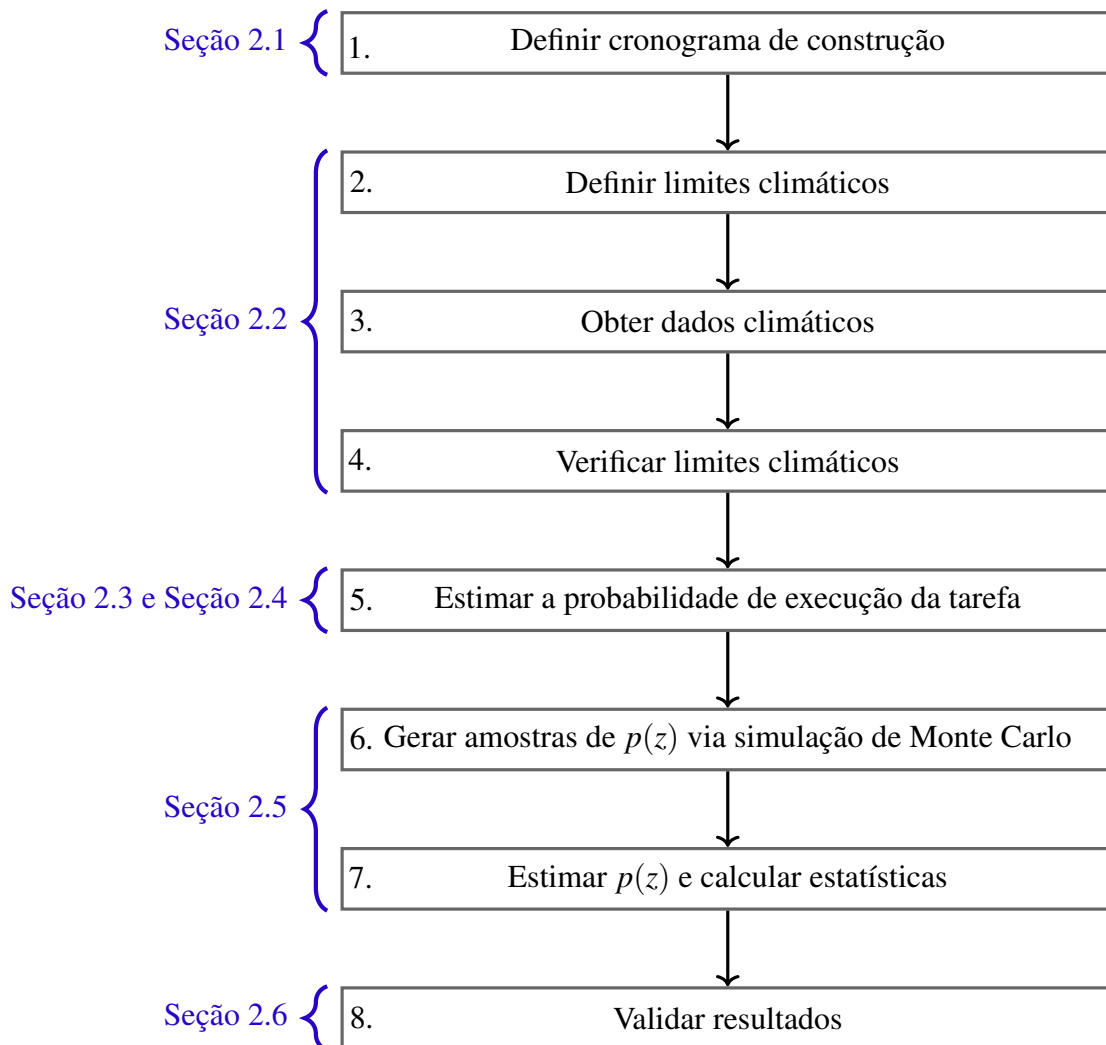
FUNDAMENTAÇÃO TEÓRICA

Iniciamos apresentando uma visão geral do nosso *framework*, detalhando-o nas subseções subsequentes. A [Figura 1](#) ilustra a sequência de etapas e seções que introduzem os conceitos necessários para estimar o tempo de execução de um projeto de construção e mensurar a qualidade preditiva do modelo.

Nosso modelo assume que cada tarefa tem uma probabilidade de execução diária, podendo ser tratada como um experimento de Bernoulli, utilizando exclusivamente dados climáticos para calcular os parâmetros necessários. Primeiramente, precisamos do cronograma de construção do projeto e dos dados climáticos disponíveis. Portanto, a primeira etapa consiste em definir o cronograma de construção a ser utilizado no projeto, conforme explicado na [Seção 2.1](#). Em seguida, estabelecemos limites climáticos para as tarefas dependentes das condições climáticas e obtemos dados climáticos históricos com as variáveis de interesse. Esses limites são aplicados aos dados históricos, permitindo estimar a probabilidade de execução de cada tarefa. Na [Seção 2.2](#), apresentamos o conceito de variáveis climáticas, bem como os limites associados a essas variáveis, e discutimos sua aplicação aos dados históricos. A partir das etapas anteriores, obtivemos o cronograma de construção e os dados históricos que indicam a viabilidade de execução de cada tarefa em cada dia do ano.

Para compreender como modelar a dinâmica temporal da realização de uma determinada tarefa ao longo do ano, levando em conta os fatores climáticos, na [Seção 2.3](#), definimos o tempo de execução como uma variável aleatória. Na [Seção 2.4](#), apresentamos como a execução de uma atividade pode ser modelada. Amostras de $p(z)$ podem ser obtidas por meio de simulação de Monte Carlo e utilizadas para estimar sua distribuição e estatísticas, conforme apresentado na [Seção 2.5](#). Por fim, na [Seção 2.6](#) apresentamos como podemos medir a qualidade preditiva do modelos e selecionar o mais adequado. As contribuições deste trabalho estão nas Etapas 5 e 8, conforme detalhado na [Seção 3.1](#) e [Seção 3.2](#), respectivamente.

Figura 1 – O diagrama apresenta a sequência de etapas para estimar o tempo de execução de um projeto de construção e avaliar a qualidade preditiva do modelo, juntamente com as seções que introduzem os conceitos necessários.



Fonte: Elaborada pelo autor.

2.1 Cronograma de construção

Um projeto de engenharia é composto por um conjunto de tarefas que precisam ser realizadas para produzir as entregas planejadas. Cada tarefa requer recursos como materiais, mão de obra, equipamentos e suprimentos. Através da estimativa dos recursos necessários, é possível determinar a quantidade de tempo que cada tarefa demandará para ser concluída (EDITION, 2008). Ademais, as tarefas geralmente apresentam dependências e relacionamentos lógicos, tais como:

- Terminar-para-Começar (TC): uma tarefa sucessora não pode começar até que uma tarefa predecessora seja concluída.

- Terminar-para-Terminar (TT): uma tarefa sucessora não pode terminar até que uma tarefa predecessora tenha terminado.
- Começar-para-Começar (CC): uma tarefa sucessora não pode iniciar até que uma tarefa predecessora seja iniciada.
- Começar-para-Terminar (CT): uma tarefa sucessora não pode terminar até que uma tarefa predecessora seja iniciada.

O cronograma de construção é crucial para o planejamento e execução bem-sucedida de um projeto. Este cronograma detalha as tarefas a serem realizadas, suas durações e dependências. Gráficos de barra, também conhecidos como gráficos de Gantt, são amplamente utilizados para representar o cronograma. No eixo vertical, as tarefas são listadas e, no eixo horizontal, as datas são apresentadas. As durações das tarefas são indicadas por barras horizontais posicionadas conforme as datas de início e término. Este planejamento é elaborado pela equipe do projeto, utilizando diversas técnicas de gerenciamento para auxiliar nas estimativas.

O gráfico de Gantt, ilustrado na [Figura 2](#), apresenta a distribuição de um conjunto de tarefas fictícias ao longo de quatro semanas, proporcionando uma visualização clara do cronograma de construção. Cada barra horizontal representa a duração planejada de uma tarefa. Por exemplo, a "Tarefa 1" tem uma duração de três dias, enquanto a "Tarefa 6" dura oito dias. As setas indicam as dependências entre as tarefas, como a relação Terminar-para-Começar ("TC") entre a "Tarefa 1" e a "Tarefa 2", onde a "Tarefa 2" começa após a conclusão da "Tarefa 1". Outro exemplo é a relação Começar-para-Começar ("CC") entre a "Tarefa 3" e a "Tarefa 4", que devem iniciar ao mesmo tempo. A relação Terminar-para-Terminar ("TT") entre a "Tarefa 3" e a "Tarefa 5" significa que a "Tarefa 5" só pode terminar após a conclusão da "Tarefa 3".

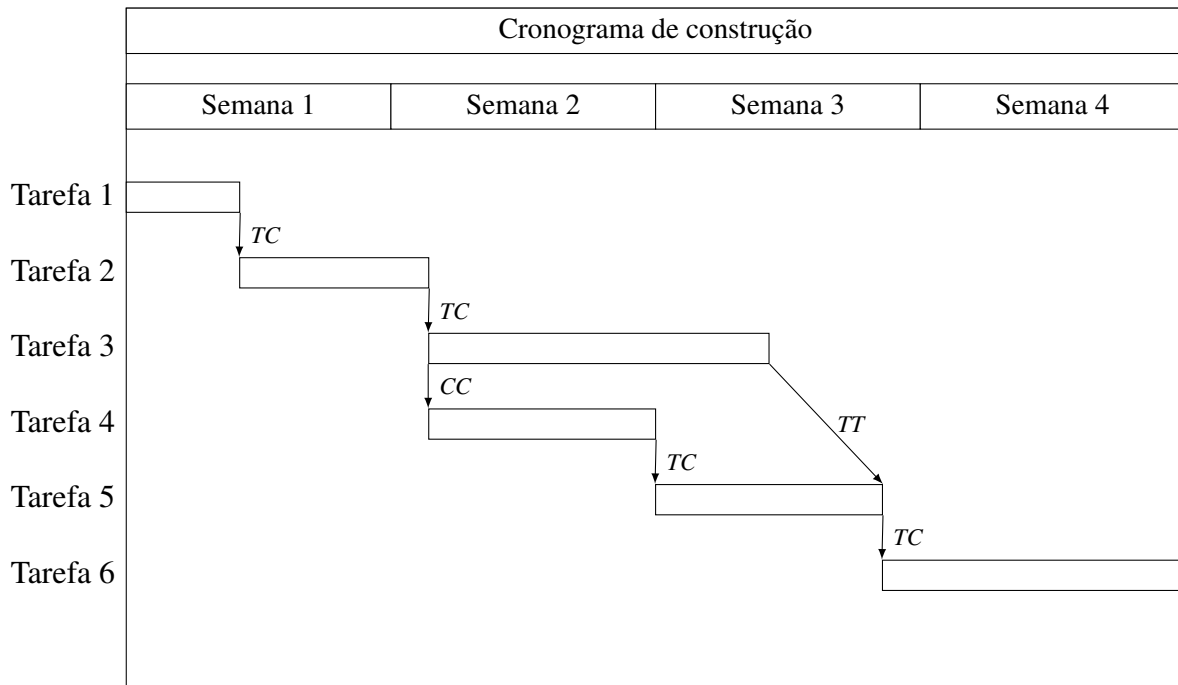
As condições climáticas são um fator externo que pode causar atrasos na execução das tarefas, tornando difícil para os contratantes estimar o impacto em cada tarefa. O cronograma de construção proposto não considera a previsão do tempo, e métodos de modelagem podem ser aplicados para estimar possíveis atrasos devido a condições climáticas, assim como o tempo de execução, levando em conta esses fatores.

2.2 Variáveis climáticas

O clima é um fenômeno complexo, caracterizado por várias variáveis, como temperatura, precipitação, umidade e vento. Essas variáveis são interdependentes e variam ao longo do tempo, tornando a modelagem de suas correlações um desafio significativo ([BALLESTEROS-PÉREZ et al., 2017a](#)).

Na construção civil, há uma vasta literatura dedicada à compreensão de como o clima e outros fatores influenciam a produtividade do trabalho ([GOLNARAGHI et al., 2019](#)). Depen-

Figura 2 – Cronograma de construção com a sequência de tarefas distribuídas ao longo de quatro semanas. O comprimento das barras horizontais é a duração planejada de cada tarefa. As dependências entre as tarefas são indicadas por setas, com as abreviações "TC" (Terminar-para-Começar), "CC" (Começar-para-Começar) e "TT" (Terminar-para-Terminar).



Fonte: Elaborada pelo autor.

dendo das condições climáticas e das tarefas a serem realizadas, condições adversas podem reduzir a produtividade dos trabalhadores ou até mesmo impedir a realização das tarefas planejadas. Além disso, esses impactos podem se prolongar por dias subsequentes. Por exemplo, uma quantidade acumulada de chuva entre 13 e 25 mm impede a realização da pavimentação em um determinado dia e deixa o solo encharcado nos dias seguintes, inviabilizando a execução da tarefa de terraplanagem, já que os equipamentos não podem ser operados de forma eficaz em condições de solo saturado (SCHULDT *et al.*, 2021). Neste trabalho, consideramos que os limites climáticos para a realização das tarefas são conhecidos e que o impacto não se estende para os dias subsequentes. Nesse contexto, a modelagem direta da produtividade implica em estimar a probabilidade de conclusão de cada tarefa em cada dia.

O efeito de um conjunto específico de variáveis climáticas sobre uma tarefa de construção pode variar significativamente, influenciado pelas tecnologias adotadas, pela configuração do canteiro de obras, pelos equipamentos utilizados, pelos materiais empregados e pelos procedimentos adotados (BALLESTEROS-PÉREZ *et al.*, 2017a). Dessa forma, a seleção da combinação adequada de variáveis e intensidades climáticas que viabilizam a realização de determinadas tarefas é específica para cada projeto e tem um impacto direto nos resultados.

A Tabela 1 apresenta os limites climáticos que inviabilizam a execução de cada tipo de

tarefa, uma realidade comum a muitos projetos de construção, como terraplanagem, concretagem, cofragem, siderurgia, pinturas exteriores e pavimentação de asfalto (BALLESTEROS-PÉREZ *et al.*, 2017a). Por exemplo, a tarefa de terraplanagem, representado por τ_1 , não pode ser iniciada se a temperatura média for inferior a 0°C ($T_{mean} \leq 0^\circ\text{C}$) ou se a precipitação for superior a 10mm ($P \geq 10$ mm). É importante ressaltar que a combinação de variáveis apresentada é considerada suficientemente representativa para o Reino Unido, contudo, pode não ser apropriada para outros países ou contextos.

Tabela 1 – Limites climáticos para as tarefas de terraplanagem (τ_1), concretagem (τ_2), cofragem (τ_3), siderurgia (τ_4), pinturas exteriores (τ_5) e pavimentação de asfalto (τ_6), considerando as variáveis climáticas temperatura mínima (T_{min}), temperatura média (T_{mean}), temperatura máxima (T_{max}), precipitação (P), ocorrência de granizo (P_{hail}), ocorrência de neve (P_{snow}), ocorrência de tempestade elétrica (E) e velocidade do vento (V).

Variável climática	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6
$T_{min} \leq 0^\circ\text{C}$		x				x
$T_{mean} \leq 0^\circ\text{C}$	x			x	x	
$T_{max} \geq 40^\circ\text{C}$		x		x		x
$P \geq 1$ mm					x	x
$P \geq 10$ mm	x	x				
$P \geq 30$ mm				x		
P_{hail} é true					x	x
P_{snow} é true					x	x
E é true			x	x		
$V \geq 30$ knots		x	x	x	x	

Fonte: Ballesteros-Pérez *et al.* (2017a).

É necessário obter dados climáticos históricos para as variáveis climáticas de interesse e, em seguida, aplicar os limites climáticos estabelecidos. A Tabela 2 apresenta um exemplo de dados históricos coletados com medições diárias das variáveis climáticas. Em seguida, a Tabela 3 ilustra o resultado após a aplicação dos limites climáticos descritos na Tabela 1, indicando a viabilidade de execução de cada tarefa em cada dia. Geralmente, as medições são realizadas por estações climáticas, que não estão localizadas no local exato onde o projeto será executado, demandando a aplicação individual dos limites climáticos em cada estação, seguida pela combinação dos dados resultantes.

Eventualmente, pode ocorrer a ausência de medições devido a falhas nos equipamentos, como exemplificado pelo valor nulo de T_{mean} no dia 3 de fevereiro de 2022. Durante a etapa de verificação, observa-se que, para algumas variáveis climáticas, como τ_1 , τ_4 e τ_5 , os valores são \emptyset , indicando a ausência de dados, pois essas variáveis dependem de T_{mean} . Para lidar com essas lacunas, podem ser aplicados métodos de imputação. Espera-se que dados climáticos apresentem certa continuidade ou similaridade entre dias próximos. Por exemplo, a temperatura, umidade, pressão atmosférica e outros parâmetros climáticos costumam variar suavemente ao longo do tempo em muitas regiões, possibilitando a utilização de técnicas como interpolação linear e

médias móveis para preencher esses espaços vazios. Na Seção 3.1, serão apresentadas formas de realizar o agrupamento das estações climáticas e de lidar com valores ausentes.

Tabela 2 – Exemplos de dados diários das condições meteorológicas, incluindo temperatura mínima (T_{min}), temperatura média (T_{mean}), temperatura máxima (T_{max}), precipitação (P), precipitação de granizo (P_{hail}), precipitação de neve (P_{snow}), tempestade elétrica (E) e velocidade do vento (V).

Data	T_{min} (°C)	T_{mean} (°C)	T_{max} (°C)	P (mm)	P_{hail}	P_{snow}	E	V (knots)
...
2022-01-30	20	25	36	0	<i>false</i>	<i>false</i>	<i>false</i>	12
2022-01-31	22	27	41	0	<i>false</i>	<i>false</i>	<i>false</i>	10
2022-02-01	18	21	25	31	<i>true</i>	<i>false</i>	<i>false</i>	13
2022-02-02	16	20	25	24	<i>false</i>	<i>false</i>	<i>true</i>	15
2022-02-03	14	<i>Nulo</i>	23	17	<i>false</i>	<i>false</i>	<i>false</i>	15
...

Fonte: Elaborada pelo autor.

Tabela 3 – Exemplos de dados diários indicando se foi possível realizar as tarefas de terraplanagem (τ_1), concretagem (τ_2), cofragem (τ_3), siderurgia (τ_4), pinturas exteriores (τ_5) e pavimentação de asfalto (τ_6) no ano y e dia do ano i .

Data	y	i	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6
...
2022-01-30	2022	30	1	1	1	1	1	1
2022-01-31	2022	31	1	0	1	0	1	0
2022-02-01	2022	32	0	0	1	0	0	0
2022-02-02	2022	33	0	0	0	0	0	0
2022-02-03	2022	34	\emptyset	0	1	\emptyset	\emptyset	0
...

Fonte: Elaborada pelo autor.

2.3 Modelagem de Z : Tempo de Execução do Projeto

Considere um projeto composto por um conjunto de tarefas, denotado por \mathcal{T} , contendo p elementos. Cada tarefa é representada por $t \in \mathcal{T}$, onde $t = \{1, \dots, p\}$. Seja Z a variável aleatória que descreve o tempo de execução total do projeto, ou seja, o número de dias necessários para concluir todas as tarefas de \mathcal{T} .

Para avaliar o tempo de execução do projeto, é necessário considerar o tempo de execução de cada tarefa t , levando em conta fatores climáticos, representados por $D^{(t)}$. Esta variável contabiliza o número de tentativas para completar a tarefa t até alcançar a duração planejada de $d^{(t)}$ dias, considerando que a tarefa teve início no dia $i^{(t)}$.

A execução de uma tarefa t em um determinado dia qualquer i é indicada pela variável $X_i^{(t)}$, que assume o valor 1 se a tarefa foi realizada nesse dia e 0 caso contrário. Modelando a

variável $X_i^{(t)}$, somos capazes de calcular $D^{(t)}$ e, conseqüentemente, Z . No restante da seção, cada variável apresentada será definida de maneira rigorosa.

Seja a variável aleatória $X_i^{(t)} : \Omega \rightarrow \{0, 1\}$ que indica se a tarefa t foi executada no dia i , levando em conta os fatores climáticos:

$$X_i^{(t)} = \begin{cases} 1 & \text{se a tarefa } t \text{ foi executada no dia } i, \\ 0 & \text{caso contrário,} \end{cases}$$

onde Ω é o espaço amostral do experimento aleatório que consiste em tentar realizar a tarefa t . A modelagem da variável $X_i^{(t)}$ será detalhada na próxima seção.

Se uma tarefa t começa no dia $i^{(t)}$, a quantidade de dias que ela pode ser trabalhada nos primeiros $n + 1$ dias após o dia $i^{(t)}$ é dada por:

$$Y_n^{(t)} := \sum_{i=i^{(t)}}^{i^{(t)}+n} X_i^{(t)}.$$

Seja $D^{(t)}$ uma variável aleatória que representa o tempo de execução da tarefa t , considerando fatores climáticos, com duração planejada de $d^{(t)}$ dias, expressa como:

$$D^{(t)} = \min\{n \in \mathbb{N}_+ : Y_n^{(t)} = d^{(t)}\}.$$

Se uma tarefa t não depende de outras tarefas, então o dia de início $i^{(t)}$ é fixo e especificado de acordo com o cronograma de construção. No entanto, se houver dependências entre tarefas, $i^{(t)}$ dependerá das durações das tarefas predecessoras. Denotamos por $\mathcal{T}'^{(t)}$ o conjunto de tarefas predecessoras da tarefa t . Todas as tarefas $t' \in \mathcal{T}'^{(t)}$ precisam ser finalizadas para que a tarefa t possa iniciar. Assumimos, portanto, que $i^{(t)}$ é uma função das durações $D^{(t')}$ e dos dias de início $i^{(t')}$ das tarefas predecessoras $t' \in \mathcal{T}'^{(t)}$, ou seja, $i^{(t)} \sim f(D^{(t')}, i^{(t')} | t' \in \mathcal{T}'^{(t)})$. Em outras palavras, a distribuição de $i^{(t)}$ é determinada pelas distribuições das durações das tarefas predecessoras, dependendo das relações lógicas entre t e cada t' .

A variável Z , representando o tempo total do projeto, é uma função das variáveis $D^{(t)}$. Entretanto, devido às interdependências entre as tarefas, as variáveis $D^{(t)}$ são dependentes, de modo que $i^{(t)}$ depende das distribuições das tarefas predecessoras. Portanto, a distribuição de Z não pode ser obtida facilmente. No entanto, a simulação de Monte Carlo pode ser empregada para estimar essa distribuição.

2.4 Modelagem de $X_i^{(t)}$: A tarefa t foi executada no dia i

A variável aleatória $X_i^{(t)} : \Omega \rightarrow 0, 1$ indica se a tarefa t foi executada no dia i , considerando os fatores climáticos. A modelagem da sequência de variáveis $X_i^{(t)}, X_{i+1}^{(t)}, X_{i+2}^{(t)}, \dots, X_{i+n}^{(t)}$

envolve um processo estocástico não estacionário. Especificamente, estamos interessados em modelar a dinâmica temporal da realização de uma determinada tarefa ao longo do tempo. Consideramos que o comportamento anual se repete ao longo dos anos, de modo que $X_i^{(t)}$ e $X_{i+365}^{(t)}$ têm a mesma distribuição. Para modelar a sazonalidade anual, utilizamos $X_j^{(t)}$, onde j representa o dia do ano correspondente ao dia i , ou seja, $j = g(i)$, onde $g : \mathbb{N}^+ \rightarrow \{1, 2, 3, \dots, 365\}$ é uma função não injetora que mapeia um dia qualquer i para o dia do ano correspondente j . Essa modelagem pode ser realizada utilizando uma distribuição de Bernoulli ou uma cadeia de Markov de tempo discreto. Na [Seção 3.1](#), serão apresentados os métodos para calcular os parâmetros associados a esses modelos.

2.4.1 Distribuição de Bernoulli independentes

A distribuição de Bernoulli é um modelo probabilístico usado para representar experimentos aleatórios com dois resultados possíveis, sendo útil em processos estocásticos simples. Supondo independência entre os dias, podemos considerar que $X_j^{(t)}$ segue uma distribuição de Bernoulli com probabilidade de sucesso $\mathbb{P}(X_j^{(t)} = 1) = \theta_j^{(t)}$. Aqui, $\theta_j^{(t)}$ representa a probabilidade de executar a tarefa t no dia do ano j , o que pode variar conforme as variações sazonais.

Portanto, ao conhecermos $\theta^{(t)} = (\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_{365}^{(t)})$, podemos estimar a distribuição de Z utilizando a simulação de Monte Carlo. No entanto, é importante ressaltar que a suposição de independência da execução da tarefa entre os dias é bastante restritiva.

Se a probabilidade $\theta_j^{(t)}$ for constante para qualquer dia do ano j , denotada como $\theta^{(t)'} = \theta_j^{(t)'} \forall i$, então $Y_n^{(t)} \sim \text{Bin}(n, \theta^{(t)'})$ e $D^{(t)} \sim \text{BinNeg}(d_t, \theta^{(t)'})$.

2.4.2 Cadeia de Markov

Uma cadeia de Markov de tempo discreto e de primeira ordem é uma sequência de eventos estocásticos que possui a propriedade Markoviana, significando que a distribuição de probabilidade do próximo estado depende apenas do estado presente. No nosso contexto, o tempo i corresponde ao dia, enquanto os estados representam a execução de uma determinada tarefa, com o espaço de estados dado por $S = \{0, 1\}$.

As probabilidades de transição podem variar conforme o dia do ano, refletindo as variações sazonais. Formalmente, podemos representar a transição de estado como:

$$\mathbb{P}(X_i^{(t)} = 1 | X_{i-1}^{(t)} = s) = P_{g(i)}^{(t)}(s) = P_j^{(t)}(s),$$

onde $P_j^{(t)}(s)$ é a probabilidade de transição de $X_{i-1}^{(t)} = s$ para $X_i^{(t)} = 1$ no dia i para a tarefa t , e j representa o dia do ano correspondente ao dia i . Cada probabilidade de transição da tarefa t é específica para cada dia do ano j porque a cadeia de Markov não é homogênea.

Conhecendo $\mathbf{P}^{(t)}(\mathbf{0}) = (P_1^{(t)}(0), P_2^{(t)}(0), \dots, P_{365}^{(t)}(0))$ e $\mathbf{P}^{(t)}(\mathbf{1}) = (P_1^{(t)}(1), P_2^{(t)}(1), \dots, P_{365}^{(t)}(1))$, podemos estimar a distribuição de Z utilizando a simulação de Monte Carlo.

2.5 Simulação Monte Carlo

Os métodos de Monte Carlo representam uma ampla categoria de algoritmos computacionais empregados em otimização, integração numérica e geração de amostras a partir de distribuições de probabilidade, um processo conhecido como simulação de Monte Carlo. A ideia central desses métodos é utilizar a aleatoriedade para resolver problemas que poderiam ser determinísticos. O princípio fundamental da simulação de Monte Carlo é que o comportamento de uma estatística pode ser avaliado por meio de um processo empírico, obtido ao gerar e analisar diversas amostras aleatórias (MOONEY, 1997).

O custo computacional associado à simulação de Monte Carlo pode ser consideravelmente elevado. Em geral, o método requer um grande número de amostras para alcançar uma boa aproximação, o que pode resultar em um tempo de execução significativo. Em resumo, o método baseia-se na amostragem aleatória repetida para obter resultados numéricos úteis, especialmente em situações onde outras abordagens se mostram difíceis ou inviáveis.

A estimativa de $p(z)$ é obtida implicitamente por meio de uma amostragem de Monte Carlo que utiliza os valores estimados de $\theta^{(t)}$, $\mathbf{P}^{(t)}(\mathbf{0})$ ou $\mathbf{P}^{(t)}(\mathbf{1})$. Especificamente, geramos uma amostra de $p(z)$ ao percorrer o cronograma dia a dia e realizando sorteios de variáveis aleatórias de Bernoulli com os parâmetros $\hat{\theta}_j^{(t)}$, $\hat{P}_j^{(t)}(0)$ ou $\hat{P}_j^{(t)}(1)$, para determinar se a tarefa t pode ser realizada no dia do ano j no local onde o projeto será implementado. Acompanhamos as tarefas que não puderam ser realizadas e ajustamos a programação de acordo. Indexando as simulações como $b = 1, \dots, B$, onde Z_b é, por definição, uma amostra de $p(z)$, representando o tempo necessário para a conclusão do projeto na simulação em questão. O [Algoritmo 1](#) detalha nosso procedimento para um conjunto de tarefas com dependências do tipo terminar-para-começar, mas outras dependências podem ser facilmente incorporadas neste *framework*.

A amostra de Monte Carlo Z_1, \dots, Z_B pode ser utilizada para estimar $p(z)$ e calcular suas estatísticas. Uma abordagem possível consiste em aplicar um estimador de densidade do kernel em Z_1, \dots, Z_B para aproximar Z por uma distribuição contínua e usar um padrão mais suave (NADARAYA, 1965). Alternativamente, é possível optar por versões discretas (RAJAGOPALAN; LALL, 1995). Além disso, o α -quantil de $p(z)$ pode ser aproximado calculando o α -quantil das amostras. Da mesma forma, a média de Z , dada por $\mathbb{E}[Z] = \sum_z zp(z)$, pode ser aproximada por $B^{-1} \sum_{b=1}^B Z_b$.

Algoritmo 1 – Estimativa do tempo de execução do projeto

Requer: conjunto de tarefas \mathcal{T} , data de início do projeto $data_inicio$, duração d_t de cada tarefa t , probabilidades de execução $\hat{\theta}^{(t)}$, ou $\mathbf{P}^{(t)}(\mathbf{0})$ e $\mathbf{P}^{(t)}(\mathbf{1})$, número de amostras de monte carlo B

Assegure: Conjunto de amostras Z_b , que representa o numero de dias que levou para realizar o projeto na b -ésima amostras de Monte Carlo

para b in $\{1, \dots, B\}$ **faça**

enquanto o conjunto \mathcal{T} não foi finalizado **faça**

para t in \mathcal{T} que podem iniciar **faça**

$data_inicio_t \leftarrow$ data de início da tarefa t

$sucessos \leftarrow 0$

enquanto $sucessos < d_t$ **faça**

$j \leftarrow$ dia do ano de $data_inicio_t$

se $mc = True$ **então**

 Sorteie $X \sim \text{Ber}(P_j^{(t)}(X_{-1}))$

senão

 Sorteie $X \sim \text{Ber}(\hat{\theta}_j^{(t)})$

fim se

se $X = 1$ **então**

$sucessos \leftarrow sucessos + 1$

fim se

se $sucessos < d_t$ **então**

$data_inicio_t \leftarrow data_inicio_t + 1$ dia

fim se

se $mc = True$ **então**

$X_{-1} \leftarrow X$

fim se

fim enquanto

fim para

fim enquanto

$data_fim \leftarrow \max\{data_inicio_t \text{ para cada tarefa } t \text{ em } \mathcal{T}\}$

$Z_b \leftarrow data_fim - data_inicio$

fim para

return Z_b

2.6 Comparação de modelos

O objetivo dos métodos propostos é desenvolver modelos computacionais com alta capacidade preditiva. A seleção de modelos consiste em escolher um modelo estatístico dentre um conjunto de modelos candidatos para um determinado conjunto de dados. Para dados temporais, a divisão de dados por conveniência é um método determinístico amplamente utilizado e eficiente (REITERMANOVA *et al.*, 2010).

Para iniciar o processo, diversas datas de início do projeto são selecionadas ou sorteadas aleatoriamente. Para cada data de início, os modelos candidatos são ajustados aos dados históricos disponíveis para estimar a duração do projeto. A função de perda é utilizada para avaliar a capacidade do algoritmo em modelar os dados, comparando as previsões com os valores observados. O risco é definido como a esperança da função de perda, sendo estimado por meio da divisão dos dados. O modelo que apresenta o menor risco estimado é, então, selecionado para ser utilizado no projeto.

A escolha da função de perda é crucial para avaliar adequadamente a capacidade preditiva dos modelos. Métricas de aprendizado de máquina podem ser empregadas para mensurar a precisão das previsões pontuais. Contudo, essas métricas não são projetadas para avaliar quão bem uma estimativa de $p(z)$ descreve a incerteza total acerca de Z . Por exemplo, o erro quadrático pode ser utilizado para quantificar a proximidade da estimativa \hat{z} em relação ao valor real z , expresso por:

$$SE = (\hat{z} - z)^2.$$

Entre as medidas usuais, destacam-se o erro absoluto e o erro percentual absoluto. É importante ressaltar que a distribuição de $p(z)$ é frequentemente resumida em um único valor pontual \hat{z} , geralmente representado pela média ou valor esperado da distribuição. Consequentemente, essas métricas não avaliam adequadamente as estimativas das caudas e dos quantis da distribuição. Na [Seção 3.2](#) apresentaremos medidas para quantificar distribuições.

METODOLOGIA

3.1 Estimando $\theta^{(t)}$, $\mathbf{P}^{(t)}(\mathbf{0})$ e $\mathbf{P}^{(t)}(\mathbf{1})$

Nesta seção, apresentaremos o procedimento para estimar $\theta^{(t)}$, $\mathbf{P}^{(t)}(\mathbf{0})$ e $\mathbf{P}^{(t)}(\mathbf{1})$. Primeiramente, descreveremos a metodologia para estimar $\theta^{(t)}$ e demonstraremos como calcular $\mathbf{P}^{(t)}(\mathbf{0})$ e $\mathbf{P}^{(t)}(\mathbf{1})$ de forma análoga.

Conforme definido na [Seção 2.4](#), $\theta^{(t)} = (\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_{365}^{(t)})$, onde $\theta_j^{(t)}$ representa a probabilidade de executar a tarefa t no dia do ano j . Para isso, pressupomos que todas as etapas anteriores foram realizadas. Assim, temos conhecimento do cronograma de construção do projeto, desconsiderando a previsão meteorológica, com um conjunto de tarefas \mathcal{T} , em que $|\mathcal{T}| = p$. Além disso, nosso conjunto de dados é representado por $\mathbf{H}^{(1)}, \dots, \mathbf{H}^{(p)}$, onde $\mathbf{H}^{(t)}$ é uma matriz binária tridimensional que indica a execução da tarefa t . O elemento $H_{l,y,j}^{(t)}$ indica a execução da tarefa t no local $l \in \mathcal{L}$, dia do ano j e ano $y \in \mathcal{Y}$, sendo \mathcal{L} o conjunto de estações climáticas disponíveis e \mathcal{Y} o conjunto de anos disponíveis. A [Tabela 4](#) apresenta um exemplo dos valores de $H_{l,y,j}^{(t)}$ para a tarefa $t = 1$, definida na [Tabela 1](#), verificando se a temperatura média foi inferior a 0°C e a precipitação foi superior a 10 mm. Por exemplo, $H_{1,2022,30}^{(1)} = 1$ indica que os limites climáticos não foram ultrapassados em 30 de janeiro de 2022 na estação 1, enquanto $H_{1,2022,33}^{(1)} = 0$ porque $P = 20 > 10$ em 2 de fevereiro de 2022, e $H_{1,2022,34}^{(1)} = 0$ porque T_{mean} é *Nulo* em 3 de fevereiro de 2022.

Para estimar $\theta_j^{(t)}$, utilizamos uma técnica de suavização nos valores de $H^{(t)}$ através de uma função $f^{(t)} : \mathbb{R}^{|\mathcal{L}| \times |\mathcal{Y}| \times 365} \rightarrow \mathbb{R}$. Especificamente, utilizamos o seguinte estimador:

$$\widehat{\theta}_j^{(t)} = f^{(t)}(\mathbf{H}^{(t)}). \quad (3.1)$$

A função $f^{(t)}$ é uma média ponderada, onde cada dimensão da matriz $H^{(t)}$ é suavizada individualmente. Para isso, descompomos a função em três funções de ponderação distintas w_{loc} , w_{dia} e

Tabela 4 – Exemplos de dados diários da temperatura média (T_{mean}), precipitação (P), realização de tarefas de terraplanagem (τ_1) e o elemento $H^{(t)}_{l,y,j}$ para o ano y , dia do ano j e local l .

Data	y	j	l	T_{mean} (°C)	P (mm)	τ_1	$H^{(t)}_{l,y,i}$
...
2022-01-30	2022	30	1	25	0	1	$H^{(1)}_{1,2022,30}$
2022-01-31	2022	31	1	27	0	1	$H^{(1)}_{1,2022,31}$
2022-02-01	2022	32	1	21	31	0	$H^{(1)}_{1,2022,32}$
2022-02-02	2022	33	1	20	24	0	$H^{(1)}_{1,2022,33}$
2022-02-03	2022	34	1	<i>Nulo</i>	17	\emptyset	$H^{(1)}_{1,2022,34}$
...

Fonte: Elaborada pelo autor.

w_{ano} , que operam em cada dimensão da matriz, indicando a proximidade de cada informação $H^{(t)}_{l,y,j'}$ a $\theta_j^{(t)}$. Com essa decomposição, reescrevemos a função $f^{(t)}$ como:

$$\hat{\theta}_j^{(t)} = f^{(t)}(\mathbf{H}^{(t)}) = \frac{\sum_{l \in \mathcal{L}} \sum_{j'=1}^{365} \sum_{y \in \mathcal{Y}} w_{loc}(l_0, l) w_{dia}(j, j') w_{ano}(y_0, y) H^{(t)}_{l,y,j'} \mathbb{I}(H^{(t)}_{l,y,j'} \neq \emptyset)}{\sum_{l \in \mathcal{L}} \sum_{j'=1}^{365} \sum_{y \in \mathcal{Y}} w_{loc}(l_0, l) w_{dia}(j, j') w_{ano}(y_0, y) \mathbb{I}(H^{(t)}_{l,y,j'} \neq \emptyset)},$$

onde a função $w_{dia}(j, j')$ mede a proximidade entre os dias do ano j e j' , enquanto $w_{ano}(y_0, y)$ mede a proximidade entre os anos y e y_0 (o ano em que o projeto será executado), e $w_{loc}(l_0, l)$ mede a proximidade entre a localização l e l_0 (o local onde o projeto será executado). O denominador garante que a soma dos pesos seja igual a 1. A função indicadora $\mathbb{I}(H^{(t)}_{l,y,j'} \neq \emptyset)$ garante que os valores nulos ($H^{(t)}_{l,y,j'} = \emptyset$) sejam ignorados no cálculo. Essa abordagem é não paramétrica, pois f não faz suposições fortes sobre o comportamento dos dados.

O vetor $\hat{\theta}^{(t)}$ representa as probabilidades de executar a tarefa t ao longo do ano. Após suavizar os valores, podemos presumir uma forma paramétrica $f_s^{(t)}$ para a sazonalidade anual, como, por exemplo, ondas senoidais, e ajustá-la aos valores de $\hat{\theta}^{(t)}$.

