ARTIGO 3

MODELO DE PREVISÃO DO LIMITE DE UMIDADE TRANSPORTÁVEL PARA MINÉRIO DE FERRO UTILIZANDO REGRESSÃO LINEAR MÚLTIPLA

Carlos Leonardo Ramos Póvoa
UENF - Universidade Estadual do Norte Fluminense

Alice Siqueira Sales
Anglo American

RESUMO

Este estudo propõe o desenvolvimento de modelos preditivos para estimar o Limite de Umidade Transportável (TML) de minério de ferro, parâmetro essencial para a segurança do transporte marítimo de granéis sólidos. A metodologia adotada baseou-se em regressão linear múltipla aplicada a um banco de dados com variáveis físico-químicas, granulométricas e mineralógicas. Utilizando técnicas de regressão linear múltipla, foi analisado um conjunto de dados composto por variáveis físico-químicas, granulométricas e mineralógicas, obtidos da produção da Anglo American no Brasil. Foram elaborados seis modelos utilizando diferentes métodos de seleção de variáveis e critérios estatísticos como AICc e BIC. Os modelos foram avaliados com base nos coeficientes de determinação (R^2 e R^2 ajustado) e nos erros absolutos e percentuais entre os valores estimados e observados. Os modelos M2 e M4 destacaram-se por apresentarem alto desempenho preditivo, com menor complexidade e maior robustez. A aplicação prática desses modelos permite otimizar a gestão logística, reduzir riscos operacionais e garantir a conformidade com as normas de segurança marítima, contribuindo para maior eficiência e previsibilidade nas operações de exportação de minério.

Palavras-chave: Transporte marítimo, Limite de Umidade Transportável (TML), modelagem preditiva.

ABSTRACT

This study presents the development of predictive models to estimate the Transportable Moisture Limit (TML) of Iron Ore Fine (IOF) - a critical parameter for ensuring the safety of solid bulk maritime transport. The proposed methodology is based on multiple linear regression applied to a comprehensive dataset comprising physical, chemical, granulometric, and mineralogical variables derived from Anglo American's iron ore production in Brazil. Six models were constructed using various variable selection techniques and statistical criteria, including the corrected Akaike Information Criterion (AICc) and the Bayesian Information Criterion (BIC). Model performance was evaluated through coefficients of determination (R^2 and adjusted R^2) and by analyzing absolute and percentage errors between predicted and observed TML values. Among the models, M2 and M4 demonstrated superior predictive accuracy, lower complexity, and greater robustness. The implementation of these models offers a strategic tool for optimizing logistical planning, mitigating operational risks, and ensuring compliance with international maritime safety standards—ultimately enhancing the efficiency and reliability of iron ore export operations.

Keywords: Maritime transport, Transportable Moisture Limit (TML), predictive modeling.

1. INTRODUÇÃO

O transporte marítimo consolida-se como a alternativa mais econômica para o deslocamento de grandes volumes de carga, especialmente granéis sólidos e líquidos, destacando-se pelas tarifas de frete competitivas. No cenário internacional, essa modalidade logística demonstra superior eficiência, uma vez que o aumento da capacidade das embarcações tem viabilizado o transporte econômico de matérias-primas geograficamente dispersas. Como resultado, esses insumos tornam-se disponíveis a preços competitivos mesmo em mercados distantes (Adriano; Robles; Cutrim, 2013).

Conforme Holmes *et al.* (2016), o transporte marítimo de minérios e minerais a granel é uma prática consolidada há mais de um século. O minério de ferro, principal carga granel sólida global, com volume anual superior a 1 bilhão de toneladas, é tipicamente transportado como granulado, pelotas ou finos. Em 2021, o comércio marítimo internacional apresentou recuperação, registrando crescimento de 3,2% (totalizando 11 bilhões de toneladas) após queda de 3,8% em 2020. Contudo, esse desempenho permaneceu ligeiramente abaixo dos níveis pré-pandemia, reflexo dos persistentes desafios sanitários e do congestionamento logístico global decorrente da alta demanda e das limitações na capacidade de oferta. O crescimento foi puxado principalmente pelo transporte de contêineres, enquanto os segmentos de gás e granéis sólidos expandiram-se, e o petróleo bruto registrou declínio (UNCTAD, 2022).

No contexto nacional, as cargas de granéis sólidos – especialmente minérios brutos e concentrados – representam aproximadamente 33% da movimentação portuária anual. Desse total, o minério de ferro responde por cerca de 31%. A movimentação concentra-se em cinco terminais das regiões Nordeste e Sudeste: Ponta da Madeira (MA), Tubarão (ES), Ilha Guaíba (RJ), Porto do Açu (RJ) e Porto Sudeste (RJ) (IBRAM, 2023).

O Porto do Açu é o terminal portuário localizado em São João da Barra, no norte do Rio de Janeiro, e é administrado pela Ferroport, uma *joint venture* entre a Anglo American e a Prumo Logística. A Anglo American opera no Brasil com dois principais negócios: minério de ferro e ferroníquel. Sua produção de minério de ferro, concentrada em Conceição do Mato Dentro e Alvorada de Minas (MG), possui capacidade licenciada de 29,1 milhões de toneladas por ano. O material é transportado por um dos maiores minerodutos do mundo, com 529 km de extensão, atravessando 33 municípios entre Minas Gerais e Rio de Janeiro até ser filtrado e exportado pelo Porto do Açu (Anglo American, 2023).

O transporte marítimo de minérios é uma das fases da cadeia produtiva da mineração, um elo essencial na economia mundial, sendo regulamentado em âmbito internacional pela Organização Marítima Internacional (IMO), cuja estrutura regulatória inclui leis que visam garantir a segurança das operações marítimas. Segundo Ferreira *et al.* (2017) nos últimos 30 anos ocorreram pelo menos 24 acidentes marítimos tendo como causa atribuída

a liquefação da carga, somando mais de 177 vítimas fatais.

O Limite de Umidade Transportável (*Transportable Moisture Limit* – TML) é um teste que determina o teor máximo de umidade de uma carga a ser transportada por via marítima, um parâmetro de segurança que pode paralisar legalmente um embarque de minério ou mesmo toda uma operação portuária caso a umidade embarcada não seja menor do que o TML (Ferreira e Lima, 2023).

Diante desse cenário, este trabalho tem como objetivo desenvolver um modelo preditivo para estimar o TML, empregando técnicas de regressão linear múltipla. A proposta visa não apenas reduzir riscos operacionais associados ao transporte de minérios, mas também otimizar processos logísticos, gerando ganhos de eficiência e redução de custos para as empresas do setor.

A crescente demanda global por minério de ferro impõe a necessidade de um transporte marítimo eficiente e seguro desse insumo estratégico. O Limite de Umidade Transportável (TML) constitui um parâmetro crítico, uma vez que influencia diretamente a segurança operacional, a eficiência logística e os custos associados ao transporte de granéis sólidos. Sua determinação precisa é essencial para prevenir colapsos de carga, derramamentos e acidentes que podem gerar impactos significativos à saúde humana e ao meio ambiente.

Entretanto, a variabilidade intrínseca das propriedades do minério, como composição química, mineralógica e granulométrica, afeta de forma direta o valor do TML, representando um desafio adicional para a indústria de mineração. A ausência de modelos preditivos confiáveis que permitam estimar esse limite antes do processo de embarque limita a otimização das operações logísticas e aumenta o risco de não conformidade com as normas de segurança internacionais.

Dessa forma, este trabalho justifica-se pela necessidade de desenvolver modelos preditivos capazes de estimar o TML com maior precisão e antecedência, oferecendo suporte à tomada de decisão, à mitigação de riscos e à melhoria da eficiência operacional no transporte marítimo de minério de ferro.

2. CONTEXTO E REVISÃO DA LITERATURA

2.1. Aspectos regulatórios do transporte marítimo

A Organização Marítima Internacional (IMO) é um testemunho da evolução do cenário marítimo global ao longo dos últimos anos. Fundada em 1948 como uma agência especializada das Nações Unidas, a IMO desempenhou um papel fundamental na promoção da segurança marítima, na prevenção da poluição dos oceanos e na facilitação do comércio internacional. A IMO desenvolveu o Código Internacional de Cargas Sólidas a Granel (IMSBC), adotado em 1965. O Código IMSBC foi atualizado regularmente desde então e é mantido sob revisão contínua pelo Subcomitê de Mercadorias Perigosas, Cargas Sólidas e Contêineres (DSC). O código é a principal legislação

internacional que rege a segurança do transporte de cargas sólidas a granel a bordo de navios, tendo por propósito estabelecer procedimentos padronizados a serem aplicados nas diversas fases das operações marítimas (IMO, 2018). O Brasil é um dos Estados-membros, sendo representado na IMO pela Marinha (Ferreira, 2019).

As práticas contidas no código IMSBC são destinadas a serem recomendações para governos, operadores de navios e capitães. Seu objetivo é chamar a atenção para um método internacionalmente aceito para lidar com os riscos à segurança que podem ser encontrados ao transportar cargas a granel, sendo o código mais relevante para a segurança dos navios graneleiros de minério de ferro. O código IMSBC destaca os riscos associados ao transporte de certos tipos de cargas a granel; fornece orientações sobre vários procedimentos a serem adotados; lista produtos típicos que são transportados a granel; dá conselhos sobre suas propriedades e como devem ser manipulados; e descreve vários procedimentos de teste que devem ser empregados para determinar as propriedades características da carga (Holmes *et al*., 2016). A classificação em grupos (A: carga com risco de liquefação; B: carga com risco químico; C: carga sem risco de liquefação ou risco químico) foi estabelecida como parte integrante do código IMSBC, fornecendo uma estrutura clara para identificar e gerenciar os riscos associados ao transporte marítimo de cargas a granel.

2.2. Liquefação

A liquefação tem sido estudada minuciosamente na literatura de mecânica do solo, principalmente em relação a terremotos. A liquefação pode ocorrer tipicamente em solos granulares não coesivos, que são relativamente finos, soltos e saturados com água (Paassen *et al*., 2015). Mas esse fenômeno se tornou uma preocupação significativa no transporte marítimo de granéis sólidos, especialmente para cargas classificadas no grupo A pelo Código IMSBC.

A liquefação ocorre quando a umidade presente no material faz com que ele se comporte como um líquido, aumentando consideravelmente o risco de estabilidade do navio. Um dos principais parâmetros para avaliar esse risco é o Limite de Umidade Transportável (TML), que determina o teor de umidade máximo seguro para o transporte da carga. Este limite é crucial para garantir a segurança da operação e evitar acidentes graves, como o naufrágio de navios devido à instabilidade causada pela liquefação da carga (Wu *et al*., 2022).

Quando ocorre a liquefação de uma carga sólida a granel, sua estabilidade é comprometida, transformando-a em um líquido denso e viscoso, comportando-se de maneira intermediária entre uma carga líquida e uma carga a granel seca. Esse fenômeno resulta na formação de uma superfície livre, caracterizada pela diminuição da estabilidade devido às variações no centro de gravidade

dos líquidos que ocupam parcialmente um tanque ou porão no navio. A carga liquefeita se desloca livremente no porão, muitas vezes em uma fase diferente em relação à embarcação: enquanto o navio se inclina para um lado, a carga se desloca na direção oposta (Ferreira, 2019).

2.3. Limite de umidade transportável – TML

O Limite de Umidade Transportável (TML) de minérios de ferro é uma realidade regulatória que passou a vigorar no dia a dia das operações, tanto no planejamento dos produtos quanto nas operações de embarque, que devem levar em consideração o atendimento a esse parâmetro de segurança (Ferreira, *et al.* 2017). Diante desse cenário, o TML passou a ser mais discutido pela Organização Marítima e pelas empresas responsáveis.

É importante destacar que o limite de umidade transportável pode variar de acordo com o tipo de carga e as condições específicas do transporte, e, portanto, deve ser avaliado caso a caso. O controle da umidade é um aspecto fundamental para garantir a eficiência e segurança do transporte de cargas a granel, especialmente em materiais como o minério de ferro fino, que são amplamente utilizados na indústria siderúrgica e de construção civil. O limite de umidade transportável, definido com precisão e atualizado regularmente, é essencial para prevenir acidentes e proteger a vida humana e o meio ambiente durante o transporte marítimo de granéis sólidos (Munro e Mohajerani, 2016).

