UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
DEPARTAMENTO DE ENERGIA NUCLEAR
PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIAS ENERGÉTICAS E NUCLEARES

PEDRO PAULO DE MEDEIROS ALVES

SELEÇÃO E AVALIAÇÃO DE COMPONENTES TEMPORAIS NO *DOWNSCALING* ESTATÍSTICO

Recife
2021

PEDRO PAULO DE MEDEIROS ALVES

SELEÇÃO E AVALIAÇÃO DE COMPONENTES TEMPORAIS NO *DOWNSCALING* ESTATÍSTICO

Dissertação apresentada ao Programa de Pós-Graduação em Tecnologias Energéticas e Nucleares da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Ciências.

Área de Concentração: Fontes Renováveis de Energia.

Orientador: Prof. Dr. Fernando Roberto de Andrade Lima.

Coorientador: Prof. Dr. Alexandre Carlos Araújo da Costa.

Recife

2021

Catalogação na fonte
Bibliotecária Margareth Malta, CRB-4 / 1198

A474m Alves, Pedro Paulo de Medeiros.
Seleção e avaliação de componentes temporais no *downscaling* estatístico / Pedro Paulo de Medeiros Alves - 2021.
77 folhas, il., gráfs., tabs.

Orientador: Prof. Dr. Fernando Roberto de Andrade Lima.
Coorientador: Prof. Dr. Alexandre Carlos Araújo da Costa.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Tecnologias Energéticas e Nucleares, 2021.
Inclui Referências e Apêndices.

1. Energia Nuclear. 2. Recurso eólico. 3. *Downscaling* estatístico. 4. Componentes temporais. 5. Autocorrelação. I. Lima, Fernando Roberto de Andrade (Orientador). II. Costa, Alexandre Carlos Araújo da (Coorientador). III. Título

UFPE

621.042 CDD (22. ed.) BCTG/2021-127

PEDRO PAULO DE MEDEIROS ALVES

SELEÇÃO E AVALIAÇÃO DE COMPONENTES TEMPORAIS NO *DOWNSCALING* ESTATÍSTICO

Dissertação apresentada ao Programa de Pós-Graduação em Tecnologias Energéticas e Nucleares da Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, como requisito parcial para a obtenção do título de Mestre em Ciências. Área de Concentração: Fontes Renováveis de Energia.

Aprovada em: 22/04/2021.

BANCA EXAMINADORA

_____
Prof. Dr. Fernando Roberto de Andrade Lima (Orientador)
Universidade Federal de Pernambuco

_____
Dr. Alexandre Carlos Araújo da Costa (Coorientador)
Universidade Federal de Pernambuco

_____
Profa. Dra. Olga de Castro Vilela (Examinadora Interna)
Universidade Federal de Pernambuco

_____
Profa. Dra. Doris Regina Aires Veleda (Examinadora Interna)
Universidade Federal de Pernambuco

_____
Prof. Dr. Tsang Ing Ren (Examinador Interno)
Universidade Federal de Pernambuco

AGRADECIMENTOS

Agradeço primeiramente aos meus pais, pelo apoio e confiança durante a trajetória de toda minha vida.

A toda a equipe do Centro de Energias Renováveis (CER-UFPE) por todo suporte prestado para que a conclusão deste trabalho fosse possível.

Ao meu tio, Nilson Medeiros, e seus inúmeros questionamentos sobre prazos e expectativa de finalização e as inúmeras vezes que auxiliou a resolver pendências com o PROTEN.

Aos meus amigos e brilhantes pesquisadores Valentin Perruci, por todo conhecimento e ideias compartilhadas comigo, e Janis Joplim por sua paciência, conselhos e sugestões de melhoria.

Aos Professores Fernando Lima e Alexandre Costa, pela orientação e confiança.

Ao Projeto IBITU.INTELIPREV (no âmbito do Programa de P&D ANEEL) pela confiança e oportunidade a mim empregadas.

Às seguintes entidades cujo apoio financeiro se deu durante a realização da Graduação e/ou Mestrado: ao CNPq pelo apoio financeiro durante os meses iniciais do mestrado, à RNP pelo apoio no âmbito do projeto HPC4E (https://hpc4e.eu/the-project/work-plan/wp4).

A minha esposa e companheira, Lisandra Medeiros, por sua paciência, apoio, incentivo e noites de sono mal dormidas devido a suas incansáveis revisões.

RESUMO

O Brasil vem se destacando internacionalmente pelo aumento na geração a partir da fonte eólica. Dado o fato desta fonte ser intermitente, este aumento traz consigo a necessidade do desenvolvimento de ferramentas mais acuradas para estimar o recurso eólico com vistas a auxiliar o planejamento da matriz elétrica brasileira. Existem modelos que estimam a velocidade do vento sobre o globo, entretanto, esta informação está disponível em baixa resolução espacial (macroescala). Estes modelos são conhecidos como GCMs (*Global Circulation Models*). O comportamento do vento no local de interesse (microescala) sofre a influência de diversos fatores que não são considerados na modelagem macroescalar. O recurso eólico na microescala pode ser estimado utilizando os dados do GCM como entrada das técnicas de aumento de resolução espacial (*downscaling*). Particularmente, este estudo utiliza técnicas de *downscaling* estatístico para estimar a velocidade do vento de 8 estações anemométricas. Usualmente as estimativas do vento local (preditando) são feitos utilizando instantes de tempo concomitantes entre macroescala e microescala. Neste trabalho, algumas técnicas utilizam em seus dados de entradas (preditores) instantes de tempos anteriores ao momento avaliado, nomeados neste texto como componentes temporais (CTs). Em uma avaliação global notou-se que a adição de componentes temporais melhorou as estimativas. Foi desenvolvida uma metodologia objetiva para seleção das CTs a partir das funções de autocorrelação e autocorrelação parcial. Após comparação dos resultados com ambas as metodologias, notou-se que para a maioria dos modelos, a metodologia baseada na autocorrelação parcial apresentou melhores resultados. Em adição ao mencionado, foram desenvolvidos modelos que utilizaram técnicas de *clustering* para melhorar as estimativas e concluiu-se que existe um ganho significativo nestas ao unir a adição de CTs e técnicas de agrupamento. Cada um dos modelos citados anteriormente utilizaram como dados de entrada cada um dos 16 níveis do GCM empregados neste trabalho. Em acréscimo a isto foram desenvolvidos modelos que além das componentes temporais realizaram seleções espaciais avaliando todos os níveis do GCM. Estes modelos apresentaram bom desempenho, especialmente nas regiões mais complexas, com destaque para o modelo RTCTP (Regressão, Teste de hipótese, *Clustering*, Teste de hipótese, Análise de componentes principais).

Palavras-chave: recurso eólico; *downscaling* estatístico; componentes temporais; autocorrelação.

ABSTRACT

Brazil has been standing out internationally for the increase in generation from wind power. Given that this source is intermittent, this increase brings the need to develop more accurate tools to estimate the wind resource to assist the planning of the Brazilian's electrical matrix. Some models estimate the wind speed over the globe. However, this information is only available in low spatial resolution (macroscale). These models are called GCMs (Global Circulation Models). The wind modeling in the place of interest (microscale) is affected by several factors that are not considered in the GCMs modeling. The microscale's wind resource can be estimated using the GCM data as inputs of strategies of increasing spatial resolution (*downscaling*). In particular, this study uses statistical *downscaling* techniques to estimate wind speed from 8 met masts. Usually, local wind estimates use concomitant moments between macroscale and microscale. In this study, some techniques used in their input data instants of times before the evaluated moment, named in this text as temporal components (CTs). The addition of temporal components improved the quality of the estimates. An objective methodology for the selection of CTs was developed based on the autocorrelation and partial autocorrelation functions. After evaluating both methodologies' results, the methodology based on partial autocorrelation was better. Some models developed use clustering techniques to improve the estimates. There is a significant gain in the results when combining CTs and clustering techniques in one model. Each of the mentioned models uses as input data for each of the 16 GCM levels employed in this study. Some of the developed models use, together with the addition of temporal components, a spatial selection evaluating all levels of the GCM. These models have a good performance, especially in more complex regions, emphasizing the RTCTP model (Regression, Hypothesis testing, Clustering, Hypothesis testing, Principal component analysis).

Keywords: wind resource; statistical downscaling; temporal components; autocorrelation.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE SÍMBOLOS

M	magnitude da velocidade do vento
$\bar{y}$	Valor médio
yi	dados observacionais
ϕ_i	Coeficiente da equação de autocorrelação parcial
$\hat{\sigma}$	Desvio Padrão
R	Coeficiente de correlação
βj	parâmetros da regressão linear múltipla
ε	erro
X	Matriz dos Preditores
Y	Matriz das observações
$\tilde{R}$	Matriz das componentes principais preservadas
Δt	*time lag* (Defasagem)
$\bar{x}$	Média da série temporal
t	instante de tempo
E_R	Erro de reconstrução
R_{txm}	Base Ortogonal formada por m preditores e t instantes de tempo
$\Sigma m \times m$	Autovetores da matriz de correlação
T	*timesteps*
C	número de agrupamentos
$\hat{Y}$	Velocidade do vento estimada no local de interesse.