3.1.1 Escolhendo $w_{ano}(y_0, y)$ - Decaimento exponencial

Se acreditarmos ser plausível assumir que a distribuição das condições climáticas ao longo dos anos é estacionária, podemos atribuir pesos constantes nessa dimensão:

$$w_{ano}(y_0, y) = 1 \forall y \in \mathcal{Y}.$$

Por outro lado, ponderar os anos de maneira diferente pode ser vantajoso para capturar tendências temporais. Por exemplo, se suspeitarmos que mudanças climáticas afetaram as condições ao longo do tempo, ou que os últimos anos são mais representativos do ano em que o projeto será implementado, podemos utilizar pesos com decaimento exponencial:

$$w_{\text{ano}}(y_0, y) = \alpha^{y_0 - y},$$

onde y_0 é o ano em que o projeto está planejado para iniciar e $\alpha > 0$ define o grau de decaimento, sendo que $\alpha = 1$ resulta na distribuição estacionária. Essa abordagem permite capturar efeitos de mudanças climáticas, atribuindo maior peso aos anos mais recentes. Por exemplo, se uma tarefa é afetada pelas temperaturas médias e estas estão aumentando ao longo dos anos, pode ser adequado atribuir maior peso aos anos mais recentes.

3.1.2 Escolhendo $w_{\text{dia}}(j, j')$ - Média móvel

Adotando uma abordagem de média móvel, podemos definir os pesos como:

$$w_{\text{dia}}(j, j') = \begin{cases} 1 & , \text{ se } (|j - j'| \bmod [365 - k]) < k \\ 0 & , \text{ caso contrário} \end{cases}$$

onde k representa o tamanho da janela móvel. Note que o final e o início do ano estão conectados e próximos. Se $k = 0$, utilizamos apenas as informações daquele dia, resultando em um estimador com alta variância; com $k \rightarrow \infty$, reduzimos essa variância, mas aumentamos seu viés.

3.1.3 Escolhendo $w_{\text{loc}}(l_0, l)$ - Distância euclidiana

Para modelar os pesos associados à localização da construção, podemos utilizar medidas de distância ou métodos de interpolação. Por exemplo, podemos atribuir pesos maiores a locais mais próximos à localização l_0 , onde o projeto será executado. Isso pode ser feito da seguinte maneira:

$$w_{\text{loc}}(l_0, l) = \frac{1}{d(l_0, l)},$$

onde $d(l_0, l)$ representa a distância entre as localizações l_0 e l . A distância euclidiana entre latitude e longitude é comumente utilizada, mas outras medidas, como a distância geodésica, que considera melhor a topologia do globo, podem ser adotadas conforme a necessidade. Quando $|\mathcal{L}| = 1$, não é necessário utilizar pesos, então $w_{\text{loc}}(l_0, l) = 1$.

3.1.4 Escolhendo $f_s^{(t)}$ - Ondas senoidais

Podemos considerar um comportamento paramétrico nos dados, como expressões de onda senoidal (BALLESTEROS-PÉREZ *et al.*, 2017a):

$$\widehat{\theta}_j^{(t)'} = K^{(t)} + A^{(t)} \cos\left(2\pi f^{(t)} \left[\frac{j}{365} - \varphi^{(t)}\right]\right),$$

onde $K^{(t)}$ é o deslocamento vertical, $A^{(t)}$ é a amplitude da onda, $f^{(t)}$ é a frequência (com $f^{(t)} = 1$ correspondendo a uma oscilação completa) e $\varphi^{(t)}$ corresponde ao dia do ano em que a onda sinusoidal atinge o seu valor máximo. Nesse caso, obtemos os valores de $K^{(t)}$, $A^{(t)}$ e $f^{(t)}$, ajustando o modelo em $\hat{\theta}^{(t)}$ por mínimos quadrados.

3.1.5 Cadeia de Markov

Os valores de $P_j^{(t)}(0)$ e $P_j^{(t)}(1)$ podem ser determinados de forma análoga à obtenção de $\theta_j^{(t)}$, ao extrair informações de $\mathbf{H}^{(t)}$ com base no estado do dia anterior. Em outras palavras,

$$P_j^{(t)}(s) = f^{(t)}(\mathbf{H}^{(t),s}),$$

onde

$$H_{l,y,j}^{(t),s} = \begin{cases} H_{l,y,j}^{(t)} & \text{se o dia anterior tem estado } s, \\ \emptyset & \text{caso contrário.} \end{cases}$$

Utilizando o exemplo da [Tabela 4](#), temos que $H_{l,y,[30:34]}^{(t)} = [1, 1, 0, 0, \emptyset]$, logo $H_{l,y,[31:34]}^{(t),0} = [\emptyset, \emptyset, 0, \emptyset]$ e $H_{l,y,[31:34]}^{(t),1} = [1, 0, \emptyset, \emptyset]$. Essa formulação permite calcular a probabilidade de transição de $X_i^{(t)}$ dado o estado anterior $X_{i-1}^{(t)}$, baseando-se nas informações disponíveis em $H^{(t)}$ correspondentes ao estado s do dia anterior.

3.1.6 Hiperparâmetros

As funções propostas dependem de parâmetros que determinam o ajuste das probabilidades. A [Tabela 5](#) lista os hiperparâmetros do modelo que precisam ser selecionados. Na próxima seção, apresentaremos funções de perda que podem ser utilizadas para a seleção da melhor combinação de hiperparâmetros para o conjunto de dados.

Tabela 5 – Lista de hiperparâmetros utilizados no modelo, com seus respectivos intervalos e descrições detalhadas.

Hiperparâmetro	Intervalo	Descrição
<i>alpha</i>	[0, 1]	Grau do decaimento exponencial. Mais detalhes na Seção Subseção 3.1.1 .
<i>smoothing</i>	\mathbb{N}^+	Tamanho da janela móvel. Mais detalhes na Seção Subseção 3.1.2 .
<i>cosine</i>	Booleano	Se Verdadeiro, é ajudado uma onda senoidal. Mais detalhes na Seção Subseção 3.1.4 .
<i>mc</i>	Booleano	Se Verdadeiro, é utilizada uma cadeia de markov. Caso contrário, é utilizado a distribuição de Bernoulli. Mais detalhes na Seção Subseção 3.1.5 .

Fonte: Elaborada pelo autor.

3.2 Seleção de modelos

Suponha que $\hat{p}_1(z|i,y), \dots, \hat{p}_n(z|i,y)$ sejam n modelos diferentes para $p(z|i,y)$, onde i representa o dia e y o ano em que o projeto será implementado. Nosso objetivo é selecionar o modelo que melhor se aproxima de p . Para atingir esse objetivo, projetamos uma função de perda L capaz de quantificar adequadamente o poder preditivo de um modelo qualquer \hat{p} . Em seguida, estimamos o risco associado a essa função de perda através da divisão de dados por conveniência:

$$\hat{R}(p, \hat{p}) = \frac{1}{m} \sum_{k=1}^m L(Z_k, \hat{p}(z|I_k, Y_k))$$

onde k representa o número de datas de início em que o projeto foi simulado, e (Z_k, I_k, Y_k) é uma amostra de $p(z, i, y) := p(z|i, y)r(i, y)$, sendo $r(i, y)$ a distribuição escolhida sobre (i, y) .

Na prática, criamos os conjuntos de dados $(Z_1, I_1, Y_1), \dots, (Z_m, I_m, Y_m)$ utilizando uma abordagem de divisão temporal. Escolhemos ou sorteamos diferentes datas de início (I_k, Y_k) para o projeto e treinamos os modelos utilizando apenas dados anteriores a cada data de início. Em seguida, obtemos os valores de Z_k aplicando o cronograma de construção aos dados posteriores à data de início, que contêm as condições climáticas reais observadas durante esse período.

Consideramos duas funções de perda distintas: a primeira mede diretamente o desempenho de \hat{p} , enquanto a segunda leva em conta que, em muitos problemas, estamos mais interessados em estimar as caudas superiores da distribuição $p(z)$.

3.2.1 Perda de densidade

A função de perda $L(p, \hat{p})$ compara diretamente a distribuição estimada com a distribuição verdadeira:

$$L(p, \hat{p}) = \sum_z (p(z|i, y) - \hat{p}(z|i, y))^2,$$

em que L depende da distribuição verdadeira $p(z)$, o que a torna impossível de ser calculada. No entanto, podemos decompor L da seguinte forma:

$$L(p, \hat{p}) = \sum_z \hat{p}^2(z|i, y) - 2\hat{p}(Z|i, y) + K,$$

onde K é uma constante que não depende de \hat{p} . Dessa forma, K não precisa ser calculado para ranquear o modelo \hat{p} de acordo com L . Assim, o risco pode ser estimado - até a constante K - utilizando a expressão abaixo:

$$\widehat{R}(p, \widehat{p}) = \frac{1}{m} \sum_{k=1}^m \left(\sum_z \widehat{p}^2(z|I_k, Y_k) - 2\widehat{p}(Z_k|I_k, Y_k) \right).$$

O termo $\sum_z \widehat{p}^2(z|I_k, Y_k)$ leva em consideração o formato da distribuição, penalizando distribuições muito concentradas. Por outro lado, o termo $\widehat{p}(Z_k|I_k, Y_k)$ representa a probabilidade associada ao valor observado, e é desejável que essa probabilidade seja alta para o valor observado. É importante ressaltar que, caso a densidade esteja mal estimada e não contenha o valor observado, apenas o formato da distribuição é considerado na função de perda. Nesse caso, a medida de perda não quantifica o quão distante a previsão incorreta está do valor observado.

3.2.2 Perda pinball ponderada

A perda proposta é baseada na perda de pinball $L_\alpha(z, \widehat{z}_\alpha)$, frequentemente utilizada na estimação da regressão α -quantil (KOENKER; HALLOCK, 2001; STEINWART; CHRISTMANN, 2011):

$$L_\alpha(z, \widehat{z}_\alpha) = \begin{cases} (z - \widehat{z}_\alpha)\alpha, & \text{se } z \geq \widehat{z}_\alpha \\ (\widehat{z}_\alpha - z)(1 - \alpha), & \text{cao contrário,} \end{cases}$$

onde z é a duração real de construção e \widehat{z}_α é o α -quantil estimado. Nossa abordagem difere da perda pinball padrão ao combinar diversos valores de α para estimar a incerteza dos modelos. Essa combinação é realizada por meio de uma função de ponderação $w : [0, 1] \rightarrow \mathbb{R}^+$, que atribui um peso para cada valor de α e garante que a integral da função no intervalo $[0, 1]$ seja igual a 1. A perda ponderada $L^w(Z, \widehat{p})$ é definida da seguinte forma:

$$L^w(Z, \widehat{p}) = \int w(\alpha) L_\alpha(Z, Q(\alpha)) d\alpha,$$

onde $Q(\alpha)$ é o quantil α de \widehat{p} , dado por $Q_j(\alpha) = \inf\{x \in \mathbb{N} : \sum_{z \leq x} \widehat{p}(z) \geq \alpha\}$. O risco é estimado por:

$$\widehat{R}_w(p, \widehat{p}) = \frac{1}{m} \sum_{k=1}^m \int w(\alpha) L_\alpha(Z_k, Q_j(\alpha)) d\alpha.$$

A escolha da função de ponderação $w(\alpha)$ pode ser feita para ponderar diferentemente as caudas, dependendo das especificidades da aplicação. No nosso caso, nos preocupamos mais em estimar bem as caudas superiores de $p(z)$, porque queremos, com uma probabilidade alta, que o valor observado esteja dentro da densidade. Por isso, a ponderação da cauda direita é mais adequada, em que $w(\alpha) \propto \alpha^2$. Isso significa que os valores de $L_\alpha(z, \widehat{z}_\alpha)$ para os valores de α mais altos são mais importantes do que para os valores de α mais baixos. Além da ponderação da cauda direita, há outras opções para a função de ponderação $w(\alpha)$ que podem ser utilizadas para

dar ênfase nos quantis da distribuição de diferentes maneiras. Por exemplo, é possível considerar a cauda esquerda da distribuição usando pesos $w(\alpha) \propto (1 - \alpha)^2$. Também podemos usar pesos que ponderam as duas caudas igualmente, como $w(\alpha) \propto (\alpha - 0.5)^2$.

3.2.3 Ilustração

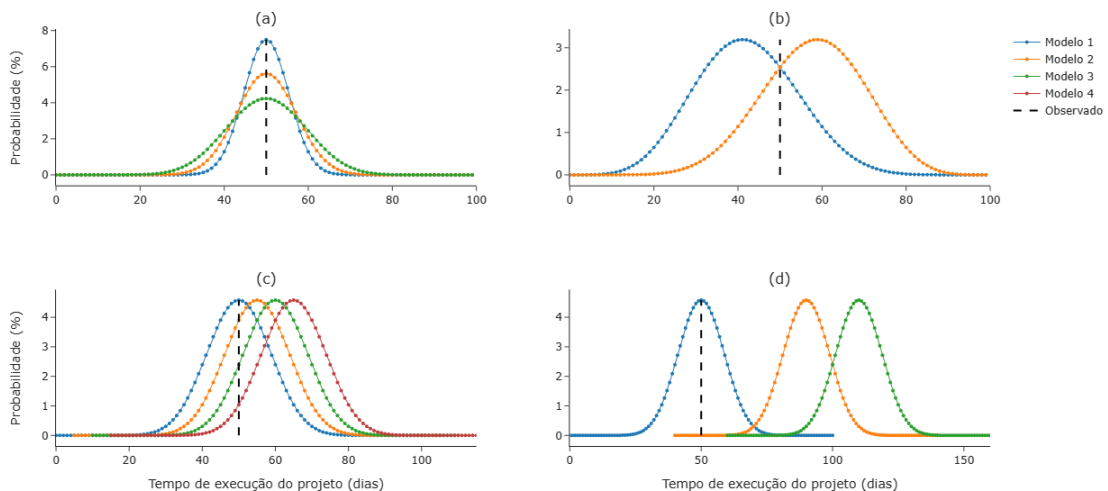
Para compreender melhor as perdas propostas, elas foram aplicadas a distribuições com diferentes formas. A função de perda quadrática foi calculada para fins de comparação e é dada por:

$$L(Z, p_j) = (Z - \mathbb{E}(\hat{p}))^2,$$

onde $\mathbb{E}(\hat{p})$ é a média de \hat{p} .

A Figura 3 apresenta a comparação das funções densidade de probabilidade para diferentes modelos. Nestes exemplos, cada modelo possui uma distribuição Beta-binomial dado por $c + \text{BetaBin}(100, a, b)$. A variável aleatória não é importante aqui, pois o interesse está em avaliar quão bem cada distribuição de um modelo proposto se ajusta ao valor observado. O parâmetro c representa a distância da média da distribuição ao valor observado, enquanto os parâmetros a e b controlam o formato da distribuição. Se a e b são diferentes, a distribuição é assimétrica, e esses parâmetros determinam se a maior parte da distribuição está localizada à direita ou à esquerda de sua média.

Figura 3 – Comparação das funções densidade de probabilidade de diferentes modelos para (a) distribuições com mesma média e variâncias diferentes, (b) distribuições simétricas em relação ao valor observado, (c) distribuições com mesma variância e médias mais distantes do valor observado e (d) distribuições com probabilidade zero para o valor observado.



Fonte: Elaborada pelo autor.

A [Tabela 6](#) apresenta as medidas de perda para as distribuições selecionadas. Na [Figura 3-\(a\)](#), as distribuições possuem a mesma média e variâncias diferentes. A perda quadrática é igual a zero para todas as distribuições, pois a média é igual ao valor observado. Analisando outras medidas de perda, observamos que as distribuições dos modelos 1, 2 e 3 apresentam desempenhos progressivamente piores à medida que a variância aumenta, demonstrando que o uso de métricas que resumem toda a distribuição em um único valor, como a média, não é adequado para medir toda a incerteza da distribuição.

Na [Figura 3-\(b\)](#), as distribuições são simétricas em relação ao valor observado e estão concentradas em lados diferentes. Observamos que a perda quadrática, a densidade, a pinball sem peso e a pinball ponderada pelas duas caudas apresentam resultados iguais, pois essas medidas são invariantes à rotação em torno do valor observado e não consideram se a densidade está concentrada abaixo ou acima do valor observado. A distribuição do modelo 1 apresenta quantis inferiores mais distantes do valor observado, resultando em um maior valor da perda pinball ponderada pela cauda esquerda. Por outro lado, a distribuição do modelo 2 possui os quantis superiores mais distantes do valor observado, resultando em um maior valor da perda pinball ponderada pela cauda direita. Logo, a escolha de pesos na perda pinball pode ser feita para ponderar de maneira diferente os quantis da distribuição.

Na [Figura 3-\(c\)](#), as distribuições possuem variâncias iguais e as médias se afastam do valor observado. Observamos que as distribuições mais distantes apresentam desempenho pior para todas as métricas, pois suas densidades estão mais distantes do valor observado, conforme esperado.

Por fim, na [Figura 3-\(d\)](#), o valor observado está fora da distribuição dos modelos 2 e 3. É importante ressaltar que, embora a distribuição do modelo 3 esteja mais distante do valor observado, a perda de densidade é igual para ambas as distribuições. Isso ocorre porque essa perda leva em consideração o formato da distribuição e se a previsão está correta com alta probabilidade. Caso o valor observado não esteja dentro da densidade, a métrica não é alterada para translações, desde que o valor continue fora da densidade. Portanto, essa perda não quantifica o quão distantes estamos prevendo incorretamente, o que pode não ser adequado em algumas situações.

Tabela 6 – Valores de perda quadrática, de densidade e de pinball para diferentes grupos e modelos. Cada grupo contém modelos caracterizados por uma distribuição específica $c + BetaBin(100, a, b)$ com parâmetros c , a e b .

Grupo	Modelo	Distribuição			Qadrática	Densidade	Pinball			
		c	a	b			Sem peso	Duas caudas	Cauda Esquerda	Cauda Direita
(a)	1	0	400	400	0	-0.150	0.62	0.05	0.20	0.20
	2	0	50	50	0	-0.112	0.82	0.07	0.27	0.27
	3	0	20	20	0	-0.084	1.09	0.09	0.36	0.36
(b)	1	0	8	11	62	-0.051	2.47	0.13	0.99	0.50
	2	0	11	8	62	-0.051	2.47	0.13	0.50	0.99
(c)	1	0	25	25	0	-0.091	1.01	0.08	0.34	0.34
	2	5	25	25	25	-0.078	1.57	0.09	0.31	0.65
	3	10	25	25	100	-0.048	3.09	0.14	0.77	1.05
	4	15	25	25	225	-0.021	5.21	0.26	1.67	1.46
(d)	1	0	25	25	0	-0.091	1.01	0.08	0.34	0.34
	2	40	25	25	1600	0.000	17.57	1.27	7.79	3.55
	3	60	25	25	3600	0.000	27.57	2.11	12.78	5.21

Fonte: Elaborada pelo autor.

EXPERIMENTOS

Nesta seção, apresentamos a otimização de hiperparâmetros, a seleção de modelos e aplicações do modelo. Para demonstrar a eficiência da escolha adequada de hiperparâmetros na captura de padrões em dados, conduzimos experimentos simulados. Além disso, aplicamos o método a um conjunto de dados reais, onde realizamos a otimização de hiperparâmetros, comparamos com outro modelo e estimamos a distribuição do tempo de execução do projeto para diferentes datas de início.

4.1 Dados simulados

Para demonstrar o impacto da escolha dos hiperparâmetros na captura de diferentes padrões nos dados, realizamos quatro experimentos distintos. Cada experimento simula uma única tarefa com duração planejada de 90 dias, apresentando características sazonais e tendências únicas. Os dados históricos de realização da tarefa, representados pela matriz $\mathbf{H}^{(1)}$, foram simulados dia a dia a partir de uma distribuição de Bernoulli. As probabilidades de geração para cada experimento são detalhadas na [Tabela 7](#), exibindo uma variedade de padrões sazonais e tendências ao longo de um período de 30 anos ($\mathcal{Y} = \{1, 2, \dots, 30\}$). A [Figura 4](#) ilustra essas probabilidades associadas a cada dia. Nos Experimentos 1 e 2, foi aplicado um padrão anual senoidal, com o Experimento 2 adicionando uma tendência linear crescente. Nos Experimentos 3 e 4, as sazonalidades são mais complexas, caracterizadas por uma combinação de ondas senoidais.

Para esses experimentos, o conjunto de validação consiste em datas de início para cada mês ao longo do próximo ano. Consideramos as seguintes combinações de hiperparâmetros para suavizar a sazonalidade anual: o parâmetro *smoothing* com valores de 0 ou 7, e o parâmetro *cosine* com *smoothing* igual a 0. Cada combinação foi testada com o parâmetro *alpha* variando entre 0.9, 0.95 ou 1. Não utilizamos o parâmetro *mc* para simplificar a análise dos gráficos, uma vez que os dados não apresentam dependência temporal. O tempo de execução do projeto foi

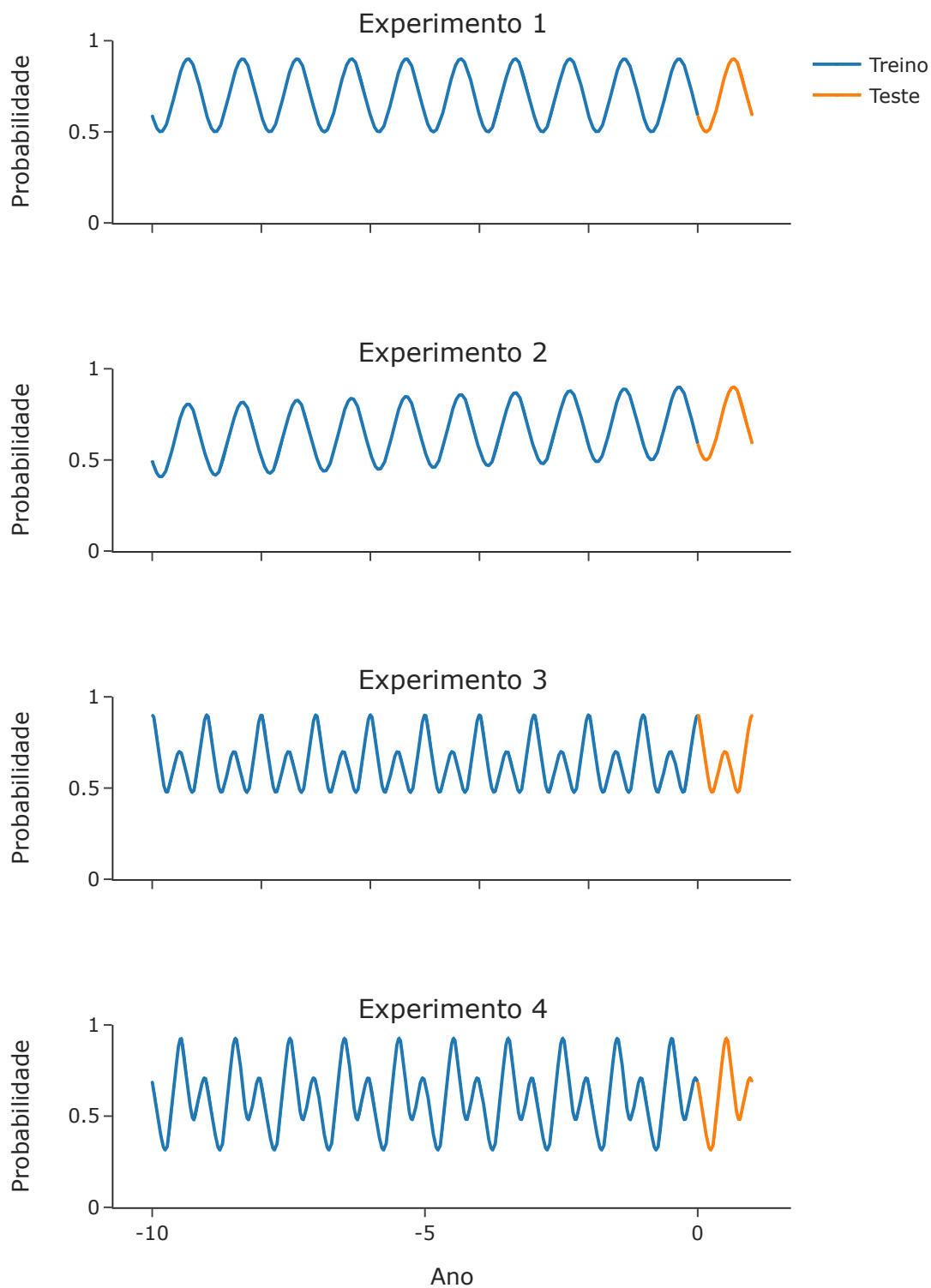
estimado e a perda pinball ponderada pela cauda direita foi calculada.

Tabela 7 – Descrição e probabilidade de geração para cada experimento. Cada experimento é descrito em termos do padrão anual (senoidal ou complexo) e da presença ou ausência de tendência. As probabilidades de geração são expressas em função do dia do ano j e ano y .

Exp.	Descrição	Probabilidade de geração
1	Padrão anual senoidal. Sem tendência.	$0.7 + 0.2\cos\left(2\pi\frac{j-240}{365}\right)$
2	Padrão anual senoidal. Tendência linear.	$0.7 + 0.2\cos\left(2\pi\frac{j-240}{365}\right) - (31 - y)0.01$
3	Padrão anual complexo. Sem tendência.	$\left(0.8 + 0.1\cos\left(2\pi\frac{j-1}{365}\right)\right)\left(0.8 + 0.2\cos\left(4\pi\frac{j-1}{365}\right)\right)$
4	Padrão anual complexo. Sem tendência.	$\left(0.8 + 0.2\cos\left(2\pi\frac{j-240}{365}\right)\right)\left(0.75 + 0.25\cos\left(4\pi\frac{j-1}{365}\right)\right)$

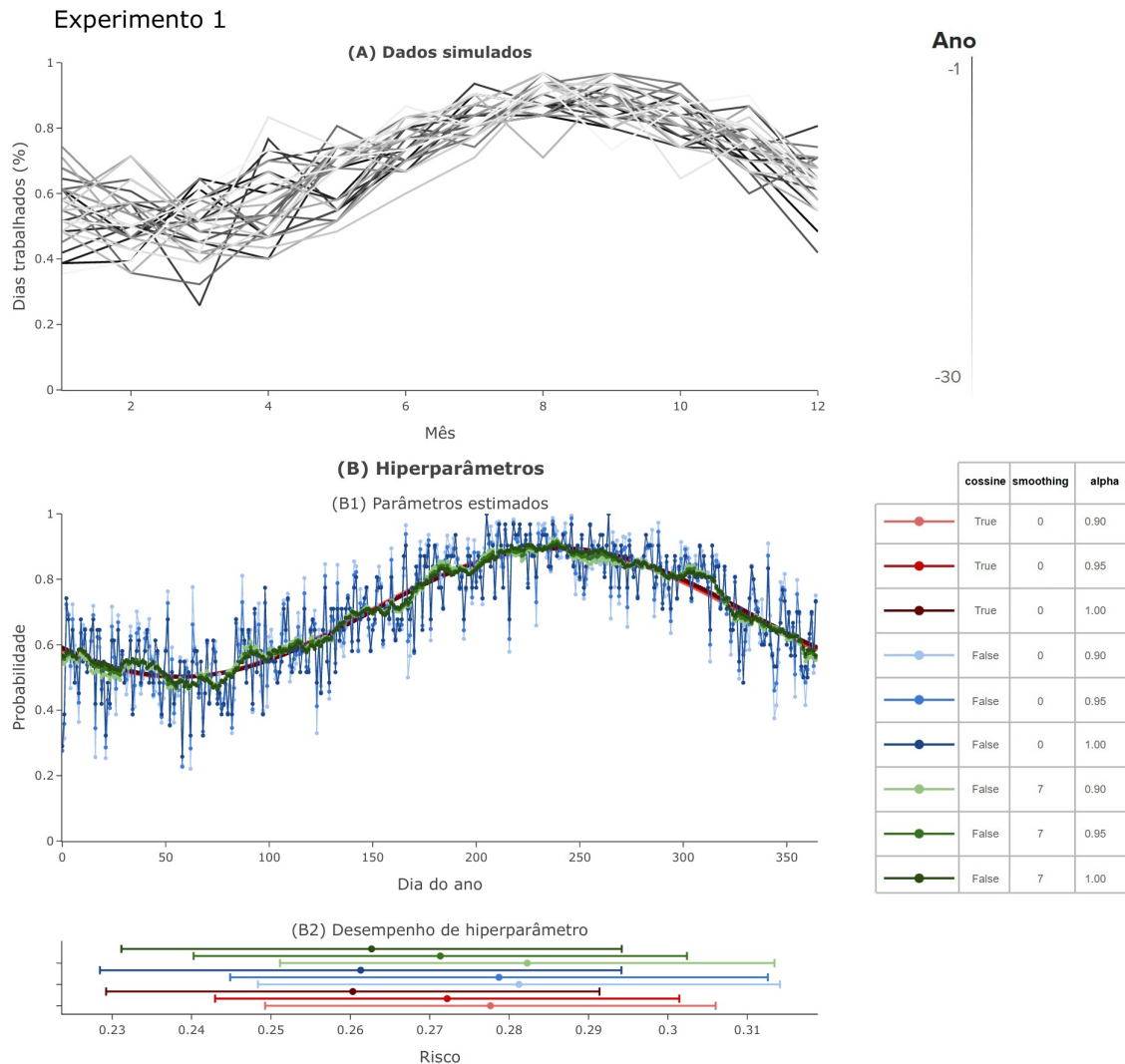
Fonte: Elaborada pelo autor.

Figura 4 – Probabilidade de geração ao longo do tempo para cada experimento. Cada gráfico está dividido em duas partes: a fase de treino (linha azul) e a fase de teste (linha laranja).



Fonte: Elaborada pelo autor.

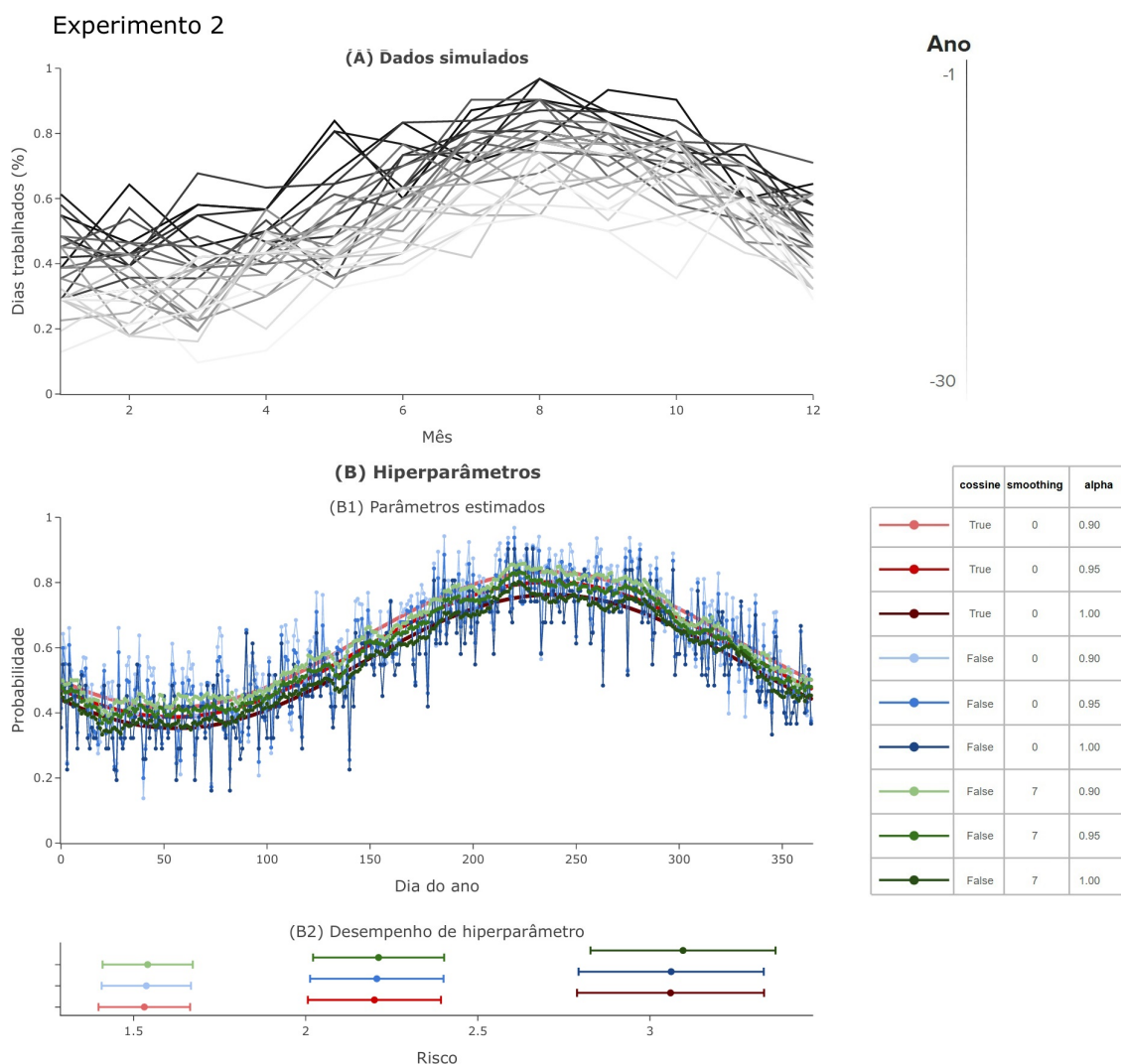
Figura 5 – Resultados do Experimento 1. (A) Porcentagem de dias trabalhados ao longo dos 12 meses, com cada linha representando um ano no histórico de dados. (B) Hiperparâmetros: (B1) Parâmetros Estimados: Probabilidade ao longo do ano com diferentes configurações de hiperparâmetros. (B2) Desempenho de Hiperparâmetro: Risco associado às combinações de hiperparâmetros, medido pela perda pinball da cauda direita.



Fonte: Elaborada pelo autor.

A Figura 5 apresenta os resultados do Experimento 1. Na Figura 5-(A) observamos a porcentagem de dias trabalhados ao longo dos 12 meses, com cada linha representando um ano no histórico de dados. A sazonalidade anual apresenta um comportamento senoidal, e as linhas se sobrepõem devido à ausência de tendência nos dados. Na Figura 5-(B1), são apresentadas as probabilidades de execução da tarefa ao longo do ano para diferentes combinações de hiperparâmetros, mostrando que todos capturam o comportamento dos dados com facilidade. Na Figura 5-(B2), o desempenho de cada combinação de hiperparâmetros é avaliado, indicando que todas as combinações apresentaram desempenho próximo. Isso sugere que tanto a abordagem paramétrica de onda senoidal quanto a de média móvel são adequadas quando os dados seguem o comportamento esperado.

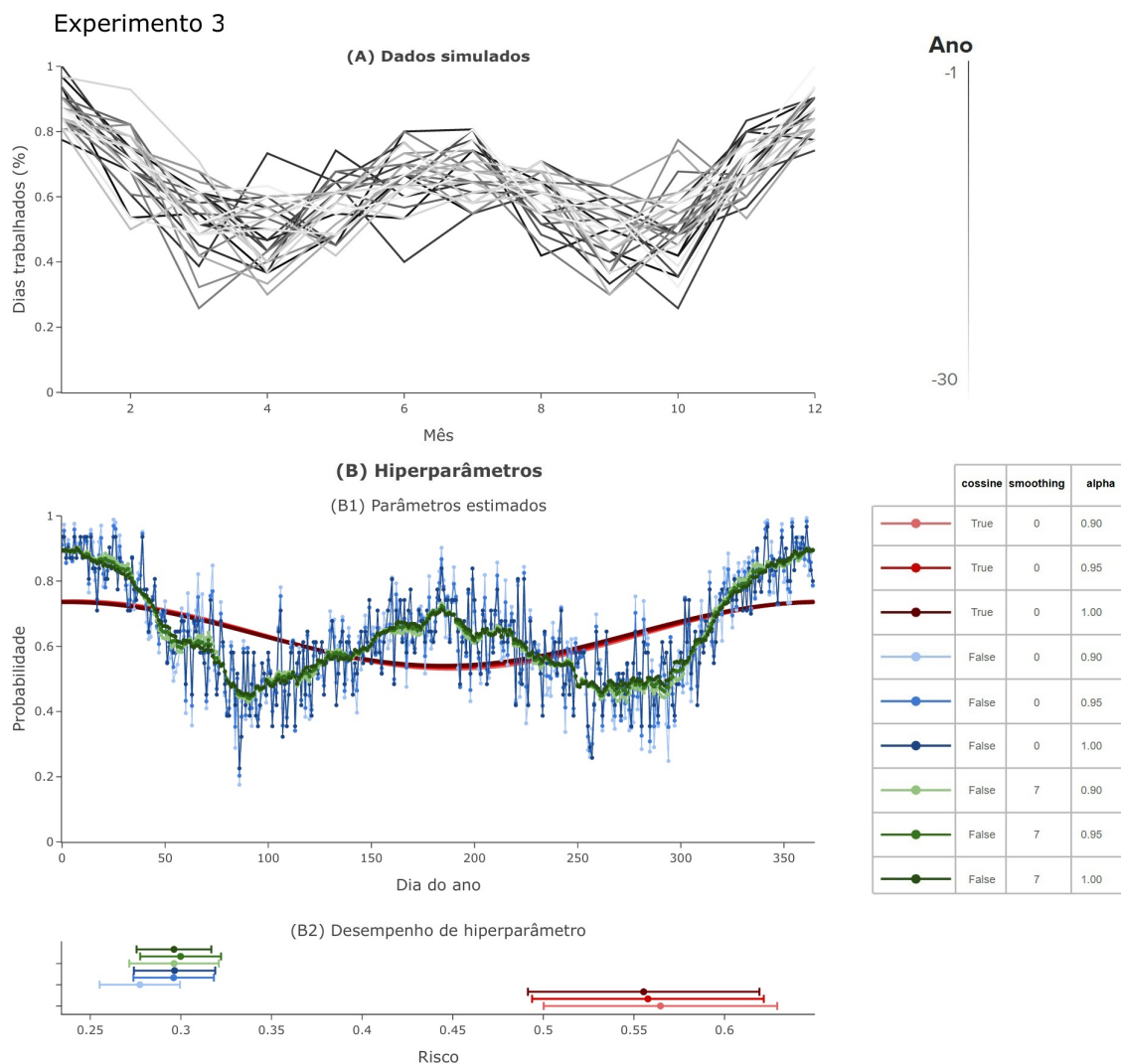
Figura 6 – Resultados do Experimento 2. (A) Porcentagem de dias trabalhados ao longo dos 12 meses, com cada linha representando um ano no histórico de dados. (B) Hiperparâmetros: (B1) Parâmetros Estimados: Probabilidade ao longo do ano com diferentes configurações de hiperparâmetros. (B2) Desempenho de Hiperparâmetro: Risco associado às combinações de hiperparâmetros, medido pela perda pinball da cauda direita.



Fonte: Elaborada pelo autor.

A [Figura 6](#) apresenta os resultados do Experimento 2. Na [Figura 6-\(A\)](#), observamos uma tendência nos dados, com a porcentagem de dias trabalhados sendo maior nos anos mais recentes, representados pelas linhas mais escuras. Na [Figura 6-\(B1\)](#), todas as combinações de hiperparâmetros capturam a sazonalidade anual, com as probabilidades sendo maiores para menores valores de α , pois dão mais peso aos anos mais recentes. Na [Figura 6-\(B2\)](#), o desempenho melhora à medida que α diminui, um resultado esperado devido à tendência nos dados e à escolha de dar mais peso aos anos mais recentes.

Figura 7 – Resultados do Experimento 3. (A) Porcentagem de dias trabalhados ao longo dos 12 meses, com cada linha representando um ano no histórico de dados. (B) Hiperparâmetros: (B1) Parâmetros Estimados: Probabilidade ao longo do ano com diferentes configurações de hiperparâmetros. (B2) Desempenho de Hiperparâmetro: Risco associado às combinações de hiperparâmetros, medido pela perda pinball da cauda direita.

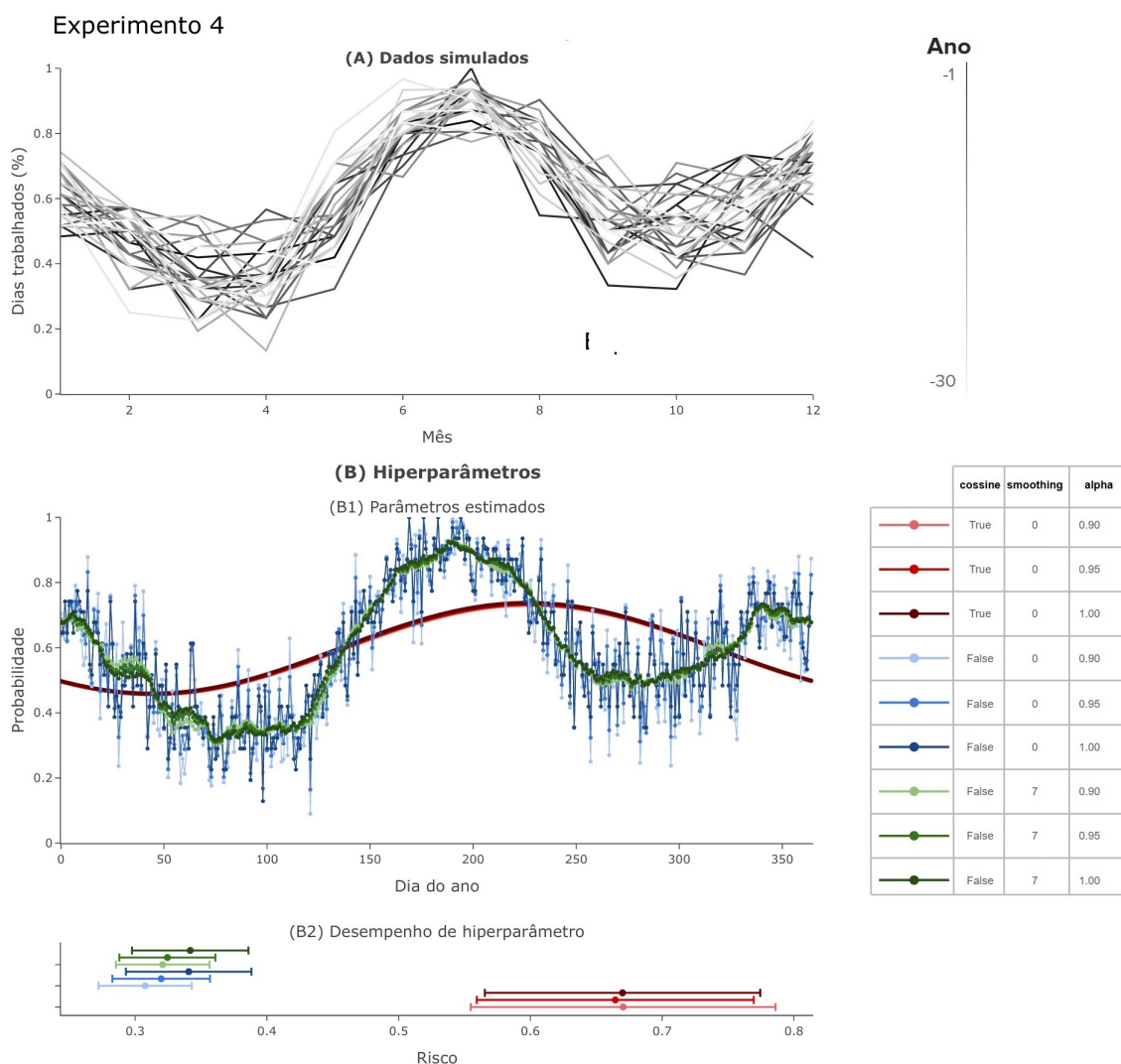


Fonte: Elaborada pelo autor.

v

A [Figura 7](#) apresenta os resultados do Experimento 3. Na [Figura 7-\(A\)](#), os dados mostram

Figura 8 – Resultados do Experimento 4. (A) Porcentagem de dias trabalhados ao longo dos 12 meses, com cada linha representando um ano no histórico de dados. (B) Hiperparâmetros: (B1) Parâmetros Estimados: Probabilidade ao longo do ano com diferentes configurações de hiperparâmetros. (B2) Desempenho de Hiperparâmetro: Risco associado às combinações de hiperparâmetros, medido pela perda pinball da cauda direita.



Fonte: Elaborada pelo autor.

sazonalidades mais complexas e sem tendência. Na [Figura 7-\(B1\)](#), a suavização se ajusta bem ao comportamento dos dados, enquanto o método paramétrico não consegue se ajustar adequadamente. Na [Figura 7-\(B2\)](#), observa-se que a média móvel demonstra um desempenho superior, indicando que o método paramétrico empregado é insuficiente para modelar padrões mais complexos. A [Figura 8](#) apresenta os resultados do Experimento 4, onde os dados possuem sazonalidades complexas e sem tendência, similar ao Experimento 3. As conclusões obtidas são as mesmas, com a média móvel mostrando-se superior ao método paramétrico para modelar padrões mais complexos.

4.2 Dados reais

Utilizamos dados meteorológicos diários do Reino Unido, seguindo critérios análogos ao estudo de [Ballesteros-Pérez *et al.* \(2017a\)](#). O conjunto de dados MIDAS (Met Office Integrated Data Archive System) ([MET OFFICE, 2012](#)) reúne todas as observações meteorológicas no Reino Unido e nas áreas marítimas adjacentes, com registros diários de precipitação, temperatura e velocidade do vento, abrangendo o período de 1985 a 2020. Aplicamos critérios semelhantes de inclusão e exclusão de estações climáticas. Para minimizar o problema de valores ausentes, utilizamos a imputação de dados com uma média móvel de 7 dias, desde que houvesse ao menos 3 observações disponíveis. Excluimos estações com grandes lacunas de dados nos conjuntos de validação e teste, resultando na seleção de 65 estações para análise.

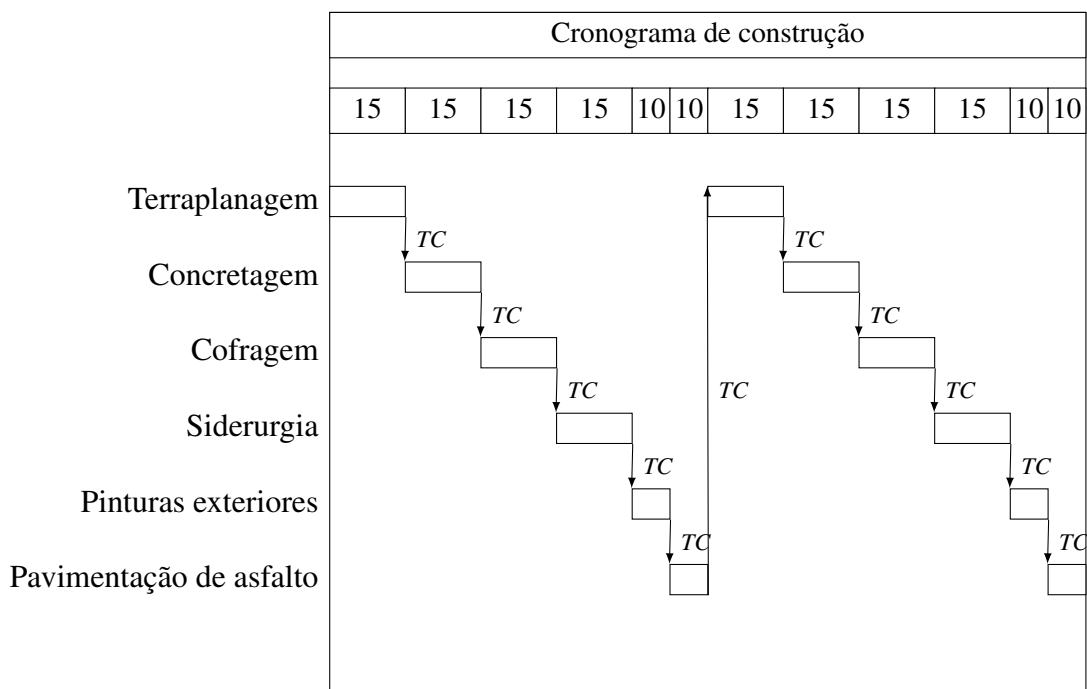
A [Figura 9](#) apresenta o cronograma fictício utilizado, onde todas as tarefas dependem das condições meteorológicas, afetando a duração do projeto. Sem considerar as condições climáticas regionais, a duração fixa do projeto seria de 160 dias. A [Tabela 1](#) especifica os limites climáticos que impedem a execução de cada tipo de tarefa ([BALLESTEROS-PÉREZ *et al.*, 2017a](#)). Devido ao grande número de valores ausentes para precipitação de granizo, precipitação de neve e tempestades elétricas, utilizamos um conjunto reduzido de variáveis. Para calcular a probabilidade de dias trabalháveis, consideramos apenas os dias com dados completos para todas as variáveis selecionadas.

Realizamos a seleção de modelos para cada estação climática. O conjunto de teste inclui os anos de 2019 e 2020, com dados semanais, totalizando 105 amostras para as estações com todos os dias disponíveis. O conjunto de validação corresponde ao ano de 2017, também com dados semanais, resultando em 53 amostras. Excluimos o ano de 2018 dos conjuntos de validação e teste devido à sua distribuição discrepante do comportamento histórico, conforme discutido no [Apêndice B](#). Consideramos todos os dias úteis, independentemente de feriados e fins de semana. No entanto, se houvesse ao menos um valor ausente no período de teste, a data de início correspondente foi descartada.

Na [Subseção 4.2.1](#), discutimos os resultados da otimização de hiperparâmetros, utilizando a perda pinball da cauda direita. Na [Subseção 4.2.2](#), apresentamos os resultados da seleção

de modelos, onde a perda pinball da cauda direita, a pinball não ponderada e a perda de densidade foram conjuntamente utilizadas para a seleção do modelo mais adequado para cada estação climática. Na [Subseção 4.2.3](#) discutimos a aplicação prática do nosso modelo, simulando diferentes datas de início do projeto. Finalmente, na [Subseção 4.2.4](#) introduzimos uma abordagem alternativa para a escolha da combinação ótima de hiperparâmetros.

Figura 9 – Cronograma de construção com a sequência de tarefas distribuídas ao longo de 160 dias, com suas respectivas durações indicadas no topo do cronograma. As tarefas têm dependências "TC"(Terminar-para-Começar), indicadas pelas setas.



Fonte: Elaborada pelo autor.

4.2.1 Otimização de hiperparâmetros

A otimização de hiperparâmetros foi realizada para cada estação climática, aplicando o cronograma semanalmente durante o ano de 2017, resultando em 53 amostras para cada estação. Os dados anteriores a cada data inicial foram utilizados como conjunto de treinamento, que serviu para calcular $\hat{\theta}(t)$, enquanto os dados posteriores a cada data inicial foram utilizados como conjunto de validação, onde a duração real do projeto Z_k foi verificada e a perda pinball ponderada pela cauda direita foi calculada e utilizada como métrica de avaliação. As combinações de hiperparâmetros consideradas incluíam:

- $alpha = \{0.9, 0.95, 1\}$;
- $smoothing = \{0, 7\}$;

- $cossine = \{False, True\}$;
- $cossine = \{False, True\}$.

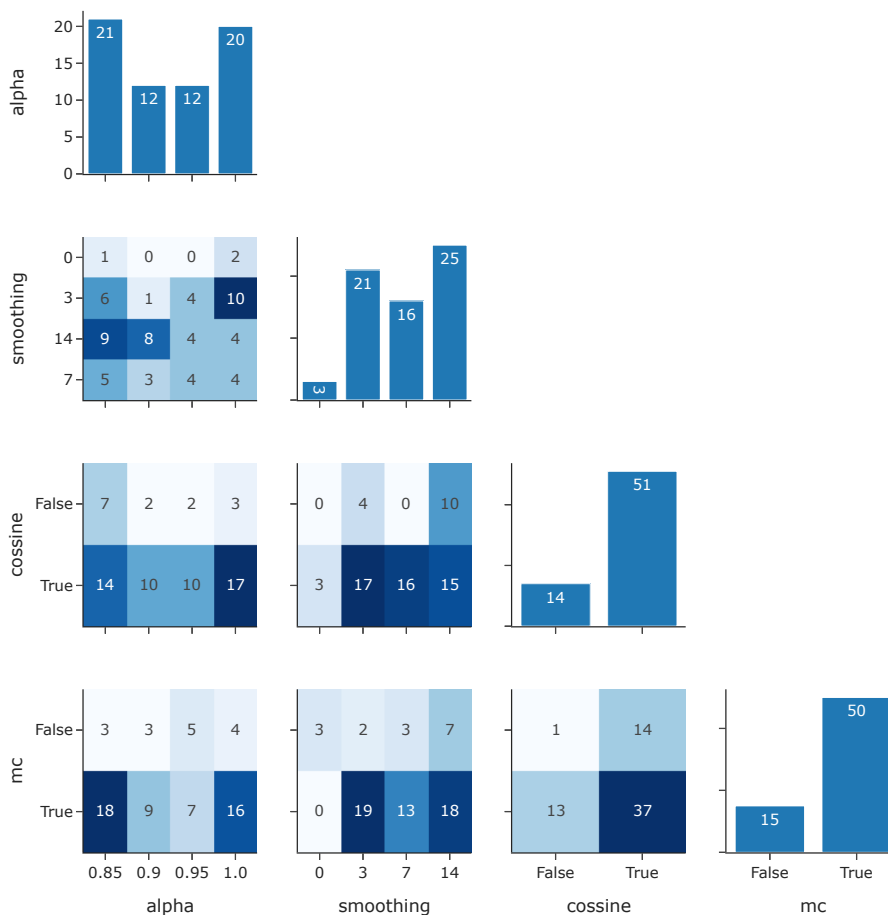
A [Figura 10](#) apresenta a contagem de estações para cada combinação de hiperparâmetros selecionada. Os resultados indicaram que $\alpha = 1$ teve o melhor desempenho em 31% (20) das estações, enquanto foi dada mais importância aos anos mais recentes nas outras estações. Em 77% (50) das estações, as cadeias de Markov foram utilizadas na modelagem, sugerindo que a adição de dependência temporal melhora o modelo. Em 78% (51) das estações, as ondas senoidais foram ajustadas após a suavização, indicando que os dados seguem o comportamento paramétrico esperado. O uso combinado de suavização com ondas senoidais foi aplicado em 94% dos casos. Em 57% (37) das estações, também foram combinadas as ondas senoidais com as cadeias de Markov, mostrando que as probabilidades condicionais possuem sazonalidades anuais modeláveis via ondas senoidais. Esses resultados permitiram uma comparação rigorosa das distribuições estimadas para escolher o melhor modelo utilizando métricas que consideram toda a incerteza. Os resultados para outras perdas para cada estação climática podem ser encontrados no [Apêndice A](#).

4.2.2 Comparação de modelos

O modelo proposto foi comparado com a abordagem convencional de duas etapas, que emprega um modelo de geração climática para simular valores climáticos futuros diariamente e projetar o impacto no cronograma de construção. Para a comparação, optou-se pelo modelo de [Apipattanavis et al. \(2007\)](#), que implementa um gerador climático semiparamétrico aplicado com sucesso em projetos de construção. Neste modelo, o estado da precipitação é inicialmente modelado por uma cadeia de Markov de primeira ordem com três estados (seco, úmido ou extremamente chuvoso). Posteriormente, o método k-NN é utilizado para aproximar localmente a PDF condicional e simular a partir dela. Os vizinhos são dias que possuem a mesma sequência gerada pela cadeia de Markov. Em seguida, a distância euclidiana entre as variáveis climáticas é utilizada como medida de similaridade, onde os vizinhos mais próximos recebem um peso maior com base na ordenação. A normalização dos pesos é feita para criar uma métrica de probabilidade. Com isso, as variáveis climáticas para o próximo dia são sorteadas. Com o clima simulado, aplica-se o cronograma de construção e verifica-se se as tarefas programadas foram possíveis de ser executadas.

Como critério principal para a escolha do modelo, optou-se pelo desempenho superior na perda pinball da cauda direita, focando na melhor estimação dos quantis superiores. Na ausência de diferença significativa, o modelo com melhor desempenho significativo na perda pinball é selecionado, garantindo a escolha de um modelo que estima toda a distribuição de maneira adequada. Por fim, também é considerada a perda de densidade, penalizando as distribuições muito concentradas e priorizando aquelas com alta probabilidade de acerto. O teste de Wilcoxon

Figura 10 – Contagem de estações para cada combinação de hiperparâmetros selecionada, utilizando a perda pinball ponderada pela cauda direita. Os gráficos de barras na diagonal apresentam a contagem de estações para cada hiperparâmetro individualmente, enquanto os gráficos de calor mostram a contagem para pares de hiperparâmetros.



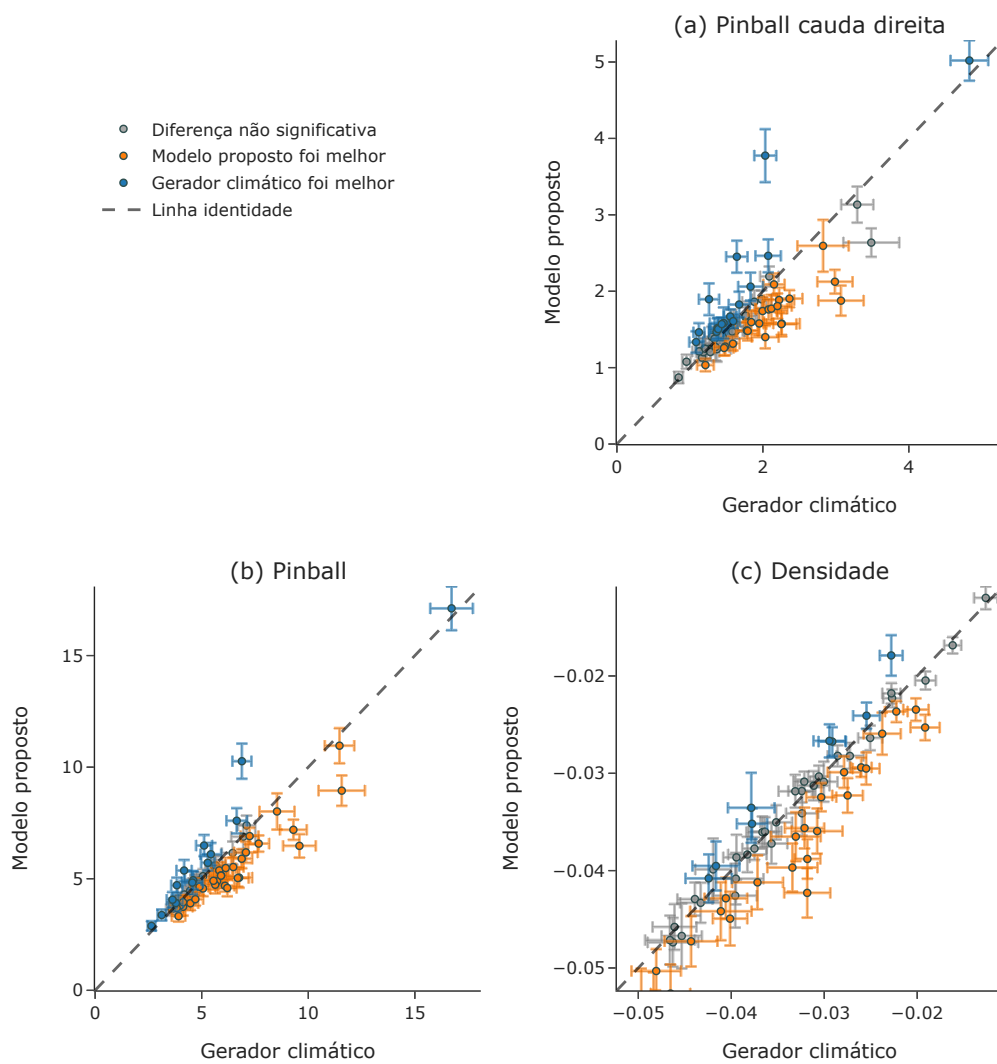
Fonte: Elaborada pelo autor.

foi utilizado para comparação das amostras, com significância de 5%, sendo um teste não paramétrico utilizado em amostras pareadas.

A Tabela 8 apresenta o resultado para cada estação climática, e a Figura 11 compara o resultado utilizando as métricas utilizadas. Os resultados mostraram que não houve diferença estatística em 17% (11) dos casos. Em 52% (34) dos casos, nosso modelo teve um desempenho superior, enquanto em 31% (20) dos casos o modelo comparado obteve um desempenho superior. As funções de perda propostas permitiram comparar os modelos e escolher o mais adequado para cada conjunto de dados. Em 57% (37) dos casos, o modelo foi escolhido pela perda pinball da cauda direita, 17% (11) pela perda pinball não ponderada e 6% (4) pela perda da densidade.

Com base nos resultados obtidos, é possível estimar a distribuição da duração do projeto, o que facilita uma melhor gestão dos projetos. Por exemplo, ao analisar datas de início e locais de projeto ideais, é viável diminuir tanto a duração quanto os custos do projeto, conforme sugerido por Ballesteros-Pérez *et al.* (2018). Além disso, examinar os caminhos críticos dentro do projeto,

Figura 11 – Comparação do desempenho do modelo proposto com o gerador climático usando três métricas de perda diferentes: (a) Pinball da cauda direita, (b) Pinball e (c) Densidade. Cada ponto representa o risco e seu intervalo de incerteza para uma estação climática. Os pontos laranjas indicam que o modelo proposto teve um desempenho melhor, os pontos azuis indicam que o gerador climático foi melhor, e os pontos cinza mostram diferenças não significativas. A linha tracejada representa a linha de identidade, onde ambos os modelos teriam o mesmo desempenho.



Fonte: Elaborada pelo autor.

conforme destacado por [Slyke \(1963\)](#), pode fornecer insights mais profundos sobre a organização de atividades de forma mais eficiente.

4.2.3 Data de início do projeto

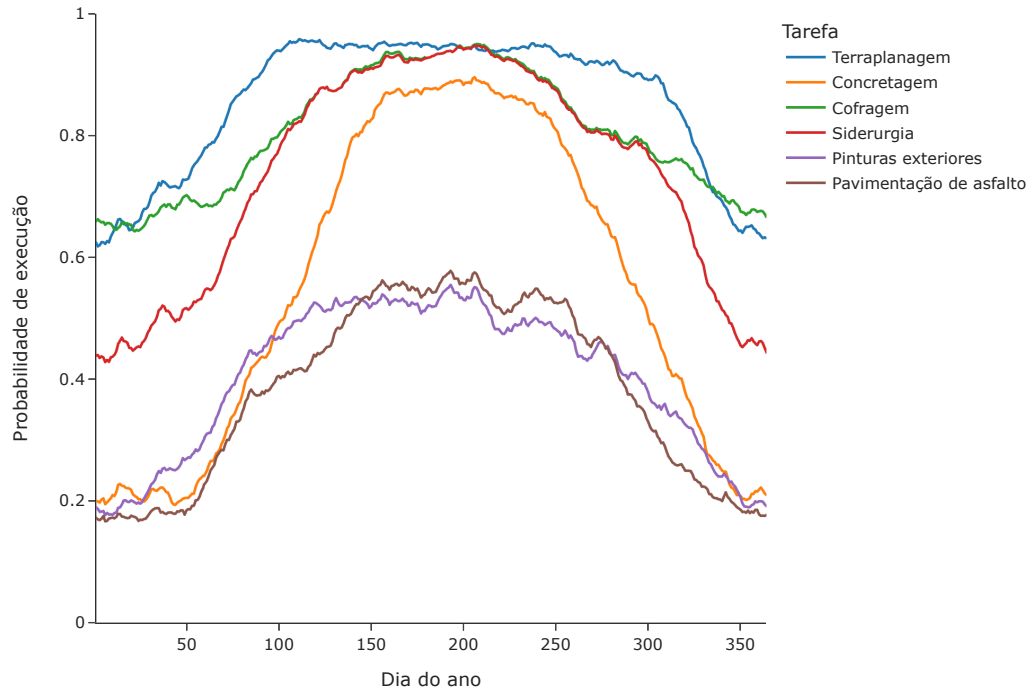
Selecionamos a estação 113 para ilustrar como o modelo pode ser utilizado. A [Figura 12](#) apresenta as probabilidades estimadas ao longo do ano, mostrando que as probabilidades tendem a ser mais altas durante o meio do ano. A [Figura 13](#) mostra a distribuição estimada do tempo de execução do projeto e o valor real observado para o início de cada mês. Observa-se uma variação da data real que vai de 230 até 316. Isso evidencia que as condições climáticas variam ao longo dos meses e a escolha adequada da data de início é fundamental para reduzir as durações e os custos dos projetos ([BALLESTEROS-PÉREZ et al., 2017a](#)). As distribuições estimadas refletem esta variação e a incerteza dos dados. Observa-se que as distribuições de março e abril possuem os menores tempos de execução, pois iniciam próximo de quando as probabilidades de execução das tarefas são maiores. Em seguida, temos as probabilidades de janeiro e fevereiro, que conseguem aproveitar o período de altas probabilidades do meio do ano. As densidades referentes ao segundo semestre do ano possuem valores maiores, pois o cronograma é executado principalmente no final do ano e, em seguida, no início do ano, períodos de mais baixas probabilidades, resultando em tempos de execução maiores. Portanto, antes de iniciar um projeto, o modelo pode ser utilizado para simular diferentes datas de início, a fim de escolher a data mais adequada.

Com a distribuição da duração do projeto, torna-se possível antecipar perdas de produtividade devido às condições climáticas antes do início do projeto. Além disso, os quantis podem ser utilizados para estabelecer limites contratuais para atrasos esperados atribuíveis às condições meteorológicas. Por exemplo, o 95º quantil pode ser utilizado para medir o atraso máximo esperado devido a condições climáticas mais adversas. Alternativamente, após a conclusão do projeto e o conhecimento da duração real, o seu quantil pode ser calculado para determinar o impacto climático. Estes quantis podem ser utilizados para facilitar negociações relacionadas à compensação com o contratante. Adicionalmente, é possível determinar a distribuição de custos do projeto a partir de funções que calculam o custo adicional de cada tarefa com base em seus atrasos. Essas funções podem ser aplicadas durante as simulações, permitindo calcular a distribuição total de custos do projeto de maneira fácil e eficiente.

4.2.4 Escolha de hiperparâmetros por tarefa

Anteriormente, os mesmos hiperparâmetros foram aplicados a todas as tarefas. No entanto, uma abordagem alternativa envolve a seleção de hiperparâmetros específicos para cada tarefa individual. Para ilustrar essa metodologia, utilizamos a estação 9, onde o desempenho do modelo proposto foi inferior ao modelo de gerador climático, e conduzimos uma otimização de

Figura 12 – Probabilidade de execução de cada tarefa ao longo dos dias do ano, para a estação 113. Cada linha representa uma tarefa específica.

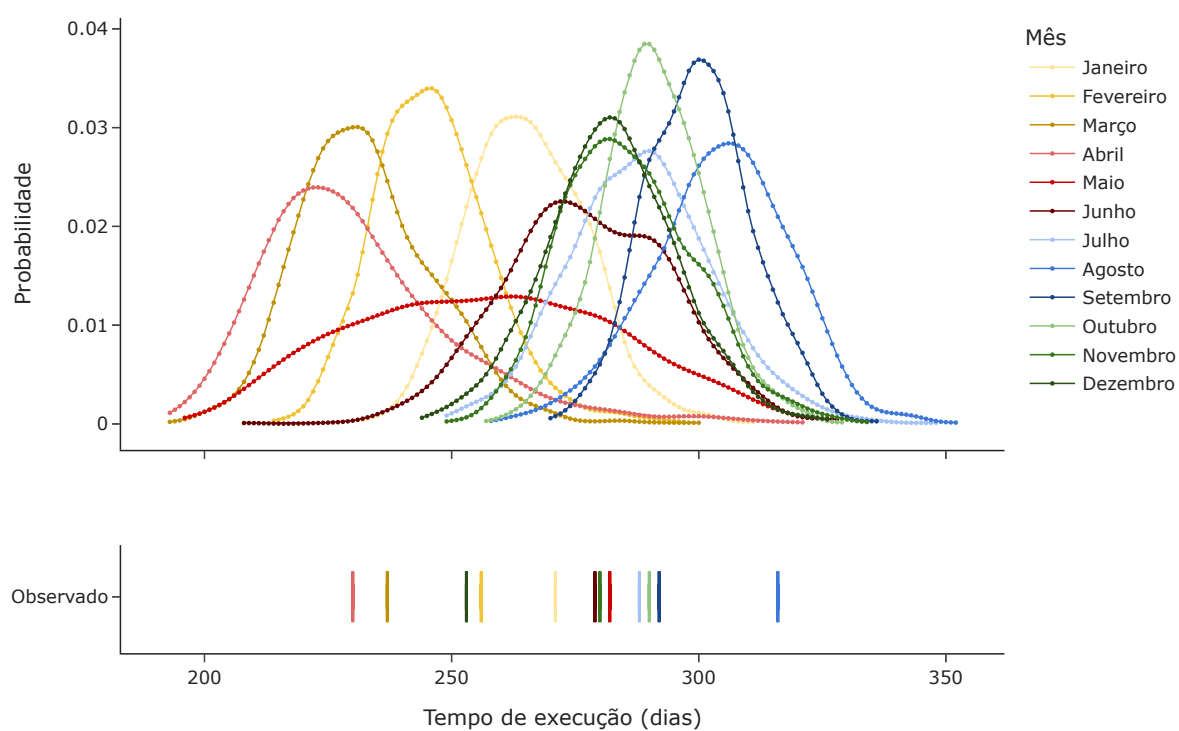


Fonte: Elaborada pelo autor.

hiperparâmetros direcionada a cada tarefa. Dessa maneira, o cronograma de construção possui apenas uma tarefa utilizando a mesma duração planejada desta tarefa.

A Figura 15 apresenta as séries temporais da média mensal dos dias trabalháveis para cada tarefa, revelando padrões de sazonalidade e variâncias distintos para cada uma. A Tabela 10 exhibe os hiperparâmetros selecionados para cada tarefa, utilizando a métrica de perda pinball ponderada pela cauda direita. O modelo treinado foi submetido ao conjunto de teste, resultando em um risco estimado de 3.95 ± 0.22 para a perda pinball ponderada pela cauda direita, um desempenho significativamente superior em comparação ao modelo de gerador climático. Esse resultado evidencia os benefícios que a otimização dos hiperparâmetros por tarefa pode trazer, especialmente quando estas apresentam padrões distintos.

Figura 13 – Distribuição estimada da duração do projeto, segmentada por mês de início, para a estação 113. As distribuições são calculadas usando um estimador da densidade por kernel, para aproximar por uma distribuição contínua e usando um padrão mais suave para facilitar a visualização. Na parte inferior, as linhas observadas representam a duração real do projeto.



Fonte: Elaborada pelo autor.

Tabela 8 – Desempenho comparativo do modelo proposto e do gerador climático para cada estação climática, utilizando três métricas de perda: Pinball Cauda Direita, Pinball e Densidade. Cada valor é o risco estimado e seu intervalo de incerteza. Os valores com um asterisco (*) indicam diferenças estatisticamente significativas. A coluna modelo escolhido indica qual modelo foi escolhido com base no melhor desempenho.

ID da estação	Pinball cauda direita		Pinball		Densidade		Modelo escolhido
	Modelo proposto	Gerador climático	Modelo proposto	Gerador climático	Modelo proposto	Gerador climático	
1033	1.48 ± 0.10	1.61 ± 0.11	5.37 ± 0.36 *	5.82 ± 0.38	-0.03 ± 0.00	-0.03 ± 0.00	Modelo proposto
105	1.48 ± 0.09	1.63 ± 0.12	4.69 ± 0.27 *	6.09 ± 0.44	-0.03 ± 0.00 *	-0.03 ± 0.00	Modelo proposto
1055	1.90 ± 0.11 *	2.36 ± 0.18	6.57 ± 0.37 *	7.67 ± 0.49	-0.02 ± 0.00 *	-0.02 ± 0.00	Modelo proposto
113	1.31 ± 0.09 *	1.59 ± 0.09	4.56 ± 0.34 *	5.06 ± 0.34	-0.04 ± 0.00 *	-0.03 ± 0.00	Modelo proposto
1137	1.74 ± 0.18 *	2.00 ± 0.19	5.48 ± 0.45 *	6.13 ± 0.48	-0.04 ± 0.00 *	-0.03 ± 0.00	Modelo proposto
1161	1.89 ± 0.12	1.98 ± 0.13	6.91 ± 0.37 *	7.25 ± 0.42	-0.02 ± 0.00	-0.02 ± 0.00	Modelo proposto
1171	3.13 ± 0.24	3.29 ± 0.22	10.96 ± 0.79 *	11.46 ± 0.69	-0.02 ± 0.00	-0.02 ± 0.00	Modelo proposto
1190	1.88 ± 0.20 *	3.07 ± 0.31	6.47 ± 0.53 *	9.58 ± 0.77	-0.03 ± 0.00 *	-0.02 ± 0.00	Modelo proposto
1198	2.09 ± 0.14 *	2.15 ± 0.15	6.88 ± 0.55	6.95 ± 0.57	-0.02 ± 0.00	-0.03 ± 0.00 *	Modelo proposto
1209	1.49 ± 0.10 *	1.78 ± 0.15	4.89 ± 0.30 *	5.72 ± 0.40	-0.03 ± 0.00 *	-0.03 ± 0.00	Modelo proposto
1285	1.89 ± 0.16 *	2.22 ± 0.21	6.18 ± 0.47 *	7.07 ± 0.56	-0.03 ± 0.00 *	-0.03 ± 0.00	Modelo proposto
132	2.59 ± 0.34 *	2.82 ± 0.35	8.01 ± 0.81 *	8.53 ± 0.82	-0.03 ± 0.00 *	-0.02 ± 0.00	Modelo proposto
1383	1.58 ± 0.17 *	2.26 ± 0.25	5.05 ± 0.45 *	6.76 ± 0.61	-0.04 ± 0.00 *	-0.03 ± 0.00	Modelo proposto
17176	1.24 ± 0.16	1.37 ± 0.19	3.74 ± 0.40 *	4.12 ± 0.46	-0.05 ± 0.00	-0.05 ± 0.00	Modelo proposto
1467	2.12 ± 0.16 *	2.99 ± 0.24	7.19 ± 0.45 *	9.30 ± 0.63	-0.02 ± 0.00 *	-0.02 ± 0.00	Modelo proposto
16611	1.76 ± 0.13 *	2.08 ± 0.17	5.90 ± 0.42 *	6.88 ± 0.49	-0.03 ± 0.00 *	-0.03 ± 0.00	Modelo proposto
384	1.27 ± 0.18	1.35 ± 0.19	3.81 ± 0.44 *	4.00 ± 0.46	-0.05 ± 0.00	-0.05 ± 0.00	Modelo proposto
386	1.47 ± 0.14	1.58 ± 0.16	4.64 ± 0.37 *	4.90 ± 0.42	-0.04 ± 0.00	-0.04 ± 0.00	Modelo proposto
393	1.21 ± 0.12	1.13 ± 0.09	3.64 ± 0.30	3.60 ± 0.27	-0.05 ± 0.00 *	-0.05 ± 0.00	Modelo proposto
595	1.28 ± 0.11 *	1.48 ± 0.14	3.90 ± 0.30 *	4.47 ± 0.36	-0.05 ± 0.00 *	-0.04 ± 0.00	Modelo proposto
605	1.03 ± 0.08 *	1.21 ± 0.12	3.32 ± 0.25 *	3.91 ± 0.31	-0.05 ± 0.00 *	-0.05 ± 0.00	Modelo proposto
643	1.40 ± 0.13	1.30 ± 0.10	4.29 ± 0.34	4.18 ± 0.30	-0.04 ± 0.00 *	-0.04 ± 0.00	Modelo proposto
657	1.77 ± 0.18 *	2.12 ± 0.22	5.14 ± 0.44 *	5.91 ± 0.52	-0.04 ± 0.00	-0.04 ± 0.00	Modelo proposto
669	1.57 ± 0.14 *	2.25 ± 0.21	5.03 ± 0.39 *	6.69 ± 0.51	-0.04 ± 0.00 *	-0.03 ± 0.00	Modelo proposto
67	2.64 ± 0.19	3.49 ± 0.38	8.95 ± 0.68 *	11.57 ± 1.09	-0.02 ± 0.00	-0.02 ± 0.00	Modelo proposto
708	1.38 ± 0.15	1.34 ± 0.14	3.90 ± 0.34	3.98 ± 0.35	-0.05 ± 0.00 *	-0.05 ± 0.00	Modelo proposto
709	1.25 ± 0.13	1.21 ± 0.12	3.69 ± 0.32 *	3.82 ± 0.32	-0.06 ± 0.00 *	-0.05 ± 0.00	Modelo proposto
744	1.58 ± 0.15 *	1.95 ± 0.19	4.73 ± 0.39 *	5.63 ± 0.46	-0.04 ± 0.00 *	-0.04 ± 0.00	Modelo proposto
779	1.81 ± 0.16 *	2.20 ± 0.20	5.52 ± 0.43 *	6.48 ± 0.48	-0.04 ± 0.00 *	-0.03 ± 0.00	Modelo proposto
811	1.60 ± 0.16 *	1.84 ± 0.17	4.91 ± 0.39 *	5.55 ± 0.43	-0.04 ± 0.00 *	-0.04 ± 0.00	Modelo proposto
842	1.26 ± 0.10 *	1.47 ± 0.13	4.08 ± 0.30 *	4.70 ± 0.36	-0.04 ± 0.00 *	-0.04 ± 0.00	Modelo proposto
862	1.20 ± 0.11	1.28 ± 0.12	3.95 ± 0.31 *	4.13 ± 0.33	-0.04 ± 0.00	-0.04 ± 0.00	Modelo proposto
869	1.40 ± 0.15 *	2.03 ± 0.18	4.58 ± 0.37 *	6.20 ± 0.44	-0.04 ± 0.00 *	-0.03 ± 0.00	Modelo proposto
395	1.48 ± 0.12 *	1.79 ± 0.15	5.15 ± 0.33 *	5.91 ± 0.36	-0.03 ± 0.00 *	-0.03 ± 0.00	Modelo proposto
137	1.47 ± 0.10	1.37 ± 0.10 *	4.93 ± 0.40	4.60 ± 0.39 *	-0.04 ± 0.00	-0.04 ± 0.00 *	Gerador climático
1393	3.77 ± 0.35	2.03 ± 0.15 *	10.27 ± 0.79	6.89 ± 0.45 *	-0.02 ± 0.00	-0.02 ± 0.00 *	Gerador climático
1395	2.46 ± 0.22	2.07 ± 0.17 *	7.60 ± 0.57	6.64 ± 0.48 *	-0.03 ± 0.00	-0.03 ± 0.00 *	Gerador climático
150	1.52 ± 0.13	1.38 ± 0.10 *	4.80 ± 0.34	4.70 ± 0.30	-0.04 ± 0.00	-0.04 ± 0.00	Gerador climático
161	1.46 ± 0.12	1.12 ± 0.08 *	4.71 ± 0.34	3.83 ± 0.27 *	-0.04 ± 0.00	-0.04 ± 0.00	Gerador climático
17314	1.58 ± 0.20	1.46 ± 0.19 *	4.83 ± 0.52	4.56 ± 0.49 *	-0.04 ± 0.00	-0.04 ± 0.00 *	Gerador climático
212	1.83 ± 0.17	1.67 ± 0.14 *	5.58 ± 0.42	5.57 ± 0.39	-0.03 ± 0.00	-0.03 ± 0.00	Gerador climático
370	1.25 ± 0.14	1.20 ± 0.12	3.91 ± 0.36	3.79 ± 0.33 *	-0.05 ± 0.00	-0.05 ± 0.00	Gerador climático
409	1.08 ± 0.09	0.95 ± 0.06	3.37 ± 0.25	3.12 ± 0.19 *	-0.05 ± 0.00 *	-0.05 ± 0.00	Gerador climático
44	2.06 ± 0.18	1.83 ± 0.17 *	6.10 ± 0.51	5.43 ± 0.48 *	-0.03 ± 0.00	-0.03 ± 0.00	Gerador climático
440	0.87 ± 0.07	0.85 ± 0.06	2.89 ± 0.21	2.64 ± 0.18 *	-0.05 ± 0.00	-0.05 ± 0.00	Gerador climático
556	1.59 ± 0.17	1.57 ± 0.17	4.77 ± 0.43	4.68 ± 0.43	-0.04 ± 0.00	-0.04 ± 0.00 *	Gerador climático
726	1.50 ± 0.15	1.39 ± 0.13 *	4.46 ± 0.36	4.46 ± 0.34	-0.04 ± 0.00	-0.04 ± 0.00	Gerador climático
847	1.33 ± 0.14	1.08 ± 0.09 *	4.06 ± 0.34	3.63 ± 0.27 *	-0.05 ± 0.00	-0.05 ± 0.00	Gerador climático
888	2.45 ± 0.21	1.64 ± 0.15 *	6.49 ± 0.47	5.11 ± 0.38 *	-0.03 ± 0.00	-0.04 ± 0.00 *	Gerador climático
889	1.90 ± 0.21	1.26 ± 0.14 *	5.37 ± 0.47	4.17 ± 0.36 *	-0.04 ± 0.00	-0.04 ± 0.00	Gerador climático
9	5.02 ± 0.27	4.83 ± 0.26 *	17.12 ± 0.98	16.72 ± 1.00 *	-0.01 ± 0.00	-0.01 ± 0.00	Gerador climático
908	1.61 ± 0.09	1.59 ± 0.09 *	5.11 ± 0.32	5.06 ± 0.32	-0.03 ± 0.00	-0.03 ± 0.00	Gerador climático
987	1.57 ± 0.08	1.44 ± 0.10 *	4.81 ± 0.28	4.70 ± 0.32	-0.03 ± 0.00	-0.03 ± 0.00	Gerador climático
1215	1.67 ± 0.09	1.55 ± 0.09 *	5.71 ± 0.35	5.29 ± 0.31 *	-0.03 ± 0.00	-0.03 ± 0.00 *	Gerador climático
177	1.39 ± 0.08	1.34 ± 0.07	4.90 ± 0.26	4.86 ± 0.25	-0.03 ± 0.00	-0.03 ± 0.00	
1319	1.66 ± 0.13	1.76 ± 0.14	5.48 ± 0.37	5.71 ± 0.40	-0.03 ± 0.00	-0.03 ± 0.00	
1448	1.13 ± 0.06	1.17 ± 0.06	3.74 ± 0.18	3.87 ± 0.20	-0.04 ± 0.00	-0.04 ± 0.00	
1450	1.40 ± 0.11	1.45 ± 0.12	4.55 ± 0.31	4.64 ± 0.33	-0.04 ± 0.00	-0.04 ± 0.00	
1534	1.25 ± 0.08	1.26 ± 0.10	4.15 ± 0.28	4.27 ± 0.31	-0.04 ± 0.00	-0.04 ± 0.00	
235	1.43 ± 0.11	1.41 ± 0.11	4.80 ± 0.33	4.80 ± 0.33	-0.04 ± 0.00	-0.04 ± 0.00	
373	1.20 ± 0.10	1.18 ± 0.10	3.76 ± 0.29	3.71 ± 0.29	-0.04 ± 0.00	-0.04 ± 0.00	
513	1.68 ± 0.15	1.77 ± 0.16	5.60 ± 0.44	5.65 ± 0.45	-0.03 ± 0.00	-0.03 ± 0.00	
52	1.86 ± 0.15	1.88 ± 0.17	6.15 ± 0.51	6.45 ± 0.51	-0.03 ± 0.00	-0.03 ± 0.00	
54	2.19 ± 0.13	2.09 ± 0.13	7.39 ± 0.44	7.10 ± 0.40	-0.02 ± 0.00	-0.02 ± 0.00	
1145	1.43 ± 0.11	1.44 ± 0.11	4.96 ± 0.32	5.00 ± 0.35	-0.03 ± 0.00	-0.03 ± 0.00	

Fonte: Elaborada pelo autor.

Tabela 9 – Estatísticas da distribuição estimada da duração do projeto, segmentada por mês de início para a estação 113. As estatísticas incluem a média, desvio padrão e vários quantis (5%, 25%, 50%, 75%, 95%) para cada mês do ano.

Mês	Média	Desvio padrão	Quantil				
			5	25	50	75	95
1	265.7	11.7	247	258	265	273	285
2	246.3	12.5	228	237	245	253	268
3	232.8	14.5	212	223	231	241	259
4	232.2	20.8	206	218	229	242	272
5	256.3	27.1	212	235	257	276	300
6	278.2	17.0	250	267	278	291	305
7	288.5	14.7	263	279	288	297	314
8	304.5	13.9	280	296	305	314	327
9	300.6	10.7	283	293	301	308	318
10	290.8	10.9	273	284	291	298	308
11	286.5	14.1	265	276	285	296	313
12	281.8	14.0	259	272	282	291	304

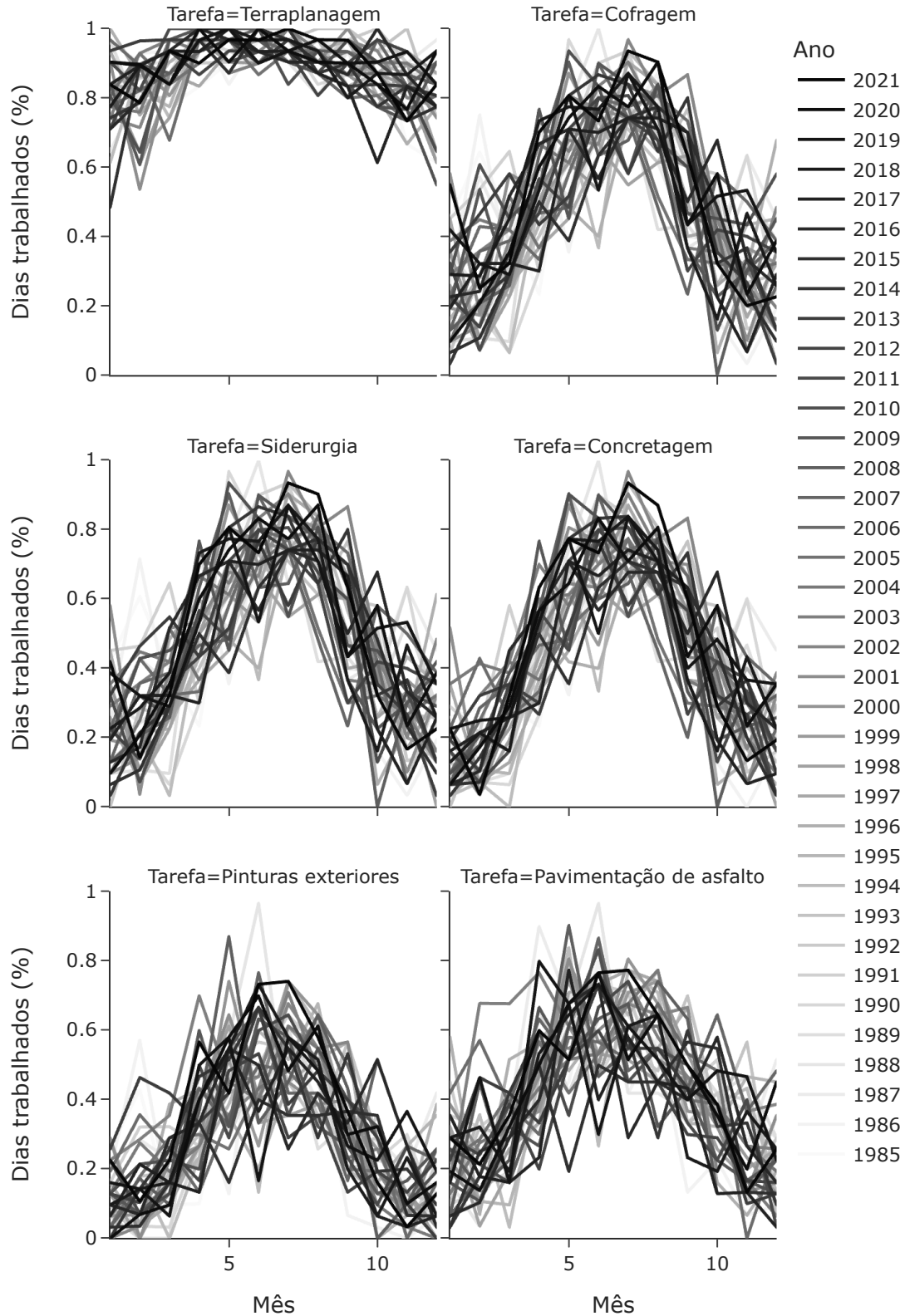
Fonte: Elaborada pelo autor.

Tabela 10 – Hiperparâmetros selecionados para cada tipo de tarefa na estação 9.

Tarefa	<i>alpha</i>	<i>smoothing</i>	<i>mc</i>	<i>cossine</i>
Terraplanagem	0.85	7	False	True
Concretagem	0.85	0	False	False
Siderurgia	1.0	7	False	False
Cofragem	1.0	7	False	False
Pavimentação de asfalto	1.0	3	False	False
Pinturas exteriores	0.95	0	False	False

Fonte: Elaborada pelo autor.

Figura 14 – Série temporal do percentual de dias trabalhados para cada mês na estação 9. Cada gráfico corresponde a uma tarefa específica do cronograma de construção. As linhas dentro de cada gráfico representam os dados de diferentes anos.



Fonte: Elaborada pelo autor.

CONCLUSÃO

Quando um projeto de construção se prolonga além das datas estipuladas no contrato, caracteriza-se um atraso, que pode ser causado por diversos fatores imprevisíveis, como incompetência das partes, falta de disponibilidade de materiais e equipamentos, e condições climáticas desfavoráveis. Eventos climáticos adversos durante a execução de projetos de construção podem afetar significativamente a produtividade e representam uma grande fonte de incerteza em relação aos atrasos. Esses atrasos impactam negativamente tanto o proprietário quanto os contratados, podendo resultar na ativação de cláusulas e responsabilidades contratuais.

Este trabalho propõe um método para modelar a probabilidade de um dia ser trabalhável e estimar a distribuição do tempo de execução de um projeto, considerando as condições climáticas. As medidas de perda propostas são utilizadas para selecionar os hiperparâmetros do modelo e escolher o melhor modelo dentro de um conjunto de candidatos. Os métodos apresentados permitem comparar as distribuições estimadas utilizando métricas que descrevem toda a incerteza, selecionando rigorosamente o modelo mais adequado.

Nos experimentos apresentados, as distribuições estimadas foram comparadas através das medidas de perda propostas. Utilizando dados simulados, verificou-se que o modelo proposto consegue capturar os padrões dos dados quando os hiperparâmetros são adequadamente escolhidos. Além disso, utilizando dados reais, demonstrou-se como a otimização de hiperparâmetros e a escolha do melhor modelo podem ser realizadas. Observou-se que as condições climáticas variam ao longo dos meses e que as distribuições estimadas acompanham essa variação, considerando a incerteza dos dados. Por fim, o modelo pode ser utilizado para escolher a data de início mais adequada para o projeto. De maneira geral, tanto a modelagem da distribuição do tempo de execução do projeto quanto a seleção de modelos foram conduzidas com êxito.

O modelo selecionado pode ser aplicado em diferentes fases do projeto de construção. Antes do início, pode-se utilizá-lo para selecionar a data mais adequada para o início do projeto, simulando diferentes datas. A estimativa para a data de término do projeto pode ser obtida

através do cálculo dos quantis da distribuição, com base no risco que o construtor está disposto a assumir. Além disso, o modelo permite simular a distribuição da data de início de cada tarefa, obtida pela simulação da duração das tarefas predecessoras. Com essas informações, é possível analisar o cronograma proposto, identificar trechos críticos e organizar as atividades de forma mais eficiente, contribuindo para a redução dos custos.

Durante a execução do projeto, o modelo pode ser atualizado para fornecer estimativas atualizadas para o término do projeto. Com os dados climáticos coletados durante a execução, é possível verificar se as condições climáticas foram um fator crítico na execução do cronograma. Além da análise exploratória, pode-se comparar a distribuição estimada com o tempo real do projeto, a fim de verificar se houve um clima atípico durante a execução. Essa informação pode ser útil para mediar negociações entre as partes envolvidas no projeto.

Consideramos que os limites climáticos impedem a realização de tarefas, enquanto as condições meteorológicas só podem reduzir a produtividade dos trabalhadores. Uma alternativa é alterar a distribuição da variável aleatória que descreve se cada tarefa pode ser executada em cada dia do ano ou dividir o dia em intervalos de tempo menores. Por exemplo, uma distribuição de probabilidade discreta pode atribuir probabilidades a percentagens de produtividade.

Além disso, consideramos que as tarefas são independentes umas das outras. Se houver tarefas que precisam ser realizadas no mesmo dia, podemos modelá-las como uma nova tarefa que consiste em completar todas as tarefas dependentes. Outra opção é modelar a probabilidade para limites climáticos, onde valores para esses limites são simulados e avaliamos quais tarefas podem ser executadas. Por exemplo, se uma tarefa depende da temperatura e da velocidade do vento, precisamos simular ambas as variáveis meteorológicas para permitir sua conclusão. Se um limite climático impedir a execução, todas as tarefas dependentes desse limite não poderão ser concluídas. Um processo de Markov com múltiplas variáveis também pode ser usado nesta fase para considerar a interação entre as variáveis climáticas.

Finalmente, a abordagem proposta pode incorporar outras extensões propostas para abordagens baseadas em PERT. Por exemplo, fatores não climáticos podem ser introduzidos tratando a tarefa planejada como uma variável aleatória, que pode ser modelada utilizando abordagens PERT tradicionais ou suas extensões. Desta forma, o conhecimento de especialistas pode ser utilizado para adicionar incertezas à duração planejada sem considerar os efeitos climáticos, e o método proposto é usado para incorporar esses efeitos climáticos na simulação, o que é difícil de prever. Além disso, é possível introduzir mais complexidade, como considerar extensões que impõem restrições de recursos e outras variáveis.

Esta investigação tem implicações que vão além dos atrasos na construção, sendo relevante para indústrias que enfrentam desafios imprevisíveis. A adaptabilidade do modelo a diferentes escalas de projetos ilustra seu potencial para melhorar o planejamento e execução estratégicos, melhorando a capacidade das empresas de responder a fatores externos. Além disso, o modelo pode simular diversos cenários, apoiando a tomada de decisões informadas e a gestão

de riscos.

Assim, o modelo proposto se mostra uma ferramenta versátil e eficiente para lidar com as incertezas presentes em diferentes fases de um projeto de construção. Sua utilização possibilita uma gestão mais precisa e eficaz das incertezas inerentes à execução do projeto.

Os códigos utilizados neste projeto estão disponíveis abertamente no GitHub:

<https://github.com/mborgesss/mestrado>.

OTIMIZAÇÃO DE HIPERPARÂMETRO

A [Tabela 11](#) apresenta os hiperparâmetros selecionados para cada estação climática utilizando perda pinball da cauda direita, pinball e densidade.

Tabela 11 – Hiperparâmetros selecionados para cada estação climática utilizando a perda pinball da cauda direta, perda pinball não ponderada e perda de densidade.

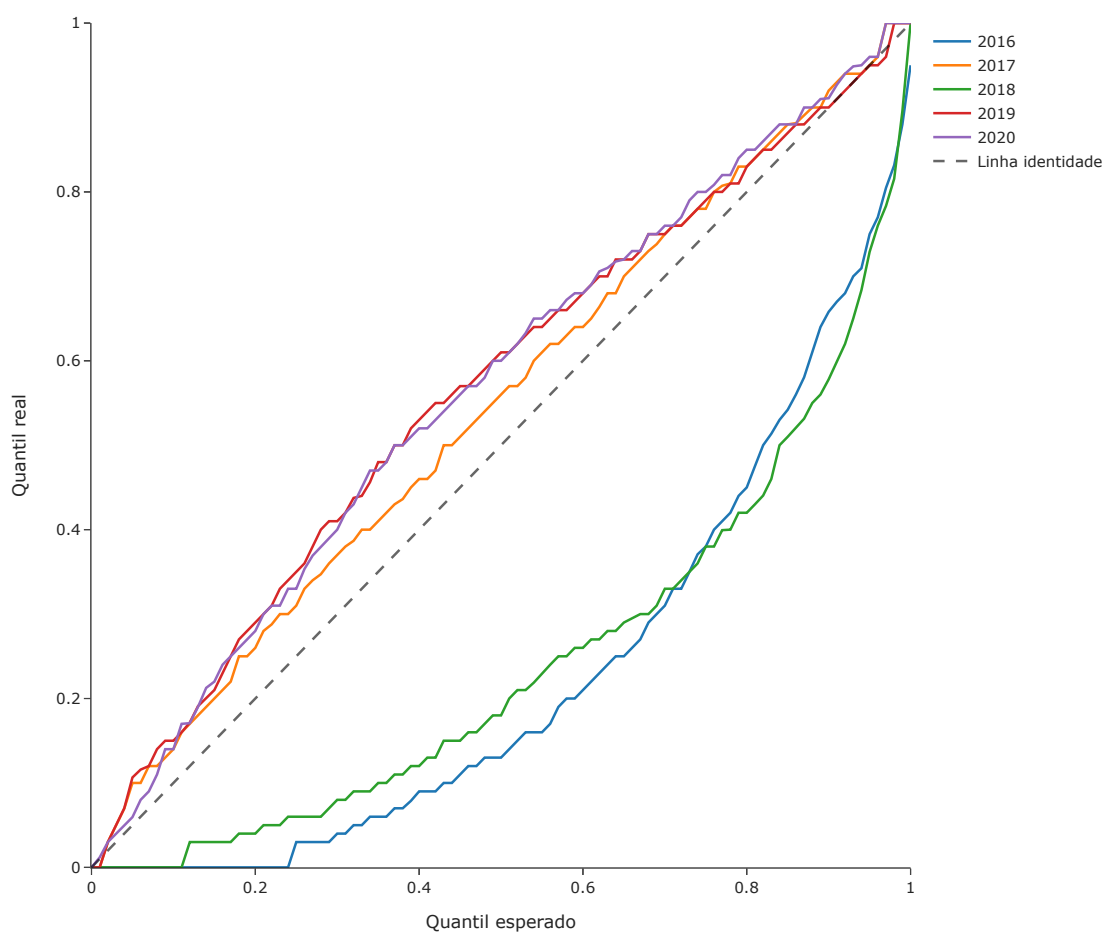
ID da estação	Pinball cauda direita				Pinball				Densidade			
	alpha	smoothing	cosine	mc	alpha	smoothing	cosine	mc	alpha	smoothing	cosine	mc
1033	0.85	7	True	True	0.85	3	True	True	0.85	0	False	False
105	0.85	7	True	True	0.85	7	True	True	0.85	7	True	True
1055	0.85	14	True	False	0.85	14	True	False	0.85	0	False	True
113	0.95	14	False	False	0.95	14	False	False	0.95	3	False	True
1137	1.0	3	True	False	1.0	3	True	False	1.0	0	False	False
1145	0.85	7	True	True	0.85	7	True	True	0.9	0	False	True
1161	0.85	3	True	True	0.85	3	True	True	0.85	3	False	False
1171	0.85	14	True	False	0.85	14	True	False	0.85	0	False	False
1190	0.85	14	True	False	0.85	14	True	False	0.9	0	False	True
1198	1.0	7	True	True	1.0	7	True	True	1.0	0	False	True
1209	0.9	14	True	False	0.9	14	True	False	0.95	0	False	True
1215	1.0	3	True	True	1.0	3	True	True	1.0	0	False	False
1285	0.85	14	True	False	0.9	3	True	False	0.85	3	False	False
1319	1.0	3	True	True	1.0	3	True	True	1.0	3	False	True
132	0.9	14	True	True	0.9	14	True	True	0.9	14	False	True
137	0.85	14	True	True	0.85	14	True	True	0.85	7	False	True
1383	1.0	7	True	True	1.0	7	True	True	1.0	3	False	True
1393	0.95	3	False	True	0.95	0	False	True	0.95	7	False	True
1395	0.95	7	True	True	1.0	3	False	True	1.0	14	False	True
1448	0.9	7	True	True	0.9	7	True	True	0.9	7	False	True
1450	0.95	14	True	False	1.0	3	True	True	1.0	3	False	True
1467	0.85	7	True	True	0.85	7	True	True	1.0	3	True	True
150	0.95	14	False	True	0.95	14	False	True	1.0	14	False	True
1534	0.85	3	True	True	0.85	3	True	True	1.0	0	False	False
161	0.85	14	False	True	0.85	14	False	True	0.85	7	False	True
16611	0.85	3	True	False	0.9	3	True	False	0.85	7	False	True
17176	0.95	3	True	True	1.0	3	True	True	1.0	0	False	True
17314	0.9	14	True	True	0.9	14	True	True	0.85	0	False	False
177	0.9	7	True	True	0.85	7	True	True	0.85	3	True	True
212	1.0	0	False	True	1.0	14	False	True	1.0	3	False	True
235	0.85	14	True	False	0.9	7	True	False	0.85	3	False	False
370	0.9	14	True	True	0.9	14	True	True	0.85	14	False	True
373	1.0	14	True	True	1.0	14	True	True	1.0	0	False	False
384	0.95	3	True	True	0.95	3	True	True	0.9	7	False	True
386	0.95	14	True	True	0.95	7	True	True	0.95	14	False	True
393	1.0	14	False	True	1.0	14	False	True	1.0	14	False	True
395	0.9	14	True	True	0.9	14	True	True	0.9	0	False	True
409	0.95	7	False	True	0.95	14	False	False	0.95	0	False	True
44	0.9	14	False	True	0.85	7	False	True	1.0	0	False	True
440	0.85	3	True	True	0.85	3	True	True	0.85	3	False	True
513	1.0	14	True	True	1.0	14	True	True	0.95	0	False	True
52	0.85	14	True	True	0.85	7	True	True	0.85	14	False	True
54	0.85	7	True	True	0.85	7	True	True	0.85	3	False	True
556	0.95	3	True	True	0.95	3	True	True	0.95	0	False	True
595	1.0	3	True	True	1.0	3	True	True	1.0	14	False	True
605	1.0	3	True	True	1.0	3	True	True	1.0	3	False	False
643	0.95	7	False	True	0.95	0	False	False	0.9	0	False	False
657	1.0	7	True	True	0.95	3	True	True	0.95	7	False	True
669	1.0	3	True	False	0.95	3	True	False	0.85	0	False	False
67	0.85	14	True	False	0.85	7	True	True	0.85	14	False	True
708	1.0	0	False	True	0.95	3	False	True	1.0	0	False	False
709	1.0	7	False	True	0.85	3	False	True	0.85	3	False	True
726	0.9	14	False	True	0.9	14	False	True	1.0	0	False	False
744	1.0	3	True	True	1.0	3	True	True	0.95	0	False	True
779	1.0	3	True	True	1.0	3	True	False	1.0	3	False	False
811	0.85	3	True	True	0.9	3	True	True	0.95	0	False	False
842	1.0	3	True	True	1.0	3	True	True	0.95	14	False	True
847	0.9	14	False	True	0.9	14	False	True	0.85	0	False	False
862	0.95	7	True	True	0.95	7	True	True	0.85	0	False	False
869	0.9	3	True	False	0.85	14	True	False	0.85	0	False	False
888	0.85	0	False	True	0.85	0	False	True	0.85	0	False	True
889	0.85	3	False	True	0.85	3	False	True	0.85	3	False	True
9	1.0	3	True	True	1.0	3	True	True	0.85	3	False	False
908	1.0	14	True	False	1.0	14	True	False	1.0	7	False	True
987	0.9	7	True	True	0.85	3	True	True	1.0	7	False	False

Fonte: Elaborada pelo autor.

CONJUNTO DE DADOS

O conjunto de dados real possui informações desde 1985 até 2020. Durante as análises, foi observado que o ano de 2018 foi um período atípico, caracterizado por um tempo de execução do projeto menor em comparação com os demais anos. Para visualizar essa discrepância, foi realizado a comparação da duração do projeto em cada estação climática e mês, com os demais os valores nos períodos correspondentes em anos anteriores. A [Tabela 11](#) os quantis esperados e os quantis observados no intervalo entre 2015 e 2020, onde é possível observar que os anos de 2016 e 2018 foram atípicos. Devido a isso, o ano de 2018 foi excluído da análise, pois a distribuição dos conjuntos de dados não reflete o comportamento histórico esperado.

Figura 15 – Comparação do quantil esperado com o quantil real para cada ano.



Fonte: Elaborada pelo autor.

REFERÊNCIAS

APIPATTANAVIS, S.; PODESTÁ, G.; RAJAGOPALAN, B.; KATZ, R. W. A semiparametric multivariate and multisite weather generator. **Water Resources Research**, Wiley Online Library, v. 43, n. 11, 2007. Citado nas páginas [27](#) e [62](#).

APIPATTANAVIS, S.; SABOL, K.; MOLENAAR, K. R.; RAJAGOPALAN, B.; XI, Y.; BLACKKARD, B.; PATIL, S. Integrated framework for quantifying and predicting weather-related highway construction delays. **Journal of construction engineering and management**, American Society of Civil Engineers, v. 136, n. 11, p. 1160–1168, 2010. Citado nas páginas [27](#) e [28](#).

ASSAF, S. A.; AL-HEJJI, S. Causes of delay in large construction projects. **International journal of project management**, Elsevier, v. 24, n. 4, p. 349–357, 2006. Citado na página [25](#).

BALLESTEROS-PÉREZ, P.; ROJAS-CÉSPEDES, Y. A.; HUGHES, W.; KABIRI, S.; PELLICER, E.; MORA-MELIÀ, D.; CAMPO-HITSCHFELD, M. L. del. Weather-wise: A weather-aware planning tool for improving construction productivity and dealing with claims. **Automation in construction**, Elsevier, v. 84, p. 81–95, 2017. Citado nas páginas [28](#), [33](#), [34](#), [35](#), [45](#), [60](#) e [65](#).

_____. Weather-wise: A weather-aware planning tool for improving construction productivity and dealing with claims. **Automation in construction**, Elsevier, v. 84, p. 81–95, 2017. Citado na página [28](#).

BALLESTEROS-PÉREZ, P.; SMITH, S. T.; LLOYD-PAPWORTH, J. G.; COOKE, P. Incorporating the effect of weather in construction scheduling and management with sine wave curves: application in the united kingdom. **Construction management and economics**, Taylor & Francis, v. 36, n. 12, p. 666–682, 2018. Citado na página [63](#).

CHAKRABORTY, D.; ELHEGAZY, H.; ELZARKA, H.; GUTIERREZ, L. A novel construction cost prediction model using hybrid natural and light gradient boosting. **Advanced Engineering Informatics**, Elsevier, v. 46, p. 101201, 2020. Citado na página [27](#).

CHINNADURAI, J.; VENUGOPAL, V.; KUMARAVEL, P.; PARAMESH, R. Influence of occupational heat stress on labour productivity—a case study from chennai, india. **International Journal of Productivity and Performance Management**, Emerald Group Publishing Limited, 2016. Citado na página [27](#).

DUFFY, G.; WOLDESENBET, A.; OBERLENDER, G. D. *et al.* Advanced linear scheduling program with varying production rates for pipeline construction projects. **Automation in construction**, Elsevier, v. 27, p. 99–110, 2012. Citado na página [27](#).

DURDYEV, S.; HOSSEINI, M. R. Causes of delays on construction projects: a comprehensive list. **International Journal of Managing Projects in Business**, Emerald Publishing Limited, v. 13, n. 1, p. 20–46, 2020. Citado na página [25](#).

EDITION, F. A guide to the project management bodyof knowledge. 2008. Citado na página [32](#).

- EGWIM, C. N.; ALAKA, H.; TORIOLA-COKER, L. O.; BALOGUN, H.; SUNMOLA, F. Applied artificial intelligence for predicting construction projects delay. **Machine Learning with Applications**, Elsevier, v. 6, p. 100166, 2021. Citado na página 27.
- EL-RAYES, K.; MOSELHI, O. Impact of rainfall on the productivity of highway construction. **Journal of construction engineering and management**, American Society of Civil Engineers, v. 127, n. 2, p. 125–131, 2001. Citado na página 28.
- GOLNARAGHI, S.; ZANGENEHMADAR, Z.; MOSELHI, O.; ALKASS, S. Application of artificial neural network (s) in predicting formwork labour productivity. **Advances in Civil Engineering**, Hindawi, v. 2019, 2019. Citado nas páginas 27 e 33.
- GONDIA, A.; SIAM, A.; EL-DAKHAKHNI, W.; NASSAR, A. H. Machine learning algorithms for construction projects delay risk prediction. **Journal of Construction Engineering and Management**, American Society of Civil Engineers, v. 146, n. 1, p. 04019085, 2020. Citado nas páginas 27 e 28.
- GUO, S.-J.; CHEN, J.-H.; CHIU, C.-H. Fuzzy duration forecast model for wind turbine construction project subject to the impact of wind uncertainty. **Automation in Construction**, Elsevier, v. 81, p. 401–410, 2017. Citado na página 28.
- HANOON, M. S.; AHMED, A. N.; ZAINI, N.; RAZZAQ, A.; KUMAR, P.; SHERIF, M.; SEFELNASR, A.; EL-SHAFIE, A. Developing machine learning algorithms for meteorological temperature and humidity forecasting at terengganu state in malaysia. **Scientific Reports**, Nature Publishing Group UK London, v. 11, n. 1, p. 18935, 2021. Citado na página 27.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. H.; FRIEDMAN, J. H. **The elements of statistical learning: data mining, inference, and prediction**. [S.l.]: Springer, 2009. v. 2. Citado na página 28.
- HERAVI, G.; ESLAMDOOST, E. Applying artificial neural networks for measuring and predicting construction-labor productivity. **Journal of Construction Engineering and Management**, American Society of Civil Engineers, v. 141, n. 10, p. 04015032, 2015. Citado na página 27.
- IBBS, W.; SUN, X. Weather's effect on construction labor productivity. **Journal of legal affairs and dispute resolution in engineering and construction**, American Society of Civil Engineers, v. 9, n. 2, p. 04517002, 2017. Citado na página 27.
- JUNG, M.; PARK, M.; LEE, H.-S.; KIM, H. Weather-delay simulation model based on vertical weather profile for high-rise building construction. **Journal of construction engineering and management**, American Society of Civil Engineers, v. 142, n. 6, p. 04016007, 2016. Citado nas páginas 27 e 28.
- KOENKER, R.; HALLOCK, K. F. Quantile regression. **Journal of economic perspectives**, v. 15, n. 4, p. 143–156, 2001. Citado na página 48.
- KOKKAEW, N.; WIPULANUSAT, W. Completion delay risk management: A dynamic risk insurance approach. **KSCE Journal of Civil Engineering**, Springer, v. 18, n. 6, p. 1599–1608, 2014. Citado na página 28.
- LI, X.; CHOW, K. H.; ZHU, Y.; LIN, Y. Evaluating the impacts of high-temperature outdoor working environments on construction labor productivity in china: A case study of rebar workers. **Building and environment**, Elsevier, v. 95, p. 42–52, 2016. Citado na página 27.

LIU, J.; SHAHI, A.; HAAS, C. T.; GOODRUM, P.; CALDAS, C. H. Validation methodologies and their impact in construction productivity research. **Journal of Construction Engineering and Management**, American Society of Civil Engineers, v. 140, n. 10, p. 04014046, 2014. Citado na página 25.

MET OFFICE. **Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853-current)**. 2012. NCAS British Atmospheric Data Centre, 04/12/2022. <<http://catalogue.ceda.ac.uk/uuid/220a65615218d5c9cc9e4785a3234bd0>>. Citado na página 60.

MOONEY, C. Z. **Monte carlo simulation**. [S.l.]: Sage, 1997. Citado na página 39.

NADARAYA, E. On non-parametric estimates of density functions and regression curves. **Theory of Probability & Its Applications**, SIAM, v. 10, n. 1, p. 186–190, 1965. Citado na página 39.

PEWDUM, W.; RUJIRAYANYONG, T.; SOOKSATRA, V. Forecasting final budget and duration of highway construction projects. **Engineering, Construction and Architectural Management**, Emerald Group Publishing Limited, v. 16, n. 6, p. 544–557, 2009. Citado na página 27.

RAJAGOPALAN, B.; LALL, U. A kernel estimator for discrete distributions. **Journaltitle of Nonparametric Statistics**, Taylor & Francis, v. 4, n. 4, p. 409–426, 1995. Citado na página 39.

REITERMANOVA, Z. *et al.* Data splitting. In: MATFYZPRESS PRAGUE. **WDS**. [S.l.], 2010. v. 10, p. 31–36. Citado na página 41.

REZAAIE, K.; AMALNIK, M. S.; GEREIE, A.; OSTADI, B.; SHAKHSENIAEE, M. Using extended monte carlo simulation method for the improvement of risk management: Consideration of relationships between uncertainties. **Applied Mathematics and Computation**, Elsevier, v. 190, n. 2, p. 1492–1501, 2007. Citado na página 28.

RICHARDSON, C. W. Stochastic simulation of daily precipitation, temperature, and solar radiation. **Water resources research**, Wiley Online Library, v. 17, n. 1, p. 182–190, 1981. Citado na página 27.

SADEGHI, N.; FAYEK, A. R.; PEDRYCZ, W. Fuzzy monte carlo simulation and risk assessment in construction. **Computer-Aided Civil and Infrastructure Engineering**, Wiley Online Library, v. 25, n. 4, p. 238–252, 2010. Citado na página 28.

SANNI-ANIBIRE, M. O.; ZIN, R. M.; OLATUNJI, S. O. Causes of delay in the global construction industry: a meta analytical review. **International Journal of Construction Management**, Taylor & Francis, v. 22, n. 8, p. 1395–1407, 2022. Citado na página 25.

SCHULDT, S. J.; NICHOLSON, M. R.; ADAMS, Y. A.; DELORIT, J. D. Weather-related construction delays in a changing climate: a systematic state-of-the-art review. **Sustainability**, Multidisciplinary Digital Publishing Institute, v. 13, n. 5, p. 2861, 2021. Citado nas páginas 25, 26, 27 e 34.

SEPASGOZAR, S. M.; KARIMI, R.; SHIROWZHAN, S.; MOJTAHEDI, M.; EBRAHIMZADEH, S.; MCCARTHY, D. Delay causes and emerging digital tools: A novel model of delay analysis, including integrated project delivery and pmbok. **Buildings**, MDPI, v. 9, n. 9, p. 191, 2019. Citado na página 25.

- SHAHIN, A.; ABOURIZK, S.; MOHAMED, Y. Modeling weather-sensitive construction activity using simulation. **Journal of construction engineering and management**, American Society of Civil Engineers, v. 137, n. 3, p. 238–246, 2011. Citado nas páginas 27 e 28.
- SLYKE, R. M. V. Monte carlo methods and the pert problem. **Operations Research**, INFORMS, v. 11, n. 5, p. 839–860, 1963. Citado nas páginas 27 e 65.
- SONG, L.; ABOURIZK, S. M. Measuring and modeling labor productivity using historical data. **Journal of construction engineering and management**, American Society of Civil Engineers, v. 134, n. 10, p. 786–794, 2008. Citado na página 27.
- STEINWART, I.; CHRISTMANN, A. Estimating conditional quantiles with the help of the pinball loss. **Bernoulli**, Bernoulli Society for Mathematical Statistics and Probability, v. 17, n. 1, p. 211–225, 2011. Citado na página 48.
- THORPE, D.; KARAN, E. P. Method for calculating schedule delay considering weather conditions. In: ASSOCIATION OF RESEARCHERS IN CONSTRUCTION MANAGEMENT (ARCOM). **Proceedings of the 24th annual conference of the Association of Researchers in Construction Management (ARCOM 2008)**. [S.l.], 2008. v. 2, p. 809–818. Citado nas páginas 27 e 28.
- WAUTERS, M.; VANHOUCKE, M. Support vector machine regression for project control forecasting. **Automation in Construction**, Elsevier, v. 47, p. 92–106, 2014. Citado na página 27.
- YASEEN, Z. M.; ALI, Z. H.; SALIH, S. Q.; AL-ANSARI, N. Prediction of risk delay in construction projects using a hybrid artificial intelligence model. **Sustainability**, MDPI, v. 12, n. 4, p. 1514, 2020. Citado na página 27.