A classificação do minério de ferro como uma carga do tipo A, segundo o código IMSBC, está diretamente relacionada ao risco de liquefação associado a essa carga. Minérios de ferro finos, especialmente aqueles com teores de umidade mais elevados, têm o potencial de se comportar como um líquido durante o transporte marítimo devido à ação das forças de vibração e movimento do navio. Portanto, a classificação do minério de ferro como carga do tipo A destaca a importância crítica de avaliar cuidadosamente o teor de umidade e seguir as diretrizes específicas do Código IMSBC para garantir um transporte seguro e proteger contra os perigos potenciais de liquefação durante o transporte marítimo.

O TML não é um valor fixo para toda e qualquer carga. Como as características relacionadas ao risco de liquefação variam de um material para outro, o TML é determinado por testes de laboratório descritos no código IMSBC, ou seja, o limite varia de acordo com o material testado. O valor do parâmetro deve ser comparado com a umidade da carga para embarques em navios que não sejam preparados para carregar cargas liquefeitas (embarcações preparadas para tal são, atualmente, uma parcela irrisória da frota mundial). Caso a umidade do minério esteja acima do TML, a carga não pode ser embarcada, apesar da exigência legal de que todos os exportadores de minério de ferro

declarem a umidade e o TML de suas cargas para que possam ser embarcadas e transportadas.

Segundo Ferreira (2019), o código IMSBC estabelece que o TML deve ser determinado através de testes cujo procedimento tenha sido aprovado por uma autoridade competente, e fornece atualmente cinco opções de procedimentos laboratoriais aprovados pela IMO para determinação do parâmetro: o teste Flow Table (contido no Código desde 1965), o teste Penetration (incluso em 1994), o teste de Proctor/Fagerberg (incluso em 1994), o teste de Proctor/Fagerberg Modificado para Finos de Minério de Ferro, ou PFD80 (incluso no Código em 2015) e o teste de Proctor/Fagerberg Modificado para Carvão (incluso no Código em 2017). Os três primeiros testes são aplicados para concentrados minerais, e somente o PFD80 é aplicado para determinação do TML de Finos de Minério de Ferro (FMF), caso alguns pré-requisitos sejam atendidos. De acordo com a determinação da Marinha do Brasil, o minério exportado pela Anglo American deve ser testado pelo método Proctor/Fagerberg C70, visto que é um material que passou por um processo de beneficiamento e é caracterizado como um *Iron Concentrate* (pellet feed).

2.4. Controle de umidade no carregamento

Na indústria de mineração, o controle da umidade dos produtos é uma prática comum, motivada por considerações comerciais, legais e de controle de processos. Geralmente, minérios são manipulados enquanto úmidos, e a presença de água entre as partículas sólidas tem um impacto significativo no comportamento do material particulado em vários aspectos, como o ângulo de repouso, a massa específica aparente e a resistência, sendo que a influência da água depende da quantidade presente. A estrutura regulatória da IMO exige que o TML e a umidade de uma carga do Grupo A sejam informados ao comandante do navio antes de iniciar o carregamento. Caso a carga esteja com umidade igual ou superior ao TML, não deve ser embarcada, a não ser que o navio seja especialmente projetado para transporte de cargas liquefeitas. O limite de umidade deve ser revisado a cada 6 meses ou quando há alterações na carga (Ferreira *et al.* 2017).

A Organização Marítima Internacional possui um protocolo para prevenção de liquefação de cargas, com o intuito de mitigar o risco. Este protocolo é guiado pelo Código Marítimo Internacional para Cargas Sólidas a Granel (Código IMSBC), que é o principal instrumento regulatório internacional para mitigação do risco de liquefação de cargas sólidas a granel durante o transporte marítimo. O código incorpora diversos controles que visam garantir que as cargas do Grupo A sejam carregadas com um teor de umidade abaixo do necessário para evitar a liquefação (Rose, 2014).

Quando a umidade do material a ser embarcado excede o TML, as empresas exportadoras de minério enfrentam uma série de desafios que podem impactar negativamente suas operações, finanças e reputação no mercado. Diante dessas situações, a possibilidade de estimar o TML com base nas características do minério, por meio da proposição de modelos, traria grandes benefícios para as operações de exportação de minério. Esses benefícios incluem:

• Antecipação de ações de redução de umidade: Com o conhecimento prévio das variações do TML, as ações para redução de umidade podem ser antecipadas, evitando atrasos no embarque. Isso permite uma gestão mais eficaz dos processos logísticos,

garantindo o cumprimento dos prazos de entrega.

• Otimização do planejamento da composição do material: A capacidade de estimar o TML permite ajustes no planejamento da composição do material a ser alimentado na usina de beneficiamento. Isso resulta em alterações na composição do material que chega ao porto, proporcionando economia de tempo e reduzindo ou eliminando o risco de inviabilização da exportação devido à umidade excessiva. Essa otimização também pode levar a uma melhor alocação de recursos e redução de custos operacionais.

• Maior estabilidade nas operações: A utilização de modelos para estimar o TML proporciona maior estabilidade nas operações de embarque de minério. Isso ajuda a evitar os problemas associados ao carregamento de cargas com umidade acima do limite permitido, como atrasos, cancelamentos de embarques e danos à reputação da empresa exportadora. Além disso, contribui para garantir o cumprimento dos requisitos regulatórios do empreendimento, promovendo uma operação mais eficiente e sustentável.

• Maior planejamento na fila de navios para aumentar o tempo de residência nas pilhas, caso o material apresente uma tendência de um TML menor que a umidade.

Portanto, a aplicação de modelos para estimar o TML representa uma abordagem estratégica para mitigar os riscos associados à umidade excessiva no transporte de minério, proporcionando benefícios significativos em termos de eficiência operacional, conformidade regulatória e desempenho financeiro.

3. MATERIAIS E MÉTODOS

3.1. Base de dados

A base de dados utilizada nesta pesquisa é composta pelos resultados de análises físicas, químicas e mineralógicas do material produzido pela empresa Anglo American Minério de Ferro Brasil. As análises foram realizadas pelos laboratórios vinculados ao Projeto Minas-Rio. O laboratório localizado no Porto do Açu foi responsável pela execução dos ensaios de Teor Máximo de Umidade (TML) e umidade, além das análises químicas. As demais análises, tais como granulometria a laser e mineralogia, foram conduzidas no laboratório situado em Conceição do Mato Dentro, local onde se encontra a mina.

A base de dados foi composta por 86 amostras de minério de ferro embarcadas no Terminal de Minério de Ferro do Porto do Açu, no período de 2020 a 2024. Dessas, 63 amostras foram utilizadas para a construção dos modelos preditivos, enquanto 23 foram reservadas exclusivamente para compor o banco de testes, assegurando a validação independente dos modelos desenvolvidos.

Para a determinação do TML, adotou-se a metodologia Proctor/Fagerberg C70, conforme exigência da Marinha do Brasil, órgão regulamentador competente. Todos os ensaios foram conduzidos em amostras coletadas e preparadas de acordo com os procedimentos estabelecidos pela ABNT NBR ISO 3082 – Minérios de ferro — Procedimentos de amostragem e preparação de amostras.

As análises químicas foram realizadas através da metodologia de pastilha

fundida e analisadas em equipamento de fluorescência de Raio X, Axios Max Panalytical. Os elementos e compostos Fe, SiO2, Al2O3 e P, foram utilizados na análise exploratória, visto que, de acordo com o Menezes *et. al* (2019), esses seriam os elementos e compostos com maior interação com o TML. A perda por calcinação (PPC) foi realizada por gravimetria em mufla a 1000°C.

A granulometria foi realizada por meio do equipamento Mastersize 3000E da Malvern, pela metodologia de laser. Nessa metodologia o feixe de laser é emitido sob a partícula de minério de ferro, e o desvio que o feixe de luz sofre ao atravessar a partícula possibilita a determinação do volume da partícula, sendo possível transformá-lo em volume de uma esfera e determinar o diâmetro da partícula. A análise de superfície específica foi realizada por meio do permeabilímetro manual de Blaine, e refere-se à área superficial total das partículas de minério de ferro por unidade de massa ou volume. A superfície específica foi determinada experimentalmente pela técnica onde o ar percola a amostra de minério de ferro, e esse tempo é cronometrado e utilizado para calcular o Blaine do material. Uma alta superfície específica significa que as partículas de minério possuem uma maior área de contato e têm relação com a quantidade de finos e contaminantes presentes no minério.

A análise de TML e Blaine utiliza a densidade dos sólidos (*p*) do material nos cálculos, então essa caracterização também foi inserida no banco de dados do estudo. A massa específica dos sólidos foi determinada conforme a norma americana ASTM D5550-14 (ASTM, 2014), pelo método de picnometria a gás (hélio), nos equipamentos *Ultrapycnometer* e *Pentapycnometer* da empresa Quantachrome®. A Tabela 1 apresenta as variáveis consideradas no estudo.

Tabela 1 - Variáveis do estudo

Caracterização PFC70	Variáveis TML
Fluorescência de Raio X	Fe SiO2 Al2O3 P
Gravimetria	PPC
Mineralogia	Hematita Lamelar Hematita Granular Hematita Lobular Martita Magnetita Goehtita Quartzo Filossilicato Argilomineral Gibbsita Manganês Outro
Granulometria a laser	<10um d60 d10 Coeficiente de Uniformidade
Superfície Específica	Blaine
Picnometria	Densidade

3.2. Pré-processamento

Antes de iniciar a análise dos dados, foi realizada uma etapa de pré-processamento para garantir a qualidade e a integridade dos dados. Esse processo incluiu as seguintes etapas:

1. Tratamento de dados ausentes: As células contendo ausência de dados foram devidamente identificadas e, na sequência, procedeu-se à remoção das respectivas linhas, de modo a garantir a consistência do conjunto de dados e assegurar que todas as variáveis apresentem o mesmo peso na análise.

2. Detecção e correção de dados incorretos: Dados inconsistentes foram identificados, e casos em que os valores incorretos eram óbvios (ex.: erros de digitação), esses tiveram seus resultados confirmados na base de dados original dos laboratórios, como nos Sistemas de Gerenciamento de Informações do Laboratório (LIMS – *Laboratory Information Managment System*), nos sistemas de produção da empresa (MES – *Manufactoring Execution Systems*) e/ou nos equipamentos e planilhas de Excel da área. Após conferência, eles foram corrigidos no banco de dados do estudo.

3. Definição das amostras a serem consideradas para o estudo: Algumas das análises de TML foram realizadas em duplicatas, a fim de avaliar a repetibilidade dos resultados, controlando e garantindo a qualidade das análises. Para essa definição levou-se em conta a avaliação da curva de saturação do teste de TML. A melhor curva foi identificada e selecionada para ser mantida no banco de dados do estudo. Caso as curvas estivessem muito semelhantes, optou-se por escolher o ensaio que obteve menor valor de TML, a fim de garantir maior segurança do teste.

Esse processo permitiu a remoção de ruídos e a padronização dos dados, resultando em um conjunto mais robusto e confiável para as etapas subsequentes de análise. Após o pré-processamento, foi realizada a avaliação das variáveis de forma a remover aquelas correlacionadas entre si, que estariam redundantes na tentativa de se explicar as variações do TML. Além disso, avaliaram-se as correlações, com o objetivo de retirar as preditoras desnecessárias no modelo de regressão, evitando que ruídos fossem adicionados ao cálculo.

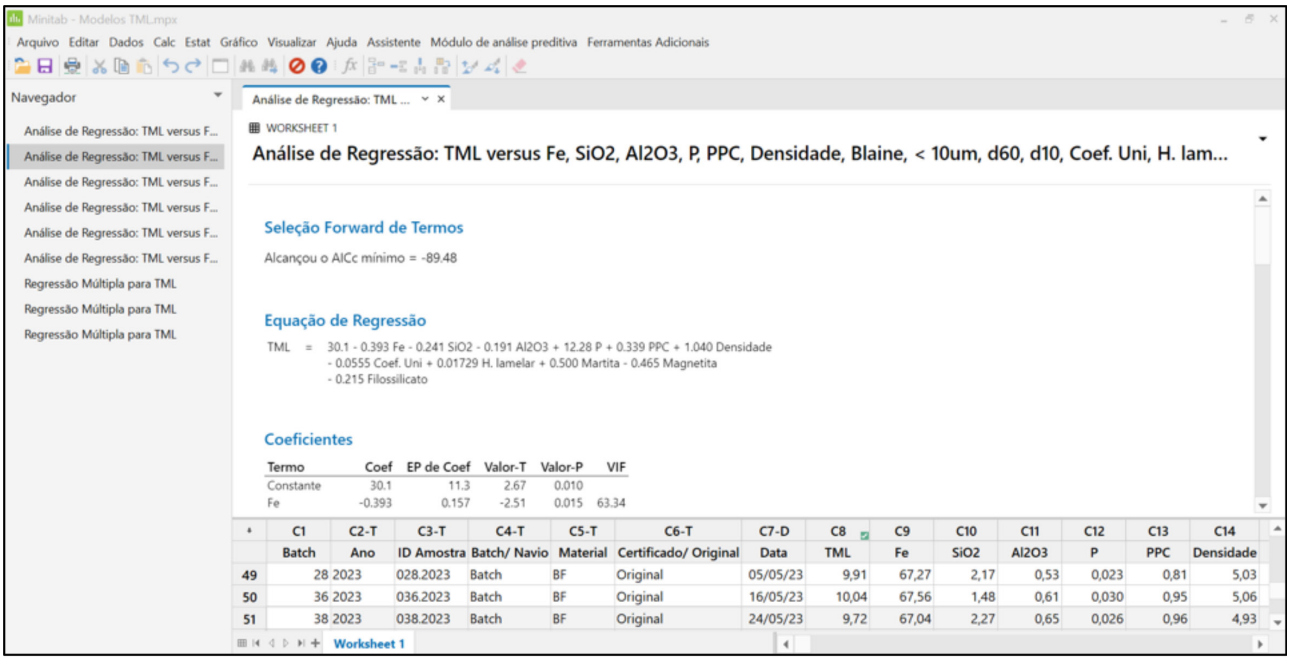
A partir do banco de dados geral, definiu-se a necessidade de extrair uma amostra específica, destinada exclusivamente às etapas de teste e validação do modelo, não sendo utilizada na sua construção. Para a composição desse banco de teste, foram selecionadas amostras que, ao longo dos anos, foram utilizadas nos processos de certificação do Teor Máximo de Umidade (TML) do material a ser embarcado, e não apenas amostras provenientes de estudos de processo. A escolha dessas amostras justificou-se pelo fato de terem sido submetidas, adicionalmente, a ensaios realizados em laboratórios externos, possibilitando a comparação dos resultados obtidos nas análises de TML, o que confere maior robustez ao processo de validação do modelo.

3.3. Modelo preditivo

Para a concepção do modelo preditivo, utilizou-se o *software* Minitab® 19.1, e optou-se pela regressão linear múltipla, técnica estatística amplamente utilizada para modelar a relação entre uma variável dependente e duas ou mais variáveis independentes. Ao contrário da regressão linear simples, que considera apenas uma variável preditora, a regressão múltipla permite incorporar diversas variáveis explicativas

simultaneamente, possibilitando uma modelagem mais complexa e realista de fenômenos onde múltiplos fatores interagem. Se um modelo de regressão se ajusta à relação entre a resposta e seus preditores, então o modelo prevê os valores de resposta com novas observações com precisão. A Figura 1 ilustra a interface do Minitab® exibindo os resultados obtidos para o modelo M2.

Figura 1 – Interface Minitab® Modelo M2



Fonte: autores (2024)

Quando o número de preditores é elevado, recomenda-se, antes de ajustar um modelo de regressão com todas as variáveis, aplicar técnicas de seleção de modelos, como *stepwise* ou seleção dos melhores subconjuntos. Esses métodos auxiliam na identificação e remoção de preditores que não apresentam associação significativa com a variável resposta. De modo geral, uma abordagem eficaz para determinar qual modelo apresenta o melhor desempenho preditivo consiste na geração de múltiplos modelos e na comparação dos respectivos desempenhos.

Entre as métricas utilizadas para avaliar modelos de regressão, destaca-se o coeficiente de determinação (R^2), amplamente empregado na análise de desempenho preditivo. O R^2 expressa o grau de ajuste do modelo, indicando a proporção da variabilidade da variável dependente que é explicada pelas variáveis independentes incluídas no modelo. Seus valores variam de 0 a 1, sendo que valores próximos de 0 indicam baixo poder explicativo do modelo, enquanto valores próximos de 1 refletem que o modelo explica praticamente toda a variação observada nos dados.

A escolha da regressão linear múltipla justifica-se pelo seu caráter exploratório e facilidade de interpretação, permitindo identificar relações significativas entre variáveis físico-químicas, granulométricas e mineralógicas e o TML. Entretanto, admite-se que a relação entre essas variáveis pode apresentar comportamentos não lineares, os quais poderiam ser mais bem capturados por modelos alternativos, como técnicas baseadas em regressão não linear, aprendizado de máquina ou métodos híbridos. Dessa forma, a utilização da regressão múltipla neste trabalho deve ser entendida como um primeiro passo, oferecendo uma base sólida para futuras pesquisas que explorem metodologias mais sofisticadas, capazes de ampliar a capacidade preditiva e a robustez dos modelos.

De acordo com Ferreira (2019), o coeficiente de determinação ajustado também pode ser um critério para avaliação de desempenho do modelo. Onde o R^2_{aj} é uma métrica alternativa ao coeficiente de determinação (R_2) quando se tem diversas variáveis, que penaliza a inclusão de variáveis regressoras pouco explicativas. Ao contrário do coeficiente de determinação, o R^2_{aj} não necessariamente aumenta sempre que uma nova variável é adicionada ao modelo, mas somente se a adição da nova variável reduzir significativamente a média quadrática dos erros, sendo, portanto, mais adequado para comparar modelos de regressão múltipla. Quanto maior R^2_{aj}, melhor o modelo. O parâmetro é calculado pela Equação 1.

$$R^2_{aj} = 1 - \frac{(n-1)}{n-(k+1)}\left(1 - \left(1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \underline{y})^2}\right)\right) \qquad (1)$$

4. RESULTADOS E DISCUSSÃO

4.1. Estatística descritiva

Na seção de resultados e discussões, é fundamental apresentar uma visão geral das características estatísticas das variáveis que compõem o modelo preditivo. Com o objetivo de proporcionar um maior entendimento sobre o comportamento dessas variáveis, a Tabela 2 sintetiza as principais estatísticas descritivas, incluindo média, mediana, desvio padrão, coeficiente de variação, além dos valores mínimo, máximo, primeiro e terceiro quartis.

A apresentação desses indicadores é essencial para a compreensão preliminar do perfil dos dados, possibilitando avaliar a variabilidade, a tendência central e a dispersão das variáveis. Essa análise fornece subsídios importantes para identificar padrões, possíveis outliers e eventuais assimetrias, além de oferecer uma base consistente para a construção, ajuste e validação do modelo preditivo.

Na Tabela 2 estão dispostas as variáveis analisadas, juntamente com o tamanho da amostra (n), a média, que representa a tendência central dos dados, e o desvio padrão, que expressa o grau de dispersão em torno da média. O coeficiente de variação (CV), calculado como a razão entre o desvio padrão e a média, multiplicado por 100%, fornece uma medida relativa da variabilidade dos dados. Além disso, são apresentados os valores mínimo e máximo observados, o primeiro quartil (Q1), que delimita os 25% menores valores, o terceiro quartil (Q3), que corresponde aos 75% menores valores, e a mediana, que representa o ponto central da distribuição, dividindo o conjunto de dados em duas partes iguais.

A variável dependente TML apresenta baixa variabilidade, com coeficiente de variação (CV) de 1,73%, indicando alta homogeneidade nos dados. A média de 10,03 é igual a mediana (10,03), o que sugere uma distribuição simétrica. De forma semelhante, o teor de Fe apresenta elevada homogeneidade (CV = 0,97%), com valores concentrados em torno da média de 67,33, reforçando a consistência dos dados. A proximidade entre Q1 (66,68), mediana (67,53) e Q3 (67,88) corrobora essa uniformidade.

Por outro lado, o teor de SiO_2 apresenta alta variabilidade relativa (CV = 38,27%), indicando ampla dispersão dos dados. O intervalo interquartil (Q1 = 1,34 e Q3 = 2,74) e o valor máximo de 3,58 sugerem presença de variabilidade significativa, que pode impactar a estabilidade do modelo. As variáveis Al_2O_3 (CV = 16,52%), P (CV = 20,65%) e PPC (CV = 23,62%) também apresentam dispersões consideráveis, o que indica que seus valores flutuam mais amplamente ao longo da amostra. A variável Densidade demonstra estabilidade (CV = 1,38%), com baixa dispersão, semelhante ao comportamento observado em TML e Fe.

Tabela 2: Informações estatísticas - variáveis de teste

Variável	*n*	Média	DesvPad	CV	Mínimo	Q1	Mediana	Q3	Máximo
TML	63	10.03	0.17	1.73	9.72	9.88	10.03	10.15	10.42
Fe	63	67.33	0.65	0.97	65.72	66.68	67.53	67.88	68.31
SiO2	63	2.02	0.77	38.27	0.97	1.34	1.78	2.74	3.58
Al2O3	63	0.62	0.10	16.52	0.43	0.53	0.62	0.70	0.88
P	63	0.023	0.005	20.65	0.014	0.020	0.023	0.026	0.035
PPC	63	0.79	0.19	23.62	0.42	0.66	0.79	0.91	1.21
Densidade	63	5.08	0.07	1.38	4.93	5.02	5.08	5.13	5.22
Blaine	63	1038	45	4	969	1011	1035	1053	1213
< 10um	63	9.48	1.21	12.71	7.60	8.70	9.30	10.10	16.00
d60	63	42.84	1.99	4.63	36.55	41.60	43.20	44.40	45.90
d10	63	10.57	1.04	9.84	5.53	10.00	10.70	11.30	12.30
Coef. Uni	63	4.09	0.47	11.58	3.32	3.82	4.02	4.27	6.93
H. lamelar	63	78.99	1.99	2.52	74.57	77.50	79.27	80.68	82.18
H. granular	63	11.88	1.19	10.02	8.78	11.04	12.03	12.45	14.68
H. lobular	63	2.05	0.64	31.12	0.90	1.58	1.92	2.44	3.48
Martita	63	0.21	0.08	39.95	0.11	0.16	0.18	0.20	0.63
Magnetita	63	0.01	0.04	344.79	0.00	0.00	0.00	0.00	0.17
Goehita	63	3.97	0.93	23.31	2.07	3.20	3.95	4.81	6.21
Quartzo	63	1.92	0.53	27.81	1.04	1.50	1.90	2.44	3.02
Filossilicato	63	0.35	0.12	32.92	0.10	0.27	0.34	0.43	0.57
Argilomineral	63	0.05	0.06	113.34	0.00	0.00	0.06	0.09	0.26
Gibbsita	63	0.30	0.14	45.74	0.06	0.18	0.28	0.42	0.56
Manganês	63	0.16	0.04	27.27	0.08	0.14	0.16	0.17	0.39
Outros	63	0.11	0.03	27.30	0.05	0.09	0.10	0.11	0.23

No que se refere às características físicas, a variável Blaine possui um CV de 4%, apontando moderada estabilidade. O parâmetro “<10µm” apresenta CV de 12,71%, indicando variação mais expressiva no percentual de partículas finas. As variáveis granulométricas d60 (CV = 4,63%), d10 (CV = 9,84%) e o Coeficiente de Uniformidade (CV = 11,58%) apresentam dispersões que refletem a heterogeneidade na distribuição de tamanhos de partículas. Quanto às fases texturais, Hematita lamelar e Hematita granular exibem relativa estabilidade (CV = 2,52% e 10,02%, respectivamente), enquanto Hematita lobular apresenta alta variabilidade (CV = 31,12%), o que pode indicar presença não uniforme desse componente.

Dentre os minerais acessórios, destaca-se a Magnetita, que possui o maior coeficiente de variação (CV = 344,79%), indicando presença extremamente esporádica e, na maioria dos casos, nula (Q1, Mediana e Q3 = 0). Comportamento similar, embora em menor escala, é observado para Argilomineral (CV = 113,34%) e Gibbsita (CV = 45,74%), cujas distribuições sugerem alta heterogeneidade na amostra.

Os minerais Goehita (CV = 23,31%), Quartzo (CV = 27,81%), Filossilicato (CV = 32,92%) e Manganês (CV = 27,27%) também apresentam dispersões relevantes, o que indica que suas participações variam significativamente entre as amostras. A categoria “Outros” mantém comportamento semelhante (CV = 27,30%), reforçando a diversidade mineralógica do material analisado.

De maneira geral, observa-se que algumas variáveis, especialmente as químicas majoritárias (Fe, TML, Densidade), apresentam alta estabilidade, enquanto variáveis associadas a fases menores ou traços (Magnetita, Argilomineral e Gibbsita) exibem elevada variabilidade. Esse comportamento deve ser considerado na etapa de modelagem, uma vez que variáveis com alta dispersão podem ter impacto relevante tanto na explicação da variabilidade da variável dependente quanto na robustez do modelo preditivo.

4.2. Modelos preditivos

Para a construção dos modelos preditivos, utilizou-se a ferramenta de regressão disponível no *software* Minitab. Foram desenvolvidos diferentes modelos, aplicando distintas metodologias, conforme detalhado nas seções subsequentes.

Os modelos de 1 a 6 foram elaborados utilizando a ferramenta de Ajuste do Modelo de Regressão, que permite maior controle sobre a seleção das variáveis, bem como a análise dos coeficientes e dos pressupostos estatísticos.

4.2.1 MODELO M1

Neste modelo, não foi aplicada nenhuma metodologia específica de pré-processamento ou tratamento dos dados. A variável TML foi definida como variável resposta, enquanto todas as demais variáveis foram inseridas como preditores contínuos no modelo. Dessa forma, todas as variáveis disponíveis foram consideradas na etapa de ajuste, buscando identificar sua contribuição na explicação da variabilidade da variável resposta. Esse modelo obteve um R2 = 75.8% e R2adj = 61.69%.

Equação de Regressão:

$$\begin{aligned}TML = {} & 769 - 0.350\,Fe - 0.205\,SiO_2 - 0.149\,Al_2O_3 + 13.24\,P + 0.373\,PPC \\ & + 1.188\,Densidade - 0.000228\,Blaine - 0.0779\,\text{“}< 10um\text{”} - 0.0206\,d60 \\ & - 0.0251\,d10 + 0.063\,Coef.\,Uni - 7.4\,H.\,lamelar - 7.4\,H.\,granular - 7.4\,H.\,lobular \\ & - 7.0\,Martita - 7.9\,Magnetita - 7.4\,Goehita - 7.4\,Quartzo - 7.6\,Filossilicato \\ & - 7.2\,Argilomineral - 7.4\,Gibbsita - 7.4\,Manganês - 6.8\,Outros\end{aligned}$$

4.2.2 MODELO M2

No segundo modelo, aplicou-se o método de seleção de variáveis Forward, com base no Critério de Informação de Akaike corrigido (AICc). Este critério avalia, de forma sequencial, a inclusão de variáveis no modelo, buscando otimizar o equilíbrio entre a qualidade do ajuste e a complexidade do modelo. A utilização do AICc tem como objetivo selecionar o modelo que maximize a capacidade preditiva, ao mesmo tempo que minimize o risco de sobreajuste — situação em que o modelo se ajusta excessivamente aos dados de treinamento, capturando ruídos e variações aleatórias, o que compromete sua capacidade de generalização (Draper; Smith, 1998; Burnham; Anderson, 2002). A adoção desse critério contribui para garantir que o modelo final seja não apenas bem ajustado aos dados observados, mas também robusto, parcimonioso e aplicável a novos conjuntos de dados. Esse modelo obteve um $R^2 = 71.93\%$ e $R^2_{adj} = 61.88\%$.

$$\begin{aligned}TML = {} & 30.1 - 0.393\,Fe - 0.241\,SiO2 - 0.191\,Al_2O_3 + 12.28\,P + 0.339\,PPC \\ & + 1.040\,Densidade - 0.0555\,Coef.\,Uni + 0.01729\,H.\,lamelar + 0.500\,Martita \\ & - 0.465\,Magnetita - 0.215\,Filossilicato\end{aligned}$$

4.2.3 MODELO M3

O terceiro modelo também aplicou o método do critério de informação Forward, porém através do Critério de Informação Bayesiano (BIC), que é uma variante do AIC, no qual avalia a inclusão progressiva de variáveis no modelo, visando encontrar aquele que oferece o melhor equilíbrio entre ajuste e complexidade, e com maior penalidade para a inclusão de variáveis adicionais ao modelo (Burnham; Anderson, 2002). Esse modelo obteve um $R^2 = 64.77\%$ e $R^2_{adj} = 60.29\%$.

$$\begin{aligned}TML = {} & 7.34 - 0.0742\,Fe + 0.187\,Al2O3 + 0.681\,PPC + 1.056\,Densidade \\ & + 0.01998\,H.\,lamelar + 0.484\,Martita - 0.704\,Magnetita\end{aligned}$$

4.2.4 MODELO M4

O método *Stepwise* é uma abordagem utilizada na seleção de variáveis em modelos de regressão múltipla, que combina procedimentos de inclusão (forward) e exclusão (*backward*) de variáveis, com base em critérios estatísticos (Draper; Smith, 1998). Durante o processo, as variáveis são inseridas ou removidas do modelo de forma sequencial, de acordo com sua significância estatística.

Um dos principais critérios adotados nesse procedimento é o valor p, que representa a probabilidade de se obter um resultado tão extremo quanto o observado, assumindo que a hipótese nula seja verdadeira. No contexto da regressão, a hipótese nula geralmente testa se o coeficiente de uma variável é igual a zero, ou seja, se essa variável não possui efeito significativo sobre a variável dependente. Valores de p menores que um nível de significância de 5% indicam que a variável possui efeito estatisticamente significativo no modelo, sendo, portanto, mantida. Por outro lado, valores de p elevados sugerem que a variável não contribui significativamente, podendo ser descartada do modelo.

A aplicação do método *Stepwise* permite, assim, a construção de um modelo preditivo mais eficiente e parcimonioso, composto apenas pelas variáveis que explicam de forma significativa a variabilidade dos dados. O modelo obtido apresentou coeficiente de determinação R^2 = 70,60% e R^2 ajustado = 65,61%, indicando que cerca de 65,61% da variabilidade da variável dependente é explicada pelas variáveis selecionadas, após o ajuste pela quantidade de preditores incluídos no modelo.

$$\begin{aligned} TML = {} & 25.06 - 0.318\ Fe - 0.193\ SiO_2 + 12.42\ P + 0.312\ PPC + 1.024\ Densidade \\ & - 0.0642\ Coef.\ Uni + 0.01653\ H.\ lamelar + 0.461\ Martita - 0.231\ Filossilicato \end{aligned}$$

4.2.5 MODELO M5

O método de seleção *Forward* é uma técnica de modelagem em regressão múltipla que inicia com um modelo vazio e adiciona variáveis uma a uma com base em critérios de significância estatística, como o valor p. A cada etapa, a variável que mais contribui para a melhoria do ajuste do modelo é selecionada e incorporada, enquanto as demais são mantidas fora (Burnham; Anderson, 2002). Esse modelo obteve R^2 = 73.7% e R^2_{adj} = 66.73%.

$$\begin{aligned} TML = {} & 33.0 - 0.425\ Fe - 0.253\ SiO_2 - 0.278\ Al_2O_3 + 12.64\ P + 0.319\ PPC \\ & + 1.046\ Densidade - 0.0577\ Coef.\ Uni + 0.0077\ H.\ lamelar \\ & - 0.0450\ H.\ lobular + 0.493\ Martita - 0.513\ Magnetita - 0.181\ Filossilicato \\ & + 0.768\ Outros \end{aligned}$$

4.2.6 MODELO M6

O método de eliminação *Backward* é uma técnica de seleção de variáveis em modelos de regressão múltipla que inicia com um modelo completo, incluindo todas as variáveis disponíveis. A partir desse ponto, o processo consiste em remover, de forma iterativa, as variáveis que têm menor impacto na explicação da variabilidade dos dados, com base em critérios estatísticos como o valor p. A cada etapa, a variável menos significativa é eliminada, e o modelo é reavaliado até que todas as variáveis restantes sejam estatisticamente significativas (Burnham; Anderson, 2002). Esse modelo obteve R^2 = 69.21% e R^2_{adj} = 64.65%.

$$\begin{aligned} TML = {} & 10.79 - 0.0983\ Fe + 17.61\ P + 0.323\ PPC + 1.383\ Densidade \\ & - 0.0364\ \text{“}< 10um\text{”} - 0.02156\ d60 - 0.0350\ H.\ granular - 0.419\ Filossilicato \end{aligned}$$

4.2.7 AVALIAÇÃO DOS RESULTADOS

Nesta seção, são apresentados os resultados obtidos a partir dos modelos desenvolvidos. A avaliação do desempenho dos modelos foi realizada por meio das métricas coeficiente de determinação (R^2) e coeficiente de determinação ajustado (R^2_a). Esses indicadores mensuram a qualidade dos modelos e sua capacidade explicativa, considerando tanto o grau de ajuste bruto (R^2) quanto o ajuste ponderado pelo número de variáveis incluídas (R^2_{aj}).

Adicionalmente, será analisado o desempenho dos modelos a partir de um banco de dados de teste, independente daquele utilizado na etapa de calibração. Essa análise tem como objetivo validar a eficácia e a robustez das abordagens propostas, assegurando que os modelos apresentem boa capacidade preditiva e confiabilidade quando aplicados a novos conjuntos de dados. Dessa forma, busca-se garantir uma avaliação abrangente da performance dos modelos desenvolvidos.

A Tabela 3 apresenta uma síntese do desempenho dos modelos desenvolvidos, com base nos indicadores coeficiente de determinação (R^2) e coeficiente de determinação ajustado (R^2_a), os quais expressam a proporção da variabilidade da variável dependente explicada pelo modelo, considerando, respectivamente, o ajuste bruto e o ajuste penalizado pelo número de variáveis.

Adicionalmente, a tabela oferece uma visão consolidada das médias das diferenças absolutas (Dif. Média) e percentuais entre os valores de TML medidos e estimados (Dif.% Média), para cada modelo (M1 a M6), calculadas a partir das 23 amostras do banco de dados de teste. Também são apresentados os desvios padrão das diferenças, os quais indicam o grau de dispersão dos erros de predição, fornecendo uma medida da consistência e estabilidade das estimativas geradas pelos modelos.

Tabela 3 - Resumo dos Modelos

Modelo	R^2	R^2_{aj}	|Difl Média	Dif % Média	DesvPad Médio (diferença)
M1	75,90	61,69	0,25	2.51	0.13
M2	71,93	65,88	0,10	1.02	0.09
M3	64,77	60,29	0,12	1.19	0.10
M4	70,60	65,61	0,10	1.03	0.09
M5	73,70	66,73	0,11	1.06	0.10
M6	69,21	64,65	0,11	1.12	0.08

O modelo 1 (M1) apresenta o maior R^2 (75,90%), indicando que, no conjunto de treinamento, é aquele que melhor explica a variabilidade dos dados. No entanto, observa-se que seu R^2_{aj} (61,69%) é significativamente menor, sinalizando possível sobreajuste (*overfitting*), ou seja, o modelo ajusta-se bem aos dados de calibração, mas perde desempenho quando penalizado pela quantidade de variáveis.

Por outro lado, os modelos M2 e M4 destacam-se por apresentarem o melhor equilíbrio entre qualidade de ajuste e simplicidade. O M2, em especial, possui o maior R^2 ajustado (65,88%) e o menor erro absoluto médio (0,10) e erro percentual médio (1,02%), além de apresentar o menor desvio padrão das diferenças (0,09), indicando elevada precisão e estabilidade nas predições. Esses resultados sugerem que o M2 é o modelo com melhor desempenho geral, tanto no ajuste quanto na capacidade preditiva sobre dados novos.

O modelo M4 exibe desempenho bastante similar ao M2, com R^2 = 70,60% e R^2 ajustado = 65,61%, além de apresentar baixos erros absolutos e percentuais (|Dif| Média = 0,10 e Dif.% Média = 1,03%). Seu desvio padrão das diferenças (0,09) também reforça a consistência das predições, tornando-o uma alternativa robusta.

Os modelos M5 e M6 possuem desempenhos intermediários, com R^2 ajustados entre 64,65% e 66,73%, e erros absolutos e percentuais próximos aos observados nos modelos M2 e M4. Contudo, apresentam desvios padrão ligeiramente superiores, sugerindo leve perda de estabilidade nas predições.

Por fim, o modelo M3 apresenta o pior desempenho, com R^2 (64,77%) e R^2 ajustado (60,29%) inferiores aos dos demais. Apesar de manter um erro absoluto médio competitivo (0,12), seus indicadores de ajuste sugerem menor capacidade explicativa dos dados.

5. CONCLUSÕES

O presente estudo teve como objetivo desenvolver modelos preditivos para estimativa do Limite de Umidade Transportável (TML) de minério de ferro, utilizando técnicas de regressão linear múltipla aplicadas a dados físico-químicos, granulométricos e mineralógicos. A proposta buscou fornecer uma ferramenta capaz de antecipar a determinação do TML, contribuindo para a otimização dos processos logísticos, redução de riscos operacionais e melhoria na segurança do transporte marítimo de granéis sólidos.

Os resultados obtidos demonstraram que os modelos desenvolvidos apresentaram desempenhos satisfatórios, com destaque para os modelos M2 e M4, que alcançaram os melhores índices de ajuste e de capacidade preditiva. Ambos apresentaram baixos erros absolutos e percentuais, além de elevados coeficientes de determinação ajustados, indicando boa robustez e estabilidade. O modelo M2, em particular, destacou-se por apresentar o melhor desempenho geral, combinando precisão, parcimônia e elevada capacidade explicativa.

A aplicação prática dos modelos desenvolvidos pode trazer benefícios significativos às operações de mineração e logística, tais como:

- Maior previsibilidade no planejamento operacional;
- Antecipação de ações corretivas no controle de umidade;
- Redução do risco de não conformidade regulatória;
- Aumento da eficiência na gestão da fila de embarques e na composição das pilhas de minério.

Por fim, destaca-se que, embora os modelos desenvolvidos apresentem desempenho satisfatório, recomenda-se que futuras pesquisas explorem abordagens baseadas em técnicas de aprendizado de máquina (*machine learning*) e bancos de dados mais extensos, visando aprimorar ainda mais a capacidade preditiva, a generalização e a aplicabilidade dos modelos em cenários operacionais diversos.

REFERÊNCIAS

Adriano, F. F.; Robles, L. T.; Cutrim, S. S. **Afretamento de navios grande porte no transporte de minério de ferro: estudo de caso da Vale S/A**. *ENAPAD*, Rio de Janeiro, RJ. 2013.

Anglo American. **Sistema Minas-Rio**, Disponível em: https://brasil.angloamerican.com/pt-pt/nossos-negocios/minerio-de-ferro. Acessado em: maio de 2024. 2024.

Astm D5550-14, **Standard test method for specific gravity of soil Solids by Gas Pycnometer**. West Conshohocken: ASTM International. 2014.

Burnham, K. P.; Anderson, D. R. **Model selection and multimodel inference: a practical information-theoretic approach**. 2. ed. New York: Springer. 2002.

Draper, N. R.; Smith, H. **Applied regression analysis**. 3. ed. New York: John Wiley.& Sons. 1998.

Ferreira, R. F. **Modelos para previsão do limite de umidade para transporte marítimo de finos de minério de ferro** – TML. Ouro Preto, MG. 2019.

Ferreira, R. F.; Lima, R. M. F. Relationship between particle size distribution and the PFD80 transportable moisture limit of iron ore fines. **Powder Technology**, v. 414, p. 118072. 2023.

Ferreira, R. F.; Policarpo, D. L. V.; Padula, V. P.; Ferreira, M. T. S. Limite de umidade transportável de minérios de ferro: aspectos regulatórios e técnicos. **Tecnologia em Metalurgia, Materiais e Mineração**, v. 14, n. 1, p. 16-23. 2017.

Holmes, R. J.; Williams, K.; Honeyands, T.; Orense, R.; Roberts, A.; Pender, M.; Mccallum, D.; Krull, T. Bulk commodity characterization for transportable moisture limit determination. P**roceedings of the XXVIII International Mineral Processing Congress**. Canadian Institute of Mining, Metallurgy and Petroleum. 2016.

IBRAM – Instituto Brasileiro De Mineração. Panorama da mineração do Brasil.

Menezes, K. B.; Russo, J. F. C.; Araújo, D. R.; Turrer, H. D. G.; Sales, A. S.; Miranda, A. A.; Silva, M. N. Entendimento sobre as variáveis que influenciam no TML. 2019. In: **20º Simpósio de Mineração - ABM Week**, São Paulo, SP. 2023.

Munro, M. C.; Mohajerani, A. Moisture content limits of Iron Ore Fines to prevent liquefaction during transport: Review and experimental study. **International Journal of Mineral Processing**, v. 148, p. 137–146. 2016.

Paassen, L. A.; Vardon, P. J.; Jeffrey, P., **Cargo liquefaction in bulk carriers: a review**. Yearbook Mijnbouwkundige Vereeniging. 2015

Rose, T. P. **Solid bulk shipping: cargo shift, liquefaction and the transportable moisture limit**. University of Oxford, Oxford. 2014.

Unctad - United Nations Conference on Trade and Development Review of Maritime Transport. UNCTAD/RMT, United Nations. 2022.

Wu, W. et al. Quantitative analysis of liquefaction risk of liquefiable solid bulk cargoes during sea transport. **Ocean Engineering**. v. 258, p. 111751. 2022.

ARTICLE 3

MAXIMUM TRANSPORTABLE MOISTURE LIMIT PREDICTION MODEL FOR IRON ORE, USING MULTIPLE LINEAR REGRESSION

Carlos Leonardo Ramos Póvoa
UENF - Universidade Estadual do Norte Fluminense

Alice Siqueira Sales
Anglo American

RESUMO

Este estudo propõe o desenvolvimento de modelos preditivos para estimar o Limite de Umidade Transportável (TML) de minério de ferro, parâmetro essencial para a segurança do transporte marítimo de granéis sólidos. A metodologia adotada baseou-se em regressão linear múltipla aplicada a um banco de dados com variáveis físico-químicas, granulométricas e mineralógicas. Utilizando técnicas de regressão linear múltipla, foi analisado um conjunto de dados composto por variáveis físico-químicas, granulométricas e mineralógicas, obtidos da produção da Anglo American no Brasil. Foram elaborados seis modelos utilizando diferentes métodos de seleção de variáveis e critérios estatísticos como AICc e BIC. Os modelos foram avaliados com base nos coeficientes de determinação (R^2 e R^2 ajustado) e nos erros absolutos e percentuais entre os valores estimados e observados. Os modelos M2 e M4 destacaram-se por apresentarem alto desempenho preditivo, com menor complexidade e maior robustez. A aplicação prática desses modelos permite otimizar a gestão logística, reduzir riscos operacionais e garantir a conformidade com as normas de segurança marítima, contribuindo para maior eficiência e previsibilidade nas operações de exportação de minério.

Palavras-chave: Transporte marítimo, Limite de Umidade Transportável (TML), modelagem preditiva.

ABSTRACT

This study presents the development of predictive models to estimate the Transportable Moisture Limit (TML) of Iron Ore Fine (IOF) - a critical parameter for ensuring the safety of solid bulk maritime transport. The proposed methodology is based on multiple linear regression applied to a comprehensive dataset comprising physical, chemical, granulometric, and mineralogical variables derived from Anglo American's iron ore production in Brazil. Six models were constructed using various variable selection techniques and statistical criteria, including the corrected Akaike Information Criterion (AICc) and the Bayesian Information Criterion (BIC). Model performance was evaluated through coefficients of determination (R^2 and adjusted R^2) and by analyzing absolute and percentage errors between predicted and observed TML values. Among the models, M2 and M4 demonstrated superior predictive accuracy, lower complexity, and greater robustness. The implementation of these models offers a strategic tool for optimizing logistical planning, mitigating operational risks, and ensuring compliance with international maritime safety standards—ultimately enhancing the efficiency and reliability of iron ore export operations.

Keywords: Maritime transport, Transportable Moisture Limit (TML), predictive modeling.

1. INTRODUCTION

Seaborne shipping is widely accepted as the lowest-cost option for transporting large cargo volumes, especially solid and liquid bulk shipments, due to its competitive freight rates. On the international stage, this logistics option offers greater efficiency, particularly since expanded vessel capacities allow budget-friendly transport of geographically dispersed raw materials. As a result, these inputs become available at competitive prices even in markets far away from their sources (Adriano; Robles; Cutrim, 2013).

According to Holmes et al. (2016), seaborne bulk shipments of ores and minerals have been commonplace for more than a century. The main global solid bulk cargo, with an annual volume of more than a billion tons, iron ore is typically transported as granules, pellets or fines. In 2021, a 3.2% uptick in international maritime trade (totalling 11 billion tons) hinted at a return to growth, after dropping 3.8% in 2020. However, this performance remained slightly below pre-pandemic levels, reflecting persistent public health challenges and global logistics jams stemming from high demands and limited supply capacities. Growth was driven mainly by container shipping, while the gas and dry bulk segments expanded, and crude oil declined (UNCTAD, 2022).

In Brazil, solid bulk cargoes, especially raw ores and concentrates, account for around 33% of annual port throughflows, with 31% being iron ores. These flows are clustered at five terminals in Northeast and Southeast Brazil: Ponta da Madeira in Maranhão State, Tubarão in Espírito Santo State; and Ilha Guaíba, Porto do Açu, and Porto Sudeste, all in Rio de Janeiro State (IBRAM, 2023).

A port terminal in São João da Barra, north of Rio de Janeiro, the Port of Açu is managed by Ferroport, a joint venture between Anglo American and Prumo Logística. In Brazil, Anglo American operates mainly in two sectors: iron ore and ferronickel. Clustered in Conceição do Mato Dentro and Alvorada de Minas, both in Minas Gerais State, the licensed capacity of its iron ore facilities reaches 29.1 million tons a year. Materials are carried by one of the world's largest pipelines (529 kilometres long), crossing 33 municipalities between Minas Gerais and Rio de Janeiro, where they are screened and exported through the Port of Açu (Anglo American, 2023).

Shipping minerals by sea is just one link in the mining production chain. A key stage in the world economy, it is regulated internationally by the International Maritime Organisation (IMO), whose regulatory framework includes laws designed to ensure the safe shipping operations. According to Ferreira et al. (2017), the causes of at least 24 maritime accidents in the past thirty years have been attributed to cargo misfunctions, resulting in more than 177 fatalities.

The Transportable Moisture Limit (TML) is a test that determines the maximum moisture content for seaborne cargoes. This safety parameter can legally halt an ore shipment or even the operations of an entire port if the shipped moisture content exceeds the TML (Ferreira e Lima, 2023).

In this context, this paper presents a predictive model for estimating the TML, using multiple linear regression techniques. Its purpose is not only to reduce the operating risks associated with ore shipments, but also to optimise logistics structures, with gains in

efficiency and lower costs for companies in the sector.

Growing global demands for iron ore require safe and efficient seaborne shipments of this strategic input. The Transportable Moisture Limit (TML) is a critical parameter that directly influences operating safety and logistics efficiency, as well as the costs associated with solid bulk shipments. Accurate definitions are essential for preventing cargo collapses, spills, and accidents with significant impacts on human health and the environment.

However, the intrinsic variability of ore properties, such as chemical and mineralogical composition, and particle size, directly affects TML values, raising an additional challenge for the mining industry. The absence of reliable predictive models that allow this limit to be estimated before loading operations begin is hampering the optimisation of logistics operations, with higher risks of non-compliance with international safety standards.

This paper is consequently justified by the need to develop predictive models able to estimate specific TML values in advance and with greater accuracy, thus supporting decision making and risk mitigation, while streamlining operating efficiency for seaborne iron ore shipments.

2. CONTEXT AND LITERATURE REVIEW

2.1. Regulatory Aspects of Seaborne Shipments

The International Maritime Organisation (IMO) is a testament to the evolution of the global maritime sector over the past few decades. Set up in 1948 as a specialised United Nations agency, the IMO has played a key role in promoting maritime safety, preventing ocean pollution, and facilitating international trade. Particularly noteworthy among its achievements is the International Maritime Solid Bulk Cargo (IMSBC) Code adopted in 1965. Updated regularly since then, the IMSBC Code is under continuous review by its Dangerous Goods, Solid Cargoes and Containers (DSC) Subcommittee. This Code is the main international legislation governing the safety of solid bulk cargo shipments by sea, establishing standardised procedures for each phase of maritime operations (IMO, 2018). An IMO member state, Brazil is represented here by its Navy (Ferreira, 2019).

The practices set forth in the IMSBC Code are designed as recommendations for governments, shipping operators and captains. Their purpose is to spotlight internationally accepted methods for dealing with bulk cargo shipping safety risks, making this the most relevant Code for iron ore bulk carrier safety. The IMSBC Code highlights the risks associated with certain types of bulk cargo shipments; it also provides guidance on procedures that should be adopted; listing typical goods that are shipped in bulk, with advice on their properties and how they should be handled. It also describes several test procedures that must be used to determine the characteristic cargo properties (Holmes et al. 2016). Classification into groups (A: cargo with liquefaction risk; B: cargo with chemical risk; C: cargo with no liquefaction / chemical risks) was established as an integral part of this Code, providing a clear framework for identifying and managing the risks associated with seaborne shipments of bulk cargoes.

2.2. Liquefaction

Liquefaction has been thoroughly studied in the soil mechanics literature, especially in relation to earthquakes. Liquefaction typically occurs in non-cohesive granular soils, which are relatively fine, loose, and saturated with water (Paassen et al. 2015). But this phenomenon has also become a significant factor for seaborne dry bulk cargoes, especially shipments classified in Group A under the IMSBC Code.

Liquefaction occurs when moisture in the material causes it to behave like a liquid, jeopardising vessel stability. One of the main parameters for assessing this risk is the Transportable Moisture Limit (TML), which defines the maximum safe moisture content for cargo shipments. This limit is crucial for ensuring operating safety and avoiding serious accidents, such as vessels sinking due to instability caused by cargo liquefaction (Wu et al. 2022).

When a solid bulk cargo liquefies, its stability is adversely affected, as it transforms into a dense, viscous liquid, whose behaviour lies halfway between a liquid cargo and a dry bulk shipment. This phenomenon results in the formation of an unstable free surface, caused by variations in the centre of gravity of liquids partially filling a tank or hold on a vessel. A liquefied cargo moves freely in the hold, often at a different phase than the vessel: when the ship tilts to one side, the cargo surges in the opposite direction (Ferreira, 2019).

2.3. Transportable Moisture Limit (TML)

The Transportable Moisture Limit (TML) of iron ores is a regulatory reality that is used in everyday operations, both for product planning and in shipping operations, which must take compliance with this safety parameter into account (Ferreira, et al. 2017). In this context, the TML came under the more intensive discussion by the International Maritime Organisation and shipping companies.

It is important to note that the Transportable Moisture Limit varies by cargo type and specific shipping conditions, and must thus be assessed on a case-by-case basis. Moisture control is a key aspect for ensuring the efficiency and safety of bulk cargo shipments, especially for goods such as iron ore fines, which are widely used in the steel and construction industries. Precisely defined and regularly updated, the Transportable Moisture Limit is an essential tool for preventing accidents and protecting human life and the environment during seaborne shipment of dry bulk goods (Munro and Mohajerani, 2016).

The classification of iron ore as a Group A cargo under the IMSBC Code is directly related to its liquefaction risks. Iron ore fines, especially those with higher moisture contents, can potentially behave like liquids when shipped by sea, due to the vessel's vibration and motion forces. Classifying iron ore as a Group A cargo highlights the critical importance of carefully assessing moisture content and following the specific guidelines of the IMSBC Code for ensuring safe transportation, thus providing protection against potential liquefaction hazards during seaborne shipping activities.

The TML cannot set a fixed value for each cargo. As liquefaction risk characteristics vary from one material to another, the TML is determined by the laboratory tests described in the IMSBC Code, with different limits for each tested material. Parameter values must be compared with the cargo's moisture content for shipments on vessels not equipped to load liquefied cargoes (carriers certified

for this type of operation currently form a negligible portion of the world fleet). If the iron ore moisture content is above the TML, the cargo cannot be loaded, despite the legal requirement that all iron ore exporters declare the moisture content and TML of their cargoes, so that they can be shipped safely.

According to Ferreira (2019), the IMSBC Code establishes that the TML must be defined through tests whose procedures are approved by a competent authority. It currently lists five options for laboratory procedures approved by the IMO to define these parameters: the Flow Table Test (included in the Code since 1965) the Penetration Test (added in 1994); the Proctor/Fagerberg Test (added in 1994); the Modified Proctor/Fagerberg test for Iron Ore Fines (PFD80, added in 2015); and the Modified Proctor/Fagerberg Test for Coal (added in 2017). The first three tests are used for mineral concentrates, and only PFD80 is used to define the TML of iron ore fines (IOF), if some prerequisites are met. As determined by the Brazilian Navy, the ores exported by Anglo American must be tested by the Proctor/Fagerberg C70 method, as this material has been processed into pellet feed, which is classified as an Iron Concentrate.

2.4. Moisture Control when Loading

In the mining industry, moisture content control is a common practice for its products, prompted by commercial, legal, and process control requirements. Generally, ores are handled while wet, and the presence of water between solid particles has significant impacts on the behaviour of particulate matter in several aspects, including angle of repose, and apparent specific mass and strength. The influence of water depends on the amount contained by the material. The IMO regulatory framework requires that the TML and moisture content of Group A cargoes be reported to the ship's captain before loading begins. If the cargo has a moisture content equal to or greater than the TML, it should not be loaded, unless the vessel is specially designed for transporting liquefied cargo. Moisture Limits must be reviewed every six months, or whenever there are changes in the cargo (Ferreira et al. 2017).

The International Maritime Organisation issued a Cargo Liquefaction Prevention Protocol that is designed to mitigate these risks. This Protocol is guided by the International Maritime Code for Solid Bulk Cargoes (IMSBC Code), which is the main international regulatory tool for mitigating liquefaction risks for solid bulk cargoes during seaborne shipments. The Code includes several controls designed to ensure that Group A cargoes are loaded with a moisture content lower than necessary to prevent liquefaction (Rose, 2014).

When the moisture content of the material to be shipped exceeds the TML, ore exporting companies face a number of challenges that can negatively impact their operations, finances, and market reputations. Consequently, the possibility of estimating the TML by analysing ore characteristics through models that would provide massive benefits for ore export operations. These benefits include:

• Anticipation of moisture content reduction actions: With prior knowledge of TML variations, moisture content reduction actions can be implemented in advance, thus, avoiding shipment delays. This allows for more effective management of logistics processes, ensuring that delivery deadlines are met.

• Optimisation of material composition

planning: The ability to estimate the TML allows adjustments in planning the composition of materials fed into the beneficiation plant. This results in changes in the composition of materials arriving at the port, saving time and reducing or even eliminating the risk of export operations halting due to excessive moisture. This optimisation can also lead to better resource allocation and lower operating costs.

• More stable operations: The use of TML estimation models provides greater stability for shipping operations. This helps avoid the problems associated with loading cargoes above the permitted Moisture Limit, including delays, shipment cancellations, and damage to exporter reputations. Furthermore, this helps ensure compliance with the regulatory requirements of the enterprise, leading to more efficient and sustainable operations.

• Better vessel line-up planning, with longer ore-heap storage times for materials whose moisture contents are higher than TML values.

The use of models to estimate the TML thus constitutes a strategic approach for mitigating the risks associated with excessive moisture content in ore shipments, with significant benefits in terms of operating efficiency, regulatory compliance, and financial performance.

3. MATERIALS AND METHODS

3.1. Database

The database used in this research project consists of the findings of physical, chemical, and mineralogical analyses of materials produced by the Anglo American Minério de Ferro Brasil mining company. These analyses were performed by laboratories linked to the Minas-Rio Project. The laboratory at the Port of Açu handled the TML and moisture tests, as well as the chemical analyses. Other analyses, such as laser granulometry and mineralogy, were conducted by the laboratory at Conceição do Mato Dentro, where the mine is located.

The database consisted of 86 iron ore samples shipped out of the Port of Açu Iron Ore Terminal between 2020 and 2024. Among them, 63 samples were used to construct the predictive models, with 23 set aside for the test bench, ensuring independent validation of the developed models.

To determine the TML, the Proctor/ Fagerberg C70 methodology was adopted, as required by the Brazilian Navy, which is the competent regulator. All tests were conducted on samples collected and prepared in accordance with the procedures established by the following standard: NBR ISO 3082 Iron ores Sampling and Sample Preparation Procedures.

The chemical analyses were performed using the fused wafer methodology and analysed by the Axios Max Panalytical X-ray fluorescence equipment. The exploratory analysis used the following elements and compounds: Fe, SiO2, Al2O3 and P. According to Menezes *et. al* (2019), these are the elements and compounds with the strongest interactions with the TML. Calcination loss (CL) was performed by gravimetry in a muffle furnace at 1000°C.

The granulometry was assessed through Malvern's Mastersize 3000E equipment, using laser methodology. In this methodology, a laser beam is emitted under the iron ore particle, and the deviation in the light beam when passing through the particle determines the particle volume, then transforming this into a sphere volume and determining the particle diameter. The specific surface analysis was performed using the Blaine manual permeability meter, measuring

the total surface area of the iron ore particles per unit mass or volume. The specific surface was determined experimentally by a technique where air percolates the iron ore sample; this is timed and used to calculate the Blaine value of the material. A high specific surface area means that the ore particles have larger contact areas, related to the amount of fines and contaminants t in the ore.

The TML and Blaine analyses use the density of the solids (ρ) in the material for their calculations, so this characterisation was also entered into the study database. The specific mass of the solids was determined according to ASTM D5550-14 (ASTM, 2014) by the gas pycnometry (helium) method, using a Quantachrome® Ultrapycnometer and Pentapycnometer. Table 1 presents the variables examined in this study.

Table 1 Study Variables

Characterisation PFC70	Variables TML
X-ray fluorescence	Fe SiO2 Al2O3 P
Gravimetry	CL
Mineralogy	Lamellar Hematite Granular Hematite Lobular Hematite Martite Magnetite Goethite Quartz Phyllosilicate Mineral Clay Gibbsite Manganese Other
Laser particle size	<10um d60 d10 Coefficient of Uniformity
Specific Surface	Blaine
Pycnometrics	Density

3.2. Pre-Processing

Before starting the data analysis, a pre-processing stage ensured data quality and integrity. This process included the following steps:

1. Processing missing data: cells containing missing data were duly identified and the respective rows were subsequently removed, in order to ensure data set consistency, with all variables having the same weight in the analysis.

2. Detection and correction of incorrect data: inconsistent data were identified, and whenever incorrect values were obvious (e.g. typing errors), their results were checked against the original laboratory databases, such as the Laboratory Information Management System (LIMS), company production systems (MES Manufacturing Execution Systems), and/or in the equipment and Excel spreadsheets for the area. Once checked, they were corrected in the study database.

3. Definition of the samples considered for the study: Some of the TML analyses were performed in duplicate, in order to assess the replicability of their findings, controlling and ensuring analysis quality. This definition used an assessment of the TML test saturation curves, identifying the best curve for inclusion in the study database. If curves were very similar, the test with the lowest TML value was selected, for greater test safety.

This process allowed for noise removal and data standardisation, resulting in a more robust and reliable set for subsequent analysis. After pre-processing, the variables were assessed in order to remove redundant correlates for explaining the TML variations.

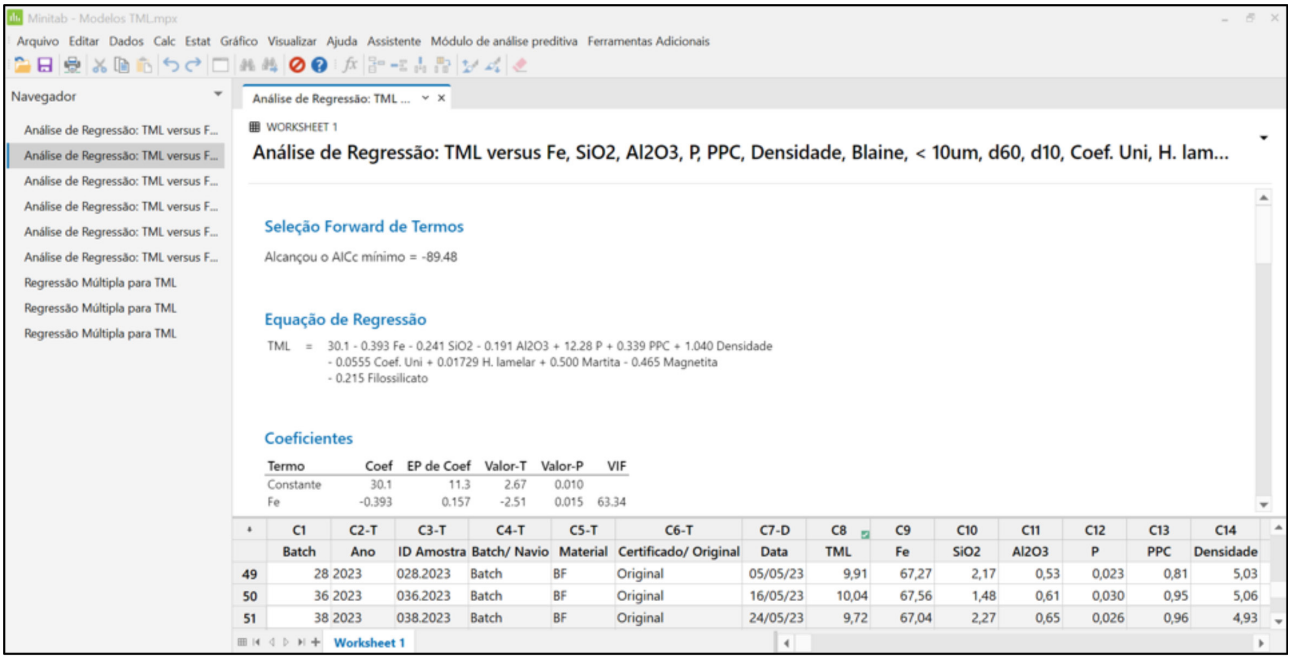
Additionally, correlations were assessed in order to remove unnecessary predictors in the regression model, preventing noise from being added to the calculation.

The general database helped define the need to extract a specific sample, intended exclusively for the model testing and validation stages, and not used for its construction. The samples selected for this test bench had been used over the years in the certification processes of the Transportable Moisture Limit (TML) of the materials amount for shipment, rather than just process study samples. The selection of these samples was justified by the fact that they were additionally submitted to tests in external laboratories, allowing comparisons with the findings of the TML analyses, in endowing the model validation process with greater robustness.

3.3. Predictive Model

To design the predictive model, the Minitab® 19.1 software was used, selecting multiple linear regression for this purpose. This statistical technique is widely used to model relationships between a dependent variable and two or more independent variables. Unlike simple linear regression, which considers only one predictor variable, multiple regression allows the incorporation of several explanatory variables simultaneously, leading to more complex and realistic modelling of phenomena where multiple factors interact. If a regression model fits the relationship between the response and its predictors, then the model accurately predicts the response values with new observations. Figure 1 illustrates the Minitab® interface displaying the results obtained for the M2 model.

Figura 1 – Interface Minitab® Modelo M2



Fonte: autores (2024)

When the number of predictors is high, model selection techniques should be applied before adjusting a regression model with all the variables, such as stepwise or selecting the best subsets. These methods help identify and remove predictors lacking significant associations with the response variable. In general, an effective approach to determining which model has the best predictive performance is to generate multiple models and compare their performances.

Among the metrics used to assess regression models, the coefficient of determination (Rp) stands out. Widely used to analyse predictive performances, Rp expresses the level of fit for the model, indicating the proportion of variability of the dependent variable, explained by the independent variables included in the model. Its values range from 0 to 1, with values close to 0 indicating a low explanatory power for the model, while values close to 1 indicate that the model explains almost all variations observed in the data.

The choice of multiple linear regression is justified by its exploratory nature and ease of interpretation, allowing the identification of significant relationships between physicochemical, granulometric, and mineralogical variables and the TML. However, it is admitted that the relationship between these variables may present nonlinear behaviours, which could be better captured by alternative models, such as techniques based on nonlinear regression, machine learning, or hybrid methods. The use of multiple regression in this work should thus be

seen as an initial step, offering a solid basis for future research that explores more sophisticated methodologies that could enhance the predictive capacity and robustness of these models.

According to Ferreira (2019), the adjusted coefficient of determination may also be a criterion for assessing model performance. The Rρaj is an alternative metric to the Rρ coefficient of determination, when there are several variables penalising the inclusion of regressive variables that are not very explanatory. Unlike the Rρ coefficient of determination, the Rρaj does not necessarily increase every time a new variable is added to the model. It changes only when the addition of the new variable significantly reduces the mean square of the errors, and is therefore more suitable for comparing multiple regression models. The higher the Rρaj, the better the model. This parameter is calculated by Equation 1.

$$R^2_{aj} = 1 - \frac{(n-1)}{n-(k+1)}\left(1 - \left(1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \underline{y})^2}\right)\right) \quad (1)$$

4. FINDINGS AND DISCUSSION

4.1. Descriptive Statistics

The Findings and Discussion section presents an overview of the statistical characteristics of the variables in this predictive model. In order to provide a better understanding of the behaviour of these variables, Table 2 summarises the main descriptive statistics, including mean, median, standard deviation, and coefficient of variation, in addition to the minimum, maximum, first and third quartile values.

The presentation of these indicators is essential for a preliminary understanding of the data profile, allowing assessments of the variability, core trends and dispersion of the variables. This analysis provides important input for identifying patterns, possible outliers, and possible asymmetries, in addition to offering a consistent basis for the construction, adjustment, and validation of the predictive model.

Table 2 shows the variables analysed, along with the sample size (n), the mean, which represents the core trend of the data, and the standard deviation, which expresses the degree of dispersion around the mean. The coefficient of variation (CV) is calculated as the ratio of the standard deviation to the mean, multiplied by 100%, providing a relative data variability measurement. Furthermore, the minimum and maximum values are presented for the first quartile (Q1), which demarcates the 25% lowest values, the third quartile (Q3), which corresponds to the 75% lowest values, and the median, which represents the central point of the distribution, dividing the data set into two equal parts.

The TML dependent variable presents low variability, with a coefficient of variation (CV) of 1.73%, indicating high data homogeneity. The 10.03 mean is equal to the median (10.03), which suggests a symmetrical distribution. Similarly, the Fe content is highly homogeneous (CV = 0.97%), with values clustered around the 67.33 mean, underscoring data consistency. The proximity between Q1 (66.68), median (67.53) and Q3 (67.88) corroborates this uniformity.

On the other hand, the SiOρ content presents high relative variability (CV = 38.27%), indicating wide data dispersion. The interquartile range (Q1 = 1.34 and Q3 = 2.74) and the maximum value of 3.58 suggest the presence of significant variability, which may impact model stability The AlρOρ (CV = 16.52%), P (CV = 20.65%) and CL (CV = 23.62%) variables also present considerable dispersions, indicating that their values fluctuate more widely throughout the sample. The Density variable shows stability (CV = 1.38%), with low dispersion, similar to the behaviour observed in TML and Fe.

Tabela 2: Informações estatísticas - variáveis de teste

Variable	*n*	Average	Sd.Dev	CV	Minimum	Q1	Median	Q3	Maximum
TML	63	10.03	0.17	1.73	9.72	9.88	10.03	10.15	10.42
Fe	63	67.33	0.65	0.97	65.72	66.68	67.53	67.88	68.31
SiO2	63	2.02	0.77	38.27	0.97	1.34	1.78	2.74	3.58
Al2O3	63	0.62	0.10	16.52	0.43	0.53	0.62	0.70	0.88
P	63	0.023	0.005	20.65	0.014	0.020	0.023	0.026	0.035
CL	63	0.79	0.19	23.62	0.42	0.66	0.79	0.91	1.21
Density	63	5.08	0.07	1.38	4.93	5.02	5.08	5.13	5.22
Blaine	63	1038	45	4	969	1011	1035	1053	1213
< 10um	63	9.48	1.21	12.71	7.60	8.70	9.30	10.10	16.00
d60	63	42.84	1.99	4.63	36.55	41.60	43.20	44.40	45.90
d10	63	10.57	1.04	9.84	5.53	10.00	10.70	11.30	12.30
Coef. Uni.	63	4.09	0.47	11.58	3.32	3.82	4.02	4.27	6.93
Lamellar Hem.	63	78.99	1.99	2.52	74.57	77.50	79.27	80.68	82.18
Granular Hem.	63	11.88	1.19	10.02	8.78	11.04	12.03	12.45	14.68
Lobular Hem.	63	2.05	0.64	31.12	0.90	1.58	1.92	2.44	3.48
Martite	63	0.21	0.08	39.95	0.11	0.16	0.18	0.20	0.63
Magnetite	63	0.01	0.04	344.79	0.00	0.00	0.00	0.00	0.17
Goethite	63	3.97	0.93	23.31	2.07	3.20	3.95	4.81	6.21
Quartz	63	1.92	0.53	27.81	1.04	1.50	1.90	2.44	3.02
Phyllosilicate	63	0.35	0.12	32.92	0.10	0.27	0.34	0.43	0.57
Mineral Clay	63	0.05	0.06	113.34	0.00	0.00	0.06	0.09	0.26
Gibbsite	63	0.30	0.14	45.74	0.06	0.18	0.28	0.42	0.56
Manganese	63	0.16	0.04	27.27	0.08	0.14	0.16	0.17	0.39
Other	63	0.11	0.03	27.30	0.05	0.09	0.10	0.11	0.23

Regarding physical characteristics, the Blaine variable has a CV of 4%, indicating moderate stability. The <10pm parameter presents a CV of 12.71%, indicating a more significant variation in the percentage of fine particles. The d60 (CV = 4.63%), d10 (CV = 9.84%) granulometric variables and the Uniformity Coefficient (CV = 11.58%) present dispersions that reflect the heterogeneity of particle size distribution. For the textural phases, lamellar hematite and granular hematite exhibit relative stability (CV = 2.52% and 10.02%, respectively), while lobular hematite presents high variability (CV = 31.12%), which may indicate the non-uniform presence of this component.

Among the ancillary minerals, Magnetite stands out with the highest coefficient of variation (CV = 344.79%), indicating an extremely sporadic and, in most cases, null presence (Q1, Median and Q3 = 0). Similar behaviour, although to a lesser extent, is noted for Mineral Clay (CV = 113.34%) and Gibbsite (CV = 45.74%), whose distributions suggest high heterogeneity in the sample.

Goethite (CV = 23.31%), Quartz (CV = 27.81%), Phyllosilicate (CV = 32.92%) and Manganese (CV = 27.27%) also show relevant dispersions, indicating that their shares vary significantly among the samples. The Others category behaves similarly (CV = 27.30%), underscoring the mineralogical diversity of the analysed material.

In general, it is noted that some variables, especially the major chemical variables (Fe, TML, Density), present high stability, while variables associated with minor phases or traces (Magnetite, Mineral Clay and Gibbsite) exhibit high variability. This behaviour should be considered in the modelling stage, as variables with high dispersions may have relevant impacts on both the explanation of the variability of the dependent variable and the robustness of the predictive model.

4.2. Predictive Models

To build the predictive models, the regression tool available in the Minitab software was used. Different models were developed, applying different methodologies, as described in detail in the following sections.

Models 1 to 6 were developed using the Regression Model Adjustment tool, which offers better control over the selection of variables, as well as analyses of coefficients and statistical assumptions.

4.2.1 MODEL M1

In this model, no specific data pre-processing or processing methodology was applied. The TML variable was defined as the response variable, while all other variables were included as continuous predictors in the model. Thus, all available variables were considered in the adjustment stage, attempting to identify their contribution to explaining the variability of the response variable. This model scores an R^2 = 75.8% and R^2_{adj} = 61.69%.

Regression Equation:

$$\begin{aligned} TML = {} & 769 - 0.350\,Fe - 0.205\,SiO_2 - 0.149\,Al_2O_3 + 13.24\,P + 0.373\,CL + 1.188\,Density - \\ & 0.000228\,Blaine - 0.0779\,\text{"}< 10um\text{"} - 0.0206\,d60 - 0.0251\,d10 + 0.063\,Uni\,Coef - \\ & 7.4\;7.4\,Granular\,Hem. - 7.4\,Lobular\,Hem. - 7.0\,Martite - 7.9\,Magnetite - 7.4 - \\ & Goethite - 7.4\,Quartz - 7.6\,Phyllosilicate - 7.2\,Mineral\,Clay - 7.4\,Gibbsite - 7.4 \\ & Manganese - 6.8\,Others \end{aligned}$$

4.2.2 MODEL M2

The second model uses the Forward variable selection method, based on the corrected Akaike Information Criterion (AICc). This criterion sequentially assesses the inclusion of variables in the model, striving to optimise the balance between fit quality fit and model complexity. The use of AICc is intended to select the model that maximises predictive capacity, while minimising the overfitting risks: a situation where the model over-adjusts to the training data, capturing noise and random variations that adversely affect its generalisation capacity (Draper; Smith, 1998; Burnham; Anderson, 2002). The adoption of this criterion helps ensure that the final model is not only well adjusted to the observed data, but also robust, lean, and applicable to new data sets. This model scored an R^2 = 71.93% and R^2_{adj} = 61.88%.

$$\begin{aligned} TML = {} & 30.1 - 0.393\ Fe - 0.241\ SiO_2 - 0.191\ Al_2O_3 + 12.28\ P + 0.339\ CL + 1.040 \\ & Density - 0.0555\ Coef\ Uni + 0.01729\ Lamellar\ Hem. + 0.500\ Martite - 0.465 \\ & Magnetite - 0.215\ Phyllosilicate \end{aligned}$$

4.2.3 MODEL M3

The third model also used the Forward information criterion method, but through the Bayesian Information Criterion (BIC). This AIC variant assesses the progressive inclusion of variables in the model, seeking those with the best balance between fit and complexity, with heavier penalties for the inclusion of additional variables in the model (Burnham; Anderson, 2002). This model scored an R^2 = 64.77% and R^2_{adj} = 60.29%.

$$\begin{aligned} TML = {} & 7.34 - 0.0742\ Fe + 0.187\ Al_2O_3 + 0.681\ CL + 1.056\ Density + 0.01998 \\ & Lamellar\ Hem. + 0.484\ Martite - 0.704\ Magnetite \end{aligned}$$

4.2.4 MODEL M4

The Stepwise method is an approach used to select variables in multiple regression models. It combines variable inclusion (forward) and exclusion (backward) procedures based on statistical criteria (Draper; Smith, 1998). During the process, variables are inserted or removed from the model sequentially, according to their statistical significance.

One of the main criteria adopted in this procedure is the p-value, which represents the probability of obtaining an outcome as extreme as the observed result, assuming that the null hypothesis is true. In a regression context, the null hypothesis usually tests whether the coefficient of a variable is equal to zero, checking whether this variable has any significant effect on the dependent variable. P values lower than a significance level of 5% indicate that the variable has a statistically significant effect on the model, and is therefore maintained. On the other hand, high p-values suggest that the variable does not contribute significantly and may be discarded from the model.

Using the Stepwise method thus allows the construction of a leaner and more

efficient predictive model, composed only of variables that significantly explain data variability. The model obtained presented a coefficient of determination of Rp = 70.60% and adjusted Rp = 65.61%, indicating that about 65.61% of the variability of the dependent variable is explained by the selected variables, after adjusting for the number of predictors included in the model.

$$TML = 25.06 - 0.318\ Fe - 0.193\ SiO_2 + 12.42\ P + 0.312\ CL + 1.024\ Density - 0.0642\ Coef\ Uni + 0.01653\ Lamellar\ Hem. + 0.461\ Martita - 0.231\ Phyllosilicate$$

4.2.5 MODEL M5

The Forward selection method is a multiple regression modelling technique that starts with an empty model and adds variables one by one, based on statistical significance criteria, such as the p-value. At each stage, the variable that most contributes to the improvement of the model fit is selected and included, while the others are left out (Burnham; Anderson, 2002). This model obtained R^2 = 73.7% and R^2_{adj} = 66.73%.

$$TML = 33.0 - 0.425\ Fe - 0.253\ SiO_2 - 0.278\ Al_2O_3 + 12.64\ P + 0.319\ CL + 1.046\ Density - 0.0577\ Coef\ Uni + 0.0077\ Lamellar\ Hem. - 0.0450\ Lobular\ Hem. + 0.493\ Martite - 0.513\ Magnetite - 0.181\ Phyllosilicate + 0.768\ Others$$

4.2.6 MODEL M6

The Backward elimination method is a technique for selecting variables in multiple regression models that starts with a complete model, including all available variables. From this point on, the process consists of iteratively removing variables with the least impact on explaining data variability, based on statistical criteria such as the p-value. At each step, the less significant variable is eliminated, and the model is reassessed until all remaining variables are statistically significant (Burnham; Anderson, 2002). This model obtained R^2 = 69.21% and R^2_{adj} = 64.65%.

$$TML = 10.79 - 0.0983\ Fe + 17.61\ P + 0.323\ CL + 1.383\ Density - 0.0364"\ <\ 10um" - 0.02156\ d60 - 0.0350\ Granular\ Hem. - 0.419\ Phyllosilicate$$

4.2.7 ASSESSMENT OF THE FINDINGS

In this section, the findings obtained from the developed models are presented. Model performances were assessed through the coefficient of determination (Rρ) and adjusted coefficient of determination (Rρaj) metrics. These indicators measure model quality and their explanatory capacities, considering both the degree of gross adjustment (Rρ) and the weighted adjustment by the number of variables included ($R\rho_{aj}$) in each of them.

Additionally, model performances are analysed from a test database, independent of the one used in the calibration stage. This analysis is intended to validate the effectiveness and robustness of the proposed approaches, ensuring that the models have good predictive capacity and reliability when applied to new datasets. Consequently, this approach strives to ensure a comprehensive performance assessment for these models.

Table 3 presents a summary of the performance of the developed models, based on the coefficient of determination (Rρ) and adjusted coefficient of determination (Rρaj) indicators, which express the proportion of variability of the dependent variable explained by the model, considering, respectively, the gross adjustment and the adjustment penalised by the number of variables.

Moreover, the following table provides a consolidated overview of the means of the absolute differences (Mean Diff) and percentages between the measured and estimated TML values (Mean Diff.%), for each model (M1 to M6), calculated from the 23 samples in the test database. The standard deviations of the differences are also presented, indicating the degree of dispersion of the prediction errors and providing a measurement of the consistency and stability of the estimates generated by the models.

Table 3 Models Overview

Model	R^2	R^2aj	Mean Diff	Mean Diff %	Sd.Dev Median (difference)
M1	75.90	61.69	0.25	2.51	0.13
M2	71.93	65.88	0.10	1.02	0.09
M3	64.77	60.29	0.12	1.19	0.10
M4	70.60	65.61	0.10	1.03	0.09
M5	73.70	66.73	0.11	1.06	0.10
M6	69.21	64.65	0.11	1.12	0.08

Model 1 (M1) has the highest Rp (75.90%), indicating that, in the training set, it provides the best explanation of data variability. However, its Rpaj (61.69%) is significantly lower, signalling a possible overfitting, i.e. the model fits well to the calibration data, but loses performance when penalised by the number of variables.

In contrast, the M2 and M4 models stand out for presenting the best balance between fit quality and simplicity. In particular, M2 has the highest adjusted Rp (65.88%) and the lowest mean absolute error (0.10) and mean percentage error (1.02%), in addition to presenting the lowest standard deviation of the differences (0.09), indicating high accuracy and stability in the predictions. These findings suggest that M2 is the model with the best overall performance, for both fitting and predicting new data.

The M4 model exhibits very similar performance to the M2, with Rp = 70.60% and adjusted Rp = 65.61%, in addition to presenting low absolute and percentage errors Mean Diff. = 0.10 and Mean Diff.% = 1.03%). Its standard deviation of the differences (0.09) also underscores the consistency of these predictions, making it a robust alternative.

The performances of the M5 and M6 models are middling, with adjusted Rp between 64.65% and 66.73%, and absolute and percentage errors close to those observed in the M2 and M4 models. However, they present slightly higher standard deviations, suggesting a minor loss of stability in their predictions.

Finally, the M3 model has the worst performance, with the lowest Rp (64.77%) and adjusted Rp (60.29%). Despite maintaining a competitive mean absolute error (0.12), its adjustment indicators suggest a weaker explanatory capacity for the data.

5. CONCLUSIONS

This study presents predictive models developed for estimating the Transportable Moisture Limit (TML) of iron ore, using multiple linear regression techniques applied to physicochemical, granulometric, and mineralogical data. The proposal sought to provide a tool able to anticipate TML definitions, thus helping optimize logistics processes, with lower operating risks and better safety for seaborne solid bulk shipments.

The findings showed that the developed models presented satisfactory performances, especially the M2 and M4 models, which achieved the best fit and predictive capacity ratings. Both presented low absolute and percentage errors, in addition to high adjusted coefficients of determination, indicating good robustness and stability. In particular, the M2 model stood out for the best overall performance, combining precision, a lean structure, and high explanatory capacity.

The practical application of the developed models could offer significant benefits for mining and logistics operations, such as:

- Greater predictability for planning operations;
- Early corrective actions for moisture control;
- Lower risks of regulatory non-compliance;
- Enhanced shipping line-up management efficiency and better ore-heap compositions.

Finally, it is stressed that, although these models performed satisfactorily, it is recommended that future research projects explore approaches based on machine learning techniques and more extensive databases, seeking further improvements in the predictive capacity, generalisation, and applicability of models in different operating contexts.

REFERENCES

CURIEL, Rafael Prieto; OSPINA, Juan P. The ABC of mobility. **Environment International**, v. 185, ed. 108541, 2024. https://doi.org/10.1016/j.envint.2024.108541.

HENSHER, David A; STANLEY, John. Performance-Based Quality Contracts in Bus Service Provision. Working Paper ITS-WP-02-11. Australia: Institute of Transport Studies, 2002. Available at: : https://ses.library.usyd.edu.au/bitstream/handle/2123/19097/ITLS-WP-02-11.pdf?sequence=1&isAllowed=y . Accessed on: October 4, 2025.

IPEA - INSTITUTO DE PESQUISA ECONÔMICA APLICADA. Efeitos da variação da tarifa e da renda da população sobre a demanda de transporte público coletivo urbano no Brasil. Brasília, DF: IPEA, 2011. 26 p. (Discussion Paper, n. 1595). ISSN 1415-4765. Available at: : https://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=14870. Accessed on: September 4, 2025

IPEA - INSTITUTO DE PESQUISA ECONÔMICA APLICADA. Tendências e desigualdades da mobilidade urbana no Brasil I: o uso do transporte coletivo e individual. Brasília, DF: Ipea, 2021. 50 p. (Discussion Paper, n. 2673). Available at: : http://dx.doi.org/10.38116/td2673. Accessed on: September 4, 2025

JOTZ, Matheus Bello. Análise da variação de demanda pagante no transporte coletivo por ônibus em grandes cidades brasileiras. 2020. 150 f. Dissertação (Mestrado em Economia do Desenvolvimento) - Programa de Pós-Graduação em Economia do Desenvolvimento, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2020. Available at: : http://tede2.pucrs.br/tede2/handle/tede/9443 . Accessed on: August 20, 2025

LEVINSON, H. S. et al. (2003). *Bus Rapid Transit: Volume 2 - Implementation Guidelines*. **Transportation Research Board**. Available at: : https://trb.org/publications/tcrp/tcrp_rpt_90v2.pdf Accessed on: October 4, 2025.

LITMAN, T. (2021). *Evaluating Transportation Equity: Guidance for Incorporating Distributional Impacts in Transport Planning*. Victoria Transport Policy Institute. Available at: : https://www.vtpi.org/equity.pdf . Accessed on: October 4, 2025.

MELO, Marcus André. Governance e reforma do Estado: o paradigma agente x principal. **Revista do Serviço Público**, Rio de Janeiro, Ano 47, v. 120, n. 1, p. 67–82, January/April 1996

METRÔ-SP. Pesquisa Origem-Destino 1977 - Região Metropolitana de São Paulo. Available at: : https://transparencia.metrosp.com.br/dataset/pesquisa-origem-e-destino. Accessed on: September 12, 2025.

NTU - Associação Nacional das Empresas de Transportes Urbanos. Anuário NTU: 2024-2025. Brasília: NTU, 2025.

PRESTON, John; WRETSTRAND, Anders. Workshop 1 report: Regulatory regimes: National and comparative regulation of public transport. **Research in Transportation Economics**, v. 103, March 2024, 101393. Available at: : https://doi.org/10.1016/j.retrec.2023.101393.

SANTANA, Felipe; FERNANDES, Danaê. Novo sistema de transporte coletivo de Cambé/PR: caminhos para a inovação regional?. In: Congresso Internacional de Política Social e Serviço Social: desafios contemporâneos; Seminário Nacional de Território e Gestão de Políticas Sociais; Congresso de Direito à Cidade e Justiça Ambiental. 2024. p. e3490-e3490.

SANTINI, Daniel. Sem Catraca: Da utopia à realidade da Tarifa Zero. Autonomia Literária, 2025.

UITP - International Association of Public Transport. Report on Urban Mobility and Public Transport Trends, 2021.

URRY, John. The 'system' of automobility. Theory, Culture & Society, [S. l.], v. 21, n. 4-5, p. 25-39, out. 2004. DOI: https://doi.org/10.1177/0263276404046059. Available at: : https://journals.sagepub.com/doi/10.1177/0263276404046059. Accessed on: October 4, 2025.

VASCONCELLOS, E. A. Mobilidade urbana e cidadania. São Paulo: Senac, 2011.