SUMÁRIO

1 INTRODUÇÃO

A permanente necessidade de produzir energia para suprir a demanda da sociedade tem aumentado ao longo dos anos, unida com a insegurança em relação à oferta e ao preço futuro de combustíveis fósseis, provocam um aumento no interesse por fontes renováveis de energia. A produção de eletricidade a partir da fonte eólica aumentou 7511 MW entre 2018 e 2019, o que corresponde a um aumento de 15,5%, e a energia a partir da dita fonte passou a representar 8,6 % da oferta interna de energia elétrica, um aumento de 13,2% se comparado a participação de 7,6% em 2018 (BEN, 2020).

Descrever o recurso eólico em lugares de interesse para a referida produção de energia em larga escala é fundamental para estimar a produção das centrais eólicas, que se tornam cada vez mais numerosas no Brasil, especialmente no Nordeste. À medida que a inserção da energia eólica na matriz energética nacional aumenta, cresce também o interesse na avaliação do recurso, visando uma estimativa mais precisa da produção a longo prazo dos parques, propiciando melhores condições para o planejamento da matriz energética nacional.

A modelagem do vento nos locais de interesse recorre em diversas ocasiões ao emprego de dados provenientes de modelos gerais da circulação atmosférica (*General Circulation Models*, GCMs – KALNAY *et al.,* 1996), modelos estes que simulam em baixa resolução o comportamento de diversas variáveis atmosféricas para todo o globo terrestre e com ampla cobertura temporal (dezenas de anos).

Devido à baixa resolução espacial dos GCMs, não é possível descrever de forma adequada o comportamento deste na escala espacial dos complexos eólicos. Informações sobre a escala local são importantes para descrever de forma mais acurada o vento, visto que, nessa resolução, o comportamento do vento é sensível a diversas variáveis, como orografia, vegetação e obstáculos (WILBY e WIGLEY, 1997; WILBY e DAWSON, 2012).

As limitações dos GCMs podem ser contornadas utilizando-se técnicas de *downscaling*, as quais relacionam variáveis em diferentes escalas espaciais. No caso do presente estudo, relacionar-se-á variáveis atmosféricas na escala sinóptica (preditores) com observações no local de interesse (preditandos). As técnicas de *downscaling* podem ser numéricas (*downscaling* dinâmico), empíricas (*downscaling* estatístico) ou híbridas. Neste trabalho, foram utilizadas técnicas estatísticas.

Há muitos estudos que utilizam técnicas de *downscaling* estatístico e habitualmente, os modelos buscam relacionar as diferentes escalas em instantes de tempo concomitantes. Adicionalmente, existem as linhas de pesquisa que buscam relacionar variáveis

temporalmente, isso é explícito no uso de modelos de séries temporais (*time series)*. Existe também uma quantidade menor de trabalhos que relacionam escalas espaciais e temporais simultaneamente como HUANG *et al.* (2020).

Buscando validar as diferentes modelagens do vento local abordadas neste estudo, utilizam-se como preditores dados de velocidade do vento fornecidas pelo o modelo geral de circulação da atmosfera (*General Circulation Models*, GCM – DEE *et al.*, 2011) disponibilizada pelo ECMWF (*European Centre for Medium-Range Weather Forecasts*) e um banco de dados com medições anemométricas em 8 locais distintos, para que a modelagem seja avaliada sob diferentes condições orográficas e regimes de ventos distintos conforme detalhado na seção METODOLOGIA E MODELOS.

O objetivo geral deste trabalho é avaliar se o uso de técnicas de *downscaling*, que levam em consideração o emprego de instantes de tempo dos preditandos não concomitantemente com a observação, apresentam uma melhora de desempenho quando comparado aos modelos que não utilizam esta informação. Tais instantes de tempo sempre antecedem a observação na escala temporal.

Os objetivos específicos deste trabalho são o desenvolvimento de um critério objetivo para a seleção dos instantes de tempo do passado do GCM em relação ao instante avaliado que serão incorporados aos modelos de *downscaling*, o desenvolvimento de um modelo, que além de utilizar um critério objetivo para a seleção de componentes temporais, também determinará de forma automática o número de componentes temporais que serão utilizadas e a avaliação preliminar da combinação de técnicas de seleção espacial e de adição de componentes temporais.

2 CONCEITOS PRELIMINARES

Esta seção tem como objetivo introduzir conceitos necessários para o entendimento do texto do presente trabalho.

A subseção 2.1 trata de modelos de circulação geral da atmosfera (GCMs), em especial, o modelo desenvolvido pelo renomado centro de previsão meteorológica ECMWF (*European Centre for Medium-Range Weather Forecasts*), ao qual será utilizado ao longo de todo o trabalho. É prática comum empregar as saídas dos GCMs em estudos relacionados com a modelagem de grandezas atmosféricas.

A subseção 2.2 aborda as técnicas de *downscaling*, que buscam acoplar a macroescala, também denominada escala sinóptica, variáveis provenientes do GCM com os dados locais (WILBY e WIGLEY, 1997). As técnicas de *downscaling* podem ser utilizadas em situações diversas, alguns exemplos são a previsão de curto prazo (e.g., COSTA *et al.*, 2008) e avaliação do recurso disponível no local de interesse, por exemplo, o recurso eólico, solar ou híbrido. Esse tipo de avaliação é fundamental para determinar a viabilidade de um empreendimento de geração de energia a partir de fontes renováveis.

A subseção 2.3 refere-se ao conceito de componentes temporais que será utilizado ao longo do estudo.

A subseção 2.4 discorre sobre os conceitos de autocorrelação e autocorrelação parcial, estes são fundamentais para os critérios de seleção de preditores, que são as variáveis de entrada dos modelos, utilizadas para estimar o preditando (medições no local de interesse).

A subseção 2.5 trata sobre testes de hipótese, que será de grande valia para a compreensão de alguns dos modelos desenvolvidos neste trabalho.

2.1 Modelos de Circulação Geral da Atmosfera (General Circulation Models, GCMs)

Os Modelos de Circulação Geral da Atmosfera são modelos matemáticos que aplicam métodos numéricos que resolvem as equações das leis de conservação (massa, energia, momento linear) de diversas variáveis atmosféricas, por exemplo, a velocidade do vento, precipitação e a temperatura em uma malha de baixa resolução espacial que envolve todo o globo terrestre

Figura *1*. As saídas dos modelos apresentam uma baixa resolução espacial com pontos de malha com distâncias para seu adjacente mais próximo da ordem de 100 quilômetros, por tratar um domínio computacional extenso, o planeta como um todo. Tal característica faz com

que alguns efeitos locais sejam suavizados, como a influência da orografia no perfil de vento. Portanto, o uso direto de dados de GCMs para representar efeitos locais não é aconselhado, tendo como alternativa o uso de técnicas de *downscaling.*

Figura 1 - Malha do GCM

Fonte: Escritório de Meteorologia do Governo de New South Wales (2003).

Os dados fornecidos pelo GCM podem ser divididos em três categorias: análise, reanálise e previsão. Os dados de análise são gerados em tempo real, por isso não são homogêneos, devido a isto existe recalibração e reparametrização dos modelos empregados ao longo dos anos. A reanálise é um procedimento em que se emprega um único modelo (mais apropriado) para todo o período de simulação, formando uma base de dados consistente, portanto, tende a apresentar erros menores que os modelos de análise. Os dados de previsão, por sua vez, são dados relativos ao estado futuro das variáveis atmosféricas (AMS, 2012). Neste trabalho, serão empregados dados de reanálise provenientes do ERA-Interim para modelagem do vento nos locais de interesse. O ERA-Interim é uma reanálise atmosférica global do ECMWF (*European Centre for Medium-Range Weather Forecasts*) com cobertura temporal na ordem de 4 décadas. Os dados são disponibilizados em 60 níveis verticais, sendo o nível 60, a aproximadamente 10 metros de altura, o mais próximo do solo. O banco de dados do ERA-Interim é gratuito para utilizações não comerciais (DEE *et al.*, 2011).

2.2 Aumento de Resolução – *Downscaling*

Os fenômenos meteorológicos ocorrem, de forma geral, em três escalas espaciais: a macroescala (sinóptica), a mesoescala, e a microescala. Na escala sinóptica, os fenômenos

meteorológicos apresentam um comprimento característico da ordem de milhares de quilômetros. A mesoescala, por sua vez, é uma escala intermediária, na qual ocorrem processos com comprimento característico da ordem de centenas de quilômetros, por exemplo, as brisas. A microescala é caracterizada pelos fenômenos mais locais, sendo mais sensíveis a variações de rugosidade e altitude. (ORLANSKI, 1975).

As técnicas de aumento de resolução, *downscaling*, são técnicas que são utilizadas para acoplar variáveis em diferentes escalas espaciais e temporais. O conjunto de técnicas pode ser dividido em dois grupos principais: as técnicas dinâmicas e as técnicas estatísticas (empíricas).

Na abordagem dinâmica, são resolvidas numericamente leis de conservação de massa, momento e energia que regem o comportamento das variáveis avaliadas.

A abordagem empírica busca descrever a preditando (microescala) a partir do preditor (macroescala), ou seja, parametrizando esta descreve-se uma relação matemática entre as duas escalas (WILBY *et al.*, 1998). As técnicas utilizadas para estabelecer tais relações, normalmente, são técnicas regressivas e técnicas de *clustering,* que buscam encontrar padrões sinópticos. O presente trabalho está inserido no campo dos modelos empíricos.

A seção 2.2.1 traz uma breve descrição da abordagem dinâmica, enquanto a seção 2.2.2 descreve as abordagens empíricas.

2.2.1 Abordagem Dinâmica

As abordagens dinâmicas podem ser divididas em dois tipos quanto ao acoplamento entre escalas. A primeira abordagem é a modelagem da camada limite planetária (PBL, *Planetary Boundary Layer*, e.g., LANDBERG e WATSON, 1994), a qual busca acoplar diretamente a macroescala com a microescala. A PBL é a porção atmosférica que está sujeita aos efeitos térmicos (trocas de calor) e mecânicos (orografia, obstáculos etc.) que ocorrem a partir da interação com a superfície. A segunda, busca acoplar as saídas dos GCMs com a mesoescala (*nested models* ou *Limited Area Models,* LAM; e.g., WILBY e WIGLEY, 1997). Uma maior descrição de tais modelagens podem ser obtidas a partir dos artigos supracitados.

2.2.2 Abordagem Empírica

A modelagem empírica, em geral, apresenta como vantagem a facilidade em ajustar os modelos para diferentes regiões, o que normalmente não se aplica para a abordagem

dinâmica, principalmente para regiões de instabilidade atmosférica devido à grande complexidade envolvida na modelagem, contudo a abordagem empírica tem como desvantagem a necessidade de dados observacionais de pelo menos um ponto no local de interesse para calibrar os modelos.

Na abordagem empírica é importante atentar aos cuidados necessários para a escolha de fatores como os preditores, o domínio empregado e o modelo apropriado para o estudo pretendido. (WILBY *et al.* 2004) propuseram um guia de boas práticas para a realização de *downscaling* estatístico.

A abordagem empírica também possui a vantagem de produzir bons resultados com menor esforço computacional. O *downscaling* estatístico pode ser dividido em três áreas principais: técnicas de regressão, classificação (*clustering*) e geradores estocásticos de tempo. O primeiro grupo emprega procedimentos de otimização de parâmetros para determinar relações lineares ou não lineares entre preditores e preditando. O segundo busca identificar padrões sinópticos, o que possibilita a parametrização de funções especializadas (locais) para as ocorrências classificadas. O terceiro grupo descreve o comportamento das variáveis atmosféricas como processos estocásticos. Este trabalho utilizará técnicas dos dois primeiros grupos, pois estes são utilizados em conjunto em outros exemplos da literatura como HART *et al.* (2015). A seção a seguir explica com um pouco mais de detalhes como as parametrizações são feitas.

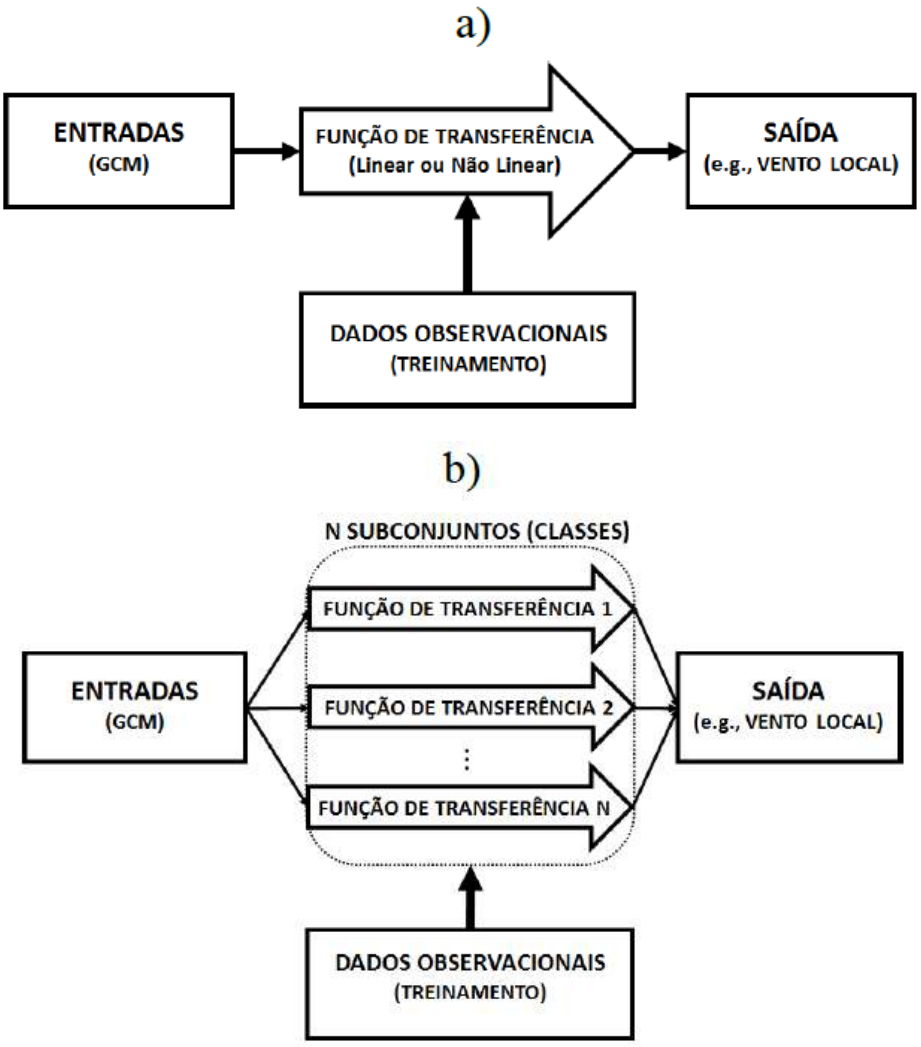
2.2.3 Parametrização de Modelos Estatísticos

Pode-se dividir as parametrizações em dois subgrupos: locais e globais. Os casos que serão abordados no presente trabalho, técnicas de *downscaling* estatístico, relacionam os dados provenientes do GCM com dados locais (observações, e.g., velocidades de vento) através de funções de transferência ajustadas em função da minimização do gradiente de erro global (calculado para todo o período disponível na série temporal). Nesse caso, considera-se que o modelo possui uma arquitetura global de parâmetros (Figura 2 a) (PERRUCI, 2018).

As parametrizações locais possuem parâmetros variáveis para diferentes padrões sinópticos. Isso quer dizer que, neste caso, ajustes são realizados em função da minimização de gradientes de erro locais (Figura 2 b), ou seja, os preditandos são divididos em subgrupos e por consequência o preditor também, o que implica que será feita uma parametrização para cada um dos subconjuntos.

Para a abordagem global e local, adota-se um conjunto específico para a calibração das funções de transferência, e outro apenas para validar os parâmetros resultantes. Após a parametrização utilizando o conjunto de calibração, os dados são validados, utilizando apenas os dados do GCM como entrada e em seguida são comparados com as observações destinadas para a validação com o intuito de verificar o desempenho do modelo empregado.

Figura 2 - Arquiteturas de Parâmetros de Modelos Estatísticos



Fonte: Perruci, 2018.

*Dados derivados de GCMs são relacionados com o vento local por meio de funções de transferência (lineares ou não lineares); a) Abordagem Global, uma função de transferência é ajustada para todo o período de treinamento; b) Abordagem Local, diversas funções de transferência são especializadas para subconjuntos, ou classes, formadas a partir do conjunto de variáveis preditoras (dados de simulação numérica).

2.3 Componentes Temporais

Usualmente quando se utiliza uma estratégia de *downscaling* estatístico para avaliação do recurso eólico local, os preditores têm instantes de tempo concomitantes com os

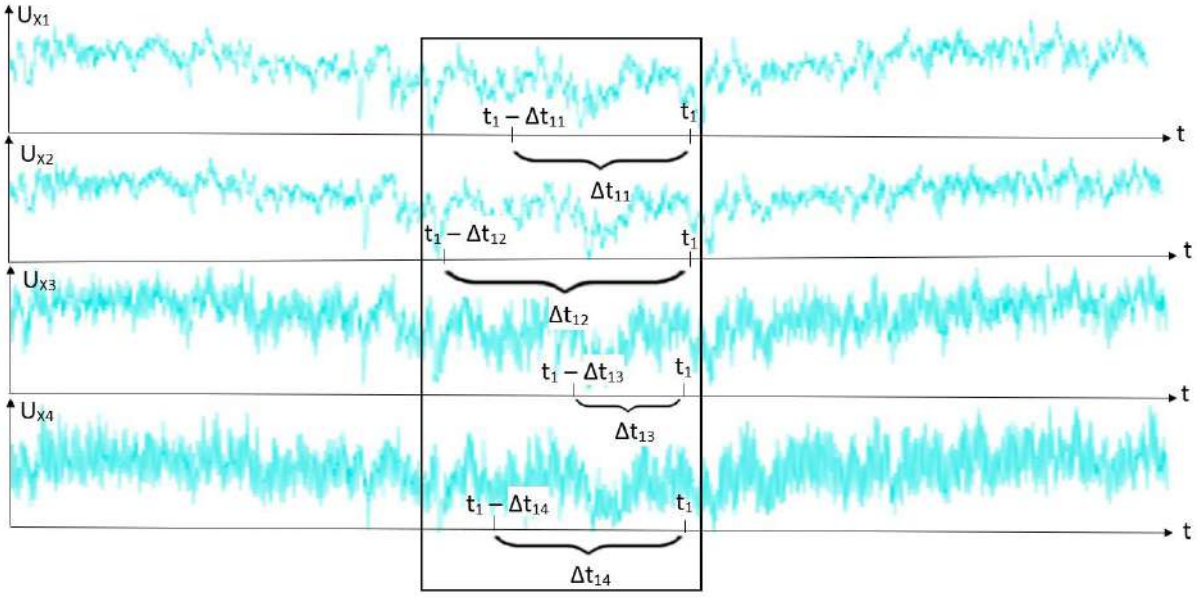
preditandos, ou seja, para a velocidade do vento, relaciona-se as variáveis da macro e microescalas no mesmo horário. As componentes temporais são obtidas a partir de instantes de tempo anterior ao que está sendo estimado. Então, quando se emprega essa metodologia, supõe-se que eventos ocorridos na macroescala podem se refletir na microescala posteriormente e, por isso, não são utilizados apenas dados concomitantes com as medições da torre de medição anemométrica (TMA), mas também dados em instantes antecessores ao momento analisado. Posteriormente, serão apresentados modelos com metodologias distintas para seleção das componentes temporais. As componentes temporais são definidas matematicamente na equação 1.

$$x'(t) = x(t-k) \quad (1)$$

Onde x é o preditor original, t é o instante avaliado e k é o valor da defasagem.

A Figura 3 representa a primeira componente temporal selecionada para os 4 pontos do GCM mais próximos do local de interesse, onde o eixo x representa o tempo e o eixo y a velocidade do vento em um ponto do GCM. A diferença entre os Δts da Figura 2.3 ocorre devido a metodologia utilizada para seleção das componentes temporais, a qual é detalhada na seção 4.

Figura 3 - Componentes Temporais



Fonte: O Autor, 2021.

2.4 Autocorrelação e Autocorrelação Parcial

A autocorrelação calcula a correlação de uma série temporal y_t com a mesma série temporal defasada um período, y_{t-k}, onde k é a defasagem da série, mais conhecida como *lag*. De acordo com BOX *et al.* (1994), a autocorrelação, r_k, é definida matematicamente através da expressão a seguir:

$$r_k = \frac{c_k}{c_0}$$

Onde c_0 é a variância amostral e c_k é definido adiante:

$$c_k = \frac{1}{T}\sum_{t=1}^{T-k}(y_t - \bar{y})(\bar{y}_{t+k} - \bar{y}) \qquad (2)$$

Onde T é o número de elementos na série e $\bar{y}$ é o valor médio.

Conforme definido formalmente por BUENO (2012), a função de autocorrelação parcial é o gráfico de $\widehat{\phi_{J,J}}$ contra j, estimado a partir das seguintes regressões em que a série original tem sua média subtraída:

$$y_t = \phi_{j,1}y_{t-1} + \phi_{j,2}y_{t-2} + \cdots + \phi_{j,j}y_{t-j} + e_t \; j = 1,2, \ldots. \qquad (3)$$

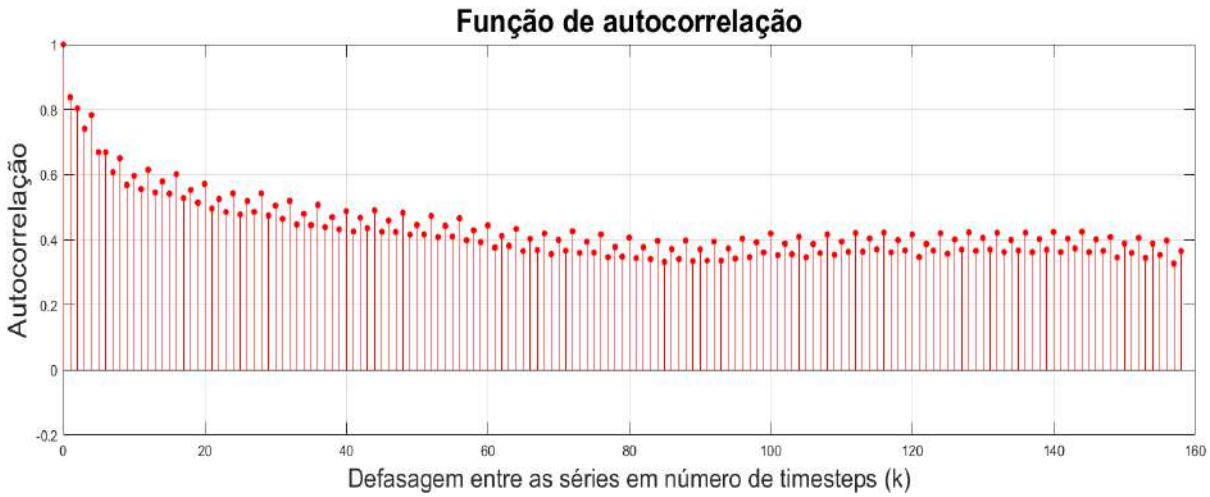
Onde y_t é a série temporal original, y_{t-1} é a série temporal deslocada em um *timestep*, y_{t-2} é a série temporal deslocada em dois *timesteps* e assim sucessivamente, já e_t é um erro.

Em outras palavras, o procedimento consiste em regredir y_t contra y_{t-1} e obter $\widehat{\phi_{1,1}}$. Em seguida deve-se regredir y_t contra y_{t-1} e y_{t-2}. São obtidos desta forma os coeficientes $\widehat{\phi_{2,1}}$ e $\widehat{\phi_{2,2}}$, dos quais interessa apenas este último; e assim por diante.

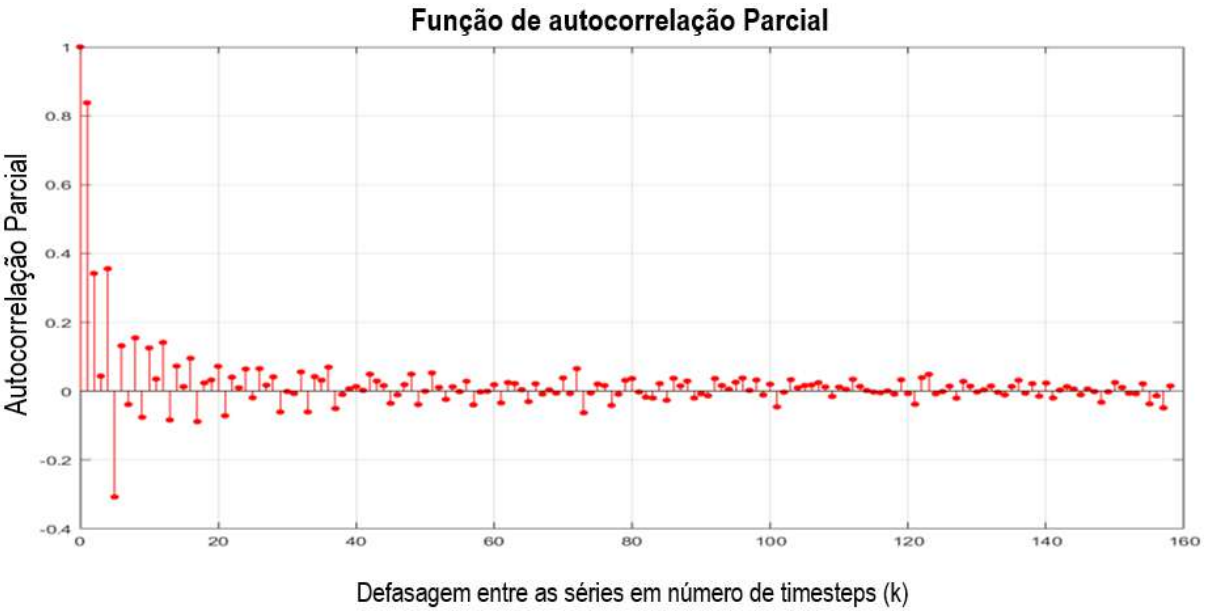
Em suma, o cálculo de autocorrelação e autocorrelação parcial são importantes na avaliação da identificação de componentes temporais significativas. Enquanto a autocorrelação costuma apresentar um decaimento suave em função de maiores defasagens (devido à relação de dependência entre as componentes temporais), a autocorrelação parcial isola a influência de uma determinada componente temporal subtraindo o efeito das outras no sinal original (BUENO, 2012). A Figura 4 a) ilustra a função de autocorrelação e a Figura *4* b) a autocorrelação parcial para um mesmo preditor.

Figura 4 - Autocorrelação e autocorrelação parcial

a)



b)



Fonte: O Autor, 2021.

2.5 Testes de Hipótese

A presente seção busca descrever brevemente o conceito de teste de hipótese, especialmente o teste de *Student*, conhecido como teste T, pois este foi o teste selecionado

para ser utilizado em alguns dos modelos desenvolvidos neste trabalho devido a sua grande difusão na literatura.

Uma hipótese é uma afirmação sobre uma propriedade da amostra. Quando é necessário decidir se uma hipótese, denominada hipótese nula (H_0), será aceita ou rejeitada, utiliza-se um procedimento denominado teste de hipótese.

Um teste de uma hipótese estatística ou teste de significância é o procedimento ou regra de decisão que nos possibilita decidir por H_0 ou uma hipótese alternativa (H_a).

Define-se região crítica (R_c) como o conjunto de valores assumidos pela variável aleatória ou estatística de teste para os quais a hipótese nula é rejeitada.

Ao se definir uma região crítica, pode-se cometer dois tipos de erros, são eles:

1) Erro tipo I: rejeita-se H_0 quando de fato H_0 é verdadeiro.
2) Erro tipo II: não se rejeita H_0 quando de fato H_0 é falso.

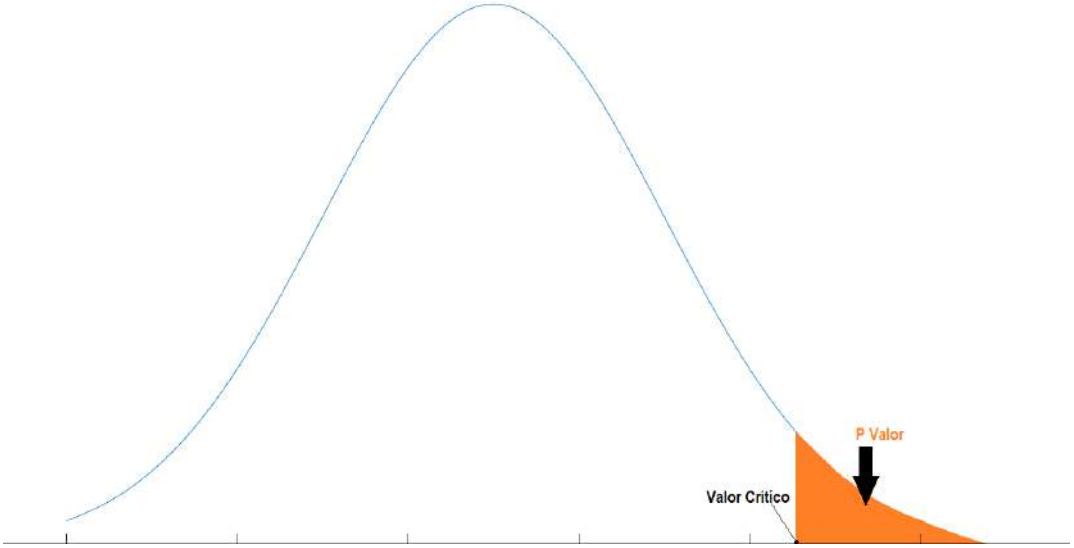
A distribuição T de *Student* é uma distribuição de probabilidade estatística. Uma grande vantagem desta distribuição é que o estatístico T da distribuição se baseia na média e desvio padrão amostral, diferentemente da distribuição normal.

Utilizar-se-á ao longo desse trabalho a distribuição T na avaliação do nível de significância dos coeficientes de regressões lineares múltiplas.

Quando realizamos um teste de hipótese sobre o coeficiente de uma regressão, temos que a hipótese nula equivale ao valor desse coeficiente ser igual a zero.

A qualquer hipótese nula se associa um estatístico denominado p-valor. O p-valor é um indicador contra a hipótese nula, ou seja, quanto menor o p-valor, maior é o indicativo que a hipótese nula deve ser rejeitada. Quando se considera que o p-valor associado ao coeficiente de uma regressão deve ser, por exemplo, menor que 0,05, significa que o preditor só será utilizado caso exista uma probabilidade menor que 5% de cometer um erro do tipo I. A Figura 5 é uma ilustração da região da curva em que a hipótese nula será rejeitada para determinado p-valor.

Figura 5 - Ilustração de uma distribuição de probabilidade e um p-valor associado a um valor crítico.



Fonte: O Autor, 2021.

3 REVISÃO BIBLIOGRÁFICA

A busca por energias de cunho renovável vem crescendo exponencialmente (REN21, 2020), com destaque para fontes eólicas e solares, o que torna cada vez mais imprescindível o aprofundamento em tais temas visando o maior aproveitamento destes recursos.

A seção 2.2.2 introduziu as técnicas de *downscaling* estatístico, estas podem ser separadas em técnicas de regressão, classificação e geradores estocásticos. WILBY E WIGLEY (1997) E XU (1999) apresentaram revisões referente ao tema abordando as técnicas conhecidas na época.

As técnicas de interpolação são muito difundidas, pois estas apresentam uma implementação de baixa complexidade. HEAP (2008) apresenta diversas abordagens e aplicações destas técnicas. No contexto do *downscaling* estatístico, a interpolação bilinear é frequentemente empregada como técnica de referência, uma vez que se dá a partir de uma média ponderada em função apenas das posições geométricas dos preditores (ACCADIA *et al.*, 2003; BERNHARDT *et al.*, 2010). Cotidianamente as ponderações se dão referentes às posições no plano horizontal, entretanto, é possível incluir outras informações relevantes para o estudo em questão. Exemplificando, têm-se os casos de modelagem do vento próximo a superfícies de orografia complexa, onde mostrou-se que a elevação é um importante fator a ser considerado (PALOMINO e MARTÍN, 1995).

As técnicas de regressão também são bastante difundidas e comumente empregadas em modelos de arquitetura global. Estas técnicas apresentam uma maior complexidade quando comparadas com as interpolações, pois possuem procedimentos de otimização de parâmetros baseados na minimização de um gradiente de erro de ajuste entre preditandos e preditor. As relações estabelecidas entre ambos os conjuntos podem ser lineares ou não lineares.

Considerando o supracitado, uma técnica usual é a regressão linear múltipla, pois, apesar de possuir implementação relativamente simples, costuma apresentar resultados satisfatórios, sendo utilizada para comparação com modelos de maior complexidade (MURPHY, 1999; CURRY *et al.*, 2012).

Outro grupo comum de técnicas são as técnicas conhecidas como classificação de padrões sinópticos (WILBY e WIGLEY, 1997). A proposição feita inicialmente por LORENTZ (1969), modelou uma técnica baseada em “análogos”, e está entre as mais simples de classificação, servindo como referência para técnicas mais complexas, sendo em alguns casos até mesmo equivalente ou superior (ZORITA e VON STORCH, 1999).

A palavra “análogos” trata do par de instantes em que os respectivos estados sinópticos estão “próximos” entre si, sendo um relativo ao instante alvo (a ser estimado), e o outro relativo ao banco de dados (PERRUCI, 2018).

As técnicas de análogos consistem em um procedimento de composição de variáveis de entradas, que são formados a partir do ranqueamento de distâncias em espaços N-dimensionais, onde N é o número de variáveis preditoras no domínio voltadas para o treinamento de modelos regressivos. O conjunto de variáveis preditoras descreve a variabilidade temporal de estados atmosféricos, os quais são amplamente conhecidos como "padrões sinópticos". Nesse sentido, a técnica de análogos se baseia no pressuposto de que entre dois instantes em que padrões sinópticos são "próximos", ou análogos (i.e., estados atmosféricos similares), o comportamento microescalar da variável também será análogo.

Entretanto as técnicas baseadas em análogos apresentam dois problemas, que estão interligados. O primeiro é que esta abordagem necessita de um esforço computacional significativo sobre um longo período de dados observacionais (GUTIÉRREZ *et al*., 2004). O segundo fator é a necessidade de bancos de dados com longos períodos de informações para que o método dos análogos tenha um bom desempenho (VAN DEN DOOL, 1994).

As técnicas de agrupamento (*clustering*) são utilizadas comumente em estruturas de arquitetura local, servindo como alternativas à técnica de análogos (GUTIÉRREZ *et al.*, 2004). O objetivo da metodologia, de forma simplificada, é a formação de subconjuntos de dados o mais semelhante internamente e o mais distinto externamente quanto for possível. Há duas categorias principais de *clustering*: Hierárquico e Não-Hierárquico (WILKS, 2011).

No *clustering* hierárquico, o agrupamento é feito iterativamente. Inicialmente cada subgrupo contém apenas um elemento. Em seguida, são agrupados os subgrupos considerados mais semelhantes entre si, assim, cada fusão reduz o número de subgrupos (HART *et al.*, 2015).

Já as técnicas de *clustering* não-hierárquico permitem a troca de elementos entre subgrupos durante o processo, o que tende a gerar uma configuração mais aperfeiçoada. Pode-se definir um número de grupos inicialmente, caracterizados pelos seus centroides e reclassificados nas etapas posteriores até que uma configuração estável seja obtida (GUTIÉRREZ *et al.*, 2004).

Por fim, os geradores estocásticos de tempo são modelos probabilísticos utilizados para descrever variáveis na escala local (BERNARDIN *et al.*, 2009; WILKS, 2010) e em estudos relacionados a mudanças climáticas (SEMENOV e BARROW, 1997; WILKS, 1999).

Normalmente, atribui-se aos modelos de *downscaling* estatísticos um conjunto de treinamento com um grande número de preditores, sendo usual nesses casos que os preditores sejam linearmente independentes, ou seja, que não possuam informações redundantes, que, além de poderem ser a causa da diminuição da acurácia do resultado de alguns modelos, exigem maior poder computacional. É possível filtrar as informações a partir da análise de componentes principais (*Principal Component Analysis* - PCA; e.g., JOLLIFFE, 1986; BORDONI e STEVENS, 2006), análise de correlações canônicas (BUSUIOC *et al.*,2008), ou ambas (HUTH, 1999). Essas técnicas determinam novas bases em que são aplicadas técnicas de regressão e classificação para relacionar preditores e preditando. Outra possibilidade para diminuir informações redundantes entre os preditores é calcular a autocorrelação parcial entre os mesmos e utilizá-la como critério para eliminar preditores redundantes (HESSAMI *et al.*, 2008) e encontrar relações importantes entre preditores e a observação (MENDES *et al.*, 2010).

Usualmente as técnicas de *downscaling* estatístico utilizam instantes de tempo concomitantes para realizar as estimativas. HEWITSON E CRANE (1996) afirmaram que a estimativa da variável local pode depender de forma significativa de eventos em instantes anteriores na escala sinóptica.

HARPHAM E WILBY (2005) obtiveram melhora nas estimativas utilizando dados do passado do preditor em relação ao instante estimado do preditando para corrigir a diferença de fase entre eles. O uso de componentes temporais melhora também a estimativa da variância (WILBY, 2008; WILBY *et al.*, 2013).

A utilização de dados do passado é uma prática comum em determinados campos de pesquisa, como as redes neurais. Elas também podem ser utilizadas para realizar técnicas de *downscaling* como as redes neurais do tipo *Time lagged feedforward neural network* (DIBIKE E COULIBALY, 2006).

O uso da correlação espacial entre preditores e preditandos e a série de autocorrelação parcial dos preditandos são critérios utilizados para determinar os preditores que serão utilizados na metodologia desenvolvida em (PICHUKA, 2016) que gerou melhora nos resultados em comparação com uma regressão comum. Em concordância com o mencionado anteriormente, a correlação espacial entre os preditores e preditando é um fator que quando selecionado de forma adequada pode contribuir significativamente para melhorar a descrição do preditando como é o caso na avaliação do comportamento da temperatura no local de interesse feito por WANG *et al*. (2020).

Apesar de não ser uma grande novidade o uso de componentes temporais, metodologias objetivas e robustas não são amplamente difundidas, porém trabalhos recentes estão trazendo resultados interessantes nesse campo. FENG *et al.* (2017) desenvolveram uma metodologia utilizando a análise de componentes principais, testes de casualidade e funções de autocorrelação para seleção de preditores que alimentam redes neurais, que são empregadas nos modelos de combinação para estimar o comportamento do vento no curto prazo.

Em adição ao supracitado, publicações recentes verificam um ganho nas estimativas da variável de interesse ao utilizar dados do passado do preditor ao momento estimado contribuindo significativamente para melhora das estimativas. Esta é uma das conclusões de (HUANG *et al.*, 2020) que notou um ganho de desempenho dos modelos ao utilizar componentes temporais e que ao realizar esta adição conseguiu estimativas com domínios menores similares a estimativas com domínios iniciais maiores, desta forma diminuindo a quantidade de preditores necessários.

Tendo em vista a revisão da literatura acima mencionada, nota-se que a utilização de componentes temporais não é algo desconhecido, porém não possui uma grande gama de metodologias associada a seleção de preditores para técnicas de *downscaling.*

Considerando todo o assunto abordado, verifica-se que algumas características em comum presentes nesses trabalhos são:

1) Uso de componentes temporais como preditores;
2) Avaliação espacial dos preditores;
3) Avaliação temporal dos preditores;
4) Seleção objetiva dos preditores;
5) Filtrar preditores;
6) Aplicação em técnicas de *downscaling;*
7) Validação da relevância das variáveis utilizando testes de hipótese;
8) Avaliação do recurso no longo prazo.

De maneira a comparar as características de alguns trabalhos na literatura e o trabalho aqui proposto, tem-se a Tabela 1, a seguir, na qual os trabalhos estão dispostos nas linhas, e as características, citadas anteriormente, nas colunas. A última linha da tabela refere-se a este trabalho.

Tabela 1 - Características de trabalhos relacionados e do presente trabalho

Trabalhos	Características							
	1	2	3	4	5	6	7	8
HUANG *et al.*, 2020	x	x		x	x	x		x
FENG *et al.*, 2017	x		x	x	x		x	
WILBY *et al.*, 2013	x		x		x	x		x
HESSAMI *et al.*, 2008				x	x	x		x
DIBIKE E COULIBALY, 2006	x		x			x		
HARPHAM *et al.*, 2005	x		x			x		
ALVES, 2021	x	x	x	x	x	x	x	x

Fonte: O Autor, 2021.

Com base nas informações Tabela 1, nota-se que o presente trabalho está alinhado com a literatura, incorporando características de diversos trabalhos e ainda se diferenciando por levar em conta a avaliação espacial e temporal dos preditores de forma simultânea em conjunto com técnicas objetivas de seleção espacial e temporal dos preditores.

4 METODOLOGIA E MODELOS

A metodologia a ser adotada possui três etapas principais (Figura 6). A primeira consiste na definição dos conjuntos de calibração, validação e teste, a etapa consequente é a parametrização do modelo e por fim é feita a avaliação dos resultados.

Figura 6 – Diagrama Geral da Metodologia

Fonte: O Autor, 2021.

O bloco na parte superior da Figura 6 representa os preditores para os modelos de *downscaling* estatístico, que são os dados do GCM adotados no estudo (ERA-Interim). A primeira caixa representa a definição dos conjuntos de calibração, validação e teste. Na segunda etapa é utilizada uma estratégia de validação cruzada, o que implica que as séries temporais de calibração e validação não podem possuir interseção para ser possível parametrizar e validar o modelo (*cross validation approach*, e.g., HUTH, 1999). É importante ressaltar que não necessariamente os dados utilizados são de instantes de tempo concomitantes, apesar de ser a prática mais comum, como ficará claro em alguns dos modelos utilizados neste trabalho.

Por fim, na última etapa temos o teste do modelo, onde é utilizada a parametrização definida na etapa anterior e utiliza-se o período do GCM destinado ao período de teste para estimar os dados observacionais deste período e fazer avaliação dos resultados.

A divisão entre os conjuntos de calibração, validação e teste é feita de forma análoga em todos os modelos. O conjunto de calibração e validação corresponde aos dois terços iniciais da série temporal, sendo o conjunto de teste o um terço restante. Os primeiros dois terços da série temporal são novamente subdivididos e o período de calibração corresponde

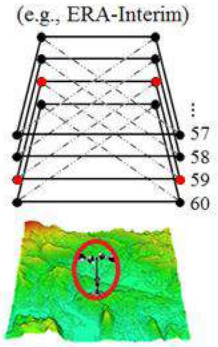
aos dois terços iniciais deste conjunto, ou seja, quatro nonos do período total, consequentemente o período de validação corresponde a um nono do período total.

O grupo multidisciplinar de técnicas de *downscaling* estatístico avaliadas neste estudo incluem interpolação bilinear (IBL), regressões lineares múltiplas (MLR), análise de componentes principais (PCA), *clustering* (CL) e variantes.

Os preditores para cada local de interesse são compostos por dados de reanálise ECMWF ERA-INTERIM, com uma resolução horizontal de 0,75 graus (para latitude e longitude). Os dados provenientes do ECMWF ERA-INTERIM também foram utilizados como preditores nos trabalhos de DANTAS (2021) e PERRUCI (2018), e mostraram-se como uma boa escolha, contribuindo para que os modelos gerassem estimativas de boa acurácia. Os domínios aplicados são compostos pelos 15 níveis de modelo mais próximos à superfície (46 a 60), bem como pelo nível superficial (altura fixa a 10 m acima do solo). As variáveis fornecidas pelo GCM nesse estudo são as componentes zonal e meridional ("U" e "V", respectivamente). A grandeza estimada nesse estudo é a magnitude da velocidade do vento, dada por $M=\sqrt{\mathrm{U}^2+\mathrm{V}^2}$

Os domínios iniciais utilizados entre os diferentes modelos de *downscaling* estatístico divergem entre si, o que significa que os modelos possuem preditores diferentes e uma quantidade distinta destes. Os tipos de malhas horizontais utilizados nos modelos podem conter os 4 ou 36 pontos mais próximos à posição de interesse. Além disso, uma quarta configuração de domínio foi composta a partir da união entre os 16 níveis mencionados, considerando as malhas horizontais com 36 pontos, totalizando assim 576 pontos (em disposição tridimensional). Os modelos que utilizam como domínio inicial uma malha horizontal com apenas 4 ou 36 pontos são avaliados 16 vezes durante a, pois o mesmo modelo é simulado mudando apenas o nível vertical considerado. A Figura 7 ilustra os níveis verticais. Cada um dos níveis verticais (altura do "paralelepípedo") é utilizado como domínio inicial uma vez, resultando em 16 diferentes resultados para estes cada um destes modelos. O Quadro 1 apresenta as configurações citadas, bem como os modelos relacionados. O significado de cada sigla e os modelos são explicados na seção 4.1.

Figura 7 - Níveis verticais de modelos de circulação geral da atmosfera



Fonte: Adaptado de PERRUCI, 2018.

Quadro 1 - Domínio para cada um dos modelos adotados

Domínios		
4 pontos (2 x 2)	36 pontos (6 x 6)	576 pontos (6 x 6 x 16)
IBL, MLR1, MLRTA1, MLRCLUSTER1	MLR2, MLRTA2, MLRCLUSTER2	PCAMLR, PCASMLR, RTCT, RTCTP

Fonte: O Autor, 2021.

Existem diversas maneiras de avaliar o desempenho de um modelo. Uma metodologia amplamente difundida é o uso do diagrama de Taylor. No mesmo artigo em que é descrito a metodologia do referido diagrama, introduz-se um índice de habilidade (*skill score*), o SS4, definido na Equação 4, que será adotado neste trabalho como principal avaliador para quantificar o desempenho geral dos modelos. Na Equação 4, R significa a correlação entre a saída do modelo e a observação e $\hat{\sigma}$ representa o desvio padrão da estimativa normalizado pelo desvio padrão das observações. Com o índice de habilidade SS4, os ajustes dos modelos sobre os dados observacionais são classificados dentre valores de zero até um, representando o pior e o melhor índice de qualidade, respectivamente (TAYLOR, 2001).

$$SS4 = \frac{(1+R)^4}{4 * (\hat{\sigma} + 1/\hat{\sigma})^2} \quad (4)$$

4.1 Técnicas empregadas nos modelos de *downscaling* estatístico

Nesta seção, são apresentadas as técnicas empregadas nos modelos de *downscaling* utilizadas neste trabalho. A seção 4.2 detalhará os modelos compostos por composições das

técnicas descritos na seção 4.1. O quadro sintetiza os diferentes tipos de técnicas descritos nesta seção e seu objetivo.

Quadro 2 - Tipos de Técnicas utilizadas nos modelos de *downscaling*

Técnica	Objetivo
IBL, MLR	Busca encontrar relações empíricas entre dois conjuntos. Esse conjunto trata das técnicas de *downscaling*
PCA	Preserva o mínimo de dados fornecendo a maior quantidade de informações do conjunto avaliado.
Clustering	Separa um conjunto em subconjuntos de forma que estes sejam o mais homogêneo possível internamente e o mais distinto possível dos demais. Esse tipo de técnica procura tornar as técnicas de associação mais eficientes
Autocorrelação/ Autocorrelação Parcial	Busca identificar semelhanças temporais dentro de uma série.

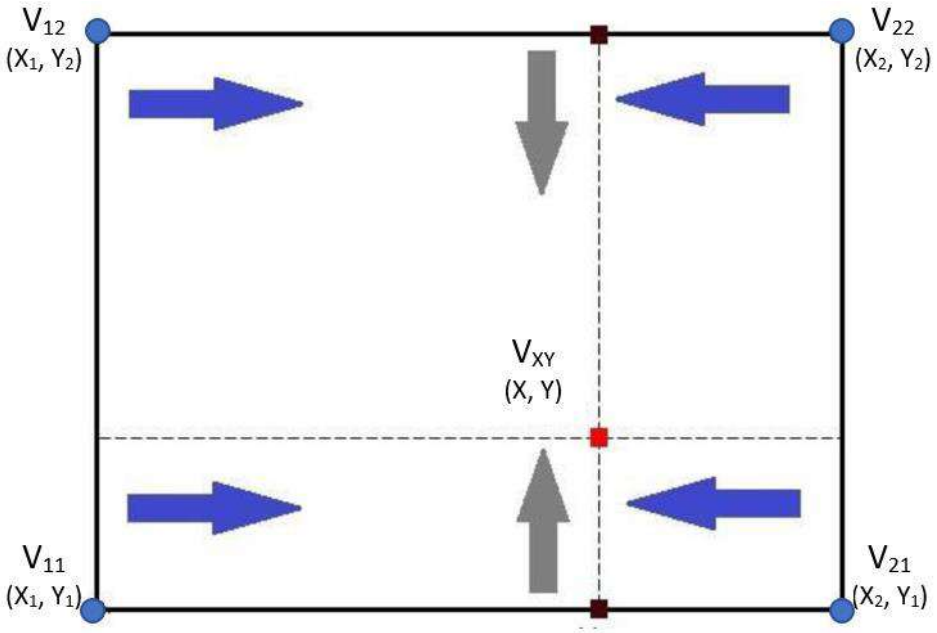
Fonte: O Autor, 2021.

4.1.1 Interpolação Bilinear (IBL)

A interpolação bilinear foi utilizada como modelo de referência e comparado com os demais modelos por sua simplicidade. A técnica tende a subestimar a variabilidade do comportamento do vento na microescala, pois trata-se de uma suavização das variáveis em questão (ACCADIA *et al.*, 2003). Isso é um resultado esperado, pois os processos que ocorrem na microescala são mais complexos do que os da macroescala, devido ao número maior de influências no vento local, como construções e variação no tipo de solo. Como consequência, no caso do vento local, a interpolação bilinear tende a elevar o valor mínimo e reduzir o valor máximo.

Ao se aplicar a interpolação bilinear utilizando os 4 pontos do GCM mais próximos ao local de interesse para determinar os valores da velocidade do vento na microescala, considera-se que o comportamento do vento é função apenas do espaço. Essa técnica consiste na realização de três interpolações lineares, sendo duas interpolações na direção das abcissas e uma na direção das ordenadas. De forma análoga pode-se realizar duas interpolações na direção das ordenadas e uma na direção das abcissas. A Figura 8 ilustra a metodologia da interpolação bilinear.

Figura 8 - Ilustração da interpolação bilinear



Fonte: O Autor, 2021.

A estimativa da variável de interesse, neste caso, a velocidade do vento, pode ser obtida através da resolução das Equações 5 a 7.

$$\left(f(V_{y1}) = f(x, y_1) = \frac{(x_2 - x)}{(x_2 - x_1)} * f(V_{11}) + \frac{(x - x_1)}{(x_2 - x_1)} * f(V_{21})\right) \quad (5)$$

$$\left(f(V_{y2}) = f(x, y_2) = \frac{(x_2 - x)}{(x_2 - x_1)} * f(V_{12}) + \frac{(x - x_1)}{(x_2 - x_1)} * f(V_{22})\right) \quad (6)$$

$$\left(f(V_{xy}) = f(x,y) = \frac{(y_2 - y)}{(y_2 - y_1)} * f(V_{y1}) + \frac{(y - y_1)}{(y_2 - y_1)} * f(V_{y2})\right) \quad (7)$$

Onde $f(V_{xy})$ representa a velocidade do vento no local de interesse, x e y representam respectivamente a latitude e longitude. A notação para os pontos da malha do GCM (pontos em azul) foi feita de forma análoga.

4.1.2 Regressão Linear Múltipla (MLR)

A regressão linear múltipla, usualmente denominado como MLR, busca estabelecer uma relação empírica entre preditores e preditandos. No trabalho em questão, a técnica parametriza funções lineares empíricas que descrevem a velocidade do vento local a partir da velocidade do vento proveniente do GCM. A representação matricial é ilustrada abaixo.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Onde,

n - quantidade de medições, k – número de variáveis preditoras, yi – dados observacionais; βj – parâmetros do modelo; xij – variáveis regressoras; εi – erro associado ao modelo.

Em notação matricial simplificada temos que o MLR (GELADI e KOWALSKI, 1986) é representado pela Equação 8.

$$Y = X\beta + \epsilon \quad (8)$$

Os parâmetros são calculados minimizando o erro quadrático médio, conhecido como método dos mínimos quadrados, ou seja, é necessário minimizar a expressão mostrada na Equação 9.

$$S = \sum_{i=1}^{n} {\varepsilon_i}^2 = {\varepsilon_1}^2 + {\varepsilon_2}^2 + \cdots {\varepsilon_n}^2 \quad (9)$$

O que pode ser escrito da seguinte forma:

$$S = \epsilon' * \epsilon = (\mathrm{Y} - \mathrm{X}\beta)\,'(\mathrm{Y} - \mathrm{X}\beta) \quad (10)$$

Pode-se encontrar o erro quadrático médio mínimo calculando a primeira derivada de S em relação a $\boldsymbol{\beta}$ e igualando a expressão a 0.

$$\mathrm{S} = 2\mathrm{X'Y} + 2\mathrm{X'X}\,\hat{\beta} \quad (11)$$

A partir da expressão acima obtém-se:

$$\hat{\hat{\beta}} = (\mathrm{X'X})^{-\mathbf{1}}\mathrm{X'Y} \quad (12)$$

A partir da equação anterior, tem-se a série estimada dada por:

$$\hat{\mathrm{Y}} = X\,\hat{\beta} \quad (13)$$

O fato de o MLR utilizar dados observacionais em sua parametrização faz com que ele leve em conta o comportamento do vento local em sua modelagem e, por isso, seu resultado normalmente é mais próximo da realidade do que o da interpolação bilinear. No estudo em questão, foram empregadas no MLR as medições da velocidade do vento realizadas por um anemômetro no local de interesse.

4.1.3 Análise de componentes principais (PCA)

A análise de componentes principais ou PCA reduz um conjunto de dados contendo uma grande quantidade de variáveis em um conjunto menor de novas variáveis. O novo conjunto é formado por variáveis que são combinações lineares das originais e são escolhidas

de forma que representem a maior fração possível da variabilidade do conjunto inicial (WILKS, 2011).

As novas variáveis são vetores ortogonais entre si denominadas componentes principais. Considerando um espaço multidimensional formado por m preditores e t instantes de tempo, R_{txm}, as componentes principais indicam as direções sobre as quais são projetados os eixos da base ortogonal de R_{txm}.

Matematicamente, as componentes principais podem ser determinadas encontrando os autovetores da matriz de correlação ($\Sigma m \times m$) do espaço em questão e a sua variância é representada pelo autovalor associado ao autovetor (JOLLIFFE, 1986). Mantém-se o menor número n de componentes ($Pm \times n$) de forma que o erro de reconstrução seja inferior a 5% (GUTIÉRREZ *et al*., 2004). A projeção do conjunto original na nova base ortogonal ($X_{t \times n}$), o espaço reconstruído ($R_t \times m_*$) e o erro de reconstrução (E_R) são definidos nas Equações de 14 a 16, respectivamente:

$$X_{t \times n} = \tilde{R}_{t \times m} \times P_{m \times n} \tag{14}$$

$$R^*_{t \times m} = \tilde{X}_{t \times n} \times P'_{n \times m} \tag{15}$$

$$E_R = \frac{1}{M}\frac{1}{T} \cdot \sum_{j=1}^{M} \sum_{i=1}^{T} \frac{\left(R^*_{i \times j} - \tilde{R}_{i \times j}\right)^2}{{R_{i \times j}}^2} \tag{16}$$

O termo $\tilde{R}$ representa a matriz das componentes principais que foram preservadas e R_{ixj} é um espaço multidimensional formado por j preditores e i instantes de tempo. É importante destacar que os preditores são “estandardizados”, ou seja, com médias nulas e desvios unitários. “M” e “T” indicam o número de variáveis regressoras e número de *timesteps*, respectivamente.

4.1.4 Agrupamento de padrões Sinópticos - *Clustering*

Esta seção apresenta as técnicas utilizadas para encontrar grupos semelhantes entre si em uma série de dados, ou seja, técnicas de agrupamentos, mais conhecidas como técnicas de *clustering*. De forma simples, pode-se dizer que estas técnicas buscam formar grupos com

valores o mais semelhante possível entre si e o mais distinto possível dos valores que se encontram nos outros grupos. Um fator muito importante nessas técnicas é determinar o critério de "semelhança".

Neste estudo, os grupos são os padrões sinópticos provenientes do GCM. Ao longo do trabalho, serão consideradas duas formas de *clustering*: (a) o agrupamento hierárquico aglomerativo (AHC, *Agglomerative Hierarquical Clustering*; SCHOOF e PRYOR, 2001; VRAC *et al.*, 2007); (b) o procedimento não hierárquico (NHC, *Non-Hierarquical Clustering*) baseado no algoritmo *k-means* (e.g., GUTIÉRREZ *et al.*, 2004). No AHC, o número de *clusters* C é estabelecido à posteriori, em função de um critério de parada de fusões. No NHC, por sua vez, C é decidido à priori, em função das classes adotadas na inicialização. Ao longo do trabalho, as classes adotadas para a inicialização do algoritmo *k-means* são obtidas através das saídas de um método hierárquico. O critério de parada para determinar o número de *clusters* será explicado de forma mais detalhada na seção dos modelos.

Após a determinação dos *clusters*, será feito uma regressão linear múltipla para cada um dos agrupamentos (MLR local) com intuito de estimar a velocidade do vento no local de interesse. O primeiro critério será o número mínimo de elementos que o *cluster* deve ter para ser utilizado. A partir de uma análise de sensibilidade determinou-se para esse estudo que cada cluster deve conter pelo menos 5% dos elementos da série temporal para que seja considerado no estudo. Além disto, em paralelo as regressões locais, é feita uma regressão linear múltipla global, ou seja, um MLR convencional. O segundo critério é comparar o resultado da saída de cada *cluster* com o MLR global, utilizando como critério de referência para comparação o cálculo do SS4. Esse critério garante que a saída estimada terá o mesmo número de *timesteps* (intervalos de tempo) que a saída dos demais modelos, tornando possível a comparação de desempenho entre os modelos.

Em relação aos métodos hierárquicos, usualmente é necessário a determinação de três fatores: (a) métrica de dissimilaridade D; (b) algoritmo de fusão, que é o procedimento adotado para definição e hierarquização das dissimilaridades entre duas classes D; (c) número de *clusters* C. Quanto ao método não hierárquico, para o *k-means*, são adotados três critérios: (a) métrica de dissimilaridade; (b) classes iniciais (C centróides); e (c) critério de parada, estabelecido aqui como um limite de 1000 iterações caso uma configuração estável não seja obtida. As métricas e algoritmos utilizados são os mesmos utilizados em (PERRUCI, 2018). Os detalhes sobre as métricas e algoritmos de hierarquização empregados estão na Tabela 2

Tabela 2 - Configuração geral de técnicas de agrupamento

a) Distâncias entre padrões (x_t e x_s)	Equações
Manhattan	$d = \sum_{j=1}^{n} \left\| x_{t_j} - x_{s_j} \right\|$
Euclideana	$d = \left[\sum_{j=1}^{n} \left(x_{t_j} - x_{s_j} \right)^2 \right]^{1/2} = \left\| x_t - x_s \right\|$
Cosseno	$d = 1 - \frac{x_t x_s'}{\sqrt{(x_t x_t')\,(x_s x_s')}}$
Correlação	$d = 1 - \frac{(x_t - \bar{x_t})(x_s - \bar{x_s})'}{\sqrt{(x_t - \bar{x_t})(x_t - \bar{x_t})'} \cdot \sqrt{(x_s - \bar{x_s})(x_s - x_s)'}}$
b) Dissimilaridades (D) entre c_p e c_q	**Algoritmos**
Vizinhos mais Próximos (VP)	$D(c_p, c_q) = \min\left(d(c_{pu}, c_{qv}) \right), i \in \{i, \ldots, n_p\}, j \in \{1, \ldots, n_q\}$
Vizinhos mais Distantes (VD)	$D(c_p, c_q) = \max\left(d(c_{pu}, c_{qv}) \right), i \in \{i, \ldots, n_p\}, j \in \{1, \ldots, n_q\}$
Vínculo Médio	$D = \frac{1}{n_p n_q} \sum_{u=1}^{n_p} \sum_{v=1}^{n_q} d(c_{pu}, c_{qv})$
Vínculo de mínima variância agregada (Método de Ward)	$D = \sqrt{\frac{2 n_p n_q}{(n_p + n_q)}} \left\| \bar{c}_p - \bar{c}_q \right\|$

(a) Métricas adotadas AHC e NHC; (b) Algoritmos de Hierarquização de Dissimilaridades. Na primeira parte (a), os padrões sinópticos x_t e x_s são vetores n-dimensionais contendo as variáveis da malha do GCM para os instantes t e s. Na segunda parte (b), c_p e c_q representam classes arbitrárias, as quais contém n_p e n_q elementos associados, respectivamente. Note que c_{pu} e c_{qv} correspondem, respectivamente, ao u-ésimo e v-ésimo elemento de c_p e c_q. Com respeito ao método de Ward, $\bar{c}$ corresponde ao centroide, ou médias, tomadas a cada dimensão j dos padrões (x) pertencentes a uma classe c (vetor n-dimensional). Note que o método de Ward apenas se aplica a distâncias euclidianas.

Fonte: Perruci, 2018

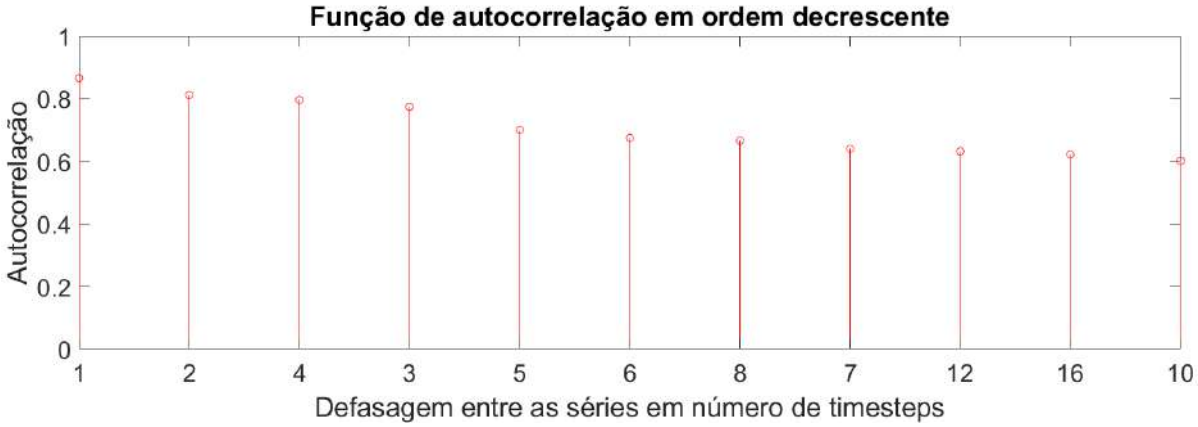
4.1.5 Adição de Componentes temporais utilizando autocorrelação ou autocorrelação parcial

Na seção 2.3 foram definidas as componentes temporais (CTs). Estas serão selecionadas de uma forma objetiva baseando-se nas funções de autocorrelação e autocorrelação parcial. É importante enfatizar que cada ponto de malha tem um comportamento diferente.

A Figura *4* a) ilustra o valor da correlação entre a série de velocidade de um ponto de malha do GCM com a própria série defasada por um número de *timesteps*, representado na abcissa do gráfico. Nota-se que o valor da correlação da série não tem uma relação linear com a defasagem aplicada, apesar de existir uma tendência global de queda com o aumento da defasagem.

O critério de seleção baseado na função de autocorrelação é ordenar as componentes temporais de acordo com o valor da função autocorrelação. O conceito pode ser mais facilmente compreendido observando a Figura 9.

Figura 9 - Correlações entre as séries defasadas organizadas em ordem decrescente e as respectivas defasagens.



Fonte: O Autor, 2021.

Observando o eixo das abcissas, é possível perceber que valores mais elevados da função de autocorrelação não acompanham fielmente a ordem crescente das defasagens. O critério de seleção de componentes temporais definido neste trabalho faz a adição de acordo com o critério de ordenamento definido anteriormente. É importante destacar que cