



Universidade Federal do Rio de Janeiro

Instituto de Matemática

Modelagem Estatística Multivariada da Relação entre Treinamento e Performance no Ciclismo

Natan Freitas Leite

Rio de Janeiro

2022



Universidade Federal do Rio de Janeiro

Modelagem Estatística Multivariada da Relação entre Treinamento e Performance no Ciclismo

Natan Freitas Leite

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro – UFRJ, como parte dos requisitos necessários para obtenção do título de Mestre em Estatística.

Orientadores: Carlos Tadeu Pagani Zanini e Hugo Tremonte de Carvalho

Rio de Janeiro

2022

CTP - Catalogação na Publicação

L533m Leite, Natan Freitas
Modelagem Estatística Multivariada da Relação
entre Treinamento e Performance no Ciclismo / Natan
Freitas Leite. -- Rio de Janeiro, 2022.
143 f.

Orientador: Carlos Tadeu Pagani Zanini.
Coorientador: Hugo Tremonte de Carvalho.
Dissertação (mestrado) - Universidade Federal do
Rio de Janeiro, Instituto de Matemática, Programa
de Pós-Graduação em Estatística, 2022.

1. Ciclismo. 2. Duração-potência. 3. Redes
neurais. 4. Gradiente estocástico. I. Zanini,
Carlos Tadeu Pagani, orient. II. Carvalho, Hugo
Tremonte de, coorient. III. Título.

Modelagem Estatística Multivariada da Relação entre Treinamento e Performance no Ciclismo

Natan Freitas Leite

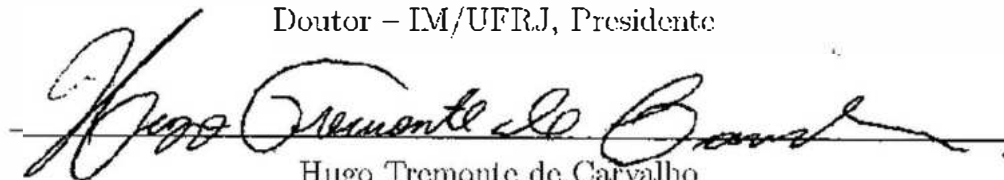
Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro – UFRJ, como parte dos requisitos necessários para obtenção do título de Mestre em Estatística.

Aprovada por:



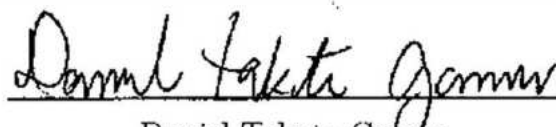
Carlos Tadeu Pagani Zanini

Doutor – IM/UFRJ, Presidente



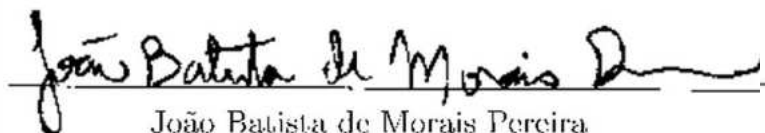
Hugo Tremonte de Carvalho

Doutor – IM/UFRJ



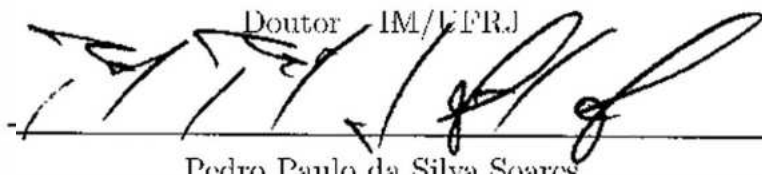
Daniel Takata Gomes

Doutor – ENCE/IBGE



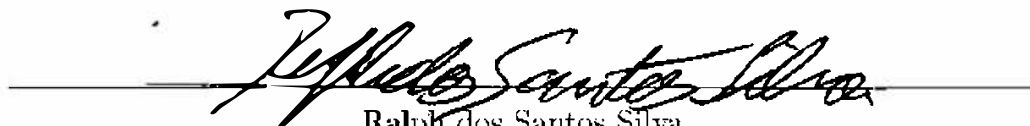
João Batista de Moraes Pereira

Doutor – IM/UFRJ



Pedro Paulo da Silva Soares

Doutor – MFL/UFF



Ralph dos Santos Silva

Doutor – IM/UFRJ (Suplente)

Rio de Janeiro

Maior 2022

Ao ciclismo,

essa atividade maravilhosa;

Aos ciclistas,

que se doam diariamente pelo esporte;

À esperança,

de que as inovações sejam sempre bem-vindas.

AGRADECIMENTOS

A jornada durante este programa de mestrado foi marcada pela pandemia de COVID-19. O cronograma reduzido, o aprendizado à distância e o isolamento social aumentaram o desafio de chegar até o fim. Se tive o desejo pelo conhecimento e pelos estudos, além de ter os privilégios que me fizeram galgar uma caminhada menos sofrida até este ponto, é porque vocês, Evaristo e Janete, meus pais, fizeram tudo por mim e me mostraram os melhores caminhos a seguir.

Os desafios da pandemia não foram fáceis, mas poder dividir com você, Fernanda, minha esposa, foram um privilégio. Além disso, ficam aqui meus agradecimentos por tanto me ouvir explicar sobre probabilidade e estatística de forma atenciosa, sem nunca demonstrar desinteresse independente da complexidade do tema.

À minha família, que junto de meus pais, sempre deram todo o suporte para seguir o caminho dos estudos, ajudando em momentos de difíceis decisões com conselhos construtivos e ponderados, sempre ouvindo com atenção e carinho aos meus sentimentos.

Aos meus amigos, do ensino médio, da faculdade, do mestrado, do trabalho, da vida. Neste período pude também aprender o quanto todos os relacionamentos são importantes e valem ser nutridos e mantidos pelo maior tempo possível.

Agradeço também a todos que trabalham na UFRJ, por manterem esse espaço incrível de aprendizado e troca de conhecimento, ciência e cultura. Em especial, agradeço aos meus orientadores, Carlos e Hugo, por todas as reuniões que culminaram nesta bela dissertação. Finalmente, mas não menos importante, agradeço à banca avaliadora pela disponibilidade de prover o rigor técnico que este trabalho necessita e suas sugestões de melhorias.

“Shut Up, Legs!”

(Jens Voigt)

RESUMO

O ciclismo é um esporte em crescente expansão de adeptos em todo o mundo. O uso de sensores para medição de diversas variáveis durante a prática o torna atrativo para análises e pesquisas, além de fornecer insumo para aplicação de metodologias de inferência estatística apropriadas para grande volume de dados. Esta dissertação tem como objetivo identificar a influência de variáveis de intensidade das atividades do treinamento na variação de parâmetros de performance, por exemplo a potência crítica (*CP – Critical power*), a partir de modelagem multivariada. Diferentemente de outros esportes cuja quantificação da performance pode ser simplificada, a dinâmica do ciclismo requer metodologias mais rebuscadas para mensurar o potencial de um atleta. Os modelos de duração-potência estudados nesta dissertação fornecem parâmetros representativos dos sistemas energéticos. O banco de dados utilizado provém do repositório *GoldenCheetah OpenData Project*, que contém atividades físicas dos usuários do programa *Golden Cheetah*. A modelagem multivariada é realizada a partir de modelos de redes neurais com estimação realizada com métodos de gradientes estocásticos, devido ao grande volume de dados. O ajuste da relação entre treinamento e as diferentes variáveis de performance é realizado simultaneamente.

Palavras-chave: Ciclismo; Duração-potência; Redes neurais; Gradiente estocástico.

ABSTRACT

Cycling is a growing sport with all over the world. The use of sensors to measure several variables during practice makes it attractive for analysis and research, in addition to providing input for the application of statistical inference methodologies appropriate for big data. This thesis aims to identify the influence of training intensity variables on the variation of performance parameters, for example critical power (CP), using multivariate modeling. Unlike other sports whose performance quantification can be simplified, cycling dynamics requires more sophisticated methodologies to measure an athlete's potential. The power-duration models studied in this thesis provide parameters that represents human energy systems. The database comes from the GoldenCheetah OpenData Project repository, which contains sports activities of Golden Cheetah software users. The multivariate modeling is built upon neural networks models with estimation performed by stochastic gradient methods, due to the large volume of data. The relationship between training and the different performance variables is modeled simultaneously.

Keywords: Cycling; Power-duration curve; Neural networks; Stochastic gradient.

LISTA DE FIGURAS

| | |
|--|----|
| Figura 1.1 – Perfil altimétrico da Etapa 11 do Giro d'Italia 2022 | 5 |
| Figura 1.2 – Perfil altimétrico da Etapa 16 do Giro d'Italia 2022 | 5 |
| Figura 2.1 – Exemplo de visualização do modelo <i>perceptron</i> | 13 |
| Figura 2.2 – Funções de ativação usuais..... | 16 |
| Figura 2.3 – Exemplo de visualização de rede neural genérica. | 17 |
| Figura 2.4 – Rede neural para modelo de regressão múltipla. | 25 |
| Figura 2.5 – Rede neural para modelo de regressão logística..... | 25 |
| Figura 3.1 – Exemplo de otimização por gradiente descendente sem e com componente de momento. | 30 |
| Figura 4.1 – Exemplo de arquivo <i>JSON</i> | 36 |
| Figura 4.2 – Exemplos de ciclocomputadores. | 37 |
| Figura 4.3 – Exemplos de sensores de velocidade. | 38 |
| Figura 4.4 – Exemplo de série temporal de elevação para um dos treinamentos de um dos indivíduos na base de dados..... | 39 |
| Figura 4.5 – Exemplos de sensores de cadência. | 40 |
| Figura 4.6 – Exemplo de série temporal de cadência para um dos treinamentos de um dos indivíduos na base de dados..... | 40 |
| Figura 4.7 – Exemplos de sensores de frequência cardíaca..... | 42 |
| Figura 4.8 – Exemplo de série temporal de frequência cardíaca para um dos treinamentos de um dos indivíduos na base de dados..... | 42 |
| Figura 4.9 – Exemplo de série temporal de potência para um dos treinamentos de um dos indivíduos na base de dados..... | 43 |
| Figura 4.10 – Exemplos de medidores de potência. | 44 |
| Figura 4.11 – Exemplo de histograma da cadência..... | 48 |
| Figura 4.12 – Exemplo de histograma da frequência cardíaca. | 48 |
| Figura 4.13 – Exemplo de histograma da potência. | 49 |

| | |
|---|----|
| Figura 5.1 – Exemplo de curva de MMP de uma atividade..... | 59 |
| Figura 5.2 – Exemplo de ajuste do modelo estendido..... | 63 |
| Figura 5.3 – Impacto do parâmetro <i>Paadec</i> no sistema ATP-PCr..... | 63 |
| Figura 5.4 – Impacto do parâmetro <i>CPdel</i> no sistema anaeróbico..... | 64 |
| Figura 5.5 – Impacto do parâmetro <i>CPdel</i> no sistema aeróbico..... | 64 |
| Figura 5.6 – Impacto do parâmetro τ no sistema anaeróbico. | 65 |
| Figura 5.7 – Impacto do parâmetro <i>CPdec</i> no sistema aeróbico..... | 65 |
| Figura 5.8 – Impacto do parâmetro <i>CPdecdel</i> no sistema aeróbico..... | 66 |
| Figura 5.9 – Visualização da rede neural para modelos de duração-potência..... | 68 |
| Figura 5.10 – Exemplo de ajuste do modelo de dois parâmetros. | 78 |
| Figura 5.11 – Exemplo de ajuste dos modelos de três parâmetros e exponencial. | 79 |
| Figura 5.12 – Exemplo de ajuste dos modelos Omni e Estendido..... | 79 |
| Figura 6.1 – Exemplo de janela móvel 5/5. | 83 |
| Figura 6.2 – Exemplo de janela móvel 8/1. | 84 |
| Figura 6.3 – Exemplo de série temporal do parâmetro <i>CP</i> | 85 |
| Figura 6.4 – Exemplo de série temporal do parâmetro <i>W'</i> | 85 |
| Figura 6.5 – Exemplo de série temporal do parâmetro <i>Pmax</i> | 85 |
| Figura 6.6 – Exemplo de série temporal do parâmetro <i>A</i> | 86 |
| Figura 6.7 – Exemplo de série temporal do parâmetro <i>TCPmax</i> | 86 |
| Figura 6.8 – Fluxograma do processo de consolidação da base de dados para modelos treinamento-performance..... | 87 |
| Figura 6.9 – Visualização da arquitetura de rede neural para modelo Banister. | 90 |
| Figura 6.10 – Visualização da arquitetura de rede neural para modelo paralelo. | 91 |
| Figura 6.11 – Visualização da arquitetura de rede neural para modelo Paralelo+. | 92 |

| | |
|--|-----|
| Figura 6.12 – Visualização da arquitetura de rede neural para modelo Completo. | 93 |
| Figura 6.13 – Visualização da partição da base de dados. | 95 |
| Figura 6.14 – <i>Boxplot</i> dos resíduos da base de teste dos modelos Banister, Paralelo, Paralelo+ e Completo, da configuração Individualizado 84/28. | 99 |
| Figura 6.15 – <i>Boxplot</i> dos resíduos da base de teste dos modelos Paralelo, Paralelo+ e Completo, da configuração Generalizado 84/84. | 99 |
| Figura 6.16 – <i>Boxplot</i> dos resíduos da base de teste dos modelos Paralelo, Paralelo+ e Completo, da configuração Generalizado 84/28. | 100 |
| Figura 6.17 – <i>Boxplot</i> dos resíduos da base de teste dos modelos Paralelo, Paralelo+ e Completo, da configuração Individualizado 84/84. | 100 |
| Figura 6.18 – <i>Boxplot</i> dos resíduos da base de teste dos modelos Paralelo, Paralelo+ e Completo, da configuração Individualizado 84/28. | 101 |
| Figura 6.19 – Distribuição empírica dos resíduos da base de treino dos modelos Paralelo, Paralelo+ e Completo, da configuração Individualizado 84/84. | 101 |
| Figura 6.20 – Distribuição empírica dos resíduos da base de validação dos modelos Paralelo, Paralelo+ e Completo, da configuração Individualizado 84/84. | 102 |
| Figura 6.21 – Distribuição empírica dos resíduos da base de teste dos modelos Paralelo, Paralelo+ e Completo, da configuração Individualizado 84/84. | 102 |
| Figura 6.22 – Relação da variável média do IF e ΔCP , no modelo Paralelo+ Generalizado 84/84. | 109 |
| Figura 6.23 – Relação da variável média do VI e $\Delta W'$, no modelo Paralelo+ Generalizado 84/84. | 109 |
| Figura 6.24 – Relação da variável tempo médio em 300+% do CP e $\Delta Pmax$, no modelo Paralelo+ Generalizado 84/84. | 110 |
| Figura 6.25 – Relação da variável tempo médio em 50% do CP e ΔA , no modelo Paralelo Generalizado 84/84. | 110 |

| | |
|--|-----|
| Figura 6.26 – Relação da variável tempo médio em 99% do CP e $\Delta TCPmax$, no modelo Paralelo Generalizado 84/84..... | 111 |
| Figura 6.27 – Relação da variável média da potência média relativa à potência crítica e ΔCP , no modelo Paralelo+ Generalizado 84/28. | 111 |
| Figura 6.28 – Relação da variável tempo médio em 107% do CP e $\Delta W'$, no modelo Paralelo+ Generalizado 84/28..... | 112 |
| Figura 6.29 – Relação da variável proporção de descanso e $\Delta Pmax$, no modelo Paralelo+ Generalizado 84/28..... | 112 |
| Figura 6.30 – Relação da variável tempo médio em 97% do CP e ΔA , no modelo Paralelo+ Generalizado 84/28..... | 113 |
| Figura 6.31 – Relação da variável tempo médio em 98% do CP e $\Delta TCPmax$, no modelo Paralelo+ Generalizado 84/28..... | 113 |

LISTA DE TABELAS

| | |
|---|-----|
| Tabela 5.1 – Valores iniciais dos parâmetros de modelos de duração-potência.. | 70 |
| Tabela 5.2 – Limite de variação dos parâmetros de modelos de duração-potência. | 71 |
| Tabela 5.3 – Análise descritiva do MAE dos modelos de duração-potência com 20% de atividades para ajuste. | 73 |
| Tabela 5.4 – Análise descritiva do RMSE dos modelos de duração-potência com 20% de atividades para ajuste. | 74 |
| Tabela 5.5 – Análise descritiva do MAE dos modelos de duração-potência com 50% de atividades para ajuste. | 74 |
| Tabela 5.6 – Análise descritiva do RMSE dos modelos de duração-potência com 50% de atividades para ajuste. | 75 |
| Tabela 5.7 – Análise descritiva do MAE dos modelos de duração-potência com 80% de atividades para ajuste. | 75 |
| Tabela 5.8 – Análise descritiva do RMSE dos modelos de duração-potência com 80% de atividades para ajuste. | 76 |
| Tabela 5.9 – Análise descritiva final do MAE dos modelos de duração-potência. | 77 |
| Tabela 5.10 – Análise descritiva final do RMSE dos modelos de duração-potência. | 77 |
| Tabela 6.1 – Métricas dos resíduos do ajuste dos modelos de performance nas bases de treino, validação e teste. | 97 |
| Tabela 6.2 – Estimativa do módulo dos coeficientes lineares da última camada, das variáveis com maior impacto no ΔCP , do modelo Paralelo+ Generalizado 84/84. | 104 |

| | |
|--|-----|
| Tabela 6.3 – Estimativa do módulo dos coeficientes lineares da última camada, das variáveis com maior impacto no $\Delta W'$, do modelo Paralelo+ Generalizado 84/84. | 104 |
| Tabela 6.4 – Estimativa do módulo dos coeficientes lineares da última camada, das variáveis com maior impacto no $\Delta Pmax$, do modelo Paralelo+ Generalizado 84/84. | 105 |
| Tabela 6.5 – Estimativa do módulo dos coeficientes lineares da última camada, das variáveis com maior impacto no ΔA , do modelo Paralelo+ Generalizado 84/84. | 105 |
| Tabela 6.6 – Estimativa do módulo dos coeficientes lineares da última camada, das variáveis com maior impacto no $\Delta TCPmax$, do modelo Paralelo+ Generalizado 84/84. | 106 |
| Tabela 6.7 – Estimativa do módulo dos coeficientes lineares da última camada, das variáveis com maior impacto no ΔCP , do modelo Paralelo+ Generalizado 84/28. | 106 |
| Tabela 6.8 – Estimativa do módulo dos coeficientes lineares da última camada, das variáveis com maior impacto no $\Delta W'$, do modelo Paralelo+ Generalizado 84/28. | 107 |
| Tabela 6.9 – Estimativa do módulo dos coeficientes lineares da última camada, das variáveis com maior impacto no $\Delta Pmax$, do modelo Paralelo+ Generalizado 84/28. | 107 |
| Tabela 6.10 – Estimativa do módulo dos coeficientes lineares da última camada, das variáveis com maior impacto no ΔA , do modelo Paralelo+ Generalizado 84/28. | 108 |
| Tabela 6.11 – Estimativa do módulo dos coeficientes lineares da última camada, das variáveis com maior impacto no $\Delta TCPmax$, do modelo Paralelo+ Generalizado 84/28. | 108 |

LISTA DE SIGLAS

| | |
|------|--|
| CP | Potência crítica, do inglês <i>Critical Power</i> |
| FTP | Limiar de potência funcional, do inglês <i>Functional Threshold Power</i> |
| IF | Fator de intensidade, do inglês <i>Intensity Factor</i> |
| MAE | Erro absoluto médio, do inglês <i>Mean Absolute Error</i> |
| MMP | Potência média máxima, do inglês <i>Maximal Mean Power</i> |
| NP | Potência normalizada, do inglês <i>Normalized Power</i> |
| RMSE | Raiz do erro quadrático médio, do inglês <i>Root Mean Squared Error</i> |
| TSS | Pontuação de estresse do treinamento, do inglês <i>Training Stress Score</i> |
| VI | Índice de Variabilidade, do inglês <i>Variability Index</i> |

SUMÁRIO

| | | |
|----------|---|-----------|
| 1 | INTRODUÇÃO..... | 1 |
| 1.1 | MOTIVAÇÃO..... | 2 |
| 1.2 | OBJETIVOS..... | 3 |
| 2 | REDES NEURAIS..... | 8 |
| 2.1 | INTRODUÇÃO..... | 8 |
| 2.2 | NOTAÇÃO | 9 |
| 2.3 | ARQUITETURA DE REDES NEURAIS | 12 |
| 2.4 | APROXIMADORES UNIVERSAIS..... | 18 |
| 2.5 | MODELOS ESTATÍSTICOS COMO REDES NEURAIS | 21 |
| 2.6 | INFERÊNCIA NÃO-PARAMÉTRICA..... | 26 |
| 3 | OTIMIZAÇÃO VIA MÉTODOS DE GRADIENTE..... | 27 |
| 3.1 | GRADIENTE DESCENDENTE..... | 28 |
| 3.2 | GRADIENTE DESCENDENTE ESTOCÁSTICO..... | 28 |
| 3.3 | GRADIENTE DESCENDENTE COM MOMENTO..... | 29 |
| 3.4 | ADAGRAD | 30 |
| 3.5 | ADAM..... | 31 |
| 3.6 | APLICAÇÃO | 32 |
| 4 | BASE DE DADOS | 34 |
| 4.1 | PRÉ-PROCESSAMENTO E VARIÁVEIS DE INTERESSE..... | 45 |
| 5 | QUANTIFICAÇÃO DA PERFORMANCE..... | 55 |
| 5.1 | SISTEMAS ENERGÉTICOS | 56 |
| 5.2 | POTÊNCIA MÉDIA MÁXIMA..... | 57 |

| | | |
|----------|---|------------|
| 5.3 | MODELO DE DOIS PARÂMETROS..... | 59 |
| 5.4 | MODELO DE TRÊS PARÂMETROS | 60 |
| 5.5 | MODELO EXPONENCIAL..... | 60 |
| 5.6 | MODELO OMNI..... | 61 |
| 5.7 | MODELO ESTENDIDO | 61 |
| 5.8 | ESTIMAÇÃO..... | 66 |
| 5.9 | MODELO FINAL | 71 |
| 6 | RELAÇÃO ENTRE TREINAMENTO E PERFORMANCE..... | 81 |
| 6.1 | SÉRIE TEMPORAL DA PERFORMANCE | 82 |
| 6.2 | MODELO BANISTER..... | 88 |
| 6.3 | MODELO PARALELO..... | 90 |
| 6.4 | MODELO PARALELO+..... | 91 |
| 6.5 | MODELO COMPLETO..... | 92 |
| 6.6 | ESTIMAÇÃO..... | 93 |
| 6.7 | RESULTADOS | 96 |
| 7 | CONCLUSÃO | 114 |
| 8 | REFERÊNCIAS BIBLIOGRÁFICAS | 118 |

1 INTRODUÇÃO

O ciclismo é um esporte praticado mundialmente que consiste na utilização de uma bicicleta em diversos terrenos, a depender da modalidade em questão (ATKINSON *et al.*, 2003). A variedade de maneiras possíveis de usufruir desta atividade a torna atrativa para pessoas dos mais diversos grupos, seja com mais adrenalina como *BMX* ou *downhill*, até de maneira mais controlada como cicloturismo ou *audax*.

A nível olímpico, o ciclismo também se mostra um esporte de bastante destaque, contabilizando ao todo 11 modalidades na competição, sendo estas distribuídas entre: ciclismo de estrada, ciclismo de estrada contrarrelógio, *mountain bike cross-country*, *BMX Racing*, *BMX Freestyle*, *Keirin*, *Madison*, *Omnium*, perseguição por equipes, velocidade individual e velocidade por equipes. Além disso, vale mencionar que o ciclismo também é um esporte que contempla diversas medalhas nas Paralimpíadas.

A diversidade do ciclismo faz com que avanços nesse universo tenham um grande impacto na sociedade, mesmo que o foco da maioria dos atletas esteja em resultados competitivos. A prática do esporte com mais recursos técnicos e equipamentos que possam proporcionar melhor desempenho durante a atividade física, seja ela decorrente da mobilidade urbana, a nível de trabalho, lazer ou exercício esportivo, individualmente, em grupo familiar ou em equipe com participantes de ampla faixa etária divididos em categorias, categoriza o ciclismo como uma mobilidade absolutamente democrática envolvendo grande diversidade de eventuais participantes, em pedaladas esporádicas ou regulares.

Entretanto, por conta de sua longa existência, o tradicionalismo é um fator predominante que ocasiona a baixa aceitação de novas tecnologias pelos

praticantes. De forma paradoxal, tal aversão faz com que pioneiros na utilização de melhorias consigam usufruir dos ganhos ocasionados pelo avanço da ciência no esporte (VISÉ, 2019).

No âmbito competitivo, mesmo sendo praticada por atletas amadores, o ciclismo requer capacidade técnica e condicionamento físico, principalmente em provas que envolvam profissionais de alto rendimento que constituem um pelotão de elite, que usufruem de diversos investimentos financeiros através de grandes patrocinadores que financiam provas continentais e de padrão internacional. Por conta do alto nível competitivo que existe em todos os níveis de performance do esporte, as análises dos desempenhos fisiológicos de cada indivíduo serão os aspectos mais importantes e escopo principal desta dissertação.

Uma das vantagens motivadoras para cientistas e pesquisadores do ciclismo é a quantidade de informação disponível para análise e compreensão dos fenômenos que impactam a performance dos atletas envolvidos. Alguns exemplos de mensurações obtidas em tempo real são: velocidade, rotação dos pedais, altitude, distância, frequência cardíaca e potência produzida pelo ciclista. A captação destas variáveis em campo, com ótimo nível de precisão, fornece um material de grande valor para o estudo do esporte.

1.1 MOTIVAÇÃO

O entendimento aprofundado da relação entre treinamento e performance, por parte de treinadores, é essencial para otimizar os ganhos ao trabalhar com atletas de elite (JOBSON *et al.*, 2009). Um dos grandes desafios para o ciclismo é compreender com mais clareza quais componentes da rotina do treinamento dos atletas fazem com que alguns se destaquem em relação a outros.

Os recentes desenvolvimentos na tecnologia de treinamento permitem a utilização de ciclocomputadores pelos atletas, possibilitando mensurar sessões de treino e de competição com um nível de acurácia antes inconcebível (KOSMIDIS; PASSFIELD, 2015). Inicialmente, os ciclocomputadores realizam somente mensuração de distância percorrida, como um hodômetro, utilizando um ímã preso na roda da bicicleta, enviando o sinal a cada rotação que, com o auxílio de uma fórmula pré-determinada, tornava possível definir quantos quilômetros foram percorridos durante uma atividade. Tal coleta gera um desafio aos que se propõem a analisar o volume de dados disponível. Atualmente, os ciclocomputadores são capazes de armazenar informação a cada segundo de uma atividade. Num ano, atletas profissionais podem treinar mais do que 1000 horas, gerando um grande volume de dados armazenados para estudo.

O ciclismo é, geralmente, menos associado com lesões decorrentes do movimento repetitivo da atividade do que a corrida a pé, desconsiderando os acidentes vistos em competições. Portanto, estudos sobre treinamento podem ser mais fáceis de implementar e menos afetados pela desistência de participantes (ATKINSON *et al.*, 2003). Assim, pesquisas que visam a otimização do treinamento para ganho de performance podem ser realizadas mais rapidamente, possibilitando o acúmulo de informações empíricas suficientes para validar a potencial proposta sugerida pelo trabalho que está sendo realizado.

1.2 OBJETIVOS

O objetivo geral desta dissertação é identificar e quantificar a influência de variáveis de intensidade do exercício físico sobre parâmetros de performance, obtidos a partir de modelos de duração-potência, utilizando metodologia multivariada como extensão de Leite (2018). Para isso, serão utilizados dados

anônimos de usuários do programa *GoldenCheetah* (LIVERSEDGE *et al.*, 2020), disponíveis no repositório *GoldenCheetah OpenData Project*¹ adquiridos em 29 de julho de 2021.

A intensidade da atividade física pode ser definida como a variação do fluxo de energia no tempo. Aplicado ao contexto do ciclismo, o fluxo de energia pode ser simplificado como a potência exercida pelo ciclista no sistema de transmissão da bicicleta. Assim, quando um determinado atleta aplica mais potência aos pedais, considerando o terreno constante, mais rápido irá se deslocar. Logo terá ocorrido um aumento na intensidade do exercício, pois mais energia será produzida pelo indivíduo.

Enquanto em outros esportes, como corrida ou natação, podem medir a intensidade que um atleta exerceu durante uma competição a partir do tempo decorrido para realizar uma certa distância, ou mesmo a velocidade instantânea ou média, singularidades do ciclismo tornam a tarefa de medir a intensidade mais complexa. A Figura 1.1 mostra o perfil altimétrico de uma etapa plana da competição Giro d'Italia 2022. Em comparação, a Figura 1.2 mostra o perfil altimétrico de uma etapa de montanha da mesma competição. A peculiaridade do ciclismo vem do fato dos mesmos competidores estarem envolvidos em ambas as etapas, que requerem diferentes capacidades fisiológicas dos atletas. Além disso, a comparação da velocidade média nos dois percursos, ambos de aproximadamente 200 km de distância, não é uma avaliação justa pois o percurso montanhoso terá um valor inferior ao percurso plano, que torna mais fácil pedalar mais rápido.

¹ Disponível em: <https://osf.io/6hfpz/>. Acesso em 29 jul. 2021.

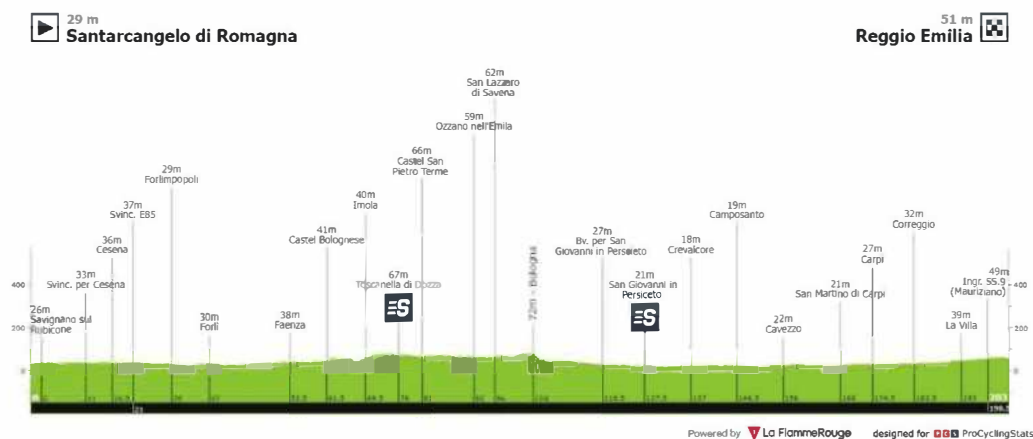


Figura 1.1 – Perfil altimétrico da Etapa 11 do Giro d'Italia 2022

Fonte: ProCyclingStats².



Figura 1.2 – Perfil altimétrico da Etapa 16 do Giro d'Italia 2022

Fonte: ProCyclingStats².

Os modelos de duração-potência, idealmente, teriam a capacidade de predição da performance de forma precisa em todos os níveis de intensidade, utilizando parâmetros que descrevem as capacidades máximas de um determinado indivíduo, além de mensurar o potencial energético dos sistemas fisiológicos que

² Disponível em <<https://www.procyclingstats.com/race/giro-d-italia/2022/gc/stages/all-stage-profiles>>. Acesso em: 16 jul. 2022

influenciam o desempenho de um atleta (PUCHOWICZ; BAKER; CLARKE, 2020).

Nesta linha, um objetivo secundário deste trabalho é a análise da qualidade de ajuste e de previsão de diferentes modelos de duração-potência, que são cruciais para a quantificação da performance do indivíduo ao longo do tempo. Do ponto de vista estatístico, será realizada a estimação da curva de potência média máxima (MMP – *Maximal Mean Power*) (QUOD *et al.*, 2010) utilizando os modelos de duração-potência, além da modelagem da relação entre treinamento (variáveis explicativas) e performance (variáveis resposta) por regressão multivariada não-paramétrica via redes neurais.

O grande desafio deste trabalho é estruturar uma metodologia de análise que consiga lidar com o grande volume de dados disponível, tal como mencionado por Passfield *et al.* (2017). Para isso, os procedimentos de inferência por mínimos quadrados serão realizados via métodos de gradientes estocásticos modificados. Ao permitir a utilização de subconjuntos dos dados para estimação dos gradientes, os métodos estocásticos trazem um ganho substancial em termos de tempo computacional para a quantidade total de dados que serão utilizados.

O trabalho está dividido em 8 capítulos: o Capítulo 1 contém a introdução ao escopo do trabalho e definição de objetivos; o Capítulo 2 tratará das definições metodológicas sobre redes neurais; o Capítulo 3 descreverá sobre a metodologia de otimização via métodos de gradiente; o Capítulo 4 descreverá a base de dados utilizada para que sejam possíveis alcançar os objetivos deste trabalho; o Capítulo 5 descreverá sobre modelos para quantificação da performance de atletas de ciclismo; o Capítulo 6 apresentará os modelos que buscam descrever a relação entre treinamento e performance, trazendo também os resultados da aplicação dos modelos discutidos; o Capítulo 7 irá concluir o trabalho, seus pontos fortes, suas

fragilidades e pesquisas futuras sobre o tema e; finalmente, o Capítulo 8 contém as referências bibliográficas utilizadas nesta dissertação.

2 REDES NEURAIS

2.1 INTRODUÇÃO

As redes neurais são modelos não-paramétricos geralmente aplicados em problemas complexos na expectativa da identificação de padrões inerentes aos dados observados. Anders (1996) mostra que as redes neurais são capazes de resolver o problema de regressão não-linear descrito como

$$\mathbf{y} = F(\mathbf{X}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon},$$

onde F é uma relação desconhecida e $\boldsymbol{\varepsilon}$ representa o erro independente e identicamente distribuído com $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \sigma^2\mathbf{I}$, $E[\boldsymbol{\varepsilon}] = \mathbf{0}$ e $E[\boldsymbol{\varepsilon}|\mathbf{X}] = \mathbf{0}$. Além disso, \mathbf{y} é o vetor coluna, com dimensão $n \times 1$, que representa o valor da variável explicativa para cada uma das n observações contidas nos dados, \mathbf{X} representa a matriz de variáveis explicativas, com dimensão $n \times p$, contendo n observações para cada uma das p variáveis, e por fim $\boldsymbol{\theta}$ é o vetor coluna, com dimensão $p \times 1$, que representa os parâmetros desconhecidos da relação F .

Ainda de acordo com Anders (1996), redes neurais se resumem a modelos de regressão não-linear, porém com a vantagem de que não é necessária uma suposição explícita sobre a forma estrutural da relação desconhecida, uma vez que a informação necessária é extraída puramente dos dados observados. Entretanto, é pressuposto que a complexidade da arquitetura é capaz de aproximar suficientemente a relação desconhecida até certo nível.

Além disso, os modelos de redes neurais possuem uma vasta gama de aplicações. Tradicionalmente, redes neurais são utilizadas para aprendizagem por reforço (*reinforcement learning*) (HENDERSON, 2018), classificação de dados de imagens e vídeos (SHARMA; JAIN; MISHRA, 2018 e REDMON, *et al.* 2016),

servindo até como ferramenta para a solução de problemas de equações diferenciais paramétricas (LI *et al.*, 2020). As aplicações também se estendem para contextos variados, como medicina (AGRAWAL; AGRAWAL, 2015) e finanças (SERRANO *et al.*, 2012), exemplos de problemas de classificação que também são desafios para modelos tradicionais da estatística. Esta classe de modelos bastante flexível será utilizada para estimação da curva de MMP, assim como na modelagem da relação não-paramétrica entre treinamento e performance.

Antes de prosseguir para uma explicação mais detalhada sobre a aplicação de redes neurais como aproximadores universais, será apresentada a notação que será utilizada ao longo desta dissertação. Vale ressaltar que, sendo a compreensão da relação treinamento-performance no ciclismo o objetivo desta dissertação, e que as relações biológicas que ocorrem no corpo humano são naturalmente complexas, a aplicação de redes neurais se torna ainda mais apropriada neste escopo.

2.2 NOTAÇÃO

A definição de uma notação clara e coerente é de grande importância, para garantir a compreensão do complexo tema que é abordado ao longo da dissertação, principalmente no que tange ao campo da matemática e estatística, na forma de equações e diagramas.

O contexto de regressão busca relacionar variáveis explicativas com variáveis respostas. Sendo assim, as variáveis explicativas serão representadas por uma matriz de dados denotada por \mathbf{X} , onde cada valor $x_{i,j}$ representa o valor observado da i -ésima observação da j -ésima variável explicativa.

De forma mais explícita,

$$\mathbf{X} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \mathbf{x}_3 \quad \dots \quad \mathbf{x}_{d_X}]_{n_X \times d_X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1d_X} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2d_X} \\ x_{31} & x_{32} & x_{33} & \dots & x_{3d_X} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n_X1} & x_{n_X2} & x_{n_X3} & \dots & x_{n_Xd_X} \end{bmatrix}_{n_X \times d_X},$$

onde \mathbf{x}_j é o vetor coluna que contém os n_X valores observados da j -ésima variável explicativa. Em geral, esta dissertação considerará caracteres minúsculos sem negrito como variáveis escalares, caracteres minúsculos com negrito como variáveis vetoriais, sejam vetores coluna ou linha, e caracteres maiúsculos com negrito como variáveis matriciais. No caso de algarismos em negrito, estes representarão a replicação do número para todas as entradas da estrutura que estão associados, por exemplo $\mathbf{x}_j = \mathbf{0}$ indica que o vetor coluna \mathbf{x}_j é um vetor com dimensão $n_X \times 1$ onde todas as entradas são zero.

As variáveis respostas serão representadas de duas formas. No contexto de regressões univariadas, a variável resposta será denotada por um vetor coluna, da forma

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{n_Y} \end{bmatrix}_{n_Y \times 1},$$

onde y_i representa o valor da variável resposta y para a i -ésima observação contida nos dados coletados. No contexto multivariado são consideradas mais do que uma variável resposta, então a representação se assemelha àquela das variáveis explicativas, com estrutura matricial, sendo

$$\mathbf{Y} = [\mathbf{y}_1 \quad \mathbf{y}_2 \quad \mathbf{y}_3 \quad \dots \quad \mathbf{y}_{d_Y}]_{n_Y \times d_Y} = \begin{bmatrix} y_{11} & y_{12} & y_{13} & \dots & y_{1d_Y} \\ y_{21} & y_{22} & y_{23} & \dots & y_{2d_Y} \\ y_{31} & y_{32} & y_{33} & \dots & y_{3d_Y} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{n_Y1} & y_{n_Y2} & y_{n_Y3} & \dots & y_{n_Yd_Y} \end{bmatrix}_{n_Y \times d_Y},$$

onde a variável \mathbf{Y} é composta por d_Y variáveis explicativas da forma univariada.

A representação utilizando caracteres gregos segue o mesmo raciocínio. No contexto univariado, serão utilizados caracteres minúsculos para denotar, por exemplo o vetor de parâmetros desconhecidos ou o vetor de erros. Considerando a regressão linear

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon},$$

$\boldsymbol{\theta}$ é o vetor coluna dos coeficientes lineares associados às variáveis explicativas, ao passo que $\boldsymbol{\varepsilon}$ é o vetor coluna dos erros aleatórios. De forma mais explícita, temos

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \vdots \\ \theta_{d_X} \end{bmatrix}_{d_X \times 1}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_{n_Y} \end{bmatrix}_{n_Y \times 1}.$$

A utilização destas quantidades, no contexto multivariado, será representada como

$$\boldsymbol{\Theta} = [\boldsymbol{\theta}_1 \quad \boldsymbol{\theta}_2 \quad \boldsymbol{\theta}_3 \quad \dots \quad \boldsymbol{\theta}_{d_X}]_{d_X \times d_Y},$$

$$\mathbf{E} = [\boldsymbol{\varepsilon}_1 \quad \boldsymbol{\varepsilon}_2 \quad \boldsymbol{\varepsilon}_3 \quad \dots \quad \boldsymbol{\varepsilon}_{d_Y}]_{n_Y \times d_Y},$$

onde ambas as matrizes são compostas pela junção das representações univariadas.

A representação de camadas como variáveis explicativas, que será posteriormente mencionada, mas de antemão é relevante explicitar que sua representação se dará de maneira similar à \mathbf{X} , porém as camadas ocultas (*hidden layers*) serão denotadas por

$$\mathbf{H}_h = [\mathbf{h}_{h1} \quad \mathbf{h}_{h2} \quad \mathbf{h}_{h3} \quad \dots \quad \mathbf{h}_{hd_h}]_{n_h \times d_h},$$

onde \mathbf{h}_{hj} é o vetor coluna que contém os valores das n_h observações da j -ésima variável que compõe a h -ésima camada oculta. Os conceitos de camada e camada oculta serão apresentados ao longo deste capítulo.

A definição da notação é importante para que o entendimento da metodologia utilizada nesta dissertação ocorra de forma clara e objetiva. Também se tratando de notações, porém específicas da metodologia utilizada nesta dissertação, a próxima Seção tratará de introduzir alguns conceitos habituais no tópico de redes neurais, por exemplo camadas ocultas, que já foi comentada nesta Seção 2.1.

2.3 ARQUITETURA DE REDES NEURAI

As redes neurais são modelos que possuem capacidade de resolver problemas de regressão não-linear, especialmente considerando a complexidade de sua arquitetura para capturar os comportamentos determinantes da relação à qual deseja-se aproximar.

A arquitetura de uma rede neural define como as variáveis explicativas podem interagir entre si, os parâmetros desconhecidos da rede e, por fim, como será aproximada a relação entre as variáveis explicativas e as variáveis respostas. Para auxiliar a introdução dos conceitos que circundam a arquitetura de redes neurais, será utilizado como exemplo inicial o modelo *perceptron*, originalmente proposto por Rosenblatt (1958).

O modelo *perceptron* tem o objetivo de realizar uma classificação binária, a partir de um conjunto de variáveis explicativas. Sua definição pode ser dada por

$$y_i = f(\mathbf{x}_i|\boldsymbol{\theta}, \beta) = \begin{cases} 1, & \text{se } \mathbf{x}_i\boldsymbol{\theta} + \beta > 0 \\ 0, & \text{caso contrário} \end{cases}$$

onde \mathbf{x}_i , $i = 1, \dots, n_X$ representa o vetor linha contendo os valores de cada uma das d_X variáveis explicativas da i -ésima observação, $\boldsymbol{\theta}$ é o vetor coluna de coeficientes lineares desconhecidos e β é um escalar desconhecido que representa o viés.

A definição deste modelo pode ser dada a partir da equação mencionada de forma simples, entretanto uma outra forma de visualização é comum no contexto de redes neurais, utilizando elementos de grafos para auxiliar a compreensão principalmente em modelos mais complexos.

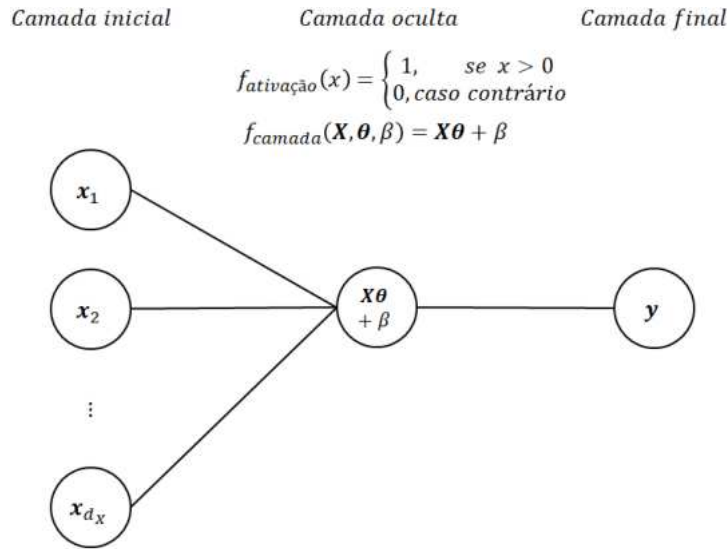


Figura 2.1 – Exemplo de visualização do modelo *perceptron*.

A Figura 2.1 apresenta a representação gráfica do modelo *perceptron*, exemplificando a arquitetura de redes neurais. Um dos elementos presentes na arquitetura são os neurônios, representados por círculos, que por sua vez representam variáveis no modelo. Os neurônios são agrupados verticalmente, nos diagramas apresentados nessa dissertação, formando camadas. As variáveis que compõem cada camada podem ser agrupadas, de acordo com a função camada (f_{camada}) e função ativação ($f_{ativação}$), para se tornarem outras variáveis na camada seguinte. Neste caso, as arestas indicam quais variáveis são combinadas para formar uma outra variável.

A primeira camada presente no diagrama é a camada inicial (*input layer*) que é composta pelas variáveis explicativas do modelo, no diagrama representadas por nós de cada um dos \mathbf{x}_j vetores coluna, $j = 1, 2, \dots, d_X$.

As camadas intermediárias da arquitetura são denominadas camadas ocultas (*hidden layers*) e podem realizar transformações das variáveis explicativas, gerando assim novas variáveis. As transformações são realizadas a partir da aplicação de duas funções: f_{camada} que atua sobre o conjunto de vetores coluna que irá compor a nova variável, de acordo com as arestas que indicam quais variáveis da camada anterior irão compor uma determinada variável na camada atual; e $f_{ativação}$ que atua sobre cada observação (*pointwise*), usualmente sendo utilizada para garantir o ajuste de relações não-lineares.

Por fim, a camada final (*output layer*) representa a variável resposta, no contexto univariado, ou variáveis respostas, no contexto multivariado. Nesta dissertação a relação entre a última camada oculta do modelo e a camada final será direta, ou seja, cada neurônio da última camada oculta terá uma aresta que conecta a somente uma variável resposta, sem a aplicação de nenhuma f_{camada} ou $f_{ativação}$.

Em geral, as funções da camada f_{camada} são definidas como funções afim, do tipo $\mathbf{X}\boldsymbol{\theta} + \boldsymbol{\beta}$, entretanto para representar alguns modelos tradicionais na literatura de modelos de duração-potência e da relação de treinamento e performance, é necessário adotar essa generalização na descrição da rede neural.

As funções de ativação $f_{ativação}$ podem ser escolhidas de acordo com comportamento desejado que as variáveis podem apresentar, entretanto são usualmente escolhidas dentre algumas opções como: função sigmoide, definida como

$$f_{sigmoide}(x): \mathbb{R} \mapsto (0,1),$$

$$f_{sigmoide}(x) = \frac{1}{1 + e^{-x}};$$

função *rectified linear unit* (ReLU) (NAIR; HINTON, 2010), definida como

$$f_{ReLU}(x): \mathbb{R} \mapsto [0, +\infty),$$

$$f_{ReLU}(x) = \max(0, x);$$

função *leaky ReLU* (*LReLU*) (MAAS; HANNUN; NG, 2013), definida como

$$f_{LReLU_\alpha}(x): \mathbb{R} \mapsto \mathbb{R},$$

$$f_{LReLU_\alpha}(x) = \max(\alpha x, x), \quad 0 < \alpha < 1;$$

função *exponential linear unit* (*ELU*) (CLEVERT; UNTERTHINER; HOCHREITER, 2015), definida como

$$f_{ELU_\alpha}(x): \mathbb{R} \mapsto (-1, +\infty),$$

$$f_{ELU_\alpha}(x) = \alpha(e^x - 1)I(x \leq 0) + xI(x > 0);$$

função *Swish* (RAMACHANDRAN; ZOPH; LE, 2017), definida como

$$f_{Swish}(x) = x \cdot f_{sigmoide}(x);$$

função tangente hiperbólica, definida como

$$f_{tanh}(x): \mathbb{R} \mapsto (-1, 1),$$

$$f_{tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}};$$

ou função identidade, definida como

$$f_{identidade}(x): \mathbb{R} \mapsto \mathbb{R},$$

$$f_{identidade}(x) = x.$$

A ilustração de cada uma das funções de ativação descritas está na Figura 2.2.

O modelo *perceptron*, como apresentado na Figura 2.1 somente possui uma única camada oculta, por isso este modelo é conhecido como modelo de uma única camada (*single layer*). As redes neurais podem conter quantas camadas ocultas sejam de interesse do pesquisador, entretanto com mais de uma camada os modelos são conhecidos como modelo de múltiplas camadas (*multilayer*).

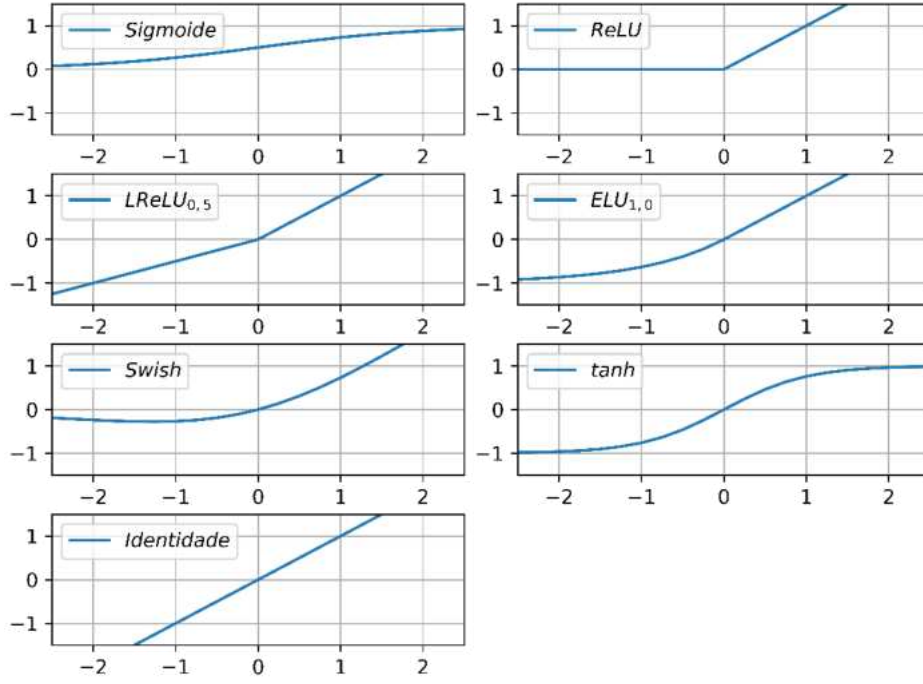


Figura 2.2 – Funções de ativação usuais.

A extensão para a representação de modelos multivariados de múltiplas camadas, mais complexos do que o modelo *perceptron*, será introduzida a seguir, dado que a arquitetura da rede é definida pelo pesquisador, sendo composta pela escolha da quantidade de camadas ocultas, as funções de cada camada e funções de ativação, além da quantidade de neurônios por camada. Uma visualização genérica está representada na Figura 2.3.

Neste caso genérico, uma camada oculta pode ser representada por uma matriz, por exemplo a h -ésima camada oculta representada pela matriz \mathbf{H}_h , definida como

$$\begin{aligned}
 \mathbf{H}_h &= [\mathbf{h}_{h1} \quad \mathbf{h}_{h2} \quad \mathbf{h}_{h3} \quad \dots \quad \mathbf{h}_{hd_h}]_{n_h \times d_h} \\
 &= \begin{bmatrix} h_{1h1} & h_{1h2} & h_{1h3} & \dots & h_{1hd_h} \\ h_{2h1} & h_{2h2} & h_{2h3} & \dots & h_{2hd_h} \\ h_{3h1} & h_{3h2} & h_{3h3} & \dots & h_{3hd_h} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{n_h h1} & h_{n_h h2} & h_{n_h h3} & \dots & h_{n_h hd_h} \end{bmatrix}_{n_h \times d_h},
 \end{aligned}$$

onde cada vetor-coluna \mathbf{h}_{hi} representa o i -ésimo neurônio da h -ésima camada oculta da rede neural, n_h representa o número de observações na matriz \mathbf{H}_h e d_h representa o número total de neurônios na camada h . A n_H -ésima camada oculta, representada pela matriz \mathbf{H}_{n_H} , possui o mesmo número de neurônios da camada final, de variáveis resposta.

Dessa forma, cada camada h possui uma função da camada, denotada por f_{camada_h} , e uma função de ativação, denotada por $f_{ativação_h}$, com o propósito de realizar a transformação dos dados da camada anterior, por exemplo

$$\mathbf{h}_{11} = f_{ativação_1}(f_{camada_1}(\mathbf{X}, \boldsymbol{\theta}_1)),$$

onde $\boldsymbol{\theta}_1$ são os parâmetros desconhecidos da f_{camada_1} , componentes da matriz paramétrica $\boldsymbol{\Theta}$. De forma genérica, considerando um específico neurônio k da camada oculta h , podemos definir como

$$\mathbf{h}_{hk} = f_{ativação_h}(f_{camada_h}(\mathbf{H}_{h-1}, \boldsymbol{\theta}_h)),$$

onde $\boldsymbol{\theta}_h$ são os parâmetros desconhecidos da f_{camada_h} . Por fim, a relação entre a n_H -ésima camada oculta \mathbf{H}_{n_H} e a matriz de variáveis respostas \mathbf{Y} é direta, da forma $\mathbf{Y} = \mathbf{H}_{n_H}$.

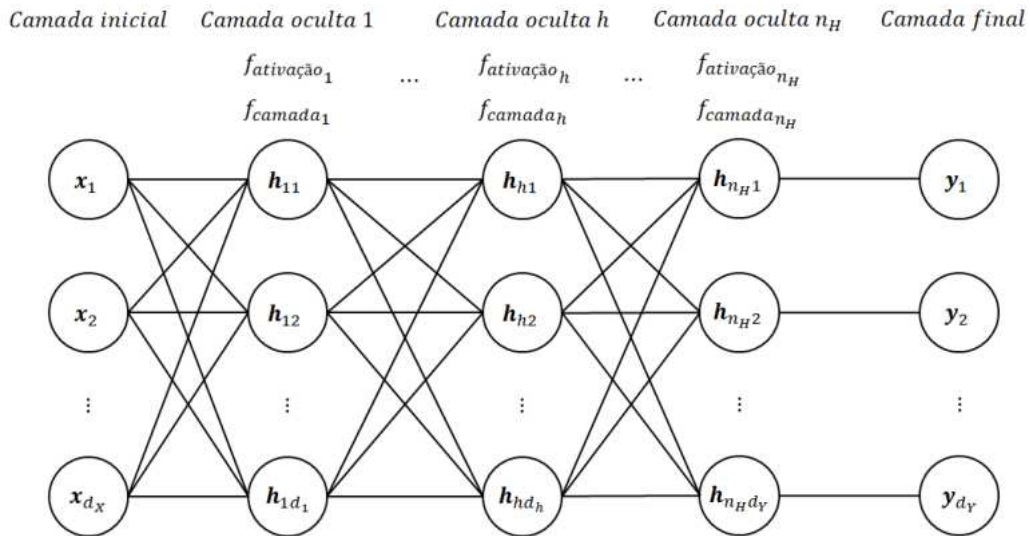


Figura 2.3 – Exemplo de visualização de rede neural genérica.

A próxima Seção retomará a discussão de como as redes neurais podem ser considerados aproximadores universais, como já foi mencionado na Seção 2.1, utilizando a notação e os conceitos que compõem a arquitetura de redes neurais anteriormente descritos.

2.4 APROXIMADORES UNIVERSAIS

As redes neurais foram consideradas nesta dissertação por sua capacidade de aproximar arbitrariamente bem qualquer função mensurável, dado que a arquitetura da rede neural seja suficientemente complexa, como explicitado por Hornik, Stinchcombe e White (1989) e Anders (1996).

O artigo de Hornik, Stinchcombe e White (1989) apresenta teoremas que demonstram que uma rede neural com uma única camada, contendo somente uma variável resposta, consegue aproximar qualquer função mensurável arbitrariamente bem, independente da função de ativação, dimensão das variáveis explicativas e do espaço de probabilidade. Entretanto, o mesmo artigo expande o resultado para redes neurais com mais de uma camada (*multilayer*) e, também, no contexto multivariado, ou seja, com mais do que uma variável resposta.

A demonstração da capacidade que rotula redes neurais como aproximadores universais, que será aqui apresentada, foi originalmente apresentada por Hornik, Stinchcombe e White (1989). Inicialmente, são necessárias algumas definições.

Seja $A^r: \mathbb{R}^r \mapsto \mathbb{R}, \forall r \in \mathbb{N}$ o conjunto de todas as funções afins, ou seja, o conjunto de todas as funções da forma $A(\mathbf{x}) = \mathbf{x} \cdot \boldsymbol{\theta} + \beta$, onde \mathbf{x} e $\boldsymbol{\theta}$ são vetores em \mathbb{R}^r e $\beta \in \mathbb{R}$ é um escalar. No contexto de redes neurais, é bastante comum a utilização de funções afins como funções da camada (f_{camada}).

Seja $\Sigma^r(G)$ a classe de funções

$$\left\{ f: \mathbb{R}^r \mapsto \mathbb{R}: f(x) = \sum_{j=1}^q \beta_j G(A_j(x)), x \in \mathbb{R}^r, \beta_j \in \mathbb{R}, A_j \in A^r, q = 1, 2, \dots \right\},$$

onde $G: \mathbb{R} \mapsto \mathbb{R}$ é qualquer função Borel-mensurável e $r \in \mathbb{N}$.

Seja também $\Sigma \Pi^r(G)$ a classe de funções

$$\left\{ f: \mathbb{R}^r \mapsto \mathbb{R}: f(x) = \sum_{j=1}^q \beta_j \prod_{k=1}^{l_j} G(A_{jk}(x)), x \in \mathbb{R}^r, \beta_j \in \mathbb{R}, A_{jk} \in A^r, l_j \in \mathbb{N}, q = 1, 2, \dots \right\},$$

onde $G: \mathbb{R} \mapsto \mathbb{R}$ é qualquer função Borel-mensurável e $r \in \mathbb{N}$. Nesta definição, redes neurais da forma Σ são um caso especial de redes $\Sigma \Pi$ em que $l_j = 1, \forall j$.

Seja C^r o conjunto de funções contínuas $f: \mathbb{R}^r \mapsto \mathbb{R}$ e seja M^r o conjunto de todas as funções $f: \mathbb{R}^r \mapsto \mathbb{R}$. A σ -álgebra de Borel em \mathbb{R}^r será denotada por B^r . Assim, as classes de funções $\Sigma^r(G)$ e $\Sigma \Pi^r(G)$ pertencem a M^r para qualquer função G Borel-mensurável. Quando G é contínua, então $\Sigma^r(G)$ e $\Sigma \Pi^r(G)$ pertencem a C^r , sendo C^r um subconjunto de M^r , que contém virtualmente todas as funções relevantes em aplicações. A proximidade de funções f e g pertencentes a C^r ou M^r pode ser dada por uma métrica, denotada por ρ .

O subconjunto S de um espaço de medida (X, ρ) é ρ -denso num subconjunto T se, para todo $\varepsilon > 0$ e todo $t \in T$, existe um $s \in S$ de forma que $\rho(s, t) < \varepsilon$. Ou seja, um elemento de S pode aproximar um elemento de T a qualquer nível desejado de acurácia. Conforme o teorema originalmente apresentado por Hornik, Stinchcombe e White (1989), T e X correspondem a C^r ou M^r , S corresponde a $\Sigma^r(G)$ e $\Sigma \Pi^r(G)$ para escolhas específicas de G e ρ é escolhido de forma conveniente.

A última definição necessária para apresentar o teorema diz que o subconjunto S de C^r é uniformemente denso no conjunto compacto em C^r se, para todo subconjunto compacto $K \subset \mathbb{R}^r$, S é ρ_K -denso em C^r , onde para $f, g \in C^r$ então $\rho_K(f, g) \equiv \sup_{\mathbf{x} \in K} |f(\mathbf{x}) - g(\mathbf{x})|$. A sequência de funções $\{f_n\}$ converge então para uma função f uniformemente no conjunto compacto se, para todo subconjunto compacto $K \subset \mathbb{R}^r$, $\rho_K(f_n, f) \rightarrow 0$ a medida que $n \rightarrow \infty$.

A partir da aplicação do teorema de Stone-Weierstrass, que garante que A , o conjunto de funções reais contínuas no conjunto compacto K , é ρ_K -denso no espaço das funções reais contínuas de K , sendo $G: \mathbb{R} \mapsto \mathbb{R}$ qualquer função contínua não-constante, então $\Sigma \Pi^r(G)$ é uniformemente densa no conjunto compacto em C^r . Ou seja, redes neurais da forma $\Sigma \Pi$ tem a capacidade de aproximar qualquer função real contínua num conjunto compacto. As suposições do conjunto compacto são mantidas sempre que os valores das variáveis explicativas são limitados ($\mathbf{x} \in K$).

Como mencionado por Hornik, Stinchcombe e White (1989), uma característica interessante deste teorema é que a função de ativação G pode ser qualquer função contínua não-constante. O resultado do teorema 2.1 do artigo mencionado é fundamental para a formulação de outros teoremas e corolários, em especial o corolário 2.7 que afirma que redes neurais multivariadas com múltiplas camadas da forma $\Sigma_l^{r,s}(\psi)$ aproximam funções em $C^{r,s}$ e $M^{r,s}$, quando $l \geq 2$, concluindo então que redes neurais $\Sigma_l^{r,s}$ são aproximadores universais de funções avaliadas sobre vetores, incluindo redes da forma $\Sigma \Pi_l^{r,s}$ pois esta classe contém a classe $\Sigma_l^{r,s}$ como um caso especial.

Por fim, é relevante mencionar também que a utilização de mais de uma camada reduz a quantidade de parâmetros necessários para atingir a qualidade desejada para a aproximação (MHASKAR; LIAO; POGGIO, 2017). Isto garante

um modelo mais parcimonioso e menos custoso computacionalmente para a estimação de todos os parâmetros desconhecidos.

A flexibilidade dos modelos de redes neurais, além de possuírem a capacidade de serem aproximadores universais, também podem representar modelos estatísticos, cujos exemplos serão mencionados na Seção a seguir.

2.5 MODELOS ESTATÍSTICOS COMO REDES NEURAI

Neste trabalho, consideraremos modelos de regressão não-linear escritos da seguinte forma. Seja \mathbf{y}_i o vetor linha de variáveis respostas para a i -ésima observação, com vetor de médias

$$\mathbb{E}(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\Theta}) = \boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\Theta}),$$

onde \mathbf{x}_i é o vetor linha de variáveis explicativas para a i -ésima observação, $\boldsymbol{\Theta}$ é matriz de parâmetros de interesse a serem estimados, e, finalmente, $\boldsymbol{\mu}$ é a relação desconhecida, parametrizada por $\boldsymbol{\Theta}$, entre \mathbf{x}_i e o vetor de médias de \mathbf{y}_i . A distribuição de \mathbf{y}_i pode ainda depender de outros parâmetros $\boldsymbol{\Psi}$ que não sejam componentes de $\boldsymbol{\Theta}$, como será mostrado em exemplos mais adiante.

As escolhas comuns para $\boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\Theta})$ em contextos de modelos de regressão são da forma $\boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\Theta}) = g(\boldsymbol{\eta}(\mathbf{x}_i, \boldsymbol{\Theta}))$, onde g representa uma função de ligação entre o preditor, $\boldsymbol{\eta}(\mathbf{x}_i, \boldsymbol{\Theta})$, e o vetor de médias $\mathbb{E}(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\Theta})$. Do ponto de vista de estimação, a dissertação irá se restringir a formas funcionais $(\boldsymbol{\Theta}, \boldsymbol{\Psi}) \mapsto P(\mathbf{y}_i | \boldsymbol{\Theta}, \boldsymbol{\Psi})$ que sejam diferenciáveis a menos de um conjunto de medida de Lebesgue nula (a fim de que seja possível calcular os gradientes de $P(\mathbf{y}_i | \boldsymbol{\Theta}, \boldsymbol{\Psi})$) e contínuas. Na estatística, diversos modelos podem ser construídos utilizando esta mesma estrutura, dentre os quais destacamos os exemplos a seguir.

Exemplo 1:

O modelo de regressão linear múltipla com variável resposta tendo distribuição Normal pode ser definido como

$$y_i \stackrel{\text{ind}}{\sim} N(\mu(\mathbf{x}_i, \boldsymbol{\theta}), \sigma^2),$$

$$\eta_i = \eta(\mathbf{x}_i, \boldsymbol{\theta}) = \mathbf{x}_i \boldsymbol{\theta},$$

$$g(\eta_i) = \eta_i,$$

$$\boldsymbol{\psi} = [\sigma^2]_{1 \times 1},$$

onde cada vetor linha \mathbf{x}_i representa a i -ésima observação contida na base de dados e σ^2 representa o parâmetro de variância da distribuição Normal.

Exemplo 2:

O modelo de regressão linear múltipla com variável resposta tendo distribuição t-Student poder ser definido como

$$y_i \stackrel{\text{ind}}{\sim} t_v(\mu(\mathbf{x}_i, \boldsymbol{\theta}), \sigma),$$

$$\eta_i = \eta(\mathbf{x}_i, \boldsymbol{\theta}) = \mathbf{x}_i \boldsymbol{\theta},$$

$$g(\eta_i) = \eta_i,$$

$$\boldsymbol{\psi} = [v, \sigma]_{2 \times 1},$$

onde v representa o parâmetro de graus de liberdade e σ representa o parâmetro de escala, ambos da distribuição t-Student.

Exemplo 3:

O modelo linear generalizado com variável resposta tendo distribuição Poisson pode ser definido como

$$y_i \stackrel{\text{ind}}{\sim} \text{Poisson}(\mu(\mathbf{x}_i, \boldsymbol{\theta})),$$

$$\eta_i = \eta(\mathbf{x}_i, \boldsymbol{\theta}) = \mathbf{x}_i \boldsymbol{\theta},$$

$$g(\eta_i) = e^{\eta_i}.$$

Exemplo 4:

O modelo de regressão logística, pode ser definido como

$$y_i \stackrel{\text{ind}}{\sim} \text{Bernoulli}(\mu(\mathbf{x}_i, \boldsymbol{\theta})),$$

$$\eta_i = \eta(\mathbf{x}_i, \boldsymbol{\theta}) = \mathbf{x}_i \boldsymbol{\theta},$$

$$g(\eta_i) = \frac{1}{1 + e^{-\eta_i}}.$$

É comum, conforme ilustrado nos exemplos acima, tomar $\eta(\mathbf{x}_i, \boldsymbol{\theta}) = \mathbf{x}_i \boldsymbol{\theta}$, representando uma combinação linear de colunas da matriz de variáveis explicativas \mathbf{X} . Nesses casos, $\eta(\mathbf{x}_i, \boldsymbol{\theta})$ é chamado de preditor linear. Ainda que η se restrinja a combinações lineares de colunas de \mathbf{X} , é possível construir relações bastante flexíveis entre $\mu(\mathbf{x}_i, \boldsymbol{\theta})$ e \mathbf{x}_i . Por exemplo,

$$\mathbf{X} = \begin{bmatrix} | & | & | & \dots & | \\ 1 & \mathbf{x} & \mathbf{x}^2 & \dots & \mathbf{x}^{d_X} \\ | & | & | & & | \end{bmatrix},$$

modela a relação polinomial de grau d_X , onde \mathbf{x} é o vetor coluna que contém os valores das diferentes observações para a variável explicativa.

Se tomarmos ainda

$$\mathbf{X} = \begin{bmatrix} | & | & | & | & \dots & | \\ 1 & \mathbf{x} & (\mathbf{x} - \xi_1)_+ & (\mathbf{x} - \xi_2)_+ & \dots & (\mathbf{x} - \xi_{d_X})_+ \\ | & | & | & | & & | \end{bmatrix},$$

onde $a_+ = \max(a, 0)$ denota a parte positiva de a , e sendo $\xi_1 < \xi_2 < \dots < \xi_{d_X}$, temos que η_i representa uma função linear por partes de \mathbf{x} , também conhecida como *spline* linear.

Caso tomemos

$$\mathbf{X} = \begin{bmatrix} | & | & | & | & | & | & \dots & | \\ 1 & \mathbf{x} & \mathbf{x}^2 & \mathbf{x}^3 & (\mathbf{x} - \xi_1)_+^3 & (\mathbf{x} - \xi_2)_+^3 & \dots & (\mathbf{x} - \xi_{d_X})_+^3 \\ | & | & | & | & | & | & & | \end{bmatrix},$$

temos a representação de *splines* cúbicos. Também é possível estender representações similares para *splines* naturais, *b-splines*, entre outros.

A depender da escolha da função η que representa o preditor linear, é possível obter diversos modelos conhecidos na estatística, com diferentes níveis de flexibilidade na modelagem da relação funcional entre a variável resposta y_i e as variáveis explicativas \mathbf{x}_i .

Os modelos de redes neurais são obtidos também com uma escolha particular das funções g e η , como casos particulares. Um exemplo é a estimação dos coeficientes no modelo de regressão múltipla

$$\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}, \beta, \sigma^2 \sim N(\mathbf{X}\boldsymbol{\theta} + \beta, \sigma^2 \mathbf{I}_n),$$

onde n representa o número de observações nos dados e \mathbf{I}_n a matriz identidade de dimensão $n \times n$. A rede neural pode ser então estruturada como na arquitetura explícita na Figura 2.4.

Um outro exemplo são os modelos de regressão logística (NELDER; WEDDERBURN, 1972), que pode ser definido como

$$y_i | p \sim \text{Bernoulli}(p),$$

$$p = f_{\text{sigmoide}}(\mathbf{x}_i \boldsymbol{\theta} + \beta) \in (0, 1)^n,$$

onde a função sigmoide é aplicada a cada combinação linear das variáveis explicativas da i -ésima observação. A rede neural pode ser então estruturada como na arquitetura explícita na Figura 2.5.

Os modelos que visam estimar a relação treinamento-performance no ciclismo propostos por essa dissertação, que serão apresentados no Capítulo 6, irão abordar o problema de regressão multivariada com arquitetura de redes neurais de forma genérica, como apresentado na Figura 2.3, cujo modelo estatístico é denotado por

$$\mathbf{y}_j | \mathbf{X}, \boldsymbol{\theta}_j, \boldsymbol{\Sigma}_j \sim N(\boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\theta}_j), \boldsymbol{\Sigma}_j), \quad j = 1, 2, \dots, d_Y,$$

onde \mathbf{y}_j é o vetor coluna que representa a j -ésima variável resposta para todas as n_Y observações, $\boldsymbol{\theta}_j$ é o vetor de parâmetros desconhecidos associados à j -ésima variável resposta que pertence à matriz $\boldsymbol{\Theta}$ de parâmetros de interesse e $\boldsymbol{\Sigma}_j$ é a matriz de covariâncias, que pertence à matriz $\boldsymbol{\Psi}$ de parâmetros não estimados.

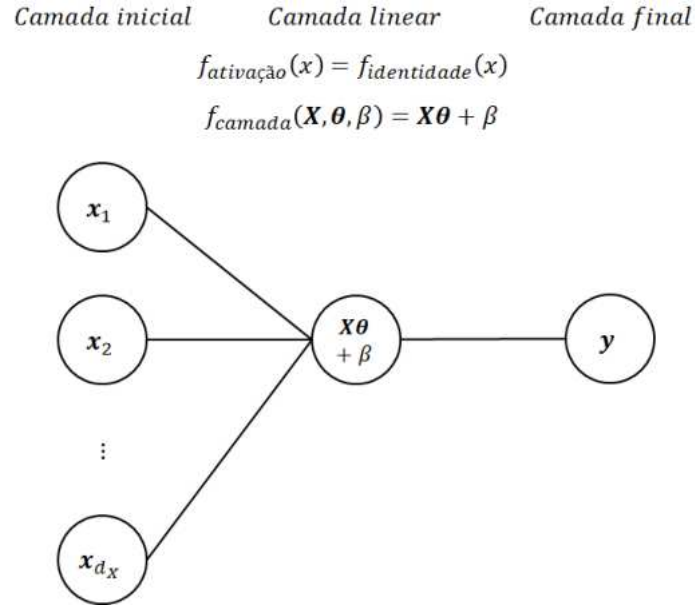


Figura 2.4 – Rede neural para modelo de regressão múltipla.

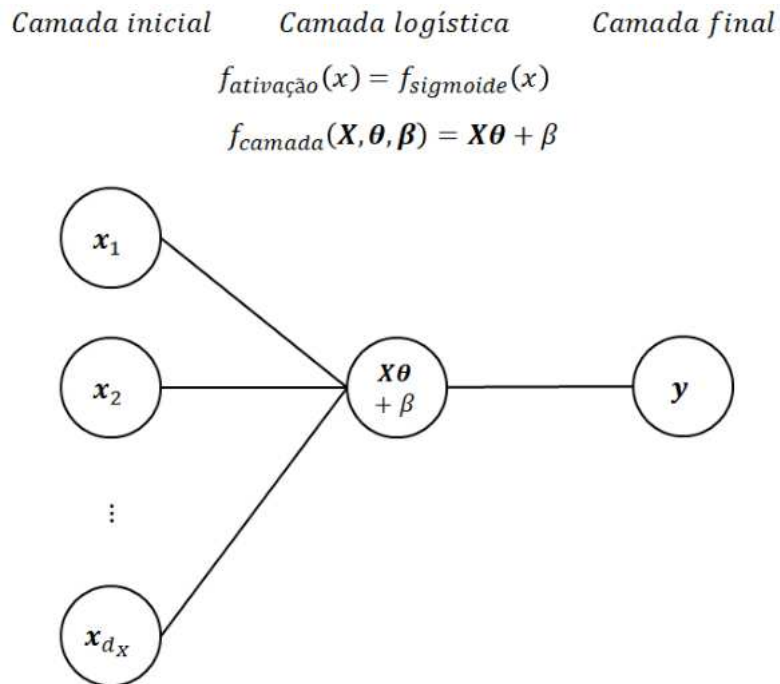


Figura 2.5 – Rede neural para modelo de regressão logística.

2.6 INFERÊNCIA NÃO-PARAMÉTRICA

As redes neurais contêm como caso particulares modelos tradicionalmente utilizados na estatística para representar relações de forma não-paramétrica, como é o caso dos *splines*, por exemplo. Portanto, são capazes de descrever com mais riqueza as possíveis relações entre treinamento e performance no ciclismo.

A estimação dos parâmetros desconhecidos, Θ , na arquitetura de redes neurais definida pelo pesquisador pode ser realizada sob a ótica de um problema de otimização, considerando a minimização de uma certa função perda $L(\Theta; \mathbf{X}, \mathbf{Y})$. Considerando o desafio de lidar com grande volume de dados, que serão apresentados mais adiante, foi escolhida a utilização de algoritmos de otimização via métodos de gradiente para gerar estimativas dos parâmetros desconhecidos nos modelos a serem definidos.

3 OTIMIZAÇÃO VIA MÉTODOS DE GRADIENTE

A otimização é um campo de estudo de grande relevância, principalmente devido às inúmeras aplicações para a estimação de parâmetros de funções não-lineares. Na estatística, observa-se o uso de métodos de otimização no cálculo de estimadores de máxima verossimilhança, máximo *a posteriori*, inferência variacional, ou mesmo em métodos de Monte Carlo via Cadeias de Markov que se utilizam da informação dos gradientes da log-verossimilhança, como Monte Carlo Hamiltoniano (DUANE *et al.*, 1987). Um exemplo é o caso de estimação clássica de parâmetros em modelos de regressão logística, onde é utilizado o método iterativo de Newton-Rhapson para o cálculo de derivadas de segunda ordem para maximizar a função log-verossimilhança (NELDER; WEDDERBURN, 1972).

O método de Newton-Rhapson, é um algoritmo de otimização de segunda-ordem pois necessita da segunda derivada da função perda em relação aos parâmetros provenientes do modelo. Na estatística são comuns as escolhas de $-\log$ -verossimilhança ou $-\log$ da função da distribuição conjunta como função perda, impondo um alto custo computacional quando, por exemplo, a quantidade de parâmetros é muito elevada, por conta da complexidade do cálculo analítico da matriz de informação de Fisher.

Os métodos de otimização de primeira ordem, que somente utilizam a primeira derivada da função perda em relação aos parâmetros, são preferíveis quando o volume de dados, ou o número de parâmetros, é muito elevado, o que ocorre em muitos casos em que se aplica a estrutura de redes neurais. Dos métodos de otimização de primeira ordem, um dos mais populares para a otimização paramétrica de redes neurais, é o método de gradiente descendente (BALDI, 1995).

3.1 GRADIENTE DESCENDENTE

O algoritmo de gradiente descendente atualiza iterativamente o valor dos parâmetros de interesse $\boldsymbol{\theta}$, considerando uma função perda $L(\boldsymbol{\theta}; \mathbf{X}, \mathbf{Y})$ a qual deseja-se minimizar, na direção oposta do gradiente da função perda $\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}; \mathbf{X}, \mathbf{Y})$ (BALDI, 1995). Considerando $\eta \in (0, 1)$ como a taxa de aprendizado, a equação de atualização de $\boldsymbol{\theta}$ pelo método de gradiente descendente é dada por

$$\boldsymbol{\theta}^{(\ell)} = \boldsymbol{\theta}^{(\ell-1)} - \eta \cdot \nabla_{\boldsymbol{\theta}^{(\ell-1)}} L(\boldsymbol{\theta}^{(\ell-1)}; \mathbf{X}, \mathbf{Y}),$$

onde ℓ indica a etapa de iteração do algoritmo e o vetor paramétrico de inicialização $\boldsymbol{\theta}^{(0)}$, pode ser definido de forma aleatória, entretanto para garantir a convergência do algoritmo e, conseqüentemente, o sucesso do método em si, a escolha ideal do vetor de valores iniciais é importante. Como é necessário calcular o gradiente para todos os dados disponíveis a cada etapa de atualização, há um grande custo computacional pela necessidade da alocação da totalidade da base de dados na memória.

O algoritmo de gradiente descendente garante a convergência para mínimos globais em superfícies convexas e para mínimos locais em superfícies não-convexas. Entretanto, a escolha de η com valor muito alto pode causar divergência, enquanto um valor muito baixo pode tornar a convergência muito lenta.

3.2 GRADIENTE DESCENDENTE ESTOCÁSTICO

Uma variação do algoritmo de gradiente descendente é o algoritmo de gradiente descendente estocástico (*SGD – Stochastic gradient descent*) (DOGO *et al.*, 2018). A principal diferença para o método dos gradientes se dá pela escolha aleatória de subconjuntos da base de dados a cada iteração do algoritmo, para o

cálculo do gradiente da função perda. A equação de atualização dos parâmetros é dada por

$$\boldsymbol{\theta}^{(\ell)} = \boldsymbol{\theta}^{(\ell-1)} - \eta \cdot \nabla_{\boldsymbol{\theta}^{(\ell-1)}} L(\boldsymbol{\theta}^{(\ell-1)}; \mathbf{X}_{S_\ell}, \mathbf{Y}_{S_\ell}),$$

onde $\mathbf{X}_{S_\ell} = \{X_i, i \in S_\ell\}$, $\mathbf{Y}_{S_\ell} = \{Y_i, i \in S_\ell\}$ e $S_\ell \subset \{1, \dots, n\}$ escolhido aleatoriamente a cada iteração. Desta forma, reduz-se consideravelmente o custo computacional do algoritmo por iteração. Como o gradiente estocástico, $\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}; \mathbf{X}_{S_\ell}, \mathbf{Y}_{S_\ell})$, é um estimador não-viesado para o gradiente completo, $\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}; \mathbf{X}, \mathbf{Y})$, o algoritmo também converge para o ponto de mínimo local da função de perda (ROBBINS; MONRO, 1951).

3.3 GRADIENTE DESCENDENTE COM MOMENTO

O algoritmo de gradiente descendente apresenta um comportamento indesejado quando a superfície de otimização possui um gradiente muito mais acentuado numa dimensão, em comparação com outra, como exemplificado na Figura 3.1. Para solucionar essa deficiência, o algoritmo de gradiente descendente com momento (QIAN, 1999), baseando-se nos aspectos provenientes da física, incorpora o conceito de momento na equação de atualização na forma

$$\mathbf{v}^{(\ell)} = \gamma \cdot \mathbf{v}^{(\ell-1)} + \eta \cdot \nabla_{\boldsymbol{\theta}^{(\ell-1)}} L(\boldsymbol{\theta}^{(\ell-1)}; \mathbf{X}, \mathbf{Y}),$$

$$\boldsymbol{\theta}^{(\ell)} = \boldsymbol{\theta}^{(\ell-1)} - \mathbf{v}^{(\ell)},$$

em que o componente de momento $\mathbf{v}^{(\ell)}$ mantém parte do momento linear da iteração anterior, $(\ell - 1)$, ponderada pelo parâmetro $\gamma \in (0, 1)$. O hiperparâmetro γ imprime uma redução sob o momento linear, atuando como uma força de atrito no processo de atualização de $\boldsymbol{\theta}$. A inclusão da componente de aceleração induzida pela equação de momento, tende a acelerar a convergência do algoritmo, além de possibilitar escapar de mínimos locais em determinadas circunstâncias.

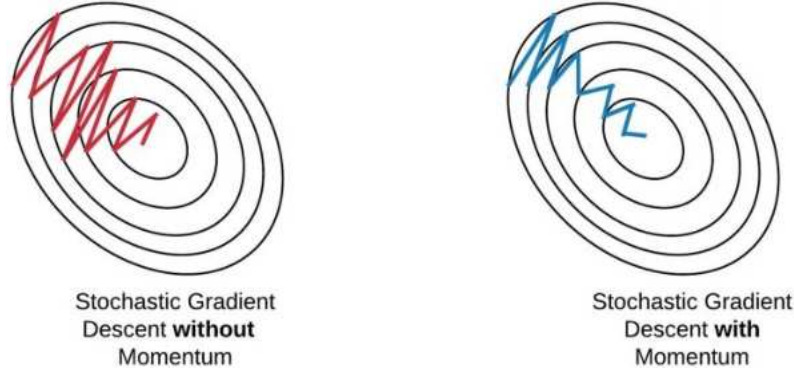


Figura 3.1 – Exemplo de otimização por gradiente descendente sem e com componente de momento.

Fonte: BISONC, 2019.

3.4 ADAGRAD

O algoritmo de gradiente adaptativo (*Adagrad – Adaptive gradient algorithm*) (DUCHI; HAZAN; SINGER, 2011) se diferencia do algoritmo de gradiente descendente com momento pela adaptação do parâmetro de aprendizado η em relação aos parâmetros de interesse, promovendo uma melhoria com a utilização de dados esparsos, como descrito por Duchi, Hazan e Singer (2011).

A função de atualização dos parâmetros no algoritmo Adagrad é definida como

$$g^{(\ell-1)} = \nabla_{\theta^{(\ell-1)}} L(\theta^{(\ell-1)}; \mathbf{X}, \mathbf{Y}),$$

$$\theta^{(\ell)} = \theta^{(\ell-1)} - \frac{\eta}{\sqrt{\text{diag}(\mathbf{G}^{(\ell-1)}) + \rho}} \odot g^{(\ell-1)},$$

onde $\mathbf{G}^{(\ell-1)} = \sum_{\tau=1}^{\ell-1} g^{\tau} \cdot g^{\tau\top}$ e $\rho > 0$ é um termo de suavização que evita a divisão por zero. O operador \odot simboliza a multiplicação vetorial elemento a elemento. Além desta operação, a operação de divisão e raiz quadrada também são aplicadas elemento a elemento.

A divisão por $\sqrt{\text{diag}(\mathbf{G}^{\ell-1}) + \rho}$ tende a reduzir o gradiente em componentes com histórico de magnitude elevada e aumentar o gradiente em componentes de magnitude baixa, reduzindo o comportamento em “zig-zag” dos gradientes de forma similar ao ilustrado pela Figura 3.1.

3.5 ADAM

O algoritmo de estimação com momento adaptativo (*Adam – Adaptive Moment Estimation*) (KINGMA; BA, 2014), utiliza dois componentes de memória dos gradientes calculados para cada parâmetro de interesse. Os componentes podem ser representados como

$$m^{(q)} = \beta_1 m^{(q-1)} + (1 - \beta_1) \cdot \nabla_{\boldsymbol{\theta}^{(\ell-1)}} L(\boldsymbol{\theta}^{(\ell-1)}; \mathbf{X}, \mathbf{Y}),$$

$$v^{(q)} = \beta_2 v^{(q-1)} + (1 - \beta_2) \cdot \nabla_{\boldsymbol{\theta}^{(\ell-1)}} L(\boldsymbol{\theta}^{(\ell-1)}; \mathbf{X}, \mathbf{Y}) \odot \nabla_{\boldsymbol{\theta}^{(\ell-1)}} L(\boldsymbol{\theta}^{(\ell-1)}; \mathbf{X}, \mathbf{Y}),$$

em que $m^{(q)}$ e $v^{(q)}$ são as estimativas do primeiro e segundo momento dos gradientes, respectivamente. Os hiperparâmetros β_1 e β_2 representam taxas de decaimento exponencial para as estimativas de momento, e podem ter valores no intervalo $[0, 1)$. Entretanto, por conta de um viés em direção a zero por conta do vetor de inicialização, a correção proposta pelos autores do método é dada por

$$\hat{m}^{(q)} = \frac{m^{(q)}}{1 - \beta_1^q},$$

$$\hat{v}^{(q)} = \frac{v^{(q)}}{1 - \beta_2^q},$$

onde β_1^q e β_2^q denotam β_1 e β_2 elevados à potência q .

As estimativas apresentadas são utilizadas para definir a função de atualização dos parâmetros de interesse ao longo das iterações do algoritmo, da forma

$$\boldsymbol{\theta}^{(q)} = \boldsymbol{\theta}^{(q-1)} - \frac{\eta}{\sqrt{\hat{v}^{(q-1)}} + \rho} \cdot \hat{\mathbf{m}}^{(q-1)}.$$

Os valores sugeridos por Kingma e Ba (2014) do método são $\beta_1 = 0,9$, $\beta_2 = 0,999$ e $\rho = 10^{-8}$, que serão mantidos e utilizados ao longo deste trabalho. O método *Adam* pode ser visto como uma modificação do método *Adagrad* onde inclui-se o conceito de momento na penalização por magnitudes elevadas ($\hat{v}^{(q-1)}$) e no cálculo da direção de variação ótima ($\hat{\mathbf{m}}^{(q-1)}$).

Os algoritmos de otimização via métodos de gradiente descendente estocástico são boas escolhas ao se trabalhar com grande volume de dados, que é o caso que será considerado neste trabalho. Ao ser considerada uma amostra dos dados para gerar a estimativa do gradiente da função perda a cada iteração, o custo computacional é reduzido, possibilitando a análise em tempo hábil da totalidade do conjunto de dados existente, que será apresentado a seguir. Além disso, somente será utilizado o método *Adam* por apresentar resultados melhores em termos de convergência quando comparado aos outros métodos.

3.6 APLICAÇÃO

O uso desta metodologia descritas nos capítulos 2 e 3, que descreveram redes neurais e o processo de otimização via métodos de gradiente, será realizado com o objetivo de realizar quantificação da performance no capítulo 5 e, posteriormente, para aproximar a relação treinamento-performance no capítulo 6, considerando o grande conjunto de dados que será introduzido no capítulo 4.

Esta dissertação fez uso dos pacotes *Tensorflow* (ABADI, 2016) e *Keras* (CHOLLET, 2015) principalmente pela funcionalidade de diferenciação automática, que facilita a etapa de otimização via métodos de gradientes, com execução eficiente mesmo com grande volume de dados. Além disso, o arcabouço

de programação auxilia a implementação de arquiteturas de redes neurais nas formas tradicionais, entretanto foram necessárias criação de camadas personalizadas para realizar a estimação dos modelos de quantificação da performance no capítulo 5, assim como o modelo Banister (BANISTER *et al.*, 1975) apresentado na Seção 6.2.

A adaptação dos modelos que serão apresentados para a arquitetura de redes neurais, além da estimação via métodos de gradiente, permitiu que esta dissertação fizesse uso de uma metodologia reduzida, mas que garante os resultados necessários para o alcance dos objetivos citados na Seção 1.2. Isto ressalta a versatilidade da metodologia apresentada.

4 BASE DE DADOS

A base de dados utilizada tem como fonte *GoldenCheetah OpenData Project* (LIVERSEDGE, 2018), em que os usuários do programa *GoldenCheetah*³ permitem o compartilhamento das informações de treinamento de diversos esportes, de forma anonimizada.

O objetivo do programa é fornecer aos usuários uma plataforma de visualização e análise de treinos de forma facilitada e intuitiva. Além disso, permite o planejamento de atividades futuras e comparação de períodos de treinamento passados. O programa permite a importação de atividades em diferentes formatos, como extensões *.wko*, *.csv*, *.fit*, *.hrm*, *.gpx*, *.srm*, entre outros. Cada um desses tipos de arquivos é proveniente de distintas marcas de aparelhos que coletam dados em tempo real, para diferentes esportes, entre eles ciclismo, corrida, natação, esqui, canoagem, patinação ou surfe, para citar alguns.

O programa utiliza os dados que foram previamente coletados a partir de aparelhos desenvolvidos para este fim, como os da marca *Garmin*⁴, por exemplo *Garmin Edge* ou *Garmin Forerunner*, da marca *Polar*⁵, por exemplo *Polar Grit* ou *Polar Vantage*, ou da marca *SRM*⁶, por exemplo *SRM PowerControl*. Certos programas permitem a coleta de informações a partir de telefones inteligentes (*smartphones*) ou relógios inteligentes (*smartwatches*), que também podem ser enviados para o *GoldenCheetah*, a fim de possibilitar análises mais aprofundadas do treinamento realizado. Entretanto, o escopo de atuação do *GoldenCheetah* não

³ Disponível em: <https://www.goldencheetah.org/>. Acesso em 19 mar. 2022.

⁴ Disponível em: <https://www.garmin.com/>. Acesso em 19 mar. 2022.

⁵ Disponível em: <https://www.polar.com/>. Acesso em 19 mar. 2022.

⁶ Disponível em: <http://www.srm.de/>. Acesso em 19 mar. 2022.

realiza a sugestão de atividades para melhora de performance ou outros objetivos do atleta.

As funcionalidades do programa também permitem a visualização da atividade de maneira analítica, com gráficos das séries temporais das variáveis medidas, gráficos de distribuições empíricas, criação de métricas de quantificação de treinamento de forma automática, visualização de métricas agregadas por períodos de treinamento, gráficos de duração-potência, entre outras informações úteis para análise do treinamento realizado por atletas de diferentes esportes.

O mercado de programas para análise de treinamento de atividades esportivas possui outras opções, como *TrainingPeaks*⁷, *WKO*⁸ ou *Strava*⁹, por exemplo. A vantagem do *GoldenCheetah* frente às outras opções disponíveis é, principalmente, o fato de ser um programa de código aberto e gratuito para os usuários. A contribuição de diferentes pesquisadores com conhecimento aprofundado em ciências do esporte, como Phil Skiba, Mark Liversedge e Michael Puchowicz, por exemplo, faz o programa ter funcionalidades, ferramentas de visualização e possibilitar análises que agregam importantes informações sobre o treinamento realizado por um indivíduo. Atualmente, qualquer sistema operacional suporta a instalação do *GoldenCheetah*.

Os dados reunidos no repositório, *GoldenCheetah OpenData Project*, no total contêm arquivos referentes a 6613 indivíduos, com mais de 2,5 milhões de atividades registradas, somando um total de 358 GB em dados antes de qualquer pré-processamento. O repositório é estruturado com diversos diretórios para cada indivíduo, sendo cada indivíduo identificado por um código randomizado. Cada diretório possui como conteúdo informações resumo no formato *JSON*, como

⁷ Disponível em: <https://trainingpeaks.com/>. Acesso em 19 mar. 2022.

⁸ Disponível em: <https://www.trainingpeaks.com/wko5/>. Acesso em 19 mar. 2022.

⁹ Disponível em: <https://www.strava.com/>. Acesso em 19 mar. 2022.

exemplificado na Figura 4.1, e as variáveis coletadas durante as atividades, no formato CSV.

```
{
  "VERSION":"1.9",
  "ATHLETE":{ "gender":"M", "yob":"1969", "id":"{0a4dfb20-0ce5-4d55-9bd9-20122b79baf5}" },
  "RIDES":[
    {
      "date":"2020/01/15 18:33:25 UTC",
      "data":"TDS-HC-AGL----R",
      "METRICS":{
        "aerobic_decoupling":"11.54375",
        "ride_count":"1.00000",
        "workout_time":"2275.00000",
        "time_recording":"2275.00000",
        "time_riding":"2262.00000",
        "time_carrying":"818.00000"
      }
    },
    {
      "date":"2014/05/19 14:57:53 UTC",
      "data":"TDS-HC-AGL-E---",
      "METRICS":{
        "ride_count":"1.00000",
        "workout_time":"11331.00000",
        "time_riding":"9069.00000",
        "time_carrying":"44.00000",
        "elevation_gain_carrying":"3.60000",
        "total_distance":"47.05136",
        "climb_rating":["13.20385","1.00000"]
      }
    }
  ]
}
```

Figura 4.1 – Exemplo de arquivo *JSON*.

Os dados das atividades são coletados a partir de sensores para medição de determinadas variáveis, sendo transmitidos a um ciclocomputador (Figura 4.2). O ciclocomputador é um aparelho instalado na bicicleta que captura dados de sensores, calcula e exibe informações referentes à atividade de ciclismo que está

sendo realizada, como velocidade (Figura 4.3), distância, tempo, frequência cardíaca, cadência dos pedais ou potência. Com o avanço da tecnologia e evolução da capacidade computacional de processadores e da transmissão de dados utilizando tecnologias sem fios, os ciclocomputadores suportam cada vez mais funcionalidades num dispositivo de tamanho e peso reduzido, gerando pouco impacto para o desempenho durante atividades e competições. O uso desta tecnologia possibilita que pesquisadores desenvolvam um conhecimento aprofundado sobre os aspectos do esporte.



Figura 4.2 – Exemplos de ciclocomputadores.

Fonte: DC Rainmaker. Disponível em <<https://www.dcrainmaker.com/>>. Acesso em: 16 jul. 2022.

A comunicação do ciclocomputador com um computador, via *USB* ou *bluetooth*, é feita após o registro de uma atividade para serem inseridos os dados

referentes à sessão de treino realizada, em programas de análise de dados de treinamento esportivo, por exemplo *GoldenCheetah*.

O ciclocomputador por si só coleta informações mais básicas, por exemplo tempo decorrido, altitude a partir de sensor barométrico contido no aparelho e distância percorrida utilizando sistema de posicionamento global. A título de exemplo, a Figura 4.4 ilustra a série temporal da altitude, em metros, de cada instante de uma específica atividade de ciclismo durante uma sessão de treinamento focada em percorrer um terreno com grande variação de elevação. Essa mesma sessão será utilizada, de forma fixa, ao longo deste capítulo para ilustrar séries temporais referentes a outras variáveis de interesse.



Figura 4.3 Exemplos de sensores de velocidade.

Um outro tipo de sensor geralmente utilizado por ciclistas é o medidor de cadência (Figura 4.5), que tem o objetivo de medir as revoluções realizadas pelo pedal acoplado à bicicleta. A Figura 4.6 ilustra a série temporal da cadência, em rotações por minuto, de cada instante da atividade de ciclismo utilizada como exemplo, ao longo da duração desta sessão de treinamento. O ato de pedalar com baixa rotação tende a colocar um estresse sobre o sistema muscular, enquanto pedalar com alta rotação coloca um estresse maior sobre o sistema cardiopulmonar.

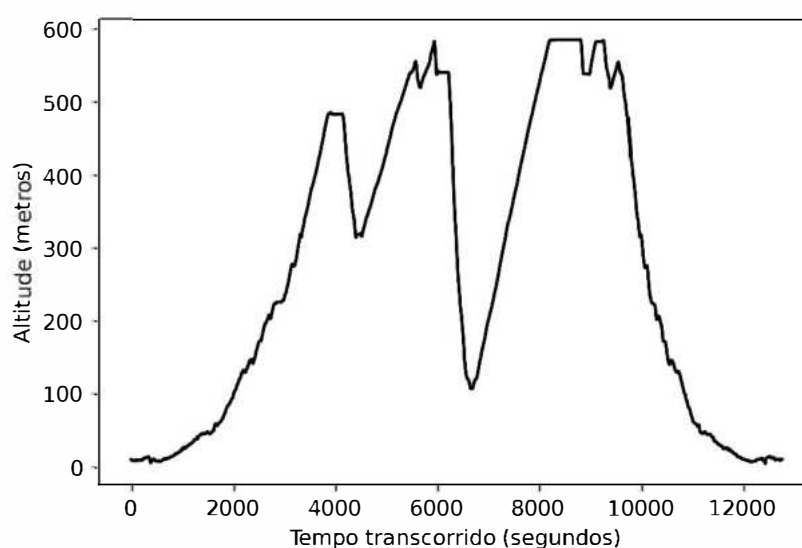


Figura 4.4 – Exemplo de série temporal de elevação para um dos treinamentos de um dos indivíduos na base de dados.



Figura 4.5 – Exemplos de sensores de cadência.

Fonte: DC Rainmaker. Disponível em <<https://www.dcrainmaker.com/>>. Acesso em: 16 jul. 2022.

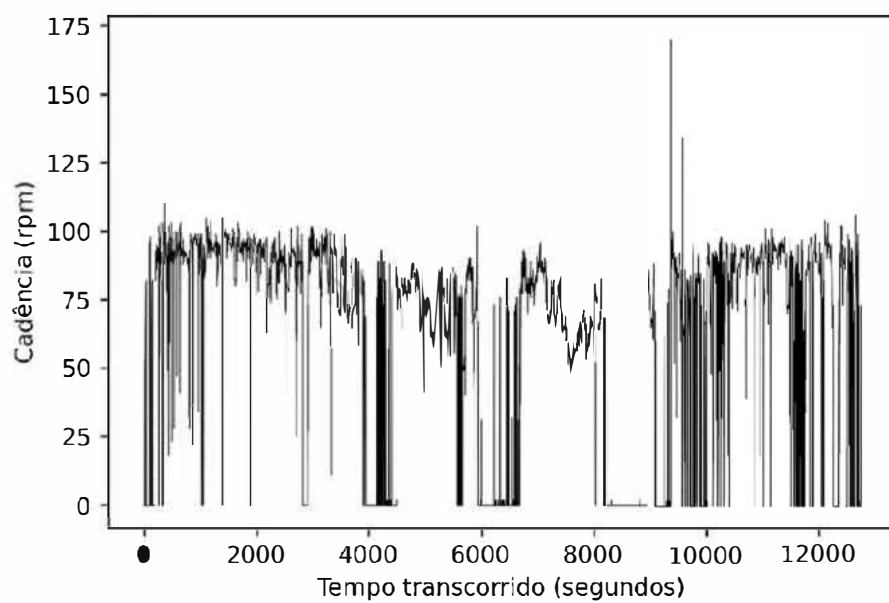


Figura 4.6 Exemplo de série temporal de cadência para um dos treinamentos de um dos indivíduos na base de dados.

A medição de intensidade do exercício é um tópico bastante relevante para o estudo dos esportes. Uma variável com alto grau de correlação com a intensidade da atividade é a frequência cardíaca do indivíduo, mensurada geralmente através da utilização de um sensor de frequência cardíaca (Figura 4.7), usualmente acoplado a uma cinta utilizada na região do peitoral. A Figura 4.8 ilustra a série temporal da frequência cardíaca, em batimentos por minuto, ao longo da sessão de treinamento. Nesta figura, é possível observar “quebras” no gráfico da série temporal que podem ser fruto de pausas manuais ou automáticas na gravação da atividade, interrompendo o armazenamento da informação dos sensores, ou mesmo falhas específicas do sensor. Entretanto, a modelagem proposta neste trabalho não requer o uso de técnicas para lidar com estas quebras, por exemplo interpolação, como será apresentado na Seção 4.1 mais à frente.

Apesar do potencial informativo desta variável, a frequência cardíaca pode ser influenciada por fenômenos externos que reduzem a correlação com a intensidade, como temperatura, umidade, hidratação, utilização de estimulantes como cafeína, entre outros.



Figura 4.7 – Exemplos de sensores de frequência cardíaca.

Fonte: DC Rainmaker. Disponível em <<https://www.dcrainmaker.com/>>. Acesso em: 16 jul. 2022.

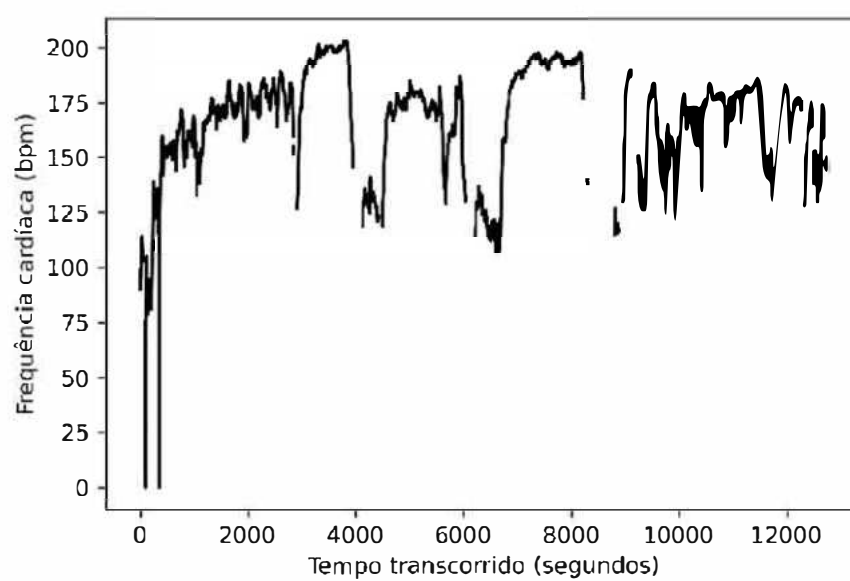


Figura 4.8 Exemplo de série temporal de frequência cardíaca para um dos treinamentos de um dos indivíduos na base de dados.

Uma das variáveis mais importantes para a mensuração da intensidade do exercício é a potência realizada pelo atleta. A Figura 4.9 ilustra a série temporal de potência, em watts, durante a sessão de treinamento. Antes dos anos 1990, este tipo de dado somente podia ser obtido através de análises laboratoriais, utilizando bicicletas ergométricas com medidores de potência com diversos fios acoplados ao equipamento para realizar a mensuração desta grandeza. Porém, avanços na tecnologia permitiram o acoplamento de sensores a alguma parte da bicicleta, como roda traseira, pedivela ou pedal, sem a utilização de fios (Figura 4.10). A partir da força exercida pelo ciclista na bicicleta, é possível medir o torque aplicado e, conjuntamente com a cadência, é possível medir a potência que está sendo realizada.

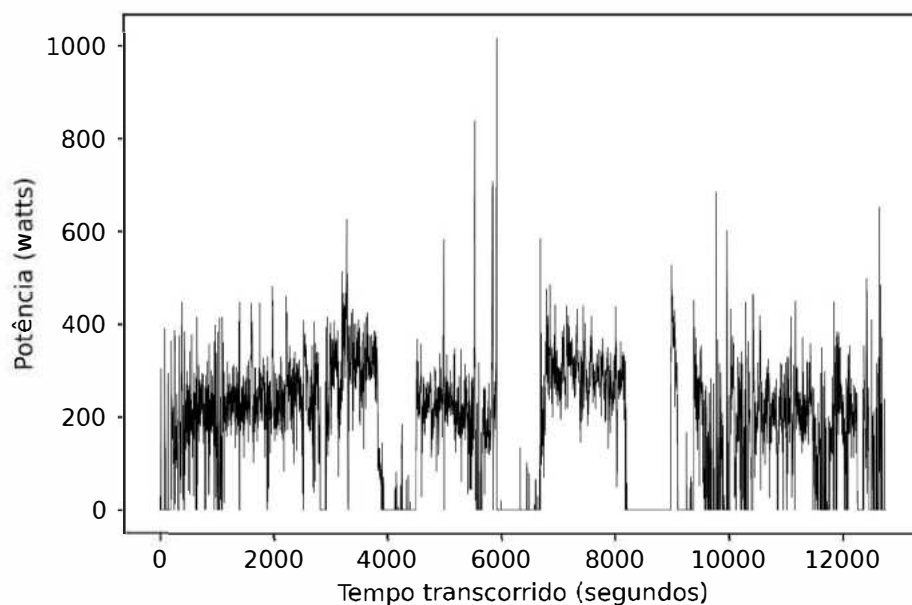


Figura 4.9 Exemplo de série temporal de potência para um dos treinamentos de um dos indivíduos na base de dados.



Figura 4.10 – Exemplos de medidores de potência.

Fonte: DC Rainmaker. Disponível em <<https://www.dcrainmaker.com/>>. Acesso em: 16 jul. 2022.

O ciclismo, diferente de outros esportes, possui percursos sem distância percorrida padronizada e com grande variação de altitude, por exemplo as edições do *Tour de France* dos anos 2019, 2020 e 2021 tiveram distância total percorrida de 3.365,8 quilômetros, 3.484,2 quilômetros e 3.414,4 quilômetros, respectivamente. Comparativamente, a corrida e a natação possuem provas específicas como maratona, com 42 quilômetros de distância, ou 50 metros nas piscinas, respectivamente. Este fenômeno faz com que a velocidade não seja uma variável tão correlacionada com a intensidade da atividade, por isso a motivação por sensores mais robustos para capturar a potência produzida pelo atleta. Os equipamentos possuem um valor agregado elevado, de certa forma selecionando ciclistas mais ávidos e comprometidos com a atividade física.

A data de referência para a coleta dos dados do repositório, que está em constante atualização, é de 29 de julho de 2021. Sendo assim, todos os dados acumulados pelo repositório desde a sua criação, 08 de abril de 2018, até tal data foram utilizados. Apesar da variedade de informações presente no repositório, somente serão analisados os usuários que realizaram treinos de ciclismo, tendo sido mensurada a variável de potência em suas atividades, descartando-se outras modalidades.

4.1 PRÉ-PROCESSAMENTO E VARIÁVEIS DE INTERESSE

O tratamento e a preparação dos dados para utilização em modelos paramétricos é uma tarefa árdua e não-trivial, por conta do grande volume de dados de fontes variadas.

A utilização de dados fora do ambiente de laboratório impõe desafios que podem ser reduzidos com a aplicação de filtros para seleção de indivíduos mais uniformes em relação ao compromisso com o ciclismo, além da remoção de dados de atividades com inconsistências significativas na mensuração das variáveis disponíveis.

O primeiro filtro aplicado é a seleção de indivíduos com mais de 100 atividades de treinamento registradas, dado que um ciclista amador treina de forma consistente ao menos 2 vezes por semana (FRIEL, 2018. Tabela 7.2), assim em um ano haverá ao menos 100 atividades de treinamento. O intuito é remover ciclistas esporádicos ou com pouca experiência de treinamento, pois o padrão de ganho de performance pode divergir do comportamento geral com atletas mais experientes. Além disso, também foram removidos indivíduos cujas atividades não ultrapassaram 1 hora de duração, considerando que um ciclista amador busca realizar atividades de treinamento com durações acima de 1 hora, assim este filtro

tem o mesmo objetivo do filtro anterior, remover ciclistas pouco experientes que podem possuir dados problemáticos ou inconsistentes.

Os outros filtros aplicados buscam identificar medições atípicas do sensor de potência, estipulando os valores limítrofes com base em números de atletas de alta performance do ciclismo, conforme descrito a seguir.

A capacidade de aceleração de um indivíduo é quantificada a partir da potência média máxima realizada por intervalos curtos de duração, em geral 5 a 10 segundos. Ao analisar as demandas físicas e distribuição da variável potência de 12 ciclistas de elite que participaram de eventos internacionais e do campeonato francês de ciclismo de pista, Dorcl *et al.* (2005) observaram uma potência máxima registrada em laboratório de 1600 W. Com base neste valor, foi aplicado um filtro de atividades de indivíduos cuja potência média máxima de 5 segundos excedeu o valor de 3000 W, pois tal valor está muito acima do alcançado por ciclistas de elite.

A capacidade de resistência de um indivíduo é quantificada a partir da potência média máxima realizada por intervalos longos de duração, em geral acima de 30 minutos. Em particular, o recorde mais notório do ciclismo é denominado recorde da hora, onde o atleta busca percorrer a maior distância possível num velódromo durante 1 hora. Apesar de não ser permitida a utilização de ciclocomputadores e sensores acoplados à bicicleta na tentativa de quebra do recorde, Bassett *et al.* (1999) estimaram para as tentativas realizadas até aquela data um máximo de 460 watts de potência durante a hora. Com base neste valor, foi aplicado um filtro de atividades cuja potência média máxima de 1 hora excedeu o valor de 500 watts, pois tal valor está muito acima do alcançado por detentores do recorde da hora.

A partir das variáveis capturadas pelos sensores existentes nas bicicletas é possível gerar outras métricas. A série temporal da mensuração da cadência (Figura 4.6) pode ser observada a partir da contagem de segundos em que o indivíduo pedalou a uma específica faixa de rotação, ilustrado na forma de um histograma na Figura 4.11 (b), obtido projetando-se a série temporal em (a) sobre o eixo y . Esta forma de visualização e consolidação dos dados facilita o trabalho do pesquisador, pois extrai as informações importantes e reduz o ruído observado nas séries temporais (KOSMIDIS; PASSFIELD, 2015), possibilitando inclusive a utilização de séries descontínuas ou com informação faltante, tais como na Figura 4.8.

A mesma transformação pode também ser aplicada nas variáveis de frequência cardíaca, Figura 4.12, e na potência, Figura 4.13. Os gráficos fornecem uma interpretação facilitada do tempo que foi investido em certas faixas das variáveis coletadas. Todos os histogramas apresentados mostram somente os valores acima de zero, para não comprometer a visualização, enquanto valores negativos para estas variáveis não são aplicáveis. Para complementar o conjunto de dados, também foram contabilizadas a duração em horas da atividade; tempo em movimento em horas, que pode ser diferente da duração da atividade; distância percorrida em quilômetros; ganho de elevação em metros; cadência média e máxima em rotações por minuto; frequência cardíaca mínima, média e máxima em batimentos por minuto; e, por fim, potência média da atividade relativa à potência crítica (CP), para cada uma das atividades.

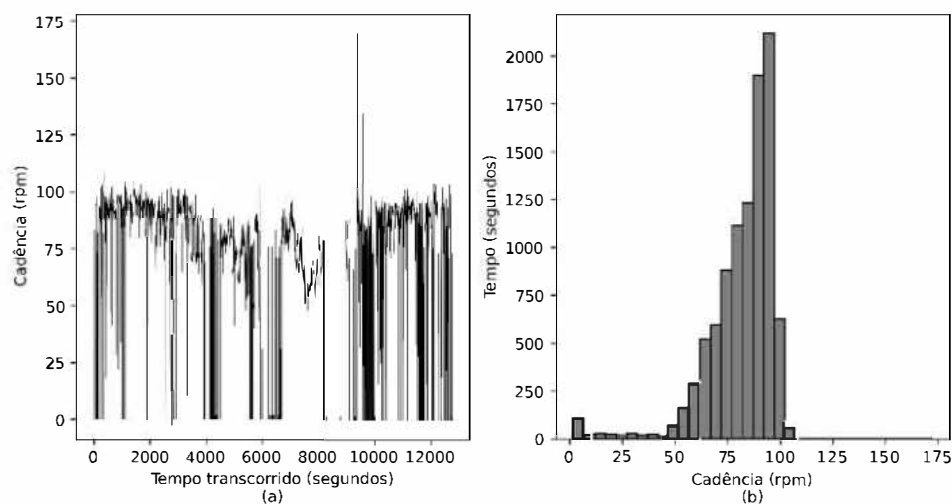


Figura 4.11 – Exemplo de histograma da cadência.

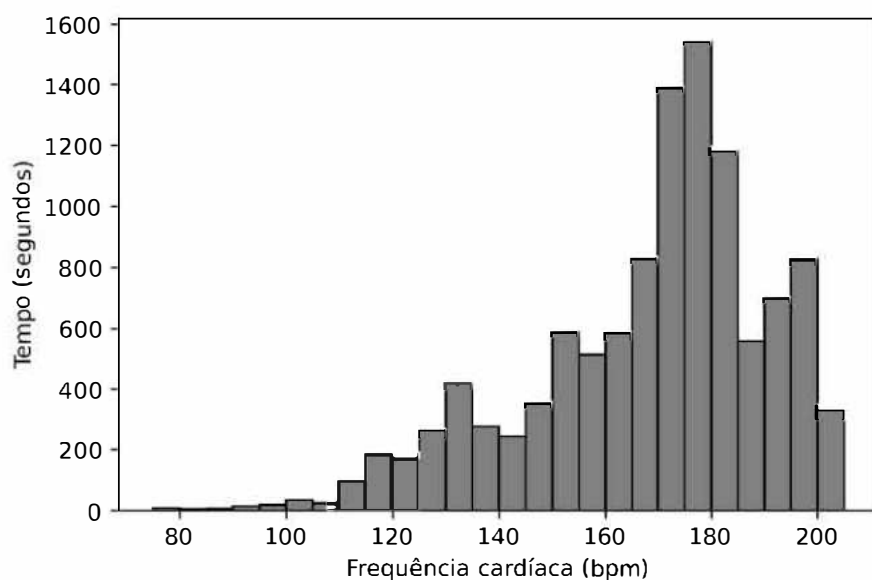


Figura 4.12 Exemplo de histograma da frequência cardíaca.

O tempo investido, por sessão de treinamento, em faixas de potência pode ser analisado de forma relativa às capacidades do indivíduo. Isto permite a fácil comparação entre atletas de diferentes níveis, se aproximando da captura da informação da intensidade que o ciclista está se exercitando. Este trabalho irá realizar a relativização da potência utilizando a potência crítica (CP – *Critical Power*) do atleta, que será calculada a partir de modelos de quantificação de performance, em breve apresentados. Desta forma, se numa determinada atividade o indivíduo passou 100 segundos a 150 watts de potência, sendo sua

potência crítica de 300 W, então a variável utilizada neste trabalho informará que o atleta passou 100 segundos a 50% da potência crítica. Este tipo de relativização é citado por Passfield *et al.* (2017) como uma prática recorrente na análise de dados similares.

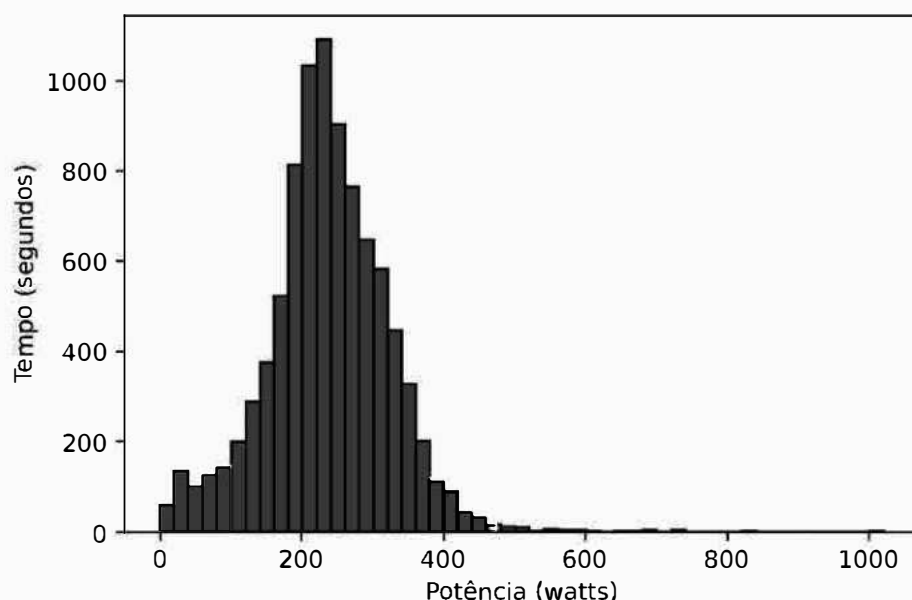


Figura 4.13 Exemplo de histograma da potência.

O tempo total investido no treinamento é uma variável simples que também é analisada por estudos do ciclismo, como realizado em Metcalfe *et al.* (2017) ao monitorar a temporada competitiva de 4 ciclistas de elite durante 12 meses. Além disso, também é importante considerar variáveis que possuem relação com a perda de performance (MUJICA; PADILLA, 2000), como a proporção de dias sem atividades num período de treinamentos. Também é contabilizada a quantidade de dias que correspondem ao período de treinamento.

O próximo conjunto de variáveis produzidas a partir da série temporal de potência são definidas por Allen e Coggan (2010): *Intensity Factor* (IF), *Training Stress Score* (TSS) e *Variability Index* (VI).

Seja a observação da série temporal de potência da atividade j do indivíduo i denotada por $\mathbf{p}_{i,j} = (p_{i,j,1}, p_{i,j,2}, \dots, p_{i,j,T_{i,j}})$, onde $T_{i,j}$ é o tempo total de duração da atividade em segundos, sendo a frequência de observação das variáveis por um ciclocomputador usualmente de 1 Hz. O vetor de médias móveis de duração d segundos, para $\mathbf{p}_{i,j}$ pode ser definido como

$$\mathbf{MA}_{i,j}(d) = [MA_{i,j,1}(d) \quad MA_{i,j,2}(d) \quad \cdots \quad MA_{i,j,(T_{i,j}-d+1)}(d)],$$

onde $MA_{i,j,k}(d)$ é a média das potências medidas do instante k até o instante $k + d - 1$ para o indivíduo i na atividade j , definido como

$$MA_{i,j,k}(d) = \frac{\sum_{t=k}^{k+d-1} p_{i,j,t}}{d}.$$

O IF é uma medida que tem o objetivo de quantificar a intensidade de uma sessão de treinamento, porém relativa às próprias capacidades do indivíduo. O cálculo faz uso da *Normalized Power* (NP) (ALLEN; COGGAN, 2010), definida como

$$NP_{i,j} = \sqrt[4]{\overline{\mathbf{MA}_{i,j}(30)}^4},$$

onde a potenciação é aplicada elemento a elemento no vetor de médias móveis de 30 segundos, onde a barra em cima de um vetor indica a média de seus valores, por exemplo seja um vetor \mathbf{x} com n observações, x_1, x_2, \dots, x_n , então a média de suas observações ao quadrado pode ser definida como $\overline{\mathbf{x}^2} = \frac{1}{n} \sum_{i=1}^n x_i^2$. A quarta potência é aplicada, segundo Allen e Coggan (2010), considerando que a regressão estimada entre a concentração de lactato sanguíneo no limiar de lactato ser igual a potência relativa à potência em limiar de lactato elevado a 3,90. Para simplificação, o valor foi arredondado para 4,00, como explicado pelos autores. Após este cálculo, o IF é definido como

$$IF_{i,j} = \frac{NP_{i,j}}{FTP_i},$$

em que o *Functional Threshold Power* (FTP) do indivíduo i , uma métrica de performance que representa a capacidade aeróbica do atleta, será obtida através dos modelos de duração-potência com a nomenclatura de CP_i , apresentados no Capítulo 5.

O TSS é uma medida que pondera a intensidade pela duração da atividade, quantificando o treinamento de forma mais completa. A métrica é definida como

$$TSS_{i,j} = \frac{T_{i,j} \times NP_{i,j} \times IF_{i,j}}{FTP_i \times 3600} \times 100,$$

onde novamente denotaremos o FTP_i por CP_i , calculado a partir de estimações dos modelos de duração-potência, conforme será visto no Capítulo 5.

Por fim, o índice de variabilidade VI busca representar a variabilidade de potência de uma determinada atividade, pois o $NP_{i,j}$ tende a ser maior para exercícios de maior variabilidade. Assim, o VI é definido como

$$VI_{i,j} = \frac{NP_{i,j}}{\overline{p}_{i,j}},$$

ou seja, a razão entre $NP_{i,j}$ e a potência média.

As métricas apresentadas são calculadas para cada atividade, entretanto o objetivo da dissertação é analisar a relação entre um período de treinamento e a variação dos parâmetros de performance.

Um período de treinamento, será apresentado de forma mais detalhada na Seção 6.1, mas de forma simples pode ser considerado um conjunto de atividades de treinamento. Sendo assim, cada observação \mathbf{x}_i da base de dados \mathbf{X} , $i = 1, 2, \dots, n_X$, apresenta valores de variáveis explicativas que contém informação sobre um conjunto de atividades. Portanto, é necessário realizar agregações das

variáveis de cada atividade para que o período de treinamento esteja representado em \mathbf{X} .

Considere a variável “frequência cardíaca média”, calculada para cada atividade. A agregação pode ser dada calculando-se a média das frequências cardíacas médias das atividades que compõem um determinado conjunto, assim formando a variável “média da frequência cardíaca média”.

Sob a ótica de um conjunto de atividades, é possível verificar em quais dias não houve registro de atividades, em outras palavras um “dia de descanso”. A variável “proporção de dias de descanso” é dada pela razão da quantidade de dias de descanso sobre a quantidade total de dias considerada no determinado período.

As métricas definidas por Allen e Coggan (2010) são agregadas de duas formas distintas. Para o IF e VI serão consolidadas as variáveis “média do IF” e “média do VI” das atividades que compõem o período de treinamento. O TSS será agregado inicialmente no escopo semanal, ou seja, o TSS de uma determinada semana que é dado pela soma dos valores de TSS das atividades que ocorreram na determinada semana, e posteriormente será considerada a média e o desvio-padrão dos valores de TSS das semanas que compõem o período de treinamento. Assim, serão consolidadas as variáveis “média do TSS semanal” e “desvio-padrão do TSS semanal”.

Por fim, as variáveis geradas a partir dos histogramas das séries temporais da cadência, frequência cardíaca e potência, de cada atividade, que representam a quantidade de segundos investida em cada faixa da variável, serão agregadas pela média, por exemplo “média do tempo investido em cada zona de cadência”.

Segue então um resumo das variáveis explicativas apresentadas neste capítulo, que serão utilizadas na estimação da relação treinamento-performance, em detalhes no capítulo 6:

- Duração do período de treinamento (em dias);
- Proporção de dias de descanso no período de treinamento;
- Média de duração das atividades no período de treinamento (em horas);
- Média de tempo em movimento das atividades no período de treinamento (em horas);
- Média de distância das atividades no período de treinamento (em quilômetros);
- Média do ganho de elevação das atividades no período de treinamento (em metros);
- Média da cadência média das atividades no período de treinamento (em rotações por minuto);
- Média da cadência máxima das atividades no período de treinamento (em rotações por minuto);
- Média da frequência cardíaca mínima das atividades no período de treinamento (em batimentos por minuto);
- Média da frequência cardíaca média das atividades no período de treinamento (em batimentos por minuto);
- Média da frequência cardíaca máxima das atividades no período de treinamento (em batimentos por minuto);
- Média da potência média das atividades no período de treinamento relativa à potência crítica (em watts);
- Média do IF das atividades no período de treinamento (adimensional);
- Média do VI das atividades no período de treinamento (adimensional);

- Média do TSS semanal das atividades no período de treinamento (adimensional);
- Desvio-padrão do TSS semanal das atividades no período de treinamento (adimensional);
- Média do tempo investido em cada zona de cadência das atividades no período de treinamento (em segundos);
- Média do tempo investido em cada zona de frequência cardíaca das atividades no período de treinamento (em segundos);
- Média do tempo investido em cada zona de potência das atividades no período de treinamento relativa à potência crítica (em segundos).

A definição das variáveis que sumarizam e representam as sessões de treinamento realizadas pelos indivíduos envolve um significativo trabalho de pré-processamento na base de dados, processo que foi descrito neste capítulo. Ao final desta etapa, restaram 3.825 indivíduos com cerca de 9 e 24 períodos de treinamento por indivíduo, em duas configurações distintas que serão apresentadas na Seção 6.1.

A quantificação do treinamento foi o tema central deste capítulo, entretanto é necessário também quantificar a performance de cada atleta ao longo do tempo, conforme o treinamento é realizado. Para isso, serão apresentados diferentes modelos que possuem como objetivo estabelecer uma relação simplificada com capacidade de descrever a capacidade de produção de potência, para uma gama de durações, para um determinado indivíduo.

5 QUANTIFICAÇÃO DA PERFORMANCE

A mensuração das capacidades de um indivíduo é crucial para o estudo de qualquer esporte. É importante quantificar os diferentes níveis que um atleta pode se encontrar ao longo de sua carreira, para assim identificar seus momentos de melhoria.

A avaliação do estado atual do indivíduo e sua evolução no esporte pode ser uma tarefa desafiadora conforme a complexidade do esporte aumenta, por exemplo onde o resultado de competições realizadas depende da interação de diversos atletas na forma de equipe. Este é o caso do ciclismo de estrada. Então, é necessária a utilização de abordagens mais rebuscadas, com o intuito de quantificar aspectos e habilidades individuais que possuam relevância prática no resultado de eventos competitivos.

A análise objetiva das capacidades de um atleta permite a aplicação de ferramentas de otimização dos treinamentos, tornando possível equipes com menos recursos alcançarem resultados similares a equipes mais abastadas em ligas e campeonatos, como é apresentado por Lewis (2004). Outra possibilidade de aplicação é o auxílio às análises antidoping, por exemplo em Puchowicz (2018) e Iljukov, Berman e Schumacher (2018).

A quantificação da performance neste trabalho utiliza os modelos de duração-potência. Estes modelos têm como objetivo estimar a potência máxima, considerando um esforço homogêneo, que o atleta é capaz de realizar ao longo de um determinado período, buscando gerar parâmetros representativos dos sistemas biológicos de produção de energia. Como a maior produção energética está diretamente associada com maior velocidade de deslocamento, então compreender

os pontos fortes de um atleta permite estabelecer estratégias que aumentem a probabilidade de vitória num evento competitivo.

5.1 SISTEMAS ENERGÉTICOS

Os sistemas energéticos envolvidos na transferência de energia durante o exercício podem ser sumarizados em três componentes: imediato, curto-prazo e longo-prazo (MCARDLE; KATCH; KATCH, 2010, cap. 7). A energia proveniente do consumo alimentar não é aproveitada diretamente para as células; antes é necessário haver a metabolização do alimento, que produz moléculas ricas em energia conhecidas como adenosina trifosfato (ATP – *Adenosine TriPhosphate*).

A via energética imediata é também denominada sistema ATP-PCr, pois envolve também moléculas de fosfocreatina (PCr – *PhosphoCreatine*). A utilização desse sistema se dá em exercícios de alta intensidade e curta duração, como competições de 100 metros rasos, 25 metros livres na natação ou levantamento de peso. Em cerca de 20 a 30 segundos de exercício de máxima duração as reservas de ATP e PCr intramuscular são depletadas, impossibilitando o fornecimento de energia por esta via.

A via energética de curto-prazo, denominada também por sistema anaeróbico ou sistema láctico, é a responsável pela maior parte da produção de ATP em atividades máximas entre 30 e 180 segundos. O processo biológico que possibilita a fosforilação das moléculas de adenosina difosfato (ADP – *Adenosine DiPhosphate*), transformando-as em ATP, é a quebra de moléculas de glicogênio presentes nos músculos sem a presença de oxigênio, conhecido como glicólise anaeróbica. Como subproduto dessa reação ocorre a formação de lactato.

A via energética de longo-prazo, é também denominada sistema aeróbico. Esta é a principal via para atividades com duração acima de 4 minutos, onde o consumo de oxigênio forma um platô, mantendo-se relativamente estável após esse período durante toda a atividade. O estado em que o consumo de oxigênio se encontra estável é conhecido como *steady state*, e nessa situação reações dependentes de oxigênio fornecem a energia para o exercício, enquanto qualquer lactato produzido é oxidado ou transformado em glicose novamente, portanto não há acúmulo do subproduto.

Em teoria, este nível de intensidade seria capaz de ser mantido de forma indefinida, a depender da vontade do atleta, entretanto perdas de fluidos e eletrólitos se mostram como limitantes para a duração. Além disso, a depleção das reservas de glicogênio, tanto muscular como hepática, também reduzem a capacidade do exercício.

Os sistemas energéticos citados possuem influência direta na capacidade de produção de potência, com diferentes graus de influência a depender da duração de um determinado esforço. No ciclismo, a ferramenta que auxilia a compreensão das capacidades do atleta em uma gama de durações, é a curva de potência média máxima.

5.2 POTÊNCIA MÉDIA MÁXIMA

Os modelos de duração-potência são ajustes paramétricos, onde cada parâmetro tem o objetivo de capturar o efeito de sistemas fisiológicos relevantes na produção energética do organismo, realizados sobre dados da curva de potência média máxima (QUOD *et al.*, 2010). Tal curva é definida por $MMP(d)$, sendo MMP a potência média máxima e d a duração em segundos, ou seja, representa a

maior potência média realizada, observada sob um conjunto de durações temporais ao longo de uma ou mais atividades.

Uma possível forma de cálculo da curva de MMP empírica, $\widehat{MMP}(d)$, a partir da série temporal de potência $p_{i,j}(t)$, $t = 1, 2, \dots, T$, de uma atividade j do indivíduo i com duração total de T segundos, pode ser obtida da seguinte forma, para uma duração $d = 1, 2, \dots, T$ específica:

1. Calcula-se a sequência de médias móveis de tamanho d para o conjunto de dados de potência disponível, obtendo o vetor $\mathbf{MA}_{i,j}(d)$, definido na Seção 4.1, $\forall d$.
2. A $\widehat{MMP}_{i,j}(d)$ é definida recursivamente como $\max(\mathbf{MA}_{i,j}(d), \widehat{MMP}_{i,j}(d+1))$, onde $\widehat{MMP}_{i,j}(T) = \mathbf{MA}_{i,j,1}(T)$, representando o máximo entre a maior média móvel observada no conjunto de dados e o valor da MMP empírica de duração maior. Dessa forma, é possível obter uma curva monótona decrescente, o que faz sentido do ponto de vista prático pois um indivíduo é capaz de manter ao menos um mesmo patamar de potência, dado que a duração do esforço é menor.

Um exemplo de visualização está apresentado na Figura 5.1.

Alguns modelos paramétricos de duração-potência estão presentes na literatura, sendo estimados a partir da curva de MMP empírica, como o modelo de dois parâmetros (MONOD; SCHERRER, 1965), modelo de três parâmetros (MORTON, 1996), modelo exponencial (HOPKINS *et al.*, 1989), modelo Omni (PUCHOWICZ; BAKER; CLARKE, 2020) e, por fim, modelo estendido (GRAUSER, 2013). A dissertação realizará a análise da capacidade preditiva destes modelos e, selecionar o mais adequado para ser utilizado como método de quantificação de performance.

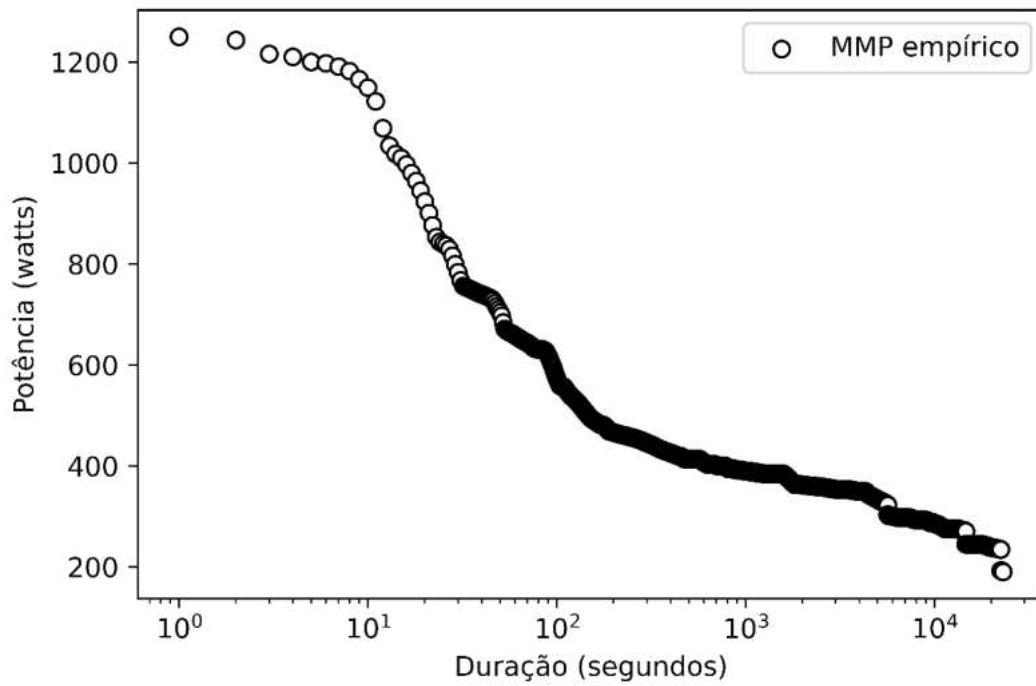


Figura 5.1 – Exemplo de curva de MMP de uma atividade.

5.3 MODELO DE DOIS PARÂMETROS

O modelo de dois parâmetros definido por Monod e Scherrer (1965), é descrito como:

$$MMP(d) = \frac{W'}{d} + CP,$$

onde CP (*Critical Power*) representa a potência crítica, a potência máxima que pode ser gerada unicamente pelo sistema aeróbico de energia, e W' é a quantidade de trabalho que pode ser realizado com potências acima do CP , representando a quantificação do sistema anaeróbico de energia. Com uma estrutura simples de realizar a estimação dos parâmetros, este modelo tem o intuito de quantificar dois sistemas energéticos importantes na geração de potência final pelo atleta.

O parâmetro CP é expresso na unidade watt e representa valores maiores do que zero. Já o parâmetro W' é expresso na unidade joule, também representando valores positivos.

5.4 MODELO DE TRÊS PARÂMETROS

O modelo de três parâmetros definido por Morton (1996) é descrito como:

$$MMP(d) = \frac{W'}{d + \left(\frac{W'}{P_{max} - CP}\right)} + CP,$$

onde a diferença em relação ao modelo de dois parâmetros é a inclusão do parâmetro P_{max} que representa a máxima potência produzida pelo atleta em durações muito curtas, representando o sistema de energia imediata ATP-PCr. Assim, os três sistemas energéticos descritos na Seção 5.1 estão sendo quantificados neste modelo. O parâmetro P_{max} é expresso na unidade watt e representa valores maiores do que zero, além da restrição $P_{max} > CP$.

5.5 MODELO EXPONENCIAL

O modelo exponencial, definido por Hopkins *et al.* (1989) é dado pela equação:

$$MMP(d) = (P_{max} - CP) \cdot \exp\left(-\frac{d}{\tau}\right) + CP,$$

onde as diferenças em relação ao modelo de dois parâmetros são a inclusão do parâmetro P_{max} , que representa a máxima potência produzida pelo atleta, assim como no modelo de três parâmetros, e τ a taxa de declínio exponencial da potência máxima até a potência crítica, ao longo do tempo. O parâmetro τ é adimensional e assume valores positivos.

É importante notar que, todos os modelos apresentados até então tratam o parâmetro CP como a assíntota da curva, ou seja, uma potência que é possível de ser sustentada por duração indeterminada, sendo exauridos os sistemas de energia imediata e anacróbico.

5.6 MODELO OMNI

O modelo Omni, definido por Puchowicz, Baker e Clarke (2020) é descrito como:

$MMP(d)$

$$= \begin{cases} \frac{W'}{d} \cdot \left(1 - \exp \left(-d \cdot \left(\frac{P_{max} - CP}{W'} \right) \right) \right) + CP, & \text{se } d \leq TCP_{max} \\ \left(\frac{W'}{d} \cdot \left(1 - \exp \left(-d \cdot \left(\frac{P_{max} - CP}{W'} \right) \right) \right) + CP - A \cdot \ln \left(\frac{d}{TCP_{max}} \right) \right), & \text{se } d > TCP_{max} \end{cases},$$

onde são apresentadas modificações em relação ao modelo de dois parâmetros, como os parâmetros P_{max} , já apresentado nos modelos de três parâmetros e exponencial, TCP_{max} , que indica o momento no tempo no qual será incluído um decaimento na potência crítica em longas durações, e A , o fator de decaimento do parâmetro CP ao longo do acréscimo da duração do esforço. O parâmetro TCP_{max} possui a unidade segundo e assume valores positivos, assim como o parâmetro A , porém este é adimensional.

O modelo, diferentemente dos demais, leva em consideração fatores importantes que afetam a capacidade de exercer a potência crítica por tempo indeterminado, como fadiga mental, desidratação, perda de eletrólitos ou depleção das reservas de glicogênio.

5.7 MODELO ESTENDIDO

O modelo estendido foi proposto por Grauser (2013) e é definido como:

$$MMP(d) = P_{ATP-PCr}(d) + P_{Anaeróbico}(d) + P_{Aeróbico}(d),$$

$$P_{ATP-PCr}(d) = P_{aa} \cdot \exp \left(P_{aa_{dec}} \cdot \frac{d}{60} \right),$$

$$P_{Anaeróbico}(d) = CP \cdot \left(1 - \exp\left(CP_{del} \cdot \frac{d}{60}\right)\right) \cdot \left(\frac{\tau}{60}\right),$$

$$P_{Aeróbico}(d) = CP \cdot \left(1 - \exp\left(CP_{del} \cdot \frac{d}{60}\right)\right) \cdot \left(1 + CP_{dec} \cdot \exp\left(\frac{CP_{dec_{del}}}{\frac{d}{60}}\right)\right),$$

onde P_{aa} é a potência máxima do componente relativo ao sistema energético APT-PCr, $P_{aa_{dec}}$ é o fator de decaimento exponencial da P_{aa} , CP_{del} é o fator de atraso (*delay*) na influência do CP como componente principal da produção de potência, τ é o fator de decaimento exponencial da influência do sistema anaeróbico, CP_{dec} é o fator de decaimento do CP no sistema aeróbico, $CP_{dec_{del}}$ é o fator de atraso do decaimento do CP ao longo do acréscimo da duração d .

A grande contribuição desta formulação é a decomposição da curva de potência máxima total em 3 componentes análogos aos sistemas energéticos já descritos na Seção 5.1, como ilustra a Figura 5.2, que utiliza os seguintes valores para os parâmetros da equação: $P_{aa} = 633 \text{ W}$, $P_{aa_{dec}} = -4,89$, $CP = 206 \text{ W}$, $CP_{del} = -2,53$, $\tau = 0,47$, $CP_{dec} = -0,38$ e $CP_{dec_{del}} = -59,36$. Isto permite a visualização do nível de contribuição de cada via para a capacidade final, para cada duração d desejada. Em compensação, há uma complexidade significativamente maior por conta do maior número de parâmetros.

O parâmetro P_{aa} é representado com a unidade watt e assume valores maiores que zero, o parâmetro $P_{aa_{dec}}$ é adimensional e assume valores negativos tendo o seu impacto ilustrado na Figura 5.3, onde todos os outros valores são mantidos como apresentado anteriormente na Figura 5.2, como todos os outros exemplos a seguir também utilizarão. O parâmetro CP_{del} é adimensional e assume valores negativos, tendo seu impacto ilustrado na Figura 5.4 no sistema anaeróbico e na Figura 5.5 no sistema aeróbico. O parâmetro τ é adimensional e assume valores positivos, tendo seu impacto ilustrado na Figura 5.6. O parâmetro

CP_{dec} é adimensional e assume valores negativos, tendo seu impacto ilustrado na Figura 5.7. E, por fim, o parâmetro $CP_{dec_{del}}$ é adimensional e assume valores negativos, tendo seu impacto ilustrado na Figura 5.8. O parâmetro CP se comporta de forma similar ao apresentado nos modelos anteriores.

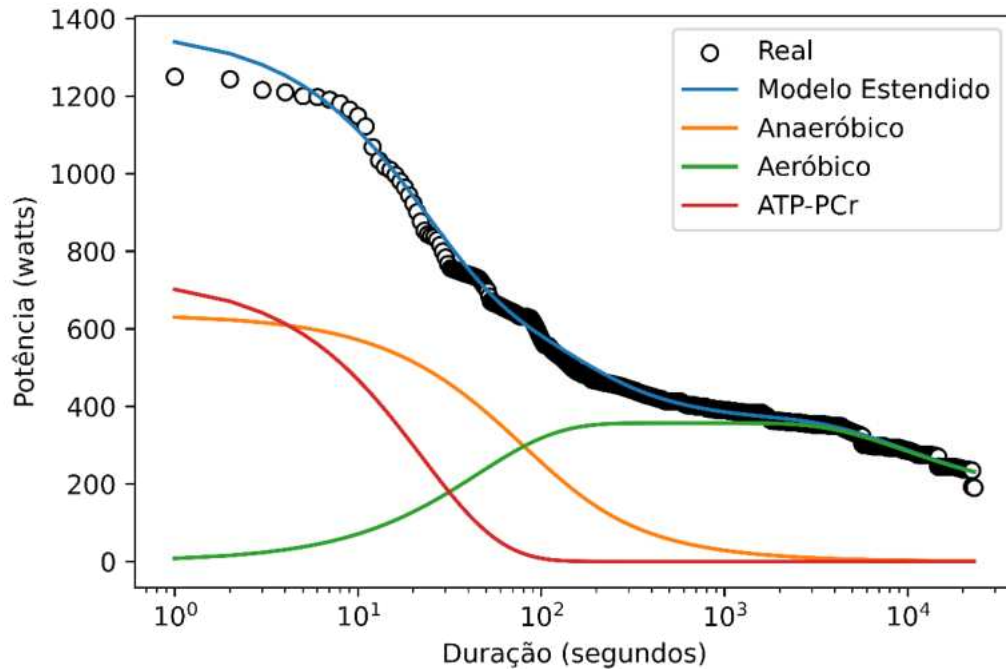


Figura 5.2 – Exemplo de ajuste do modelo estendido.

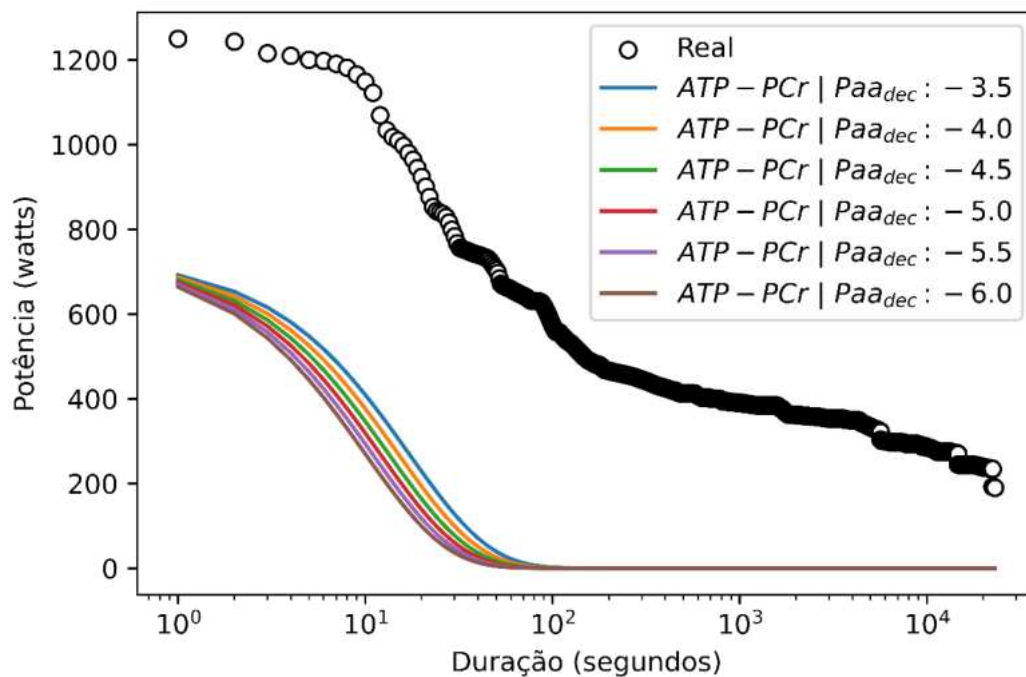


Figura 5.3 – Impacto do parâmetro $P_{aa_{dec}}$ no sistema ATP-PCr.

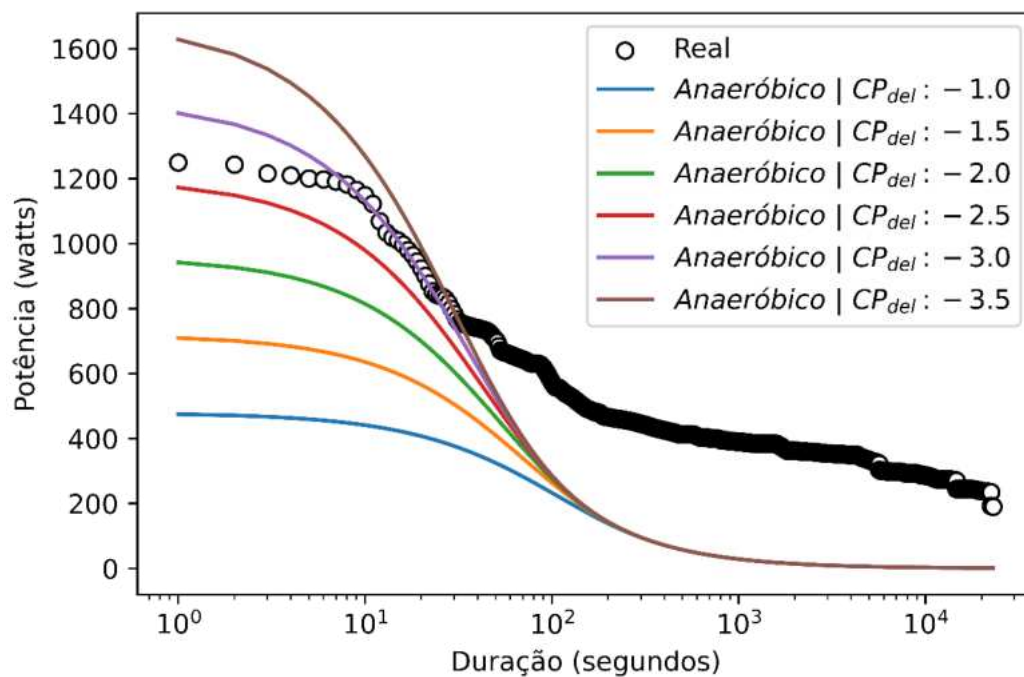


Figura 5.4 – Impacto do parâmetro CP_{del} no sistema anaeróbico.

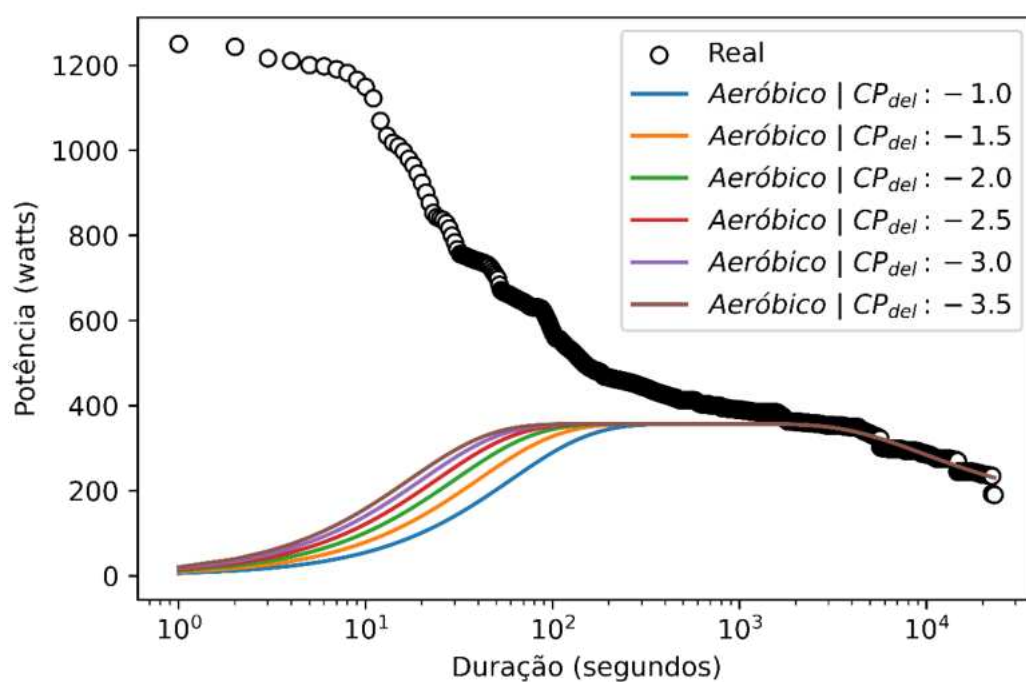


Figura 5.5 – Impacto do parâmetro CP_{del} no sistema aeróbico.

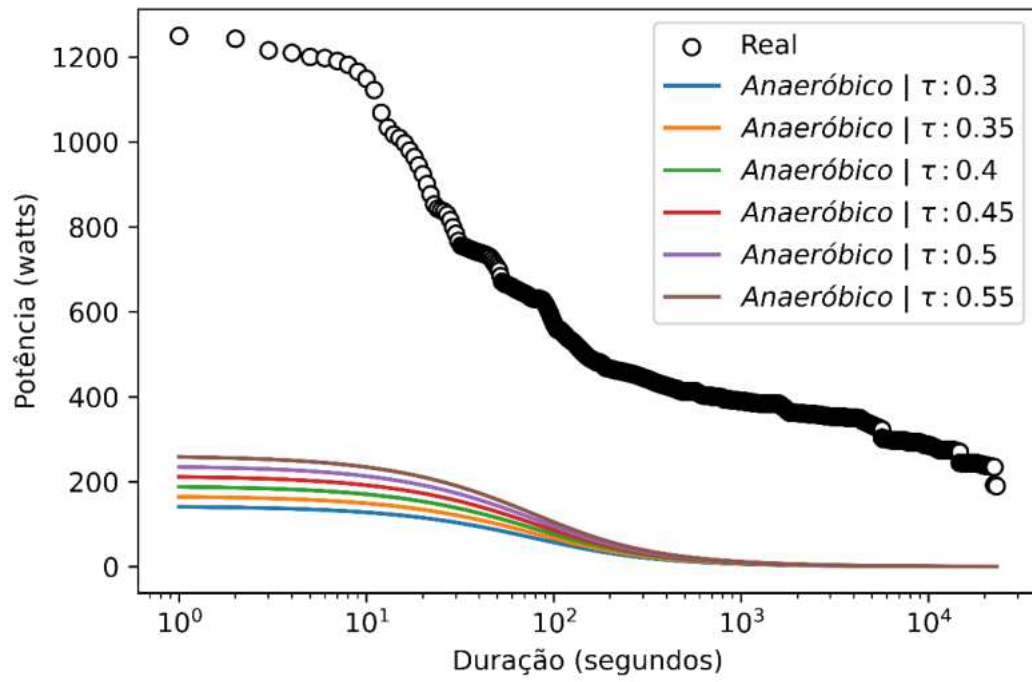


Figura 5.6 – Impacto do parâmetro τ no sistema anaeróbico.

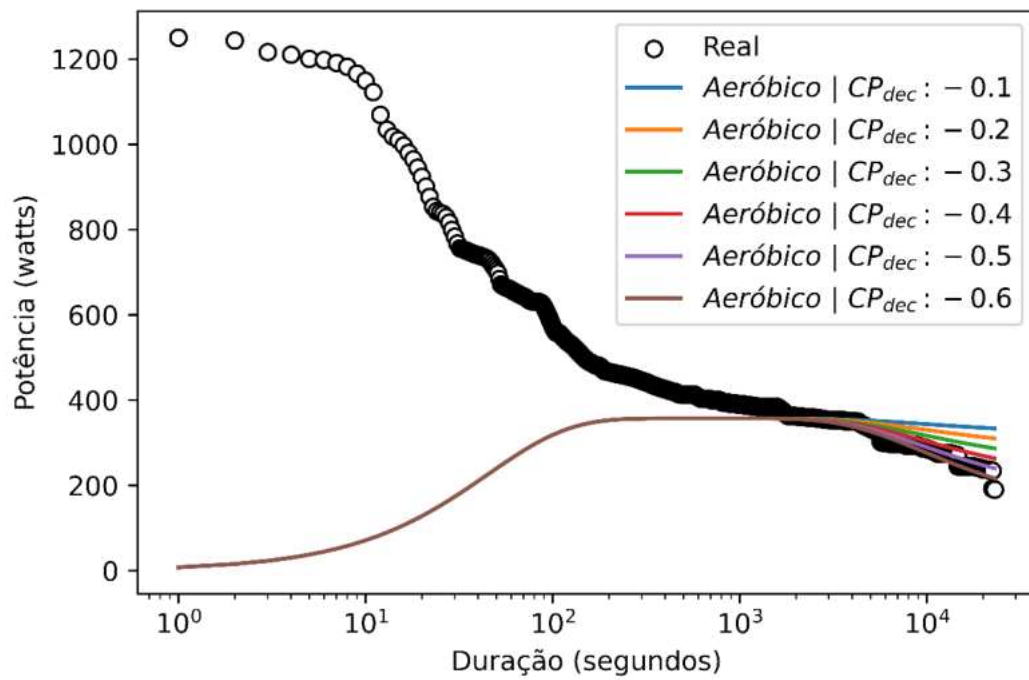


Figura 5.7 – Impacto do parâmetro CP_{dec} no sistema aeróbico.

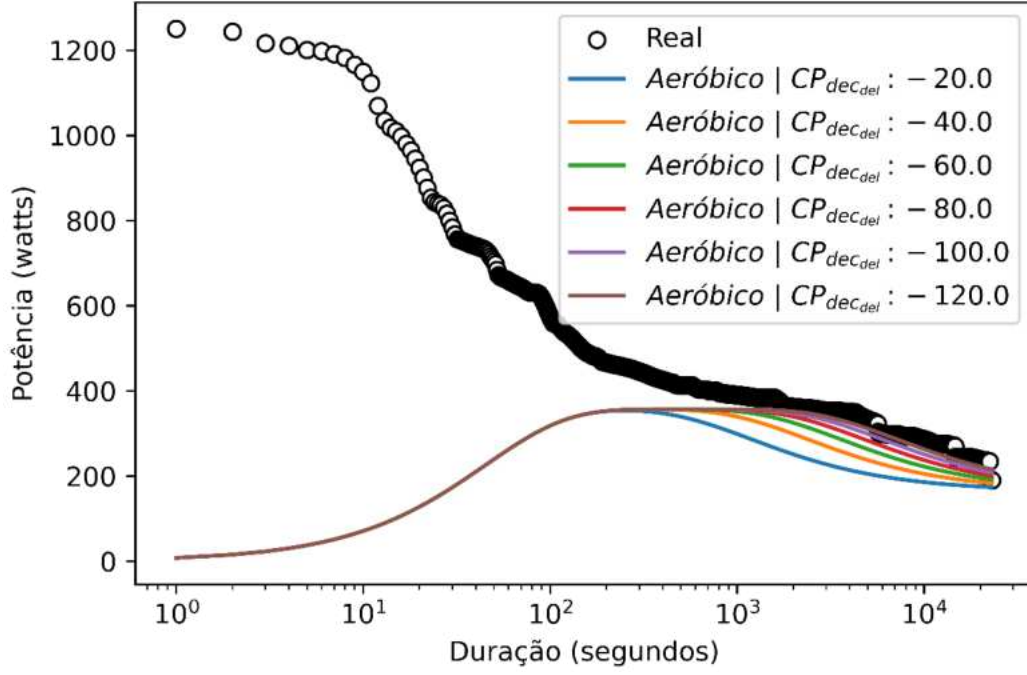


Figura 5.8 – Impacto do parâmetro CP_{dec_del} no sistema aeróbico.

5.8 ESTIMAÇÃO

A estimação dos parâmetros de cada um dos modelos de duração-potência citados, neste trabalho, será realizada seguindo uma mesma metodologia de inferência. Porém, é necessário levar em conta o grande volume de dados submáximos disponíveis, o que inviabiliza a utilização de técnicas de estimação que visam somente a redução do resíduo em relação aos dados observados.

Os dados de MMP de uma única atividade consideram a maior potência média realizada para cada duração. A agregação de um conjunto de atividades para a obtenção de uma MMP mais geral é realizada a partir do cálculo do máximo, duração a duração, da curva $\widehat{MMP}_{i,j}$. Seja um conjunto de atividades J_i do indivíduo i , a curva \widehat{MMP}_{J_i} referente a tal conjunto é dada por:

$$\widehat{MMP}_{J_i}(d) = \max \left(\max_{j \in J_i} \left(\widehat{MMP}_{i,j}(d) \right), \widehat{MMP}_{J_i}(d+1) \right),$$

$$\widehat{MMP}_{J_i}(T) = \max \left(\max_{j \in J_i} \left(\widehat{MMP}_{i,j}(T) \right) \right).$$

A informação empírica por atividade, por sua vez, possui diversos pontos submáximos pois não há possibilidade de um indivíduo realizar esforços máximos para todas as durações num único dia. Entretanto, ao agrupar diversas atividades a quantidade de pontos da curva de MMP representativos de esforços verdadeiramente máximos tende a ser maior.

O método para a estimação aplicado utiliza mínimos quadrados modificados via método de gradientes estocásticos, aproveitando a estrutura de redes neurais para garantir velocidade de execução e viabilizar o estudo com o grande volume de dados. A rede neural possui uma única camada, de certa maneira degenerada, para determinar os parâmetros dos modelos de duração-potência, que a Figura 5.9 busca exemplificar. O intuito é utilizar a metodologia já descrita no Capítulo 2, mas aproveitando a implementação robusta de algoritmos de otimização e o arcabouço tecnológico do *Tensorflow* e *Keras* descritos no Capítulo 3, neste caso o *Adam* (KINGMA; BA, 2014). Entretanto, cada um dos modelos apresentados anteriormente exigiu a criação de uma camada personalizada no ambiente *Keras*.

A taxa de aprendizado utilizada para todos os modelos é $\eta = 0,1$ por um número máximo de 1.000 épocas, entretanto caso não haja melhoria nos valores da função perda avaliada no conjunto de validação em 100 épocas, o algoritmo é interrompido e retorna as estimativas paramétricas da época com menor perda. Nesta primeira aplicação, a cada iteração do algoritmo todos os dados disponíveis são utilizados.

A estimação dos modelos de duração-potência apresentados considera como variável resposta y_d a agregação da curva de MMP empírica para um conjunto de atividades J_i , para o indivíduo i . Assim:

$$y_d = \widehat{MMP}_{J_i}(d),$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_T \end{bmatrix}_{T \times 1},$$

e o modelo ajustado trará resultados para \hat{y}_d . O ajuste é realizado de forma independente para cada indivíduo i e para cada conjunto de atividades J_i .

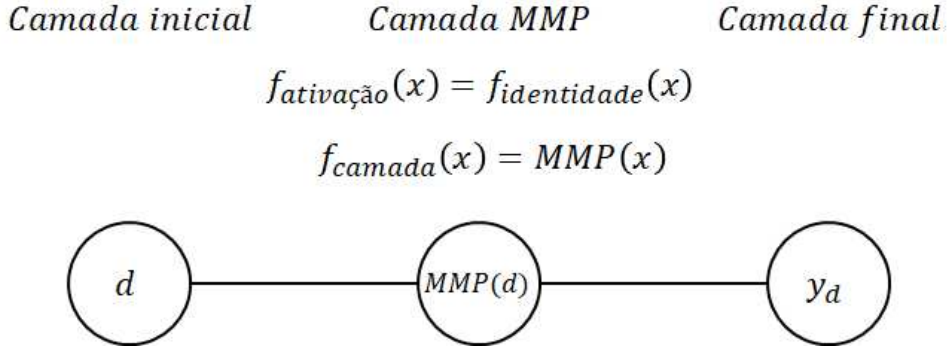


Figura 5.9 – Visualização da rede neural para modelos de duração-potência.

A função perda a ser minimizada tem como base a soma do erro quadrático, $\sum_{d=1}^T (y_d - \hat{y}_d)^2$, entretanto sofre a incorporação de mais componentes para induzir certos comportamentos desejados do ajuste, explicados a seguir.

A curva de MMP empírica traz uma informação valiosa sobre as capacidades do indivíduo: no mínimo o atleta é capaz de realizar ao menos o que foi observado. Entretanto para certas durações o atleta é capaz de produzir mais potência do que os dados informam, devido a este fato o modelo não deveria estimar abaixo da curva empírica já observada. Para isso um dos componentes da função perda $L(y, \hat{y})$ é definido por $\sum_{d=1}^T (f_{ReLU}(y_d - \hat{y}_d))^2$. Ou seja, sempre que o ajuste for abaixo da curva empírica, a componente de erro penalizará o ajuste. Do contrário, quando o ajuste for acima da curva empírica, a função $f_{ReLU}(y_d - \hat{y}_d)$ transformará a diferença $y_d - \hat{y}_d$ em zero, pois $\hat{y}_d > y_d$, $d = 1, 2, \dots, T$. Sendo assim, a penalização age apenas sobre ajustes abaixo da curva de MMP empírica.

Uma outra necessidade da curva ajustada é o comportamento monótono decrescente. Dado que o indivíduo é capaz de produzir um determinado nível de potência para uma determinada duração, isto indica que o atleta é ao menos capaz de produzir a mesma potência para durações menores. Assim, a terceira componente da função perda é definida por $\sum_{d=1}^T (f_{ReLU}(\hat{y}_d - \hat{y}_{d-1}))^2$. Desta maneira, sempre a curva estimada não respeitar o comportamento monótono decrescente, ou seja $\hat{y}_d > \hat{y}_{d-1}$, a componente de erro penalizará o ajuste. Do contrário, quando o ajuste for decrescente, a função $f_{ReLU}(\hat{y}_d - \hat{y}_{d-1})$ transformará a diferença $\hat{y}_d - \hat{y}_{d-1}$ em zero, pois $\hat{y}_d < \hat{y}_{d-1}$, $d = 1, 2, \dots, T$. Sendo assim, a penalização age apenas sobre ajustes crescentes da curva de MMP empírica.

A função perda para a estimação dos parâmetros dos modelos de duração-potência é dada por:

$$L(y, \hat{y}) = 0,10 \cdot \sum_{d=1}^T (y_i - \hat{y}_i)^2 + 0,45 \cdot \sum_{d=1}^T (f_{ReLU}(y_i - \hat{y}_i))^2 + 0,45 \cdot \sum_{d=1}^T (f_{ReLU}(\hat{y}_i - \hat{y}_{i-1}))^2,$$

onde as ponderações utilizadas de 10% para o componente de soma do erro quadrático e 45% para os outros dois componentes de erro, são utilizados para garantir que o ajuste paramétrico considere de forma mais enfática a importância dos comportamentos desejados (curva ajustada acima dos dados empíricos e monótona decrescente). A utilização de ponderações ligeiramente diferentes não afeta o resultado, porém é importante que se mantenha a prioridade da mesma maneira.

Os valores iniciais, importantes para algoritmos de otimização em geral, auxiliam o ajuste pois certos modelos são bastante sensíveis. Após diferentes

experimentações, inspeção visual e velocidade de otimização, os pontos de partida do algoritmo de otimização *Adam* foram definidos conforme a Tabela 5.1. Além disso, também se faz importante a definição de intervalos de variação dos parâmetros, principalmente para os modelos mais flexíveis, de forma a produzir resultados factíveis com a realidade, sendo estes limites explícitos na Tabela 5.2.

A utilização destes limites de variação dos parâmetros no processo de estimação é feita de forma simples. Seja x_- o limite inferior e x_+ o limite superior, então

$$x = \begin{cases} \max(x, x_-), & \text{se } x < x_- \\ \min(x, x_+), & \text{se } x > x_+. \\ x, & \text{caso contrário} \end{cases}$$

Tabela 5.1 – Valores iniciais dos parâmetros de modelos de duração-potência.

| Parâmetro | Valor inicial | Parâmetro | Valor inicial |
|----------------|---------------|---------------------|---------------|
| CP | 200,0 | P_{aa} | 1000,0 |
| W' | 15000,0 | $P_{aa_{dec}}$ | -2,0 |
| P_{max} | 1000,0 | CP_{del} | -2,0 |
| τ_{ExpCP} | 300,0 | CP_{dec} | -1,0 |
| A | 10,0 | $CP_{dec_{del}}$ | -50,0 |
| TCP_{max} | 1800,0 | $\tau_{ExtendedCP}$ | 1,0 |

Tabela 5.2 – Limite de variação dos parâmetros de modelos de duração-potência.

| Parâmetro | Limite inferior | Limite superior |
|---------------------|-----------------|-----------------|
| CP | 100,0 | 500,0 |
| W' | 5000,0 | 50000,0 |
| P_{max} | 500,0 | 2000,0 |
| τ_{ExpCP} | 1,0 | 500,0 |
| A | 1,0 | 100,0 |
| TCP_{max} | 1200,0 | 5400,0 |
| P_{aa} | 500,0 | 2000,0 |
| $P_{aa_{dec}}$ | -20,0 | -1,0 |
| CP_{del} | -20,0 | -1,0 |
| CP_{dec} | -2,0 | -0,5 |
| $CP_{dec_{del}}$ | -200,0 | -1,0 |
| $\tau_{ExtendedCP}$ | 0,1 | 2,0 |

5.9 MODELO FINAL

A quantificação da performance é decorrente da aplicação de um dos modelos mencionados aos dados de atividades de ciclismo, com medição de potência, de diversos indivíduos. O modelo final a ser utilizado foi escolhido a partir da qualidade da predição, além do tempo de execução da estimação dos parâmetros do modelo.

O maior número de dados de potência máxima de um indivíduo é obtido, de forma empírica, a partir da observação de todas as atividades disponíveis. Entretanto, para avaliar a capacidade de predição de um modelo para dados fora do conjunto de treino, é necessário restringir o número de atividades sobre as quais os parâmetros serão estimados.

O método de mensuração da qualidade preditiva dos modelos de duração-potência foi estruturado selecionando de forma aleatória percentuais específicos do total de atividades do indivíduo, sem reposição, para realização do ajuste. Com o modelo ajustado, foi avaliada a métrica raiz do erro quadrático médio (RMSE – *Root Mean Squared Error*), definida por $\sqrt{\frac{1}{T} \sum_{d=1}^T (y_d - \hat{y}_d)^2}$, sendo os dados reais, representados por $\mathbf{y} = [y_1, y_2, \dots, y_T]'$, fornecidos pela curva de MMP empírica do conjunto de todas as atividades do indivíduo, enquanto a curva ajustada é representada por $\hat{\mathbf{y}} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_T]'$, obtida com base numa fração das atividades do indivíduo. Também avaliou-se a métrica erro absoluto médio (MAE – *Mean Absolute Error*), definida por $\sqrt{\frac{1}{T} \sum_{d=1}^T |y_d - \hat{y}_d|}$.

O percentual de atividades selecionadas para realizar o ajuste foi investigado, a fim de compreender o comportamento das métricas observadas para cada modelo. Foi realizada a comparação de diferentes percentuais considerando somente 10% dos indivíduos, por conta do prolongado tempo de processamento.

Os resultados da Tabela 5.3, Tabela 5.4, Tabela 5.5, Tabela 5.6, Tabela 5.7 e Tabela 5.8 indicam que não há diferença na qualidade dos modelos conforme altera-se o número de atividades a serem utilizadas para a estimação dos parâmetros dos modelos de duração-potência. Apesar disso, é possível notar que os modelos Omni e Estendido possuem menores valores para média e mediana dos erros em ambas as métricas, MAE e RMSE, quando comparados aos demais modelos. Com isso, a análise final com todos os indivíduos, considerou somente 20% das atividades para o ajuste do modelo.

A justificativa para a pequena utilização de atividades para ajuste do modelo é decorrente do contexto em que os dados se encontram. Em geral, as atividades que contribuem para a curva de MMP empírica, a partir dos dados do total de

sessões de treino de um indivíduo, são poucas pois necessitam de esforços máximos. Tais situações ocasionalmente ocorrem devido a testes pré-determinados ou competições, conseqüentemente quanto maior a proporção de atividades utilizadas para a estimação dos parâmetros dos modelos de duração-potência, menos será avaliada a capacidade preditiva e mais será avaliado o ajuste aos dados.

Tabela 5.3 – Análise descritiva do MAE dos modelos de duração-potência com 20% de atividades para ajuste.

| Modelo | MAE | | | | | | |
|-----------------|----------------|----------------|---------------|----------------|----------------|----------------|-----------------|
| | Média | Desvio-Padrão | Mínimo | 1º Quartil | Mediana | 3º Quartil | Máximo |
| Dois Parâmetros | 38,0368 | 21,1323 | 12,23 | 25,9605 | 32,6889 | 42,3394 | 196,0530 |
| Três Parâmetros | 36,2582 | 21,4902 | 8,8818 | 23,9617 | 31,0312 | 41,2043 | 209,2735 |
| Exponencial | 37,3368 | <u>18,7269</u> | 14,3353 | 26,1068 | 32,7749 | 41,2069 | 175,5183 |
| Omni | <u>24,9982</u> | 23,7522 | <u>1,9238</u> | <u>11,0297</u> | <u>17,8766</u> | <u>29,2402</u> | 199,4203 |
| Estendido | 28,0480 | 26,3873 | 3,1500 | 12,2805 | 19,3050 | 33,0673 | <u>172,6954</u> |

Tabela 5.4 – Análise descritiva do RMSE dos modelos de duração-potência com 20% de atividades para ajuste.

| Modelo | RMSE | | | | | | |
|-----------------|----------------|----------------|---------------|----------------|----------------|----------------|-----------------|
| | Média | Desvio-Padrão | Mínimo | 1º Quartil | Mediana | 3º Quartil | Máximo |
| Dois Parâmetros | 142,4651 | 36,9619 | 40,6972 | 127,3869 | 132,1625 | 146,7107 | 332,7636 |
| Três Parâmetros | 48,2738 | 31,4751 | 13,4404 | 30,7942 | 39,7608 | 53,5885 | 317,3740 |
| Exponencial | 54,5142 | <u>25,5840</u> | 20,5314 | 40,1349 | 47,4987 | 58,7792 | <u>295,3692</u> |
| Omni | 36,4974 | 33,7364 | <u>4,1610</u> | 17,1692 | 25,8007 | 42,3037 | 304,7423 |
| Estendido | <u>35,6871</u> | 32,9777 | 5,5027 | <u>16,6106</u> | <u>24,2808</u> | <u>41,3589</u> | 304,9171 |

Tabela 5.5 – Análise descritiva do MAE dos modelos de duração-potência com 50% de atividades para ajuste.

| Modelo | MAE | | | | | | |
|-----------------|----------------|----------------|---------------|---------------|----------------|----------------|-----------------|
| | Média | Desvio-Padrão | Mínimo | 1º Quartil | Mediana | 3º Quartil | Máximo |
| Dois Parâmetros | 39,3747 | 19,5179 | 12,9602 | 27,9592 | 34,4507 | 44,1514 | 160,6830 |
| Três Parâmetros | 37,4768 | 19,9588 | 6,9838 | 25,9360 | 32,8950 | 42,8231 | 160,5987 |
| Exponencial | 38,6880 | <u>16,9954</u> | 13,8906 | 28,7132 | 33,9583 | 43,2338 | 143,4004 |
| Omni | <u>22,4748</u> | 19,8845 | 2,3650 | <u>8,7225</u> | 16,6808 | <u>28,9860</u> | <u>133,5321</u> |
| Estendido | 23,4483 | 20,7811 | <u>2,1831</u> | 9,6603 | <u>16,4047</u> | 29,4418 | 155,4603 |

Tabela 5.6 – Análise descritiva do RMSE dos modelos de duração-potência com 50% de atividades para ajuste.

| Modelo | RMSE | | | | | | |
|-----------------|----------------|----------------|---------------|----------------|----------------|----------------|-----------------|
| | Média | Desvio-Padrão | Mínimo | 1º Quartil | Mediana | 3º Quartil | Máximo |
| Dois Parâmetros | 144,0219 | 37,1377 | 36,7463 | 127,0772 | 132,1067 | 150,2436 | 330,2710 |
| Três Parâmetros | 48,6341 | 30,5290 | 11,6493 | 32,4436 | 40,7312 | 54,5247 | 314,7814 |
| Exponencial | 55,8891 | 23,3359 | 22,3008 | 41,8718 | 49,1295 | 61,9553 | <u>293,5449</u> |
| Omni | 33,9057 | 31,7169 | <u>3,8403</u> | 15,3297 | 25,4812 | 40,6185 | 303,7115 |
| Estendido | <u>30,6177</u> | <u>28,6019</u> | 5,3408 | <u>13,5819</u> | <u>21,7617</u> | <u>36,8975</u> | 297,4533 |

Tabela 5.7 – Análise descritiva do MAE dos modelos de duração-potência com 80% de atividades para ajuste.

| Modelo | MAE | | | | | | |
|-----------------|----------------|----------------|---------------|---------------|----------------|----------------|-----------------|
| | Média | Desvio-Padrão | Mínimo | 1º Quartil | Mediana | 3º Quartil | Máximo |
| Dois Parâmetros | 39,3815 | 18,4702 | 10,8595 | 28,8737 | 35,5027 | 43,4352 | 155,4973 |
| Três Parâmetros | 37,5701 | 18,9945 | 6,3250 | 26,9794 | 34,1189 | 42,2290 | 154,1374 |
| Exponencial | 38,5659 | <u>15,4286</u> | 14,1278 | 29,3991 | 35,2572 | 42,5474 | 138,1063 |
| Omni | <u>21,6314</u> | 19,3802 | <u>1,8201</u> | <u>8,2301</u> | 16,1455 | 29,9305 | 151,0583 |
| Estendido | 21,6975 | 19,4589 | 2,6458 | 8,7383 | <u>15,2338</u> | <u>25,7500</u> | <u>122,3365</u> |

Tabela 5.8 – Análise descritiva do RMSE dos modelos de duração-potência com 80% de atividades para ajuste.

| Modelo | RMSE | | | | | | |
|-----------------|----------------|----------------|---------------|----------------|----------------|----------------|-----------------|
| | Média | Desvio-Padrão | Mínimo | 1º Quartil | Mediana | 3º Quartil | Máximo |
| Dois Parâmetros | 143,6087 | 34,8716 | 48,9897 | 127,8667 | 133,1862 | 151,3453 | 303,5871 |
| Três Parâmetros | 48,7523 | 27,7481 | 11,7373 | 33,1178 | 41,7516 | 53,9942 | 277,5311 |
| Exponencial | 55,6629 | <u>20,8164</u> | 25,3340 | 43,7810 | 49,7459 | 61,9156 | <u>251,5151</u> |
| Omni | 32,9123 | 29,4494 | <u>4,4134</u> | 13,8008 | 24,3898 | 40,1301 | 276,1965 |
| Estendido | <u>28,5416</u> | 26,2088 | 5,2312 | <u>12,9578</u> | <u>19,8116</u> | <u>33,0446</u> | 263,4706 |

Os resultados da avaliação final estão dispostos na Tabela 5.9 e Tabela 5.10. Considerando os valores, o modelo escolhido como melhor com capacidade preditiva é o modelo Omni. Assim como já observado anteriormente, nas análises finais novamente os modelos Omni e Estendido se destacam dos demais. Apesar disso, levando em conta que MAE e RMSE selecionaram modelos diferentes, a simplicidade do modelo Omni em relação ao Estendido é significativa e atende aos princípios da parcimônia. Então, o modelo Omni escolhido será utilizado para o restante das análises deste trabalho, participando da geração dos parâmetros de performance dos atletas.

Tabela 5.9 – Análise descritiva final do MAE dos modelos de duração-potência.

| Modelo | MAE | | | | | | |
|-----------------|----------------|----------------|---------------|----------------|----------------|----------------|-----------------|
| | Média | Desvio-Padrão | Mínimo | 1º Quartil | Mediana | 3º Quartil | Máximo |
| Dois Parâmetros | 38,7331 | 19,9806 | 8,5881 | 27,0932 | 33,9912 | 43,3669 | 192,1343 |
| Três Parâmetros | 36,8573 | 20,3486 | 4,1612 | 25,0393 | 32,3474 | 42,0346 | 192,5060 |
| Exponencial | 38,0376 | <u>17,9993</u> | 9,4019 | 27,5360 | 33,6552 | 42,2853 | <u>179,4504</u> |
| Omni | <u>25,2237</u> | 20,9204 | <u>1,7591</u> | <u>11,8551</u> | <u>19,5878</u> | <u>31,4921</u> | 186,8829 |
| Estendido | 27,9474 | 22,6193 | 2,6309 | 13,1927 | 21,2045 | 34,8209 | 191,8664 |

Tabela 5.10 – Análise descritiva final do RMSE dos modelos de duração-potência.

| Modelo | RMSE | | | | | | |
|-----------------|----------------|----------------|---------------|----------------|----------------|----------------|-----------------|
| | Média | Desvio-Padrão | Mínimo | 1º Quartil | Mediana | 3º Quartil | Máximo |
| Dois Parâmetros | 142,3697 | 34,4395 | 24,9689 | 126,9468 | 132,2091 | 147,9092 | 335,1649 |
| Três Parâmetros | 48,0002 | 28,7596 | 8,3880 | 31,7806 | 41,0176 | 53,6241 | 328,1110 |
| Exponencial | 54,3146 | <u>23,2184</u> | 17,5989 | 41,2817 | 47,9995 | 59,3527 | <u>309,6138</u> |
| Omni | 35,7940 | 29,8290 | <u>3,9371</u> | 17,6753 | 27,6106 | 42,8714 | 325,5551 |
| Estendido | <u>34,8444</u> | 28,8103 | 4,4758 | <u>17,4249</u> | <u>26,1246</u> | <u>41,6989</u> | 311,6345 |

Os modelos de dois parâmetros, três parâmetros e exponencial possuem uma limitação pois consideram o parâmetro CP como a assíntota, presumindo uma potência que é capaz ser mantida por duração indefinida. Este comportamento não é observado nos dados empíricos, conforme Figura 5.10 e Figura 5.11.

Os modelos Omni e Estendido, por sua vez, possuem uma vantagem em relação ao demais pois consideram um comportamento decrescente do parâmetro CP a partir de determinada duração, conforme a Figura 5.12, isso faz com que se aproxime do comportamento dos dados observados aprimorando o ajuste e a predição.

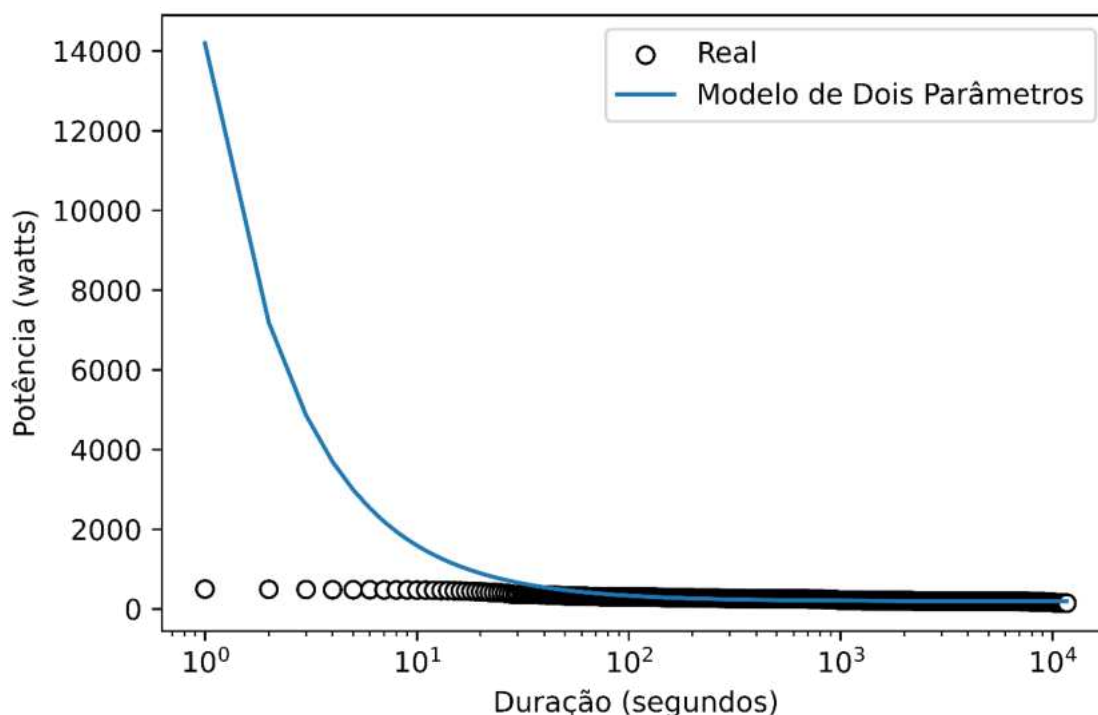


Figura 5.10 – Exemplo de ajuste do modelo de dois parâmetros.

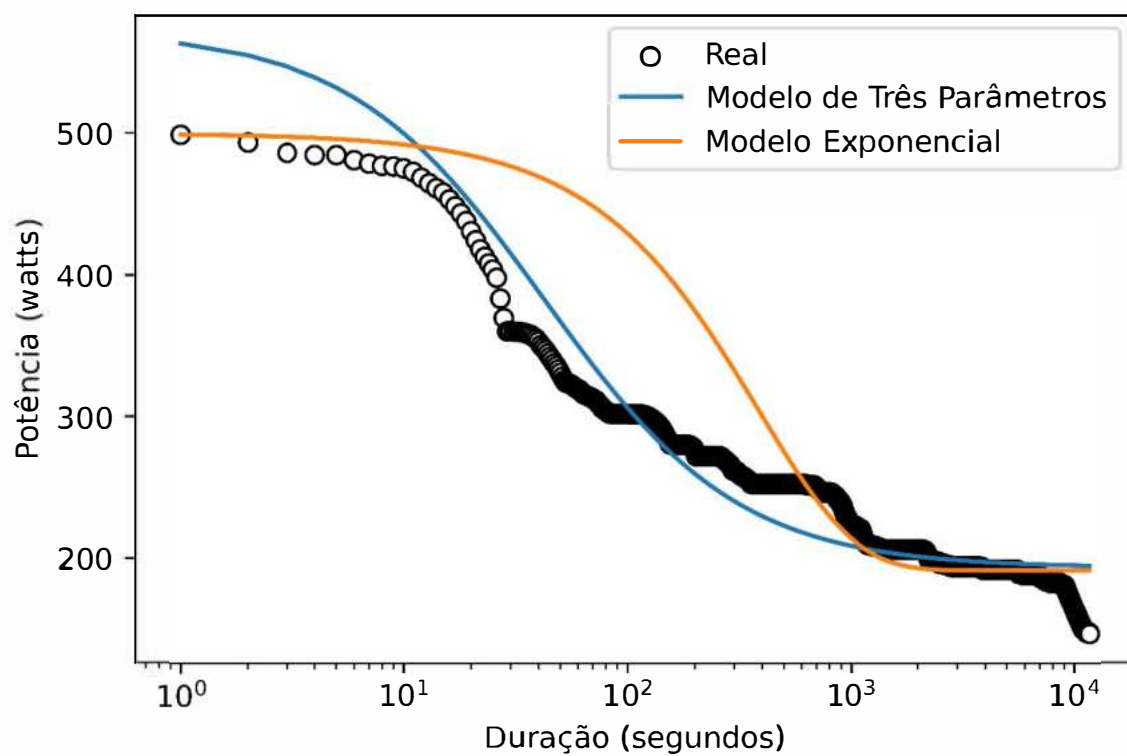


Figura 5.11 – Exemplo de ajuste dos modelos de três parâmetros e exponencial.

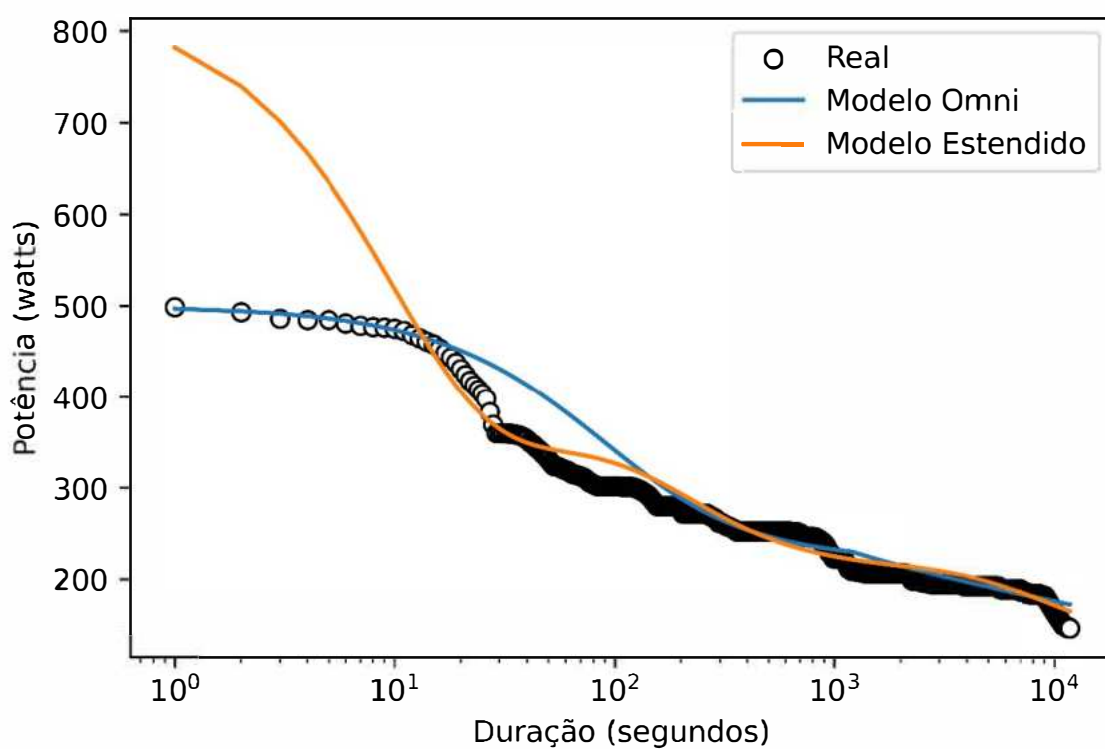


Figura 5.12 – Exemplo de ajuste dos modelos Omni e Estendido.

Os modelos de duração-potência apresentados neste capítulo possuem um grande valor pois, a partir de uma certa quantidade de parâmetros, é possível descrever as capacidades de produção de potência, por um atleta, em variadas durações.

O objetivo deste capítulo é alcançado pela definição do modelo Omni como o mais adequado para tornar possível a quantificação de atributos de performance de um ciclista, a partir dos treinamentos realizados por ele até então. O resultado da análise, tomando como mais adequado o modelo Omni, será utilizado no capítulo seguinte, para quantificar a performance dos indivíduos ao longo do tempo, possibilitando entender como o treinamento é responsável por alterações na performance ao longo dos períodos de treinamentos pelos quais o atleta passa.

6 RELAÇÃO ENTRE TREINAMENTO E PERFORMANCE

A relação entre treinamento, os estímulos que tornam o atleta mais capacitado e mais preparado, e a performance, uma observação do desempenho do indivíduo em função de suas habilidades, é um objetivo muito importante para a ciência do esporte, não só no ciclismo.

O primeiro desafio é a transformação dos dados de treinamento em variáveis mais práticas para a utilização em modelos estatísticos, assunto abordado no Capítulo 4. O segundo desafio é a definição e estimação de parâmetros que permitam avaliar diferentes aspectos que caracterizam um ciclista, conforme abordado no Capítulo 5.

A importância de estabelecer uma relação treinamento-performance de fácil interpretação, que permita a compreensão clara do impacto de diferentes padrões de treinamento na performance de um atleta, pode ser desdobrada em diferentes ganhos. De posse deste esclarecimento é possível otimizar o planejamento de treino para que se possa melhorar a performance do atleta, visando os resultados desejados numa competição específica.

Em outras palavras, a subjetividade em torno da prescrição de atividades aos atletas pode ser reduzida, uma vez que ficará mais evidente o impacto de escolhas específicas para se tornar um atleta melhor.

Este trabalho se diferencia de outros trabalhos de estatística aplicada a esportes em alguns aspectos. Alguns exemplos de aplicações de modelos, para estudo da relação entre treinamento e performance de atletas, na literatura utilizam dados provenientes de outros esportes, porém com uma quantidade de dados consideravelmente reduzida. Como exemplos podemos citar Morton, Fitz-Clarke e Banister (1990) e Kosmidis e Passfield (2015). Além disso, considerando

a simplicidade em quantificar a performance em esportes como corrida (KOSMIDIS; PASSFIELD, 2015) ou natação (AVALOS; HELLARD; CHATARD, 2003), mencionado no Capítulo 4, o foco deste trabalho em dados de ciclismo também é outro aspecto inovador.

Como inovação deste trabalho, destaca-se ainda o tamanho da base de dados utilizada, conforme o Capítulo 4, e pela estimação paramétrica de um modelo que considera o impacto do treinamento sobre distintas variáveis de performance, de forma conjunta. Por conta desses fatores, a estimação dos modelos que serão descritos a seguir utilizarão o arcabouço teórico de redes neurais, apresentado no Capítulo 2, e os algoritmos de otimização utilizados são os métodos de gradientes estocásticos (*Adam*), apresentados no Capítulo 3.

6.1 SÉRIE TEMPORAL DA PERFORMANCE

A avaliação de mudanças e impactos gerados por uma certa intervenção é usualmente precedida e sucedida pela coleta das variáveis, as quais espera-se que sofram alteração, estabelecendo assim a relação de causa-efeito. A aplicação deste racional à prescrição de um planejamento de treino, traz a necessidade de quantificar os parâmetros de performance antes e depois do período de treinamento.

Os dados coletados a partir do repositório *GoldenCheetah OpenData Project* (LIVERSEDGE, 2018) não possuem marcações temporais de início e fim de períodos de treinamento, ou seja, um conjunto específico de atividades de treinamento com foco específico para melhorar parâmetros de performance mensurados previamente ao início de tão período, então foram consideradas janelas móveis para seleccionar as atividades que serão utilizadas para o cálculo da

curva de MMP empírica, insumo para o modelo Omni para quantificação da performance, descrito na Seção 5.6.

As janelas móveis são definidas a partir de duas quantidades, o tamanho da janela e o deslocamento da janela, ambos medidos em dias. Considere o conjunto de atividades J_i do indivíduo i , onde foi realizada 1 atividade em cada um de 10 dias consecutivos totalizando 10 atividades, desta forma a atividade $j = 1$ corresponde ao primeiro dia e a atividade $j = 10$ corresponde ao décimo dia. Neste caso, j representa uma atividade contida no conjunto J_i .

Uma janela de 5 dias com deslocamento de 5 dias, para este indivíduo, gerará então 2 conjuntos de atividades, o primeiro conjunto com as atividades $j = 1$ até $j = 5$, e o segundo conjunto com as atividades $j = 6$ até $j = 10$, como ilustrado na Figura 6.1, onde cada círculo representa um dia de atividades. Entretanto, utilizando uma janela de 8 dias com deslocamento de 1 dia, gerará conjuntos de atividades que possuem sobreposição, neste caso os conjuntos com atividades $j = 1$ até $j = 8$, $j = 2$ até $j = 9$ e $j = 3$ até $j = 10$, como ilustrado na Figura 6.2.

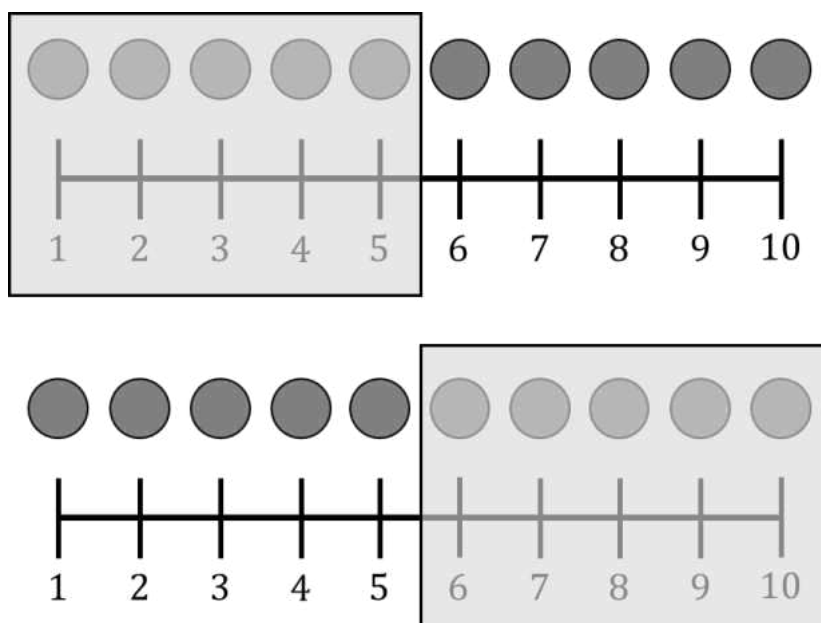


Figura 6.1 – Exemplo de janela móvel 5/5.

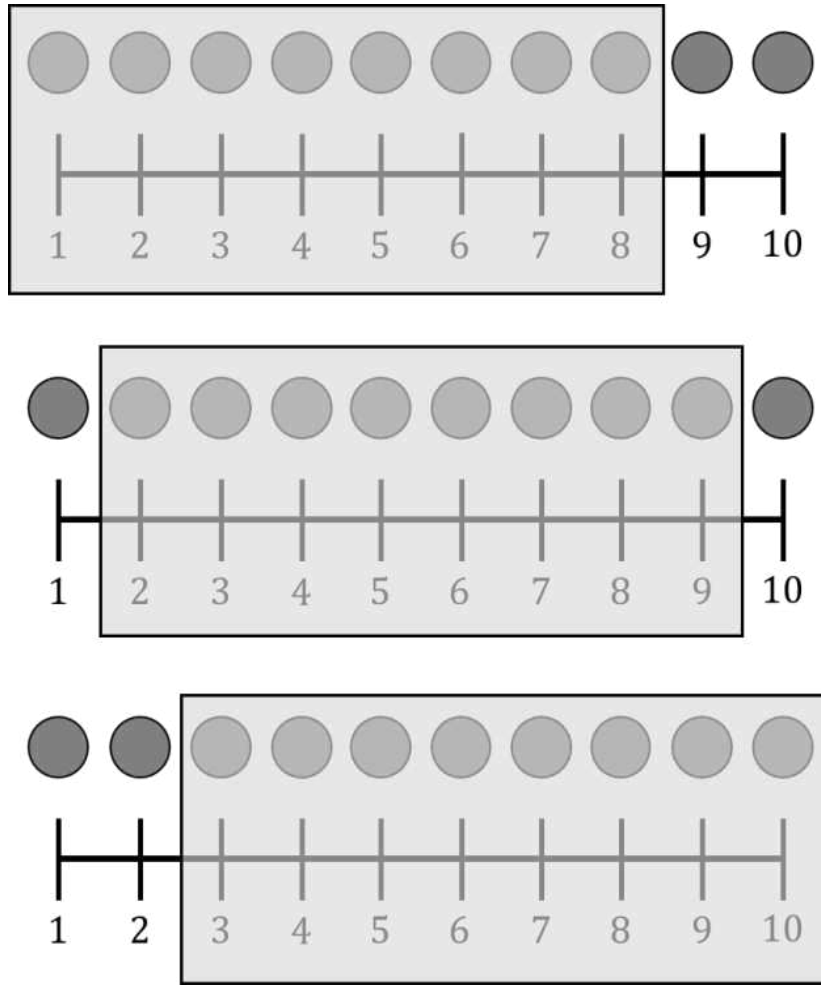


Figura 6.2 – Exemplo de janela móvel 8/1.

Considere o conjunto de atividades, A_{ip} , e a curva de MMP empírica associada a tais atividades, $\widehat{MMP}_{A_{ip}}$, onde i indica o indivíduo e p indica o período de treinamento. A partir das curvas de MMP empíricas para cada janela de atividades, é possível estimar os parâmetros do modelo Omni, como descrito no Capítulo 5. Ao utilizar as datas do final das janelas construídas como data de referência, obtém-se então a série temporal das estimativas dos parâmetros de performance CP , W' , P_{max} , A e TCP_{max} , pelo modelo Omni. As séries temporais de cada parâmetro estão ilustradas na Figura 6.3, Figura 6.4, Figura 6.5, Figura 6.6 e Figura 6.7, utilizando um mesmo indivíduo como exemplo e a configuração de janelas 84/84, onde cada ponto denota a estimativa realizada utilizando a metodologia descrita na Seção 5.8.

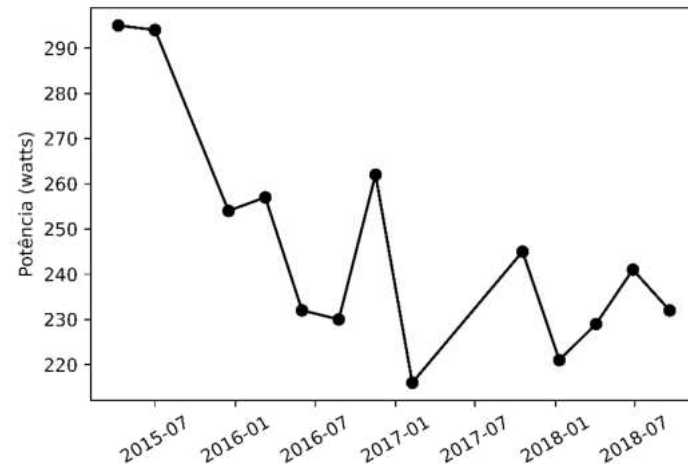


Figura 6.3 – Exemplo de série temporal do parâmetro CP .

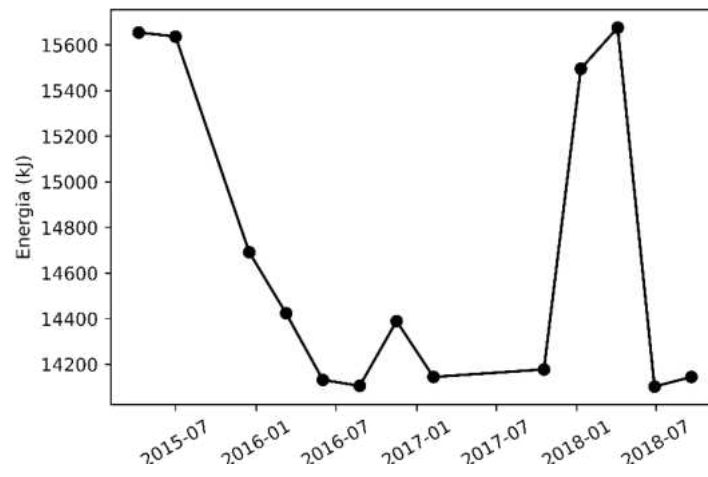


Figura 6.4 – Exemplo de série temporal do parâmetro W' .

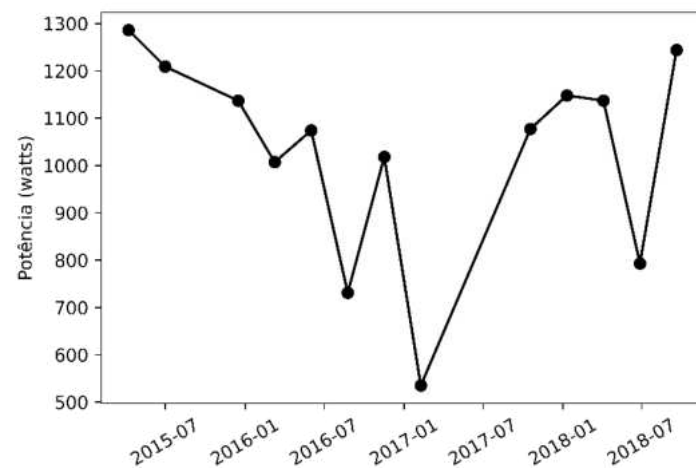


Figura 6.5 – Exemplo de série temporal do parâmetro P_{max} .

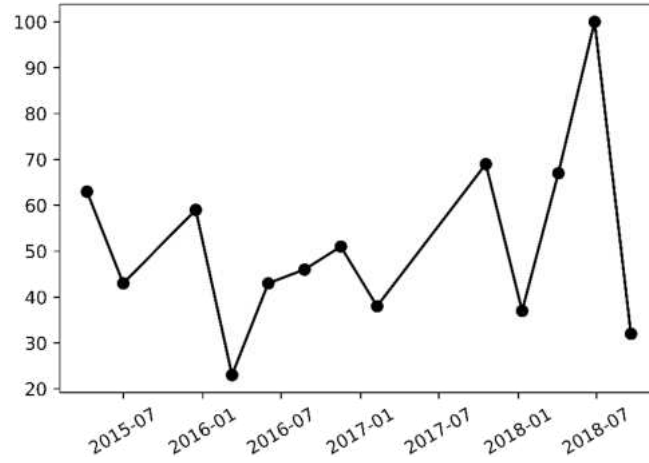


Figura 6.6 – Exemplo de série temporal do parâmetro A .

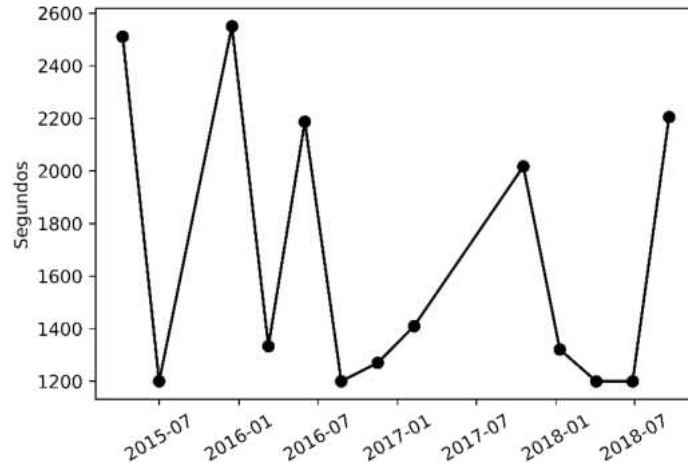


Figura 6.7 – Exemplo de série temporal do parâmetro TCP_{max} .

As configurações de janelas utilizadas neste trabalho consideraram um tamanho de 84 dias (12 semanas ou aproximadamente 3 meses) porém com duas opções de deslocamento, 84 e 28 dias, para investigar as diferenças na capacidade dos modelos em função da ocorrência ou não de sobreposição.

A configuração 84/84 não apresenta o fenômeno de sobreposição, ao contrário da configuração 84/28. A escolha do período de 84 dias é motivada pela preocupação de que a curva $\widehat{MMP}_{A_{ip}}$ contenha, para cada conjunto de atividades, pontos de esforço máximo suficientes para que a estimação paramétrica não seja prejudicada.

As variáveis-resposta, utilizadas nos modelos que serão introduzidos a seguir, são construídas a partir da diferença (Δ) entre dois períodos consecutivos. Seja por exemplo o parâmetro CP_{ip} , o valor da potência crítica para o indivíduo i no p -ésimo período de treinamento, originado a partir da estimação do modelo Omni utilizando o conjunto de atividades A_{ip} . Então a variável-resposta é dada por $\Delta CP_{ip} = CP_{ip} - CP_{i(p-1)}$. As variáveis que visam mensurar a quantidade e padrão das sessões de treinamento realizadas, são geradas considerando o conjunto de atividades A_{ip} . Os outros parâmetros de performance são manipulados do mesmo modo. Um fluxograma que representa o processo de consolidação da base dados, para atingir o objeto de estimar a relação treinamento-performance, para facilitar o entendimento da influência do treinamento na variação da performance, é apresentado na Figura 6.8.

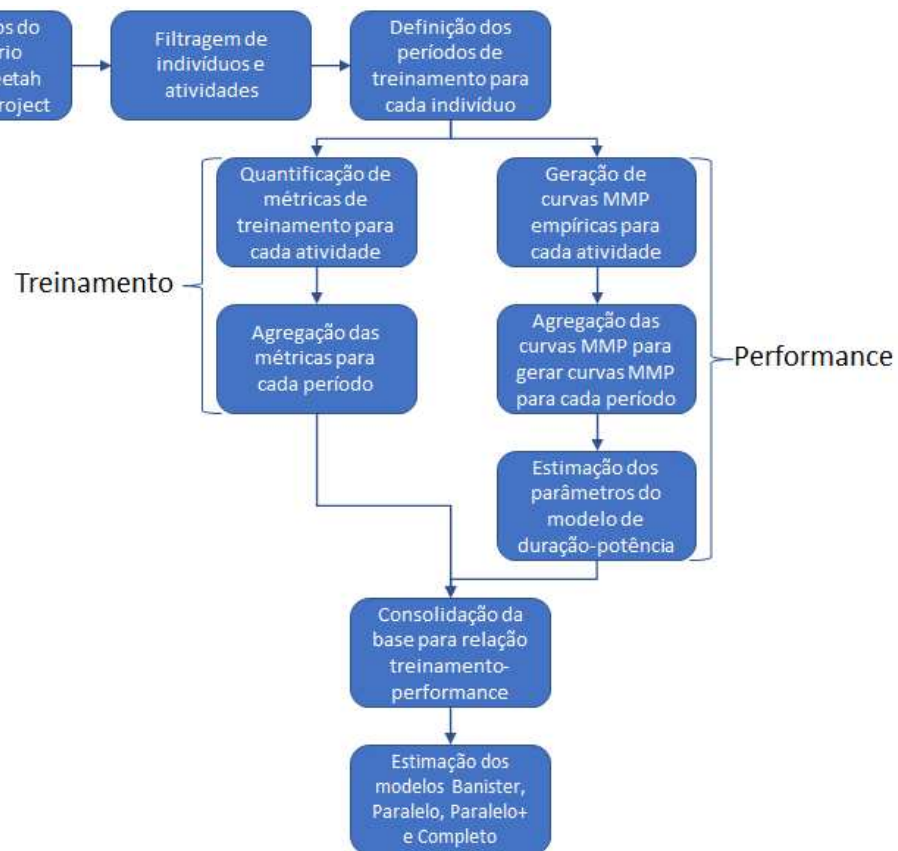


Figura 6.8 – Fluxograma do processo de consolidação da base de dados para modelos treinamento-performance

6.2 MODELO BANISTER

O modelo Banister (BANISTER *et al.*, 1975) é bastante utilizado na literatura para modelar a relação entre treinamento e performance, e por isso será utilizado por este trabalho como um modelo de base de comparação, visando trazer evidências sobre a validade da inovação apresentada neste trabalho.

O modelo pode ser definido, recursivamente, como

$$P_{p+1} = P_p + k_1 \sum_{n=0}^{n_p-1} w_n e^{-\frac{n_p-n}{r_1}} - k_2 \sum_{n=0}^{n_p-1} w_n e^{-\frac{n_p-n}{r_2}},$$

onde P_p é a quantificação da performance no período de treinamento p . As constantes k_1 e k_2 medem a influência positiva e negativa do treinamento sobre a performance, respectivamente. w_n é a quantificação do n -ésimo dia de treinamento no período analisado. n_p representa a quantidade de dias totais no período p . r_1 e r_2 representam o decaimento exponencial dos efeitos positivos e negativos do treinamento, respectivamente. Todas as quantidades apresentadas na equação são adimensionais e assumem valores positivos.

A construção do modelo se baseia no entendimento de que uma atividade possui dois tipos de efeitos, uma resposta crônica geralmente associada com os benefícios e uma resposta aguda geralmente associada com os prejuízos. A influência das atividades se torna cada vez menos significativa conforme se distancia temporalmente do momento da medição de performance ao fim do período de treinamento.

A equação, com uma simples transformação algébrica e utilizando o TSS (Seção 4.1) como quantificação do treinamento, pode também ser definida como

$$\begin{aligned}
f_{Banister}(\mathbf{TSS}_p, n_p) &= \Delta P_{p+1} = P_{p+1} - P_p \\
&= k_1 \sum_{n=0}^{n_p-1} TSS_n e^{-\frac{n_p-n}{r_1}} - k_2 \sum_{n=0}^{n_p-1} TSS_n e^{-\frac{n_p-n}{r_2}},
\end{aligned}$$

onde $\mathbf{TSS}_p = [TSS_0 \ TSS_1 \ \cdots \ TSS_{n_p-1}]$ é o vetor de TSS para cada dia que compõe o período de treinamento p , sendo o TSS diário a soma do TSS das atividades ocorridas no dia, assumindo o valor zero quando não ocorrer alguma sessão de treino. Neste caso, o modelo Banister é possível de ser aplicado a qualquer variável de performance, porém este trabalho irá utilizar o parâmetro CP como a principal métrica. Mujika e Padilla (2001), ao investigarem as características fisiológicas e de performance de ciclistas profissionais competindo na categoria masculina identificaram a importância do sistema aeróbico para se obter resultados competitivos de alto nível no ciclismo. A relevância do parâmetro CP , neste trabalho, é fundamental pois é o parâmetro que representa a capacidade do sistema aeróbico de um atleta.

A representação da estrutura de redes neurais utilizada para a estimação dos parâmetros do modelo: k_1 , k_2 , r_1 e r_2 , está ilustrada na Figura 6.9. A $f_{Banister}(\mathbf{TSS}_p, n_p)$ é utilizada como função da camada da única camada oculta, tendo como dados na camada inicial o vetor de TSS diário, para cada um dos dias do período de treinamento p , e o número total de dias contidos no período.

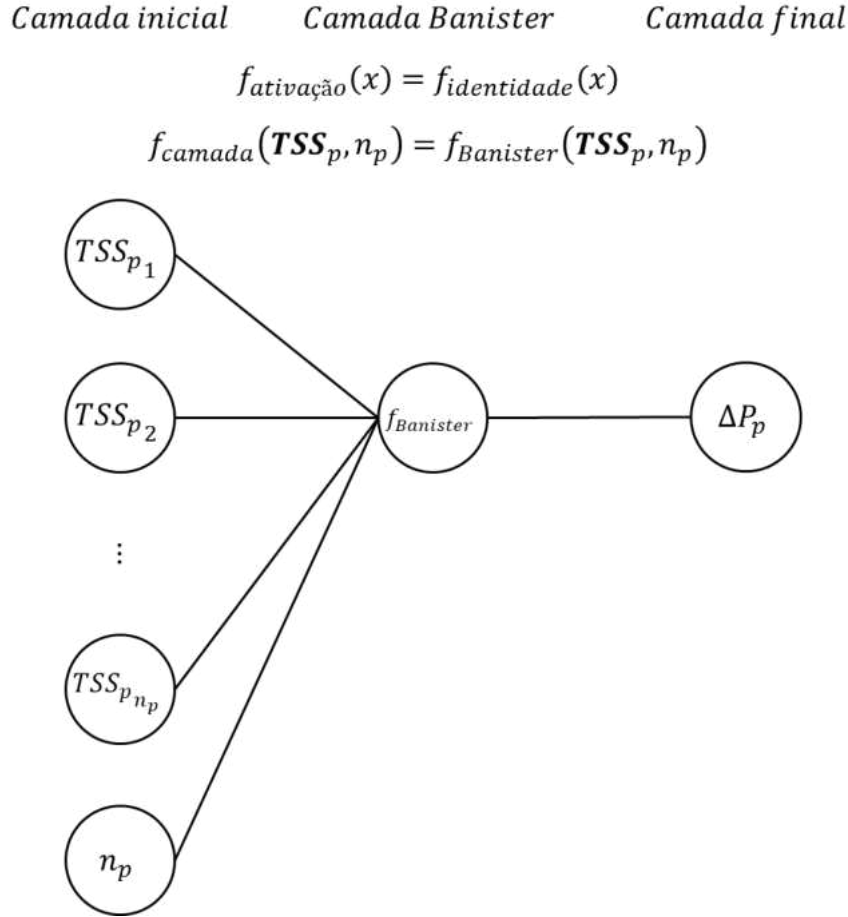


Figura 6.9 – Visualização da arquitetura de rede neural para modelo Banister.

6.3 MODELO PARALELO

Os modelos baseados na estrutura de redes neurais, apesar dos possíveis ganhos em relação à qualidade do ajuste aos dados e a capacidade de generalização, também possuem o estigma da perda da interpretabilidade, pois a interconexão dos neurônios dificulta o processo de visualizar o impacto de uma determinada covariável sobre a variável resposta.

O principal objetivo deste trabalho é compreender como métricas do treinamento se relacionam com métricas de performance. Como primeira proposta desta dissertação para estimar a relação treinamento-performance, foi definida uma arquitetura de redes neurais que mantenha o aspecto não-paramétrico da modelagem da relação entre covariáveis e variáveis resposta, sem renunciar à

interpretação das estimativas dos parâmetros. A arquitetura está ilustrada na Figura 6.10.

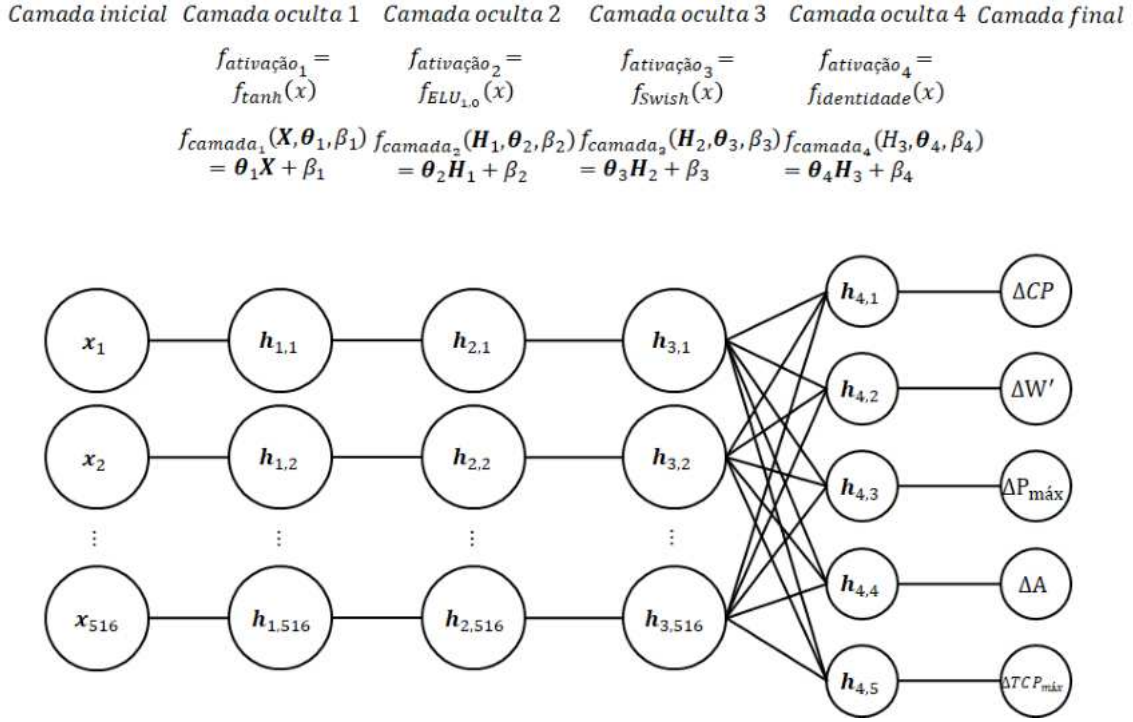


Figura 6.10 – Visualização da arquitetura de rede neural para modelo paralelo.

A estrutura de conexões paralelas, somente permitindo a interação entre variáveis na última camada oculta, facilita a interpretação. Por exemplo, o efeito da variável X_1 sobre ΔCP está representado pela aresta que liga $H_{3,1}$ a $H_{4,1}$, que é aditivo com respeito ao efeito das demais covariáveis X_2, \dots, X_{516} e independe analiticamente dos valores observados de X_2 a X_{516} . No total, são 5.681 parâmetros desconhecidos neste modelo e, diferentemente do modelo Banister, são utilizadas as diferenças de cada uma das métricas de performance do modelo Omni: ΔCP , $\Delta W'$, ΔP_{max} , ΔA e ΔTCP_{max} , gerando uma regressão multivariada.

6.4 MODELO PARALELO+

O modelo Paralelo+, segunda proposta desta dissertação, se diferencia do modelo Paralelo pois permite o aumento da dimensão de cada uma das variáveis

de treinamento, a cada camada. Dessa forma, com esse incremento na complexidade, espera-se que haja uma melhor captura das relações não-lineares entre treinamento e performance.

A interpretabilidade ainda é mantida, pois retém a interação entre variáveis explicativas somente à última camada da arquitetura, ilustrada na Figura 6.11. No total, são 11.357 parâmetros desconhecidos a serem estimados.

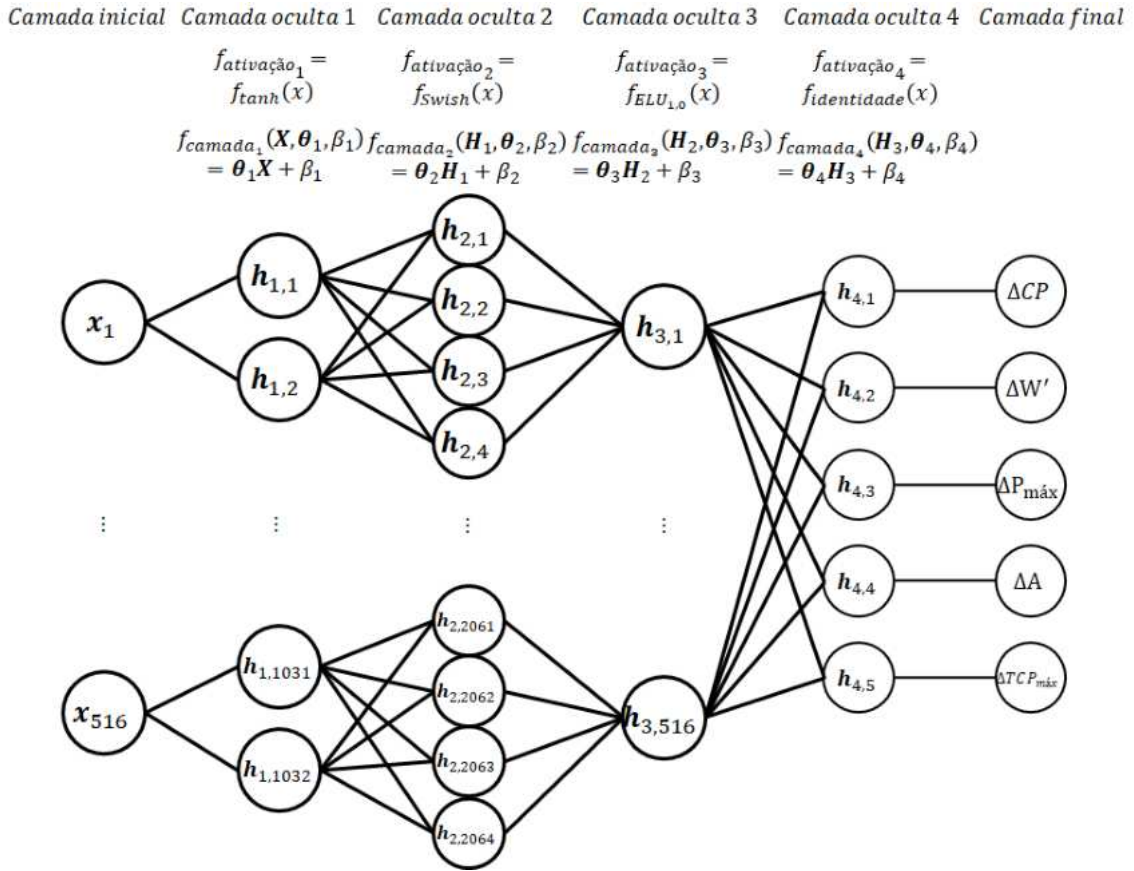


Figura 6.11 – Visualização da arquitetura de rede neural para modelo Paralelo+.

6.5 MODELO COMPLETO

Os modelos Paralelo e Paralelo+ impõem uma simplificação na arquitetura de redes neurais para facilitar a interpretação, entretanto é importante avaliar os possíveis ganhos quando a interpretação dos efeitos de covariáveis sobre a resposta

não for primordial. Assim, a terceira proposta desta dissertação, o modelo Completo tem o objetivo de mensurar os possíveis benefícios de considerar a arquitetura usual completa de redes neurais, onde as camadas são completamente conectadas, para estabelecer a relação entre treinamento e performance no ciclismo de forma mais flexível.

A arquitetura utilizada neste caso permite as interações entre os neurônios em todas as camadas, diferente dos modelos Paralelo e Paralelo+. Em compensação, o número de parâmetros desconhecidos sofre um acréscimo considerável, totalizando 78.305. A arquitetura está ilustrada na Figura 6.12.

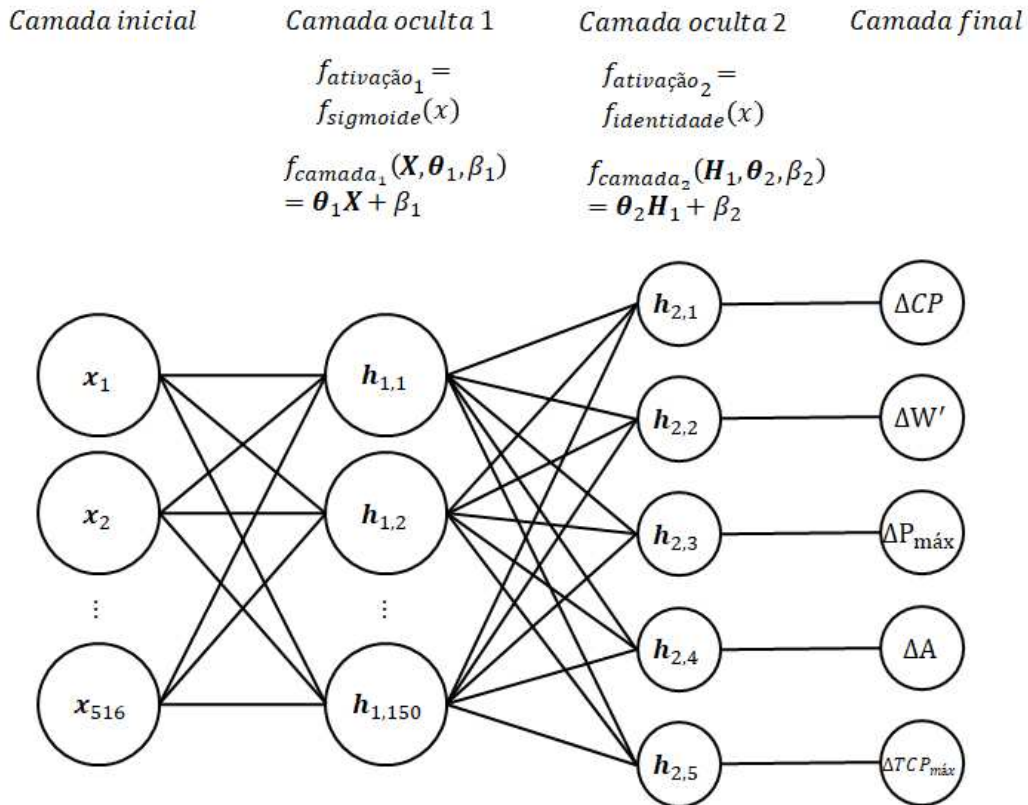


Figura 6.12 – Visualização da arquitetura de rede neural para modelo Completo.

6.6 ESTIMAÇÃO

A estimação dos modelos anteriormente apresentados é realizada utilizando duas configurações distintas de janelas móveis, 84/84 e 84/28, para assim

compreender a influência da sobreposição na avaliação da capacidade preditiva. Outra variação considerada foi a estimação dos parâmetros dos modelos de forma generalizada e individualizada.

A aplicação generalizada se apoia na premissa de que um mesmo modelo é capaz de generalizar as relações treinamento-performance para diversos indivíduos, assim uma única estimação é realizada considerando toda a base de dados. A hipótese desta configuração é de que a relação entre treinamento e performance é invariante ao indivíduo, que pode ser avaliada a depender da heterogeneidade da base de dados analisada. A presença de diferentes grupos de atletas na base de dados, forneceria evidências de que a utilização de um único modelo pode não ser a abordagem mais adequada para o respectivo conjunto de dados. Desta forma, caso este fosse a situação da amostra utilizada por este trabalho, seria esperado que tal hipótese fosse rejeitada.

A simplicidade do modelo Banister, entretanto, pode apresentar melhores resultados caso seja realizada uma estimação por indivíduo, o objetivo da aplicação individualizada. Neste caso, a hipótese é que a relação entre treinamento e performance possui nuances específicas para cada atleta, dessa forma um modelo por indivíduo pode resultar em menores erros de predição.

No fim, formam-se quatro variações para cada modelo previamente definido, generalizado 84/84, generalizado 84/28, individualizado 84/84 e individualizado 84/28.

Os percentuais definidos para selecionar as bases de treino, validação e teste são 50%, 30% e 20%, respectivamente. No caso da separação em treino, validação e teste, utilizando a aplicação generalizada, a seleção é feita com base nos indivíduos, ou seja, de todos os atletas disponíveis, 50% destes são selecionados para compor a base de treino, 30% para compor a base de validação e 20%

compondo a base de teste. Já nas separações para os modelos de aplicação individualizada, a seleção é feita com base nas datas de referência para os períodos de treinamento de cada ciclista. Assim, 50% dos períodos disponíveis são seleccionados para treino, 30% para validação e 20% para teste. Neste caso, para garantir uma boa separação, foram seleccionados somente os indivíduos com ao menos 10 períodos de treinamento disponíveis. A visualização deste esquema está ilustrada na Figura 6.13.

O algoritmo de otimização via métodos de gradiente utilizado é o *Adam* (KINGMA; BA, 2014), onde a função perda é o RMSE. A taxa de aprendizado utilizada é $\eta = 0,001$ e um total de 1.000 épocas, entretanto caso não haja melhoria nos valores da função perda avaliada no conjunto de validação ao longo de 10 épocas consecutivas, o algoritmo é interrompido e retorna as estimativas paramétricas da época com menor perda. Foram seleccionadas amostras dos dados de tamanho 32 (*batches*) para cada iteração do algoritmo de otimização estocástica.

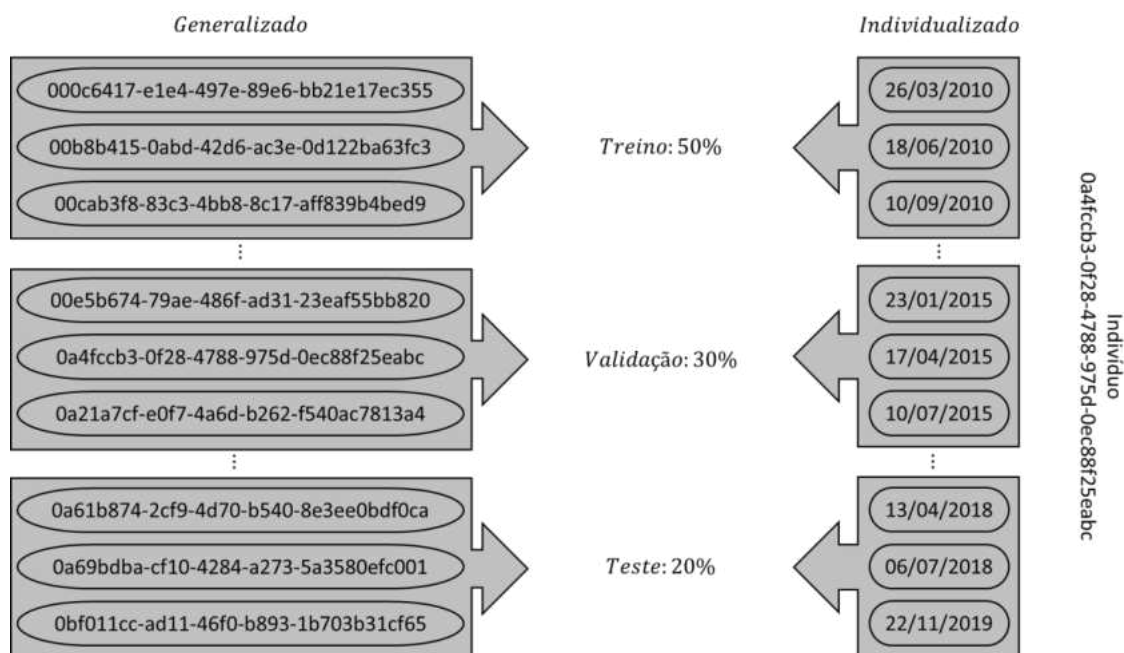


Figura 6.13 – Visualização da partição da base de dados.

6.7 RESULTADOS

Os diferentes modelos ajustados possuem características peculiares, com aumento de complexidade, notável principalmente pelo número de parâmetros a serem estimados. Tal desafio requer uma grande quantidade de dados, a fim de prover devida capacidade de generalização e extrapolação da amostra utilizada para o ajuste.

As distintas configurações apresentadas buscam evidenciar eventuais fragilidades e inconsistências de cada arquitetura. Um modelo que se comporte de maneira consistente, independente da janela que é utilizada, demonstra uma construção mais robusta para modelar o fenômeno de interesse.

A Tabela 6.1 mostra diferentes métricas dos resíduos, $r_i = y_i - \hat{y}_i$, para os diferentes ajustes, considerando as bases de dados de treino, validação e teste. A partir dos valores apresentados, torna-se evidente a melhoria de ajuste e capacidade de previsão dos modelos Paralelo, Paralelo+ e Completo, em relação ao modelo Banister. Além disso, este sofre uma grande perda de qualidade nas configurações individualizadas.

É importante notar, entretanto, que o modelo Banister possui um comportamento bastante consistente entre as bases de treino, validação e teste, uma característica que demonstra as capacidades de generalização do modelo. Tal efeito está presente também no modelo Paralelo e Paralelo+, porém o modelo Completo apresenta maior variabilidade nas métricas, principalmente comparando os resultados nas bases de treino (*in-sample*) e teste (*out-of-sample*). O grande número de parâmetros desconhecidos pode ser o fator responsável por este indício de sobreajuste (*overfitting*).

Tabela 6.1 – Métricas dos resíduos do ajuste dos modelos de performance nas bases de treino, validação e teste.

| Métricas | Modelos | Configurações | | | | | | | | | | | |
|-----------------|-----------|--------------------|--------------|---------------|--------------------|--------------|--------------|-----------------------|---------------|---------------|-----------------------|---------------|---------------|
| | | Generalizado 84/84 | | | Generalizado 84/28 | | | Individualizado 84/84 | | | Individualizado 84/28 | | |
| | | Treino | Validação | Teste | Treino | Validação | Teste | Treino | Validação | Teste | Treino | Validação | Teste |
| RMSE | Banister | 14,17 | 14,43 | 14,28 | 9,46 | 9,49 | 9,50 | >999,9 | >999,9 | >999,9 | >999,9 | >999,9 | >999,9 |
| | Paralelo | 7,45 | 8,10 | 7,63 | 6,41 | 6,65 | 6,37 | 13,74 | 13,87 | <u>14,28</u> | 9,28 | 9,42 | 9,78 |
| | Paralelo+ | 7,51 | 8,11 | 7,51 | 6,37 | 6,56 | 6,32 | 13,81 | 13,85 | <u>14,28</u> | 9,32 | <u>9,41</u> | <u>9,77</u> |
| | Completo | <u>5,91</u> | <u>7,31</u> | <u>7,32</u> | <u>4,75</u> | <u>5,40</u> | <u>5,30</u> | <u>12,06</u> | <u>13,84</u> | <u>14,28</u> | <u>8,20</u> | 9,45 | 9,89 |
| MAE | Banister | 11,64 | 11,86 | 11,74 | 6,51 | 6,55 | 6,51 | >999,9 | >999,9 | >999,9 | >999,9 | >999,9 | >999,9 |
| | Paralelo | 5,21 | 5,61 | 5,44 | 4,40 | 4,61 | 4,47 | 11,07 | 11,28 | 11,63 | 6,24 | <u>6,45</u> | 6,72 |
| | Paralelo+ | 5,17 | <u>5,51</u> | <u>5,32</u> | 4,26 | 4,44 | 4,33 | 11,16 | 11,28 | 11,63 | 6,28 | <u>6,45</u> | <u>6,71</u> |
| | Completo | <u>4,45</u> | <u>5,51</u> | 5,50 | <u>3,42</u> | <u>3,90</u> | <u>3,86</u> | <u>9,23</u> | <u>11,24</u> | <u>11,59</u> | <u>5,27</u> | 6,58 | 6,93 |
| $\sigma(r_i^2)$ | Banister | 230,50 | 238,56 | 232,92 | 165,65 | 165,70 | 168,51 | >999,9 | >999,9 | >999,9 | >999,9 | >999,9 | >999,9 |
| | Paralelo | 188,88 | 202,12 | 168,79 | 129,53 | 127,45 | 108,05 | 227,34 | 233,55 | <u>245,66</u> | 162,35 | 165,36 | 174,17 |
| | Paralelo+ | 213,13 | 224,59 | 172,86 | 147,05 | 141,88 | 119,32 | 228,52 | <u>230,19</u> | 247,76 | 162,84 | 164,73 | <u>173,93</u> |
| | Completo | <u>67,57</u> | <u>99,97</u> | <u>105,08</u> | <u>53,69</u> | <u>65,90</u> | <u>65,07</u> | <u>197,86</u> | 234,56 | 253,98 | <u>139,89</u> | <u>164,80</u> | 176,00 |

Nos casos das configurações individualizadas, na Tabela 6.1 e na Figura 6.14, observa-se uma grande diferença de magnitude nos resíduos obtidos com o modelo Banister em comparação com os modelos Paralelo, Paralelo+ e Completo. Para melhor visualização, apenas os resíduos dos modelos propostos neste trabalho serão exibidos nos gráficos. A Figura 6.15, Figura 6.16, Figura 6.17 e Figura 6.18 apresentam os gráficos de boxplot dos resíduos da base de teste. Os gráficos de boxplot mostram que a distribuição dos resíduos dos modelos Paralelo, Paralelo+ e Completo são semelhantes, principalmente nas configurações individualizadas. Os modelos propostos para ambas as configurações possuem melhor capacidade preditiva quando comparados ao tradicional modelo Banister, relacionando de maneira mais apropriada a relação treinamento-performance.

A Figura 6.19, Figura 6.20 e Figura 6.21 apresentam as distribuições empíricas dos resíduos dos modelos Paralelo, Paralelo+ e Completo nas diferentes bases de dados (treino, validação e teste) na configuração Individualizado 84/84. Além dos resultados na Tabela 6.1, é possível notar como os modelos Paralelo e Paralelo+ se comportam de maneira mais consistente em relação ao modelo Completo, em particular na configuração individualizada onde a quantidade de dados para a estimação dos parâmetros é reduzida. A simplicidade dos modelos Paralelo e Paralelo+, neste caso, demonstra-se mais adequada para descrição das relações entre treinamento e performance.

A principal vantagem da arquitetura dos modelos Paralelo e Paralelo+, propositalmente, é a interpretabilidade. Limitando a existência de interações entre as diferentes variáveis explicativas somente à última camada da rede neural, independente das funções ativação utilizadas nas camadas precedentes, é possível compreender a influência das características do treinamento nas diferentes variáveis resposta a partir dos coeficientes lineares presentes em f_{camada_4} (Figura

6.10 e Figura 6.11), tal qual ocorre em modelos de regressão linear, comuns ao campo da Estatística.

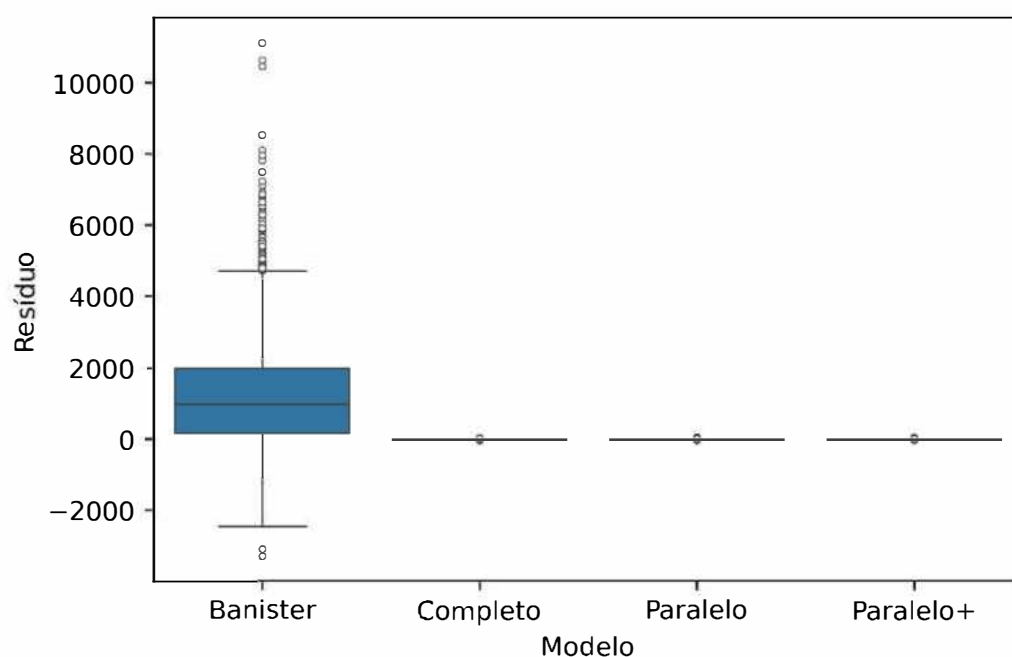


Figura 6.14 – *Boxplot* dos resíduos da base **de teste** dos modelos Banister, Paralelo, Paralelo+ e Completo, da configuração **Individualizado 84/28**.

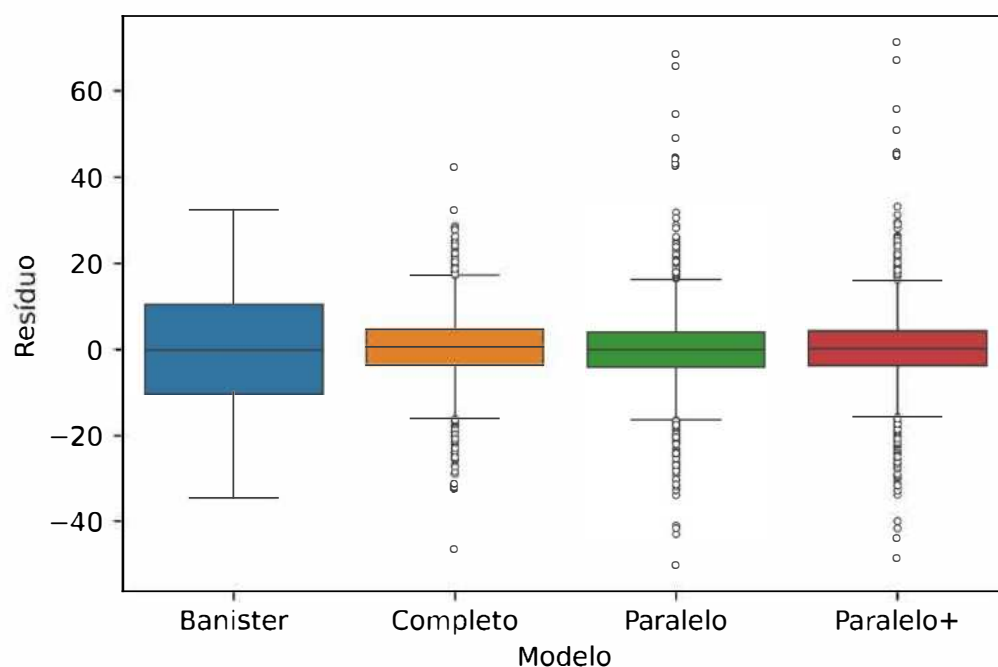


Figura 6.15 – *Boxplot* dos resíduos da base **de teste** dos modelos Paralelo, Paralelo+ e Completo, da configuração **Generalizado 84/84**.

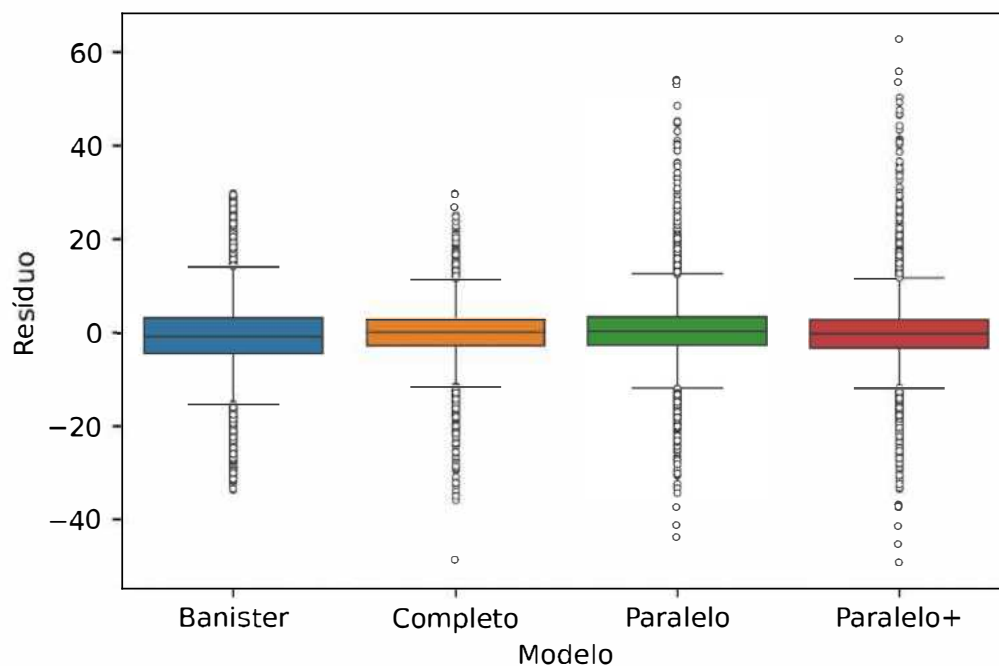


Figura 6.16 – *Boxplot* dos resíduos da base de teste dos modelos Paralelo, Paralelo+ e Completo, da configuração **Generalizado 84/28**.

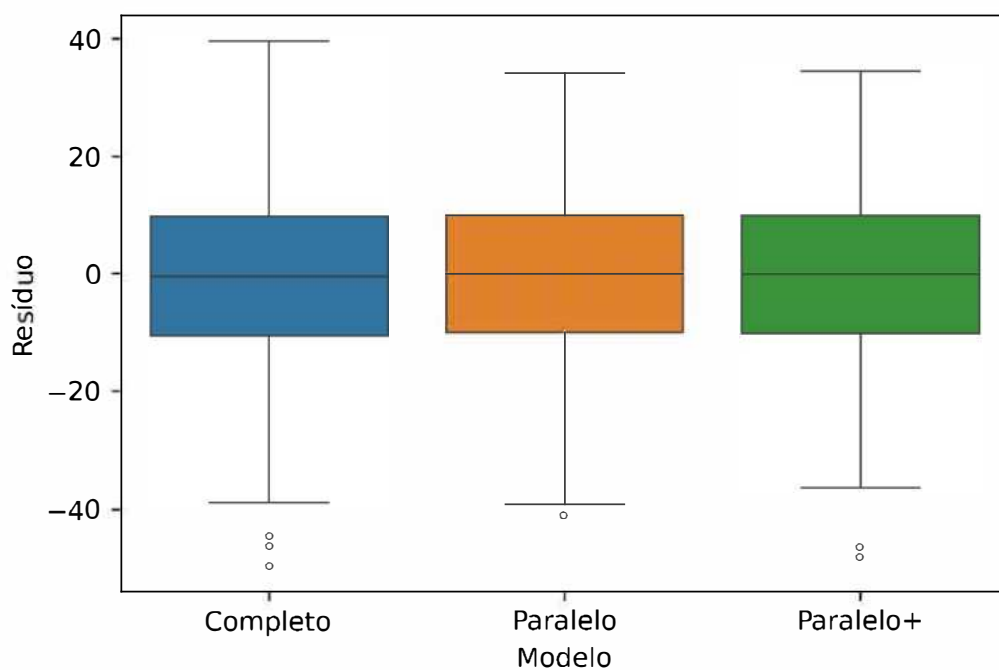


Figura 6.17 – *Boxplot* dos resíduos da base de teste dos modelos Paralelo, Paralelo+ e Completo, da configuração **Individualizado 84/84**.

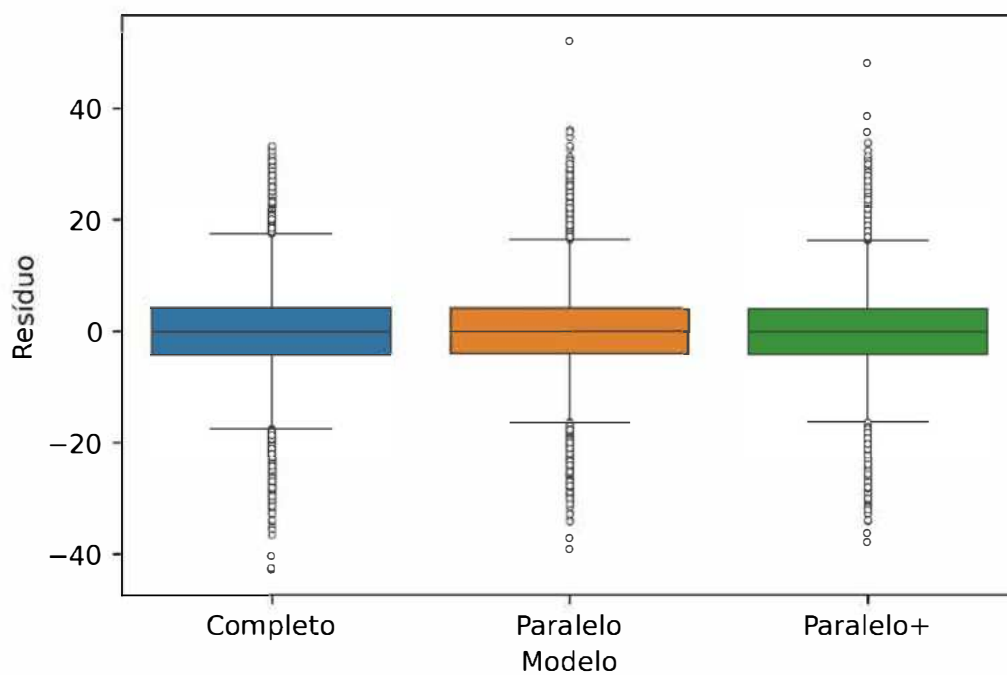


Figura 6.18 – *Boxplot* dos resíduos da base de teste dos modelos Paralelo, Paralelo+ e Completo, da configuração Individualizado 84/28.

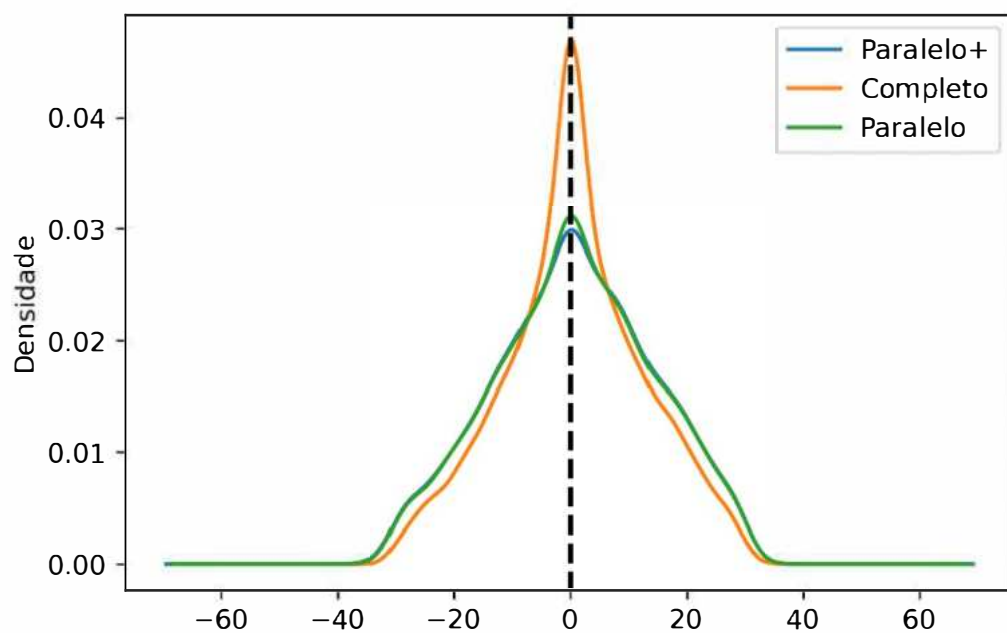


Figura 6.19 Distribuição empírica dos resíduos da base de treino dos modelos Paralelo, Paralelo+ e Completo, da configuração Individualizado 84/84.

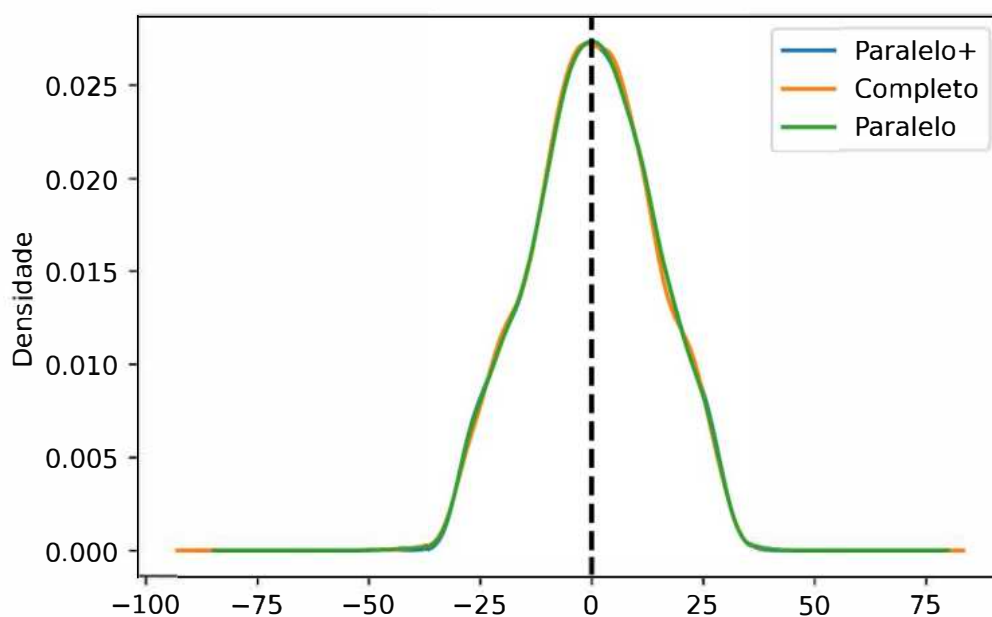


Figura 6.20 – Distribuição empírica dos resíduos da base de validação dos modelos Paralelo, Paralelo+ e Completo, da configuração Individualizado 84/84.

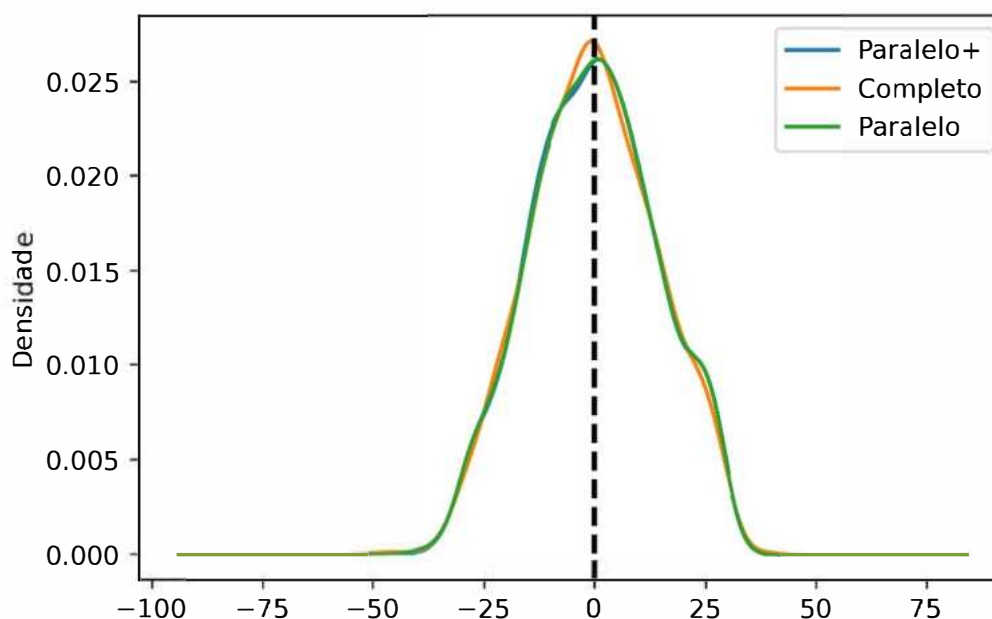


Figura 6.21 – Distribuição empírica dos resíduos da base de teste dos modelos Paralelo, Paralelo+ e Completo, da configuração Individualizado 84/84.

A Tabela 6.2, Tabela 6.3, Tabela 6.4, Tabela 6.5 e Tabela 6.6 mostram as variáveis explicativas com maior impacto na variação dos diferentes parâmetros de performance, dados pelo modelo Omni, a partir do módulo do coeficiente linear

estimado associado a cada uma das variáveis, no modelo Paralelo+ Generalizado 84/84. A interpretação dos valores das tabelas pode ser realizada considerando o impacto das variáveis explicativas, em cada uma das variações dos parâmetros de performance ΔCP , $\Delta W'$, ΔP_{max} , ΔA e ΔTCP_{max} representados respectivamente pelos coeficientes $\hat{\theta}_{41}$, $\hat{\theta}_{42}$, $\hat{\theta}_{43}$, $\hat{\theta}_{44}$ e $\hat{\theta}_{45}$. Cada tabela ordena as variáveis explicativas por ordem de impacto absoluto por uma variável de performance por vez, assim uma variável explicativa com maior coeficiente associado é interpretada como tendo maior influência nas variáveis respostas, uma vez que as covariáveis são padronizadas para o ajuste dos modelos. Isto faz com que as diferentes magnitudes das covariáveis não influencie negativamente a interpretação dos coeficientes estimados. Por fim, a utilização do módulo dos coeficientes ocorre, pois, a arquitetura da rede neural modifica as covariáveis ao longo de suas camadas, logo não é possível garantir a interpretação de que coeficientes positivos (negativos) indicam que um aumento unitário na covariável, acresça (decreça) o valor da variável resposta na magnitude do coeficiente estimado. Para garantir esta interpretação, a Figura 6.22, Figura 6.23, Figura 6.24, Figura 6.25 e Figura 6.26 ilustram as relações estimadas, entre o valor da variável explicativa e o impacto *ceteris paribus* na variável resposta, de algumas das covariáveis com maior impacto absoluto visto nas tabelas.

A Tabela 6.7, Tabela 6.8, Tabela 6.9, Tabela 6.10 e Tabela 6.11 mostram as variáveis explicativas com maior impacto na variação dos diferentes parâmetros de performance, a partir do módulo do coeficiente linear estimado associado a cada uma das variáveis, no modelo Paralelo+ Generalizado 84/28. A Figura 6.27, Figura 6.28, Figura 6.29, Figura 6.30 e Figura 6.31 ilustram as relações estimadas de algumas das variáveis com maior impacto.

As configurações individualizadas não possuem tais resultados pois preservam um modelo por indivíduo, que pela quantidade de atletas na base de dados torna inviável a apresentação dos resultados detalhados para todos.

Tabela 6.2 – Estimativa do módulo dos coeficientes lineares da última camada, das variáveis com maior impacto no ΔCP , do modelo Paralelo+ Generalizado 84/84.

| Variável explicativa | $ \hat{\theta}_{41} $ | $ \hat{\theta}_{42} $ | $ \hat{\theta}_{43} $ | $ \hat{\theta}_{44} $ | $ \hat{\theta}_{45} $ |
|---|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Média do <i>Intensity Factor</i> (IF) | 2,7624 | 0,1887 | 0,6061 | 0,6076 | 0,4148 |
| Média da potência média relativa à potência crítica | 2,4947 | 0,9993 | 2,4063 | 0,6759 | 0,0026 |
| Média do <i>Variability Index</i> (VI) | 2,0546 | 1,9551 | 1,2619 | 0,6052 | 0,4314 |
| Média da frequência cardíaca média | 1,0491 | 0,4993 | 0,2163 | 0,2477 | 0,2421 |
| Tempo médio em 39% do <i>CP</i> | 1,0015 | 0,3488 | 0,5279 | 0,3324 | 0,2687 |

Tabela 6.3 – Estimativa do módulo dos coeficientes lineares da última camada, das variáveis com maior impacto no $\Delta W'$, do modelo Paralelo+ Generalizado 84/84.

| Variável explicativa | $ \hat{\theta}_{41} $ | $ \hat{\theta}_{42} $ | $ \hat{\theta}_{43} $ | $ \hat{\theta}_{44} $ | $ \hat{\theta}_{45} $ |
|--|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Média do <i>Variability Index</i> (VI) | 0,1037 | 1,8283 | 1,3720 | 0,0109 | 1,4546 |
| Média do TSS semanal | 0,0072 | 1,5774 | 0,5107 | 0,1766 | 0,0530 |
| Tempo médio em 291% do <i>CP</i> | 0,0482 | 1,5059 | 0,1331 | 0,0475 | 0,3305 |
| Tempo médio em 226% do <i>CP</i> | 0,0232 | 1,4518 | 0,1244 | 0,1656 | 1,1867 |
| Proporção de descanso | 0,0159 | 1,4284 | 0,9446 | 0,0391 | 0,4110 |

Tabela 6.4 – Estimativa do módulo dos coeficientes lineares da última camada, das variáveis com maior impacto no ΔP_{max} , do modelo Paralelo+ Generalizado 84/84.

| Variável explicativa | $ \hat{\theta}_{41} $ | $ \hat{\theta}_{42} $ | $ \hat{\theta}_{43} $ | $ \hat{\theta}_{44} $ | $ \hat{\theta}_{45} $ |
|---|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Tempo médio em 300+% do <i>CP</i> | 0,0265 | 0,3955 | 2,9020 | 0,0459 | 0,3833 |
| Média da potência média relativa à potência crítica | 2,4947 | 0,9993 | 2,4063 | 0,6759 | 0,0026 |
| Tempo médio em 273% do <i>CP</i> | 0,0098 | 0,4879 | 2,0050 | 0,2025 | 0,4006 |
| Proporção de descanso | 0,0094 | 1,4604 | 1,7814 | 0,0193 | 1,6394 |
| Tempo médio em 279% do <i>CP</i> | 0,0773 | 0,7150 | 1,6919 | 0,3001 | 0,3796 |

Tabela 6.5 – Estimativa do módulo dos coeficientes lineares da última camada, das variáveis com maior impacto no ΔA , do modelo Paralelo+ Generalizado 84/84.

| Variável explicativa | $ \hat{\theta}_{41} $ | $ \hat{\theta}_{42} $ | $ \hat{\theta}_{43} $ | $ \hat{\theta}_{44} $ | $ \hat{\theta}_{45} $ |
|---|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Tempo médio em 86% do <i>CP</i> | 0,4064 | 0,1830 | 0,3417 | 1,0387 | 0,4929 |
| Desvio-padrão do TSS semanal | 0,3213 | 1,1411 | 0,2803 | 0,7824 | 0,1673 |
| Tempo médio em 33% do <i>CP</i> | 0,8033 | 0,2522 | 0,2848 | 0,7612 | 0,0317 |
| Tempo médio em 50% do <i>CP</i> | 0,2613 | 0,4569 | 0,2197 | 0,7418 | 0,1228 |
| Média da potência média relativa à potência crítica | 2,4947 | 0,9993 | 2,4063 | 0,6759 | 0,0026 |

Tabela 6.6 – Estimativa do módulo dos coeficientes lineares da última camada, das variáveis com maior impacto no ΔTCP_{max} , do modelo Paralelo+ Generalizado 84/84.

| Variável explicativa | $ \hat{\theta}_{41} $ | $ \hat{\theta}_{42} $ | $ \hat{\theta}_{43} $ | $ \hat{\theta}_{44} $ | $ \hat{\theta}_{45} $ |
|----------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Tempo médio em 99% do <i>CP</i> | 0,0348 | 0,5028 | 0,5870 | 0,0311 | 2,4829 |
| Tempo médio em 98% do <i>CP</i> | 0,0362 | 1,1804 | 0,0801 | 0,0772 | 2,0549 |
| Tempo médio em 72% do <i>CP</i> | 0,0565 | 0,0216 | 0,5583 | 0,1137 | 1,8673 |
| Tempo médio em 105% do <i>CP</i> | 0,0811 | 1,1737 | 0,2914 | 0,0234 | 1,8577 |
| Tempo médio em 104% do <i>CP</i> | 0,0609 | 0,5087 | 0,3743 | 0,0086 | 1,8347 |

Tabela 6.7 – Estimativa do módulo dos coeficientes lineares da última camada, das variáveis com maior impacto no ΔCP , do modelo Paralelo+ Generalizado 84/28.

| Variável explicativa | $ \hat{\theta}_{41} $ | $ \hat{\theta}_{42} $ | $ \hat{\theta}_{43} $ | $ \hat{\theta}_{44} $ | $ \hat{\theta}_{45} $ |
|---|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Média da potência média relativa à potência crítica | 2,7615 | 0,4493 | 1,7112 | 0,9560 | 0,9709 |
| Média do <i>Intensity Factor</i> (IF) | 2,3849 | 1,4839 | 0,0962 | 0,7422 | 0,0473 |
| Média do <i>Variability Index</i> (VI) | 1,5833 | 1,6942 | 1,2689 | 0,4955 | 0,5627 |
| Média da frequência cardíaca média | 1,2180 | 0,3187 | 0,1044 | 0,0575 | 0,0680 |
| Tempo médio em 97% do <i>CP</i> | 0,9111 | 0,3941 | 0,1269 | 0,5534 | 0,6451 |

Tabela 6.8 – Estimativa do módulo dos coeficientes lineares da última camada, das variáveis com maior impacto no $\Delta W'$, do modelo Paralelo+ Generalizado 84/28.

| Variável explicativa | $ \hat{\theta}_{41} $ | $ \hat{\theta}_{42} $ | $ \hat{\theta}_{43} $ | $ \hat{\theta}_{44} $ | $ \hat{\theta}_{45} $ |
|--|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Média do TSS semanal | 0,0843 | 1,7603 | 0,7672 | 0,0052 | 0,6901 |
| Média do <i>Variability Index</i> (VI) | 1,5833 | 1,6942 | 1,2689 | 0,4955 | 0,5627 |
| Tempo médio em 235% do <i>CP</i> | 0,0514 | 1,6737 | 0,6731 | 0,1428 | 0,3790 |
| Tempo médio em 107% do <i>CP</i> | 0,0219 | 1,5663 | 0,2864 | 0,0429 | 0,5304 |
| Tempo médio em 148% do <i>CP</i> | 0,0710 | 1,5243 | 0,2309 | 0,0473 | 1,2071 |

Tabela 6.9 – Estimativa do módulo dos coeficientes lineares da última camada, das variáveis com maior impacto no ΔP_{max} , do modelo Paralelo+ Generalizado 84/28.

| Variável explicativa | $ \hat{\theta}_{41} $ | $ \hat{\theta}_{42} $ | $ \hat{\theta}_{43} $ | $ \hat{\theta}_{44} $ | $ \hat{\theta}_{45} $ |
|---|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Tempo médio em 300+% do <i>CP</i> | 0,0150 | 0,5742 | 3,1266 | 0,0095 | 0,1792 |
| Proporção de descanso | 0,0084 | 1,1439 | 1,8253 | 0,0162 | 1,2912 |
| Média da potência média relativa à potência crítica | 2,7615 | 0,4493 | 1,7112 | 0,9560 | 0,9709 |
| Tempo médio em 288% do <i>CP</i> | 0,0464 | 0,0220 | 1,5073 | 0,0347 | 0,0955 |
| Tempo médio em 261% do <i>CP</i> | 0,0919 | 0,6841 | 1,3897 | 0,0370 | 0,5157 |

Tabela 6.10 – Estimativa do módulo dos coeficientes lineares da última camada, das variáveis com maior impacto no ΔA , do modelo Paralelo+ Generalizado 84/28.

| Variável explicativa | $ \hat{\theta}_{41} $ | $ \hat{\theta}_{42} $ | $ \hat{\theta}_{43} $ | $ \hat{\theta}_{44} $ | $ \hat{\theta}_{45} $ |
|---|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Média da potência média relativa à potência crítica | 2,7615 | 0,4493 | 1,7112 | 0,9560 | 0,9709 |
| Tempo médio em 33% do CP | 1,1134 | 0,1517 | 0,0665 | 0,8719 | 0,1533 |
| Média do IF | 2,3849 | 1,4839 | 0,0962 | 0,7422 | 0,0473 |
| Tempo médio em 97% do CP | 0,9111 | 0,3941 | 0,1269 | 0,5534 | 0,6451 |
| Tempo médio em 195% do CP | 0,0115 | 0,0571 | 0,7152 | 0,5163 | 0,1335 |

Tabela 6.11 – Estimativa do módulo dos coeficientes lineares da última camada, das variáveis com maior impacto no ΔTCP_{max} , do modelo Paralelo+ Generalizado 84/28.

| Variável explicativa | $ \hat{\theta}_{41} $ | $ \hat{\theta}_{42} $ | $ \hat{\theta}_{43} $ | $ \hat{\theta}_{44} $ | $ \hat{\theta}_{45} $ |
|----------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Tempo médio em 66% do CP | 0,0055 | 0,8157 | 0,1465 | 0,0285 | 1,7465 |
| Tempo médio em 40% do CP | 0,0515 | 0,6938 | 0,0422 | 0,1236 | 1,6805 |
| Tempo médio em 38% do CP | 0,0353 | 0,3801 | 0,0891 | 0,1137 | 1,6573 |
| Tempo médio em 99% do CP | 0,0264 | 1,1291 | 0,3094 | 0,0343 | 1,6164 |
| Tempo médio em 98% do CP | 0,1100 | 0,8393 | 0,3382 | 0,1764 | 1,4509 |

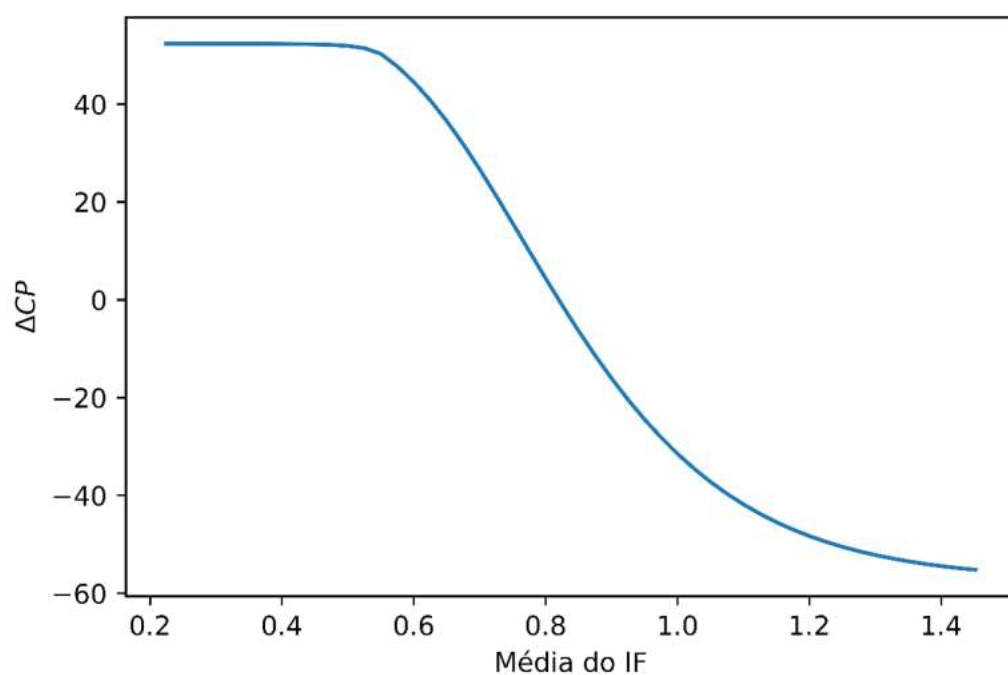


Figura 6.22 – Relação da variável média do IF e ΔCP , no modelo Paralelo+ Generalizado 84/84.

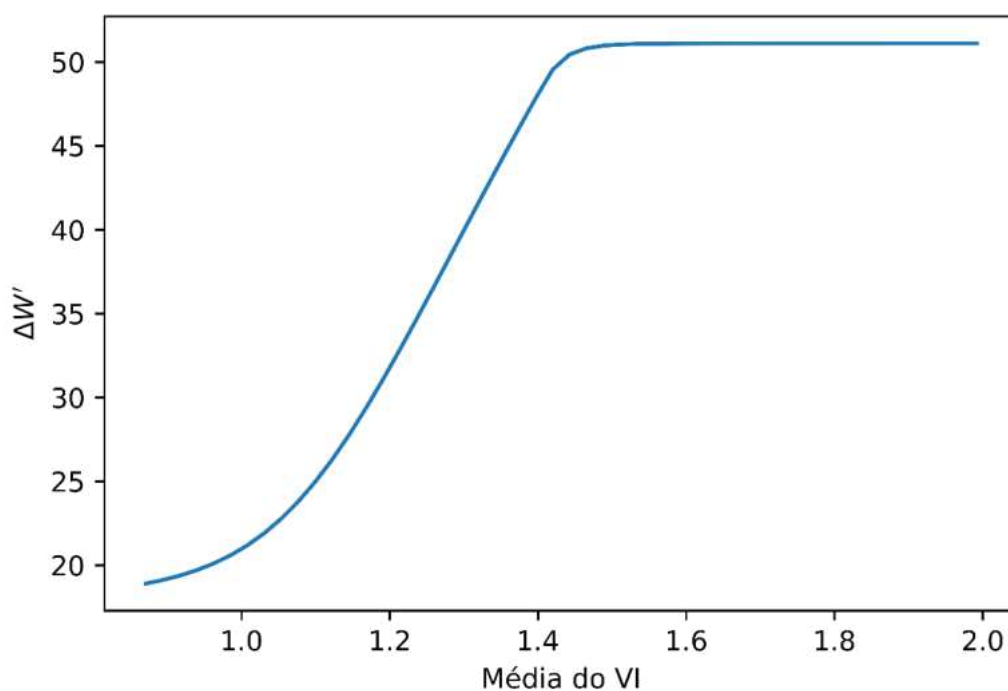


Figura 6.23 – Relação da variável média do VI e $\Delta W'$, no modelo Paralelo+ Generalizado 84/84.

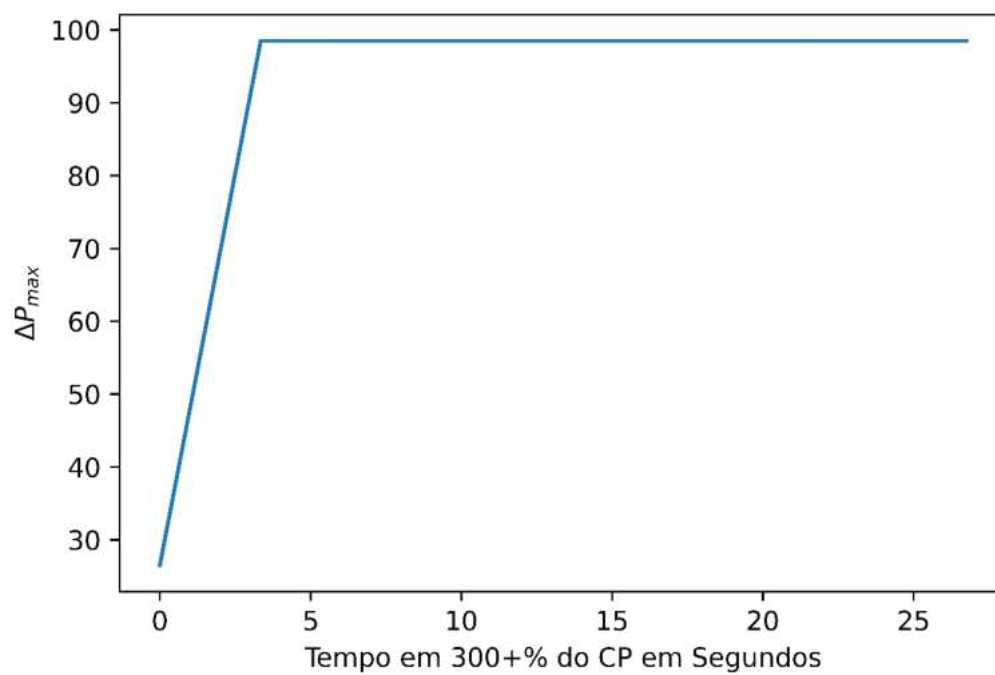


Figura 6.24 – Relação da variável tempo médio em 300+% do CP e ΔP_{max} , no modelo Paralelo+ Generalizado 84/84.

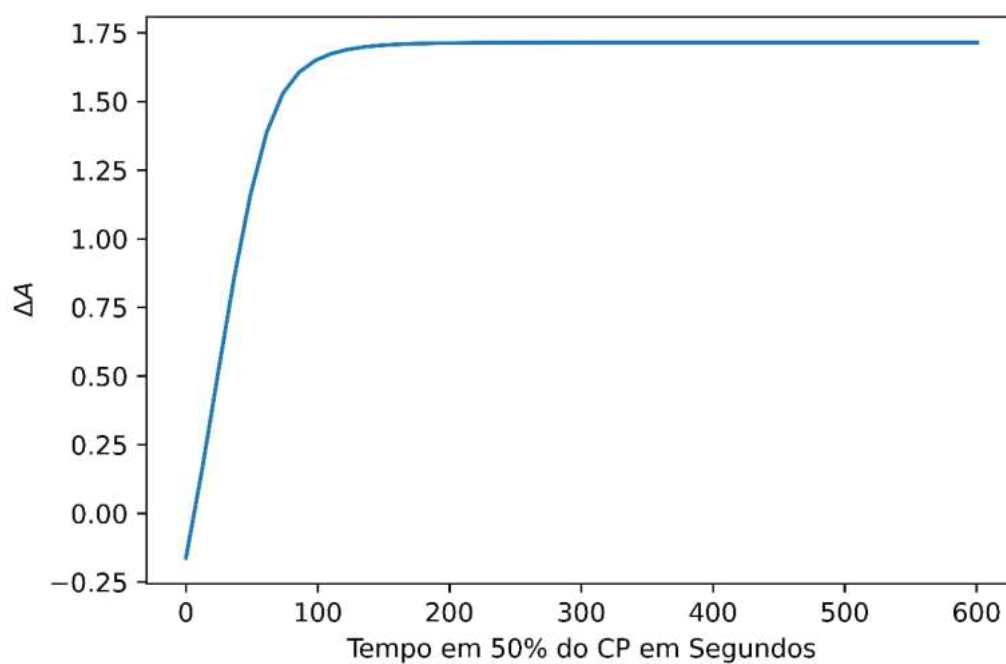


Figura 6.25 – Relação da variável tempo médio em 50% do CP e ΔA , no modelo Paralelo Generalizado 84/84.

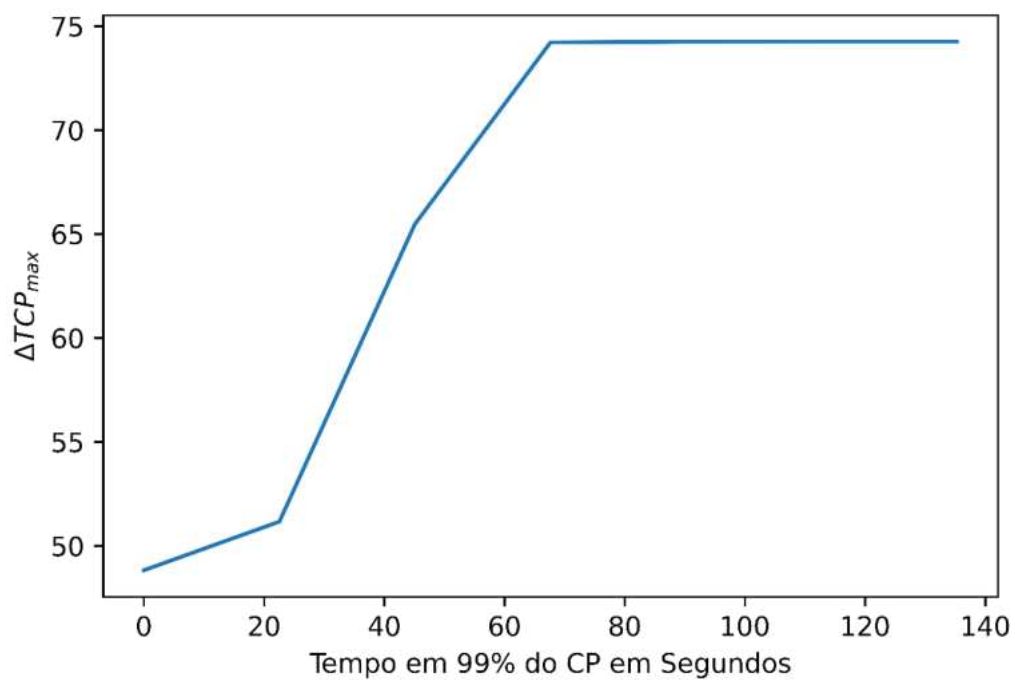


Figura 6.26 – Relação da variável tempo médio em 99% do CP e ΔTCP_{max} , no modelo Paralelo Generalizado 84/84.

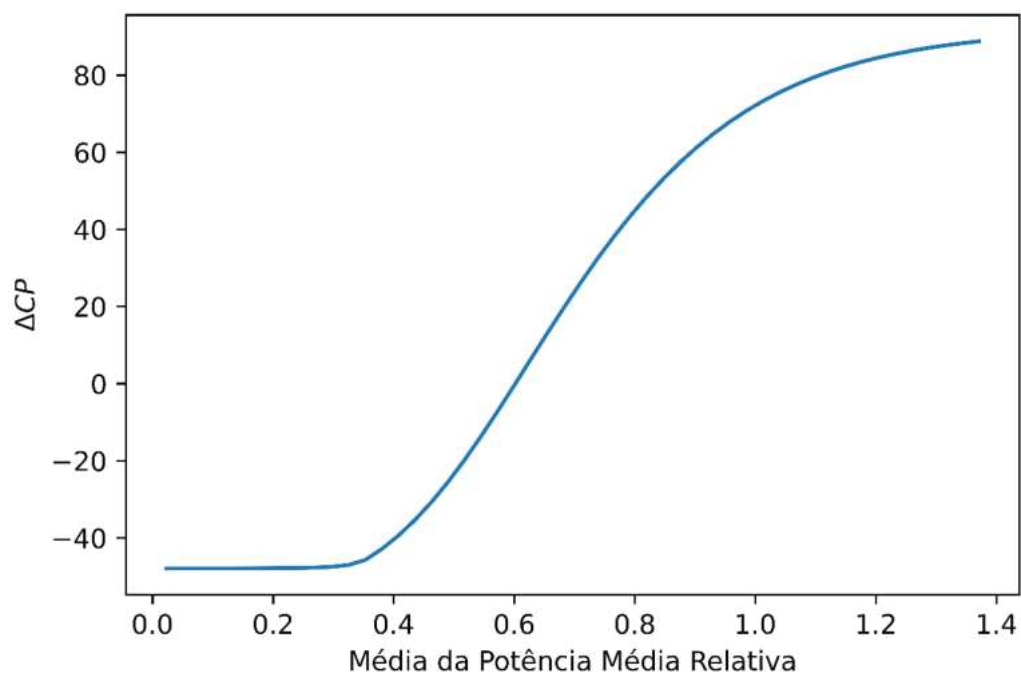


Figura 6.27 – Relação da variável média da potência média relativa à potência crítica e ΔCP , no modelo Paralelo+ Generalizado 84/28.

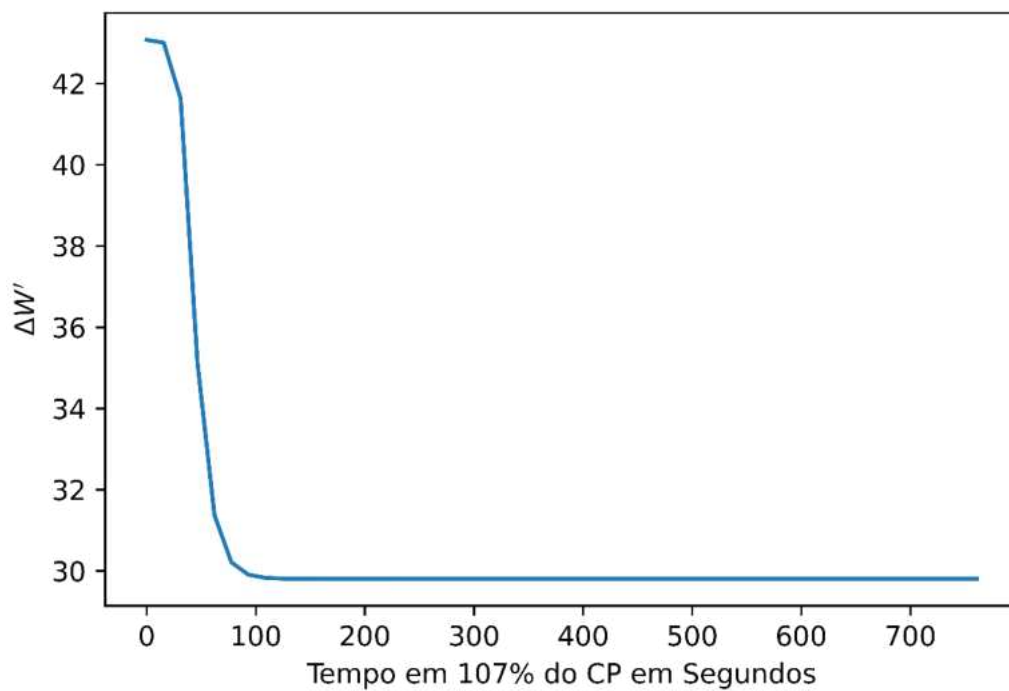


Figura 6.28 – Relação da variável tempo médio em 107% do CP e $\Delta W'$, no modelo Paralelo+ Generalizado 84/28.

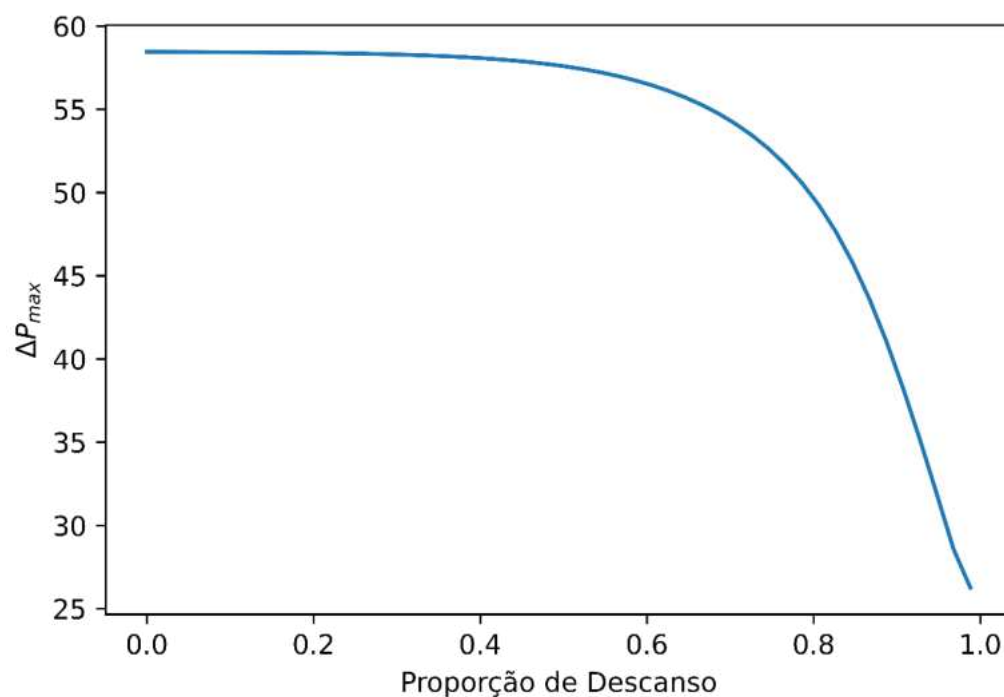


Figura 6.29 – Relação da variável proporção de descanso e ΔP_{max} , no modelo Paralelo+ Generalizado 84/28.

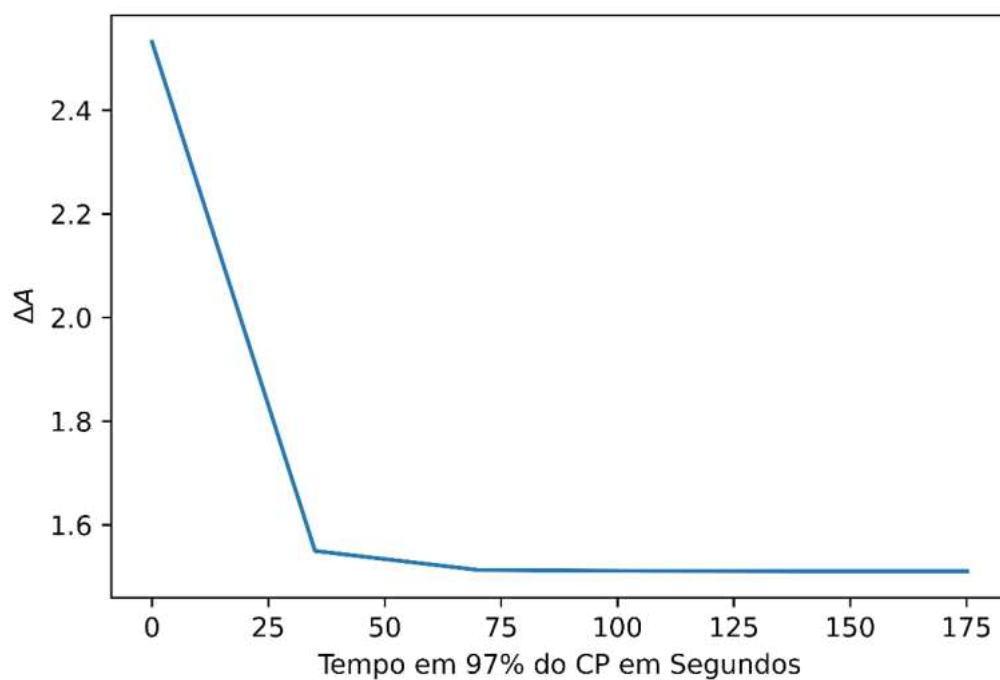


Figura 6.30 – Relação da variável tempo médio em 97% do CP e ΔA , no modelo Paralelo+ Generalizado 84/28.

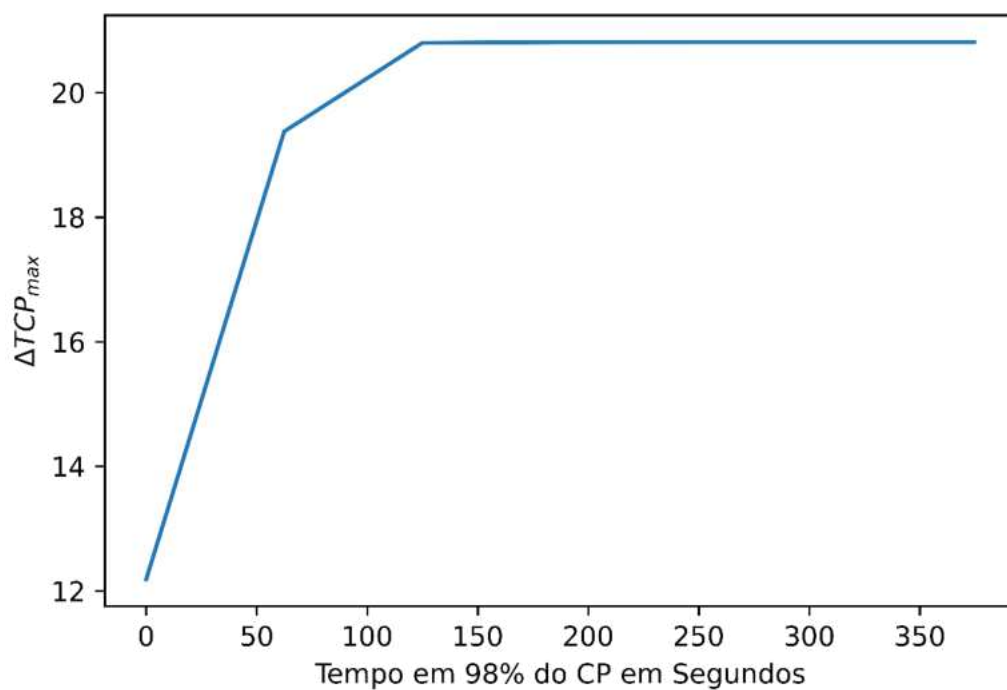


Figura 6.31 – Relação da variável tempo médio em 98% do CP e ΔTCP_{max} , no modelo Paralelo+ Generalizado 84/28.

7 CONCLUSÃO

O presente trabalho se dispôs ao desafio de extrair informação sobre a relação entre treinamento e performance no ciclismo, a partir de dados não-controlados e com as mais diversas peculiaridades encontradas fora de um ambiente controlado, como um laboratório. Este é um dos aspectos que mostra a relevância desta dissertação.

O grande volume de dados faz com que metodologias computacionalmente eficientes sejam necessárias. A versatilidade das redes neurais, aliada à implementação computacional dos algoritmos de otimização via métodos de gradientes estocásticos para estimação dos parâmetros desconhecidos, se mostrou uma ferramenta de grande valor e foi aplicada em diferentes escopos, possibilitando o aproveitamento geral da base de dados disponível, não se limitando por sua quantidade. Este foi um dos desafios superados: a aplicação de modelos estatísticos para uma base de dados de ciclismo em grande escala, já mencionado por Passfield *et al.* (2017).

A extração de informação a partir de dados tão heterogêneos tornou primordial a etapa de pré-processamento, responsável pela criação das variáveis explicativas relevantes para o contexto do trabalho e pela filtragem e limpeza de dados problemáticos, fenômeno comum ao lidar com bases de dados inéditas. A contribuição que este trabalho faz, neste sentido, também possui grande relevância para quem desejar prosseguir com estudos utilizando este conjunto de dados de treinamento de ciclismo, ou então bases de dados semelhantes disponíveis ao pesquisador.

O modelo Banister (BANISTER *et al.*, 1975) possui uma enorme importância na ciência do esporte, não só para o ciclismo, porém suas limitações

são conhecidas. Por isso, destaca-se como trabalhos futuros a comparação com outros modelos que visem capturar a complexidade da relação entre treinamento e performance no ciclismo, de maneira mais assertiva.

Os modelos propostos por esta dissertação: Paralelo, Paralelo+ e Completo, mostraram uma melhor capacidade preditiva que o modelo Banister, entretanto a grande quantidade de parâmetros nos modelos propostos torna necessário um grande volume de dados para realização da inferência. O modelo Paralelo+ se mostrou mais adequado quando comparado ao modelo Paralelo, gerando evidências de que a interpretação sobre seus resultados pode ser mais valiosa para o entendimento do ciclismo. É importante notar que, apesar do senso comum sobre a falta de interpretabilidade de modelos de redes neurais, a sua arquitetura pode ser estruturada para melhor acomodar as necessidades do pesquisador em descrever com clareza a potencial influência de um treinamento específico sobre a performance do atleta, como evidenciado na Seção 6.7.

O modelo Completo, por sua vez, apresentou resultados melhores em relação às métricas de qualidade de ajuste aos dados e previsão. Apesar do comprometimento da interpretabilidade em sua arquitetura, sendo o interesse da aplicação focado na previsão da alteração de parâmetros de performance de atletas de ciclismo, este modelo é mais recomendado do que as outras duas sugestões, Paralelo e Paralelo+, aqui apresentadas.

Uma vantagem da metodologia aplicada é a estimação conjunta dos parâmetros de performance, que por sua vez definem o modelo de duração-potência. Isto proporciona a possibilidade de compreender como o treinamento afeta a performance de um indivíduo, numa gama de durações, com relevância a depender das competições e desafios que o atleta buscar durante a sua temporada.

Os resultados apresentados reforçam a importância do treinamento de resistência para a melhoria do parâmetro CP , representativo da capacidade aeróbica do indivíduo, assim como a importância do investimento temporal em intensidades acima da potência crítica para que se obtenham ganhos em capacidade anaeróbica, representada pelo parâmetro W' , e em capacidade explosiva, representada pelo parâmetro P_{max} .

Ainda assim, há melhorias possíveis de gerar um modelo ainda mais bem ajustado aos dados de treinamento de atletas de ciclismo. A estimativa intervalar é uma informação bastante agregadora, para compreender melhor a incerteza em torno das estimações pontuais realizadas. O uso de uma metodologia que forneça este resultado, enquanto mantém a capacidade de lidar com o grande volume de dados em tempo hábil, é de grande valor. Como exemplos de abordagens neste sentido, pode ser incluído um componente de variância na arquitetura de redes neurais, como também a aplicação de modelos generalizados aditivos fazendo uso de *splines*.

A relação entre treinamento e performance também pode ser mais bem estimada pela arquitetura de redes neurais avaliando-se grupos mais homogêneos de atletas selecionados a partir de uma análise de clusterização, seja por características focadas nos aspectos genéticos, seja pelo padrão de treinamento realizado. Os histogramas que sumarizam as informações, apresentados no Capítulo 4, das séries temporais com dados dos sensores de cadência, frequência cardíaca ou potência, por exemplo, podem servir de insumos para auxiliar o agrupamento de sessões de treinamento com foco comum, seja melhorar potência crítica até a melhoria da capacidade de aceleração.

A performance também é influenciável por outros fatores, como nutrição, psicologia e genética, por exemplo. O estudo da influência de variáveis que cubram

estes aspectos é fundamental para ampliar o conhecimento em torno deste tema. A melhoria das rotinas dos atletas a partir de resultados deste tipo, apresenta potencial para alavancar o resultado em eventos competitivos.

Assim como no esporte, é vital que barreiras sejam tidas como limites temporários, mas que em algum momento são passíveis de serem transpostas com novas e constantes pesquisas, a fim de obter uma compreensão mais ampla das grandezas estudadas pela ciência.

8 REFERÊNCIAS BIBLIOGRÁFICAS

ABADI, Martín. *et al.* TensorFlow: a system for Large-Scale machine learning. **12th USENIX symposium on operating systems design and implementation (OSDI 16)**, p. 265-283, 2016.

AGRAWAL, Shikha; AGRAWAL, Jitendra. Neural network techniques for cancer prediction: A survey. **Procedia Computer Science**, v. 60, p. 769-774, 2015.

ALLEN, Hunter; COGGAN, Andrew. **Training and racing with a power meter**. 2. ed. Boulder: Velopress, 2010.

ANDERS, Ulrich. Statistical model building for neural networks. **AFIR Colloquium**, v. 6, p. 963-979, 1996.

ATKINSON, Greg. *et al.* Science and cycling: current knowledge and future directions for research. **Journal of Sports Sciences**, v. 21, n. 9, p. 767-787, set. 2003.

AVALOS, Marta; HELLARD, Philippe; CHATARD, Jean-Claude. Modeling the training-performance relationship using a mixed model in elite swimmers. **Medicine and science in sports and exercise**, v. 35, n. 5, p. 838, 2003.

BALDI, Pierre. Gradient descent learning algorithm overview: a general dynamical systems perspective. *IEEE Transactions on neural networks*, v. 6, n. 1, p. 182-195, 1995.

BANISTER, Eric W. *et al.* A systems model of training for athletic performance. *Aust J Sports Med*, v. 7, n.3, p. 57-61, 1975.

BASSETT, David R. Jr. *et al.* Comparing cycling world hour records, 1976-1996: modeling with empirical data. *Medicine and Science in Sports and Exercise*, v. 31, n. 11, p. 1665-1676, 1999.

BISONG, Ekaba. More on Optimization Techniques. In: *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. Apress, Berkeley, CA, 2019.

CHOLLET, François. *et al.* **Keras**. 2015. Disponível em: <<https://github.com/keras-team/keras>>. Acesso em 16 jun. 2022.

CLEVERT, Djork-Arné; UNTERTHINER, Thomas; HOCHREITER, Sepp. Fast and accurate deep network learning by exponential linear units (elus). *arXiv*, 2015.

DOGO, E. M. *et al.* A comparative analysis of gradient descent-based optimization algorithms on convolutional neural networks. *International*

Conference on Computational Techniques, Electronics and Mechanical Systems, p. 92-99, 2018.

DOREL, Sylvain. *et al.* Torque and power-velocity relationships in cycling: relevance to track sprint performance in world-class cyclists. **International Journal of Sports Medicine**, v. 26, n. 9, p. 739-746, 2005.

DUANE, Simon. *et al.* Hybrid monte carlo. **Physics letters B**, v. 195, n. 2, p. 216-222, 1987.

DUCHI, John; HAZAN, Elad; SINGER, Yoram. Adaptive subgradient methods for online learning and stochastic optimization. **Journal of machine learning research**, v. 12, n. 7, 2011.

FIESLER, E. Neural network classification and formalization. **Computer Standards & Interfaces**, v. 16, n. 3, p. 231-239, jul. 1994.

FRIEL, Joe. **The Cyclist's Training Bible: The World's Most Comprehensive Training Guide**. 5. ed.: VeloPress, 2018.

GRAUSER, Damien. **Extended critical power**. 2013. Disponível em: <<https://github.com/GoldenCheetah/GoldenCheetah/blob/master/src/Metrics/ExtendedCriticalPower.cpp>>. Acesso em 20 out. 2021.

HENDERSON, Peter. *et al.* Deep reinforcement learning that matters. *Proceedings of the AAAI conference on artificial intelligence*, v. 32, n. 01, 2018.

HOPKINS, W. G. *et al.* Relation between power and endurance for treadmill running of short duration. *Ergonomics*, v. 32, n. 12, p. 1565-1571, dez. 1989.

HORNIK, Kurt; STINCHCOMBE, Maxwell; WHITE, Halbert. Multilayer feedforward networks are universal approximators. *Neural networks*, v. 2, n. 5, p. 359-366, 1989.

ILJUKOV, Sergci; BERMON, Stephane; SCHUMACHER, Yorck O.. Application of the athlete's performance passport for doping control: a case report. *Frontiers in physiology*, v. 9, p. 280, 2018.

JOBSON, Simon A. *et al.* The analysis and utilization of cycling training data. *Sports Medicine*, v. 39, n. 10, p.833-844, out. 2009.

KINGMA, Diederik P.; BA, Jimmy. Adam: a method for stochastic optimization. *arXiv*, 2014.

KOSMIDIS, Ioannis; PASSFIELD, Louis. Linking the performance of endurance runners to training and physiological effects via multi-resolution elastic net. *arXiv*, 2015.

KUTNER, Michael H. *et al.* **Applied linear statistical models**. 5. ed.: McGraw-Hill/Irwin, 2005.

LEITE, Natan Freitas. **Análise Estatística da Intensidade Otimizada para Aumento de Performance no Ciclismo de Estrada**. Orientador: Larissa de Carvalho Alves, Daniel Takata Gomes. 2018. 82 f. TCC (Graduação) – Curso de Estatística, Escola Nacional de Ciências Estatísticas, Rio de Janeiro, 2018.

LEWIS, Michael. **Moneyball: The art of winning an unfair game**. WW Norton & Company, 2004.

LI, Zongyi. *et al.* Fourier neural operator for parametric partial differential equations. *arXiv*, 2020.

LIVERSEDGE, Mark. **GoldenCheetah OpenData Project**. abr. 2018. Disponível em: <<https://doi.org/10.17605/OSF.IO/6HFPZ>>. Acesso em: 29 jul. 2021.

LIVERSEDGE, Mark. *et al.* **GoldenCheetah**. v. 3.5, jan. 2020. Disponível em: <<https://www.goldencheetah.org/>>. Acesso em 10 jul. 2021.

MAAS, Andrew L.; HANNUN, Awni Y.; NG, Andrew Y.. Rectifier nonlinearities improve neural network acoustic models. *ICML*, v. 30, n. 1, p. 3, jun. 2013.

MCARDLE, William D.; KATCH, Frank I.; KATCH, Victor L.. **Exercise physiology: nutrition, energy, and human performance**. 7. ed.: Lippincott Williams & Wilkins, 2010.

METCALFE, Alan J. *et al.* Within-season distribution of external training and racing workload in professional male road cyclists. *International Journal of Sports Physiology and Performance*, v. 12, n. s2, p. 142-146, 2017.

MHASKAR, Hrushikesh; LIAO, Qianli; POGGIO, Tomaso. When and why are deep networks better than shallow ones?. *Proceedings of the AAAI conference on artificial intelligence*, 2017.

MONOD, H.; SCHERRER, J.. The work capacity of a synergic muscular group. *Ergonomics*, v. 8, n. 3, p. 329-338, jul. 1965.

MORTON, R. Hugh. A 3-parameter critical power model. *Ergonomics*, v. 39, n. 4, p. 611-619, mar. 1996.

MORTON, R. H.; FITZ-CLARKE, J. R.; BANISTER, E. W.. Modeling human performance. **Journal Of Applied Physiology**, v. 69, n. 3, p. 1171-1177, set. 1990.

MUJICA, Iñigo; PADILLA, Sabino. Detraining: loss of training-induced physiological and performance adaptations. Part I. **Sports Medicine**, v. 30, n. 2, p. 79-87, 2000.

MUJICA, Iñigo; PADILLA, Sabino. Physiological and performance characteristics of male professional road cyclists. **Sports medicine**, v. 31, n. 7, p. 479-487, 2001.

NAIR, Vinod; HINTON, Geoffrey E.. Rectified linear units improve restricted boltzmann machines. **ICML**, 2010.

NELDER, John Ashworth; WEDDERBURN, Robert W. M.. Generalized linear models. **Journal of the Royal Statistical Society**, v. 135, n. 3, p. 370-384, 1972.

PASSFIELD, Louis. *et al.* Knowledge is power: Issues of measuring training and performance in cycling. **Journal of Sports Sciences**, v. 35, n. 15, p. 1426-1434, 2017.

PUCHOWICZ, Michael J.; BAKER, Jonathan; CLARKE, David C.. Development and field validation of an omni-domain power-duration model. **Journal of Sports Sciences**, v. 38, n. 7, p. 801-813, mar. 2020.

PUCHOWICZ, Michael J. *et al.* The critical power model as a potential tool for anti-doping. **Frontiers in physiology**, v. 9, p. 643, 2018.

QIAN, Ning. On the momentum term in gradient descent learning algorithms. **Neural Networks**, v. 12, n. 1, p. 145-151, 1999.

QUOD, M. J. *et al.* The power profile predicts road cycling MMP. **International Journal Of Sports Medicine**, v. 31, n. 06, p. 397-401, mar. 2010.

RAMACHANDRAN, Prajit; ZOPH, Barret; LE, Quoc V. Searching for activation functions. **arXiv**, 2017.

REDMON, Joseph *et al.* You only look once: Unified, real-time object detection. **Proceedings of the IEEE conference on computer vision and pattern recognition**, p. 779-788, 2016.

ROBBINS, Herbert; MONRO, Sutton. A stochastic approximation method. **The annals of mathematical statistics**, p. 400-407, 1951.

ROSENBLATT, Frank. The perceptron: A probabilistic model for information storage and organization in the brain. **Psychological Review**, v.65, n. 6, p. 386, 1958.

SANDERS, Dajo; HELJBOER, Mathieu. Physical demands and power profile of different stages types within a cycling grand tour. **European Journal of Sport Science**, v. 19, n. 6, p. 736-744, 2019.

SERRANO, Antonio Manual Rubio *et al.* Neural Network Predictor for Fraud Detection. **ICoFCS 2012**, p. 61, 2012.

SHARMA, Neha; JAIN, Vibhor; MISHRA, Anju. An analysis of convolutional neural networks for image classification. **Procedia computer science**, v. 132, p. 377-384, 2018.

VISÉ, Daniel Dc. 30 years ago, Greg Lemond made the greatest comeback in modern sports history. **Bicycling**, 2019. Disponível em: <<https://www.bicycling.com/tour-de-france/a27681555/greg-lemond-1989/>>. Acesso em 10 dez. 2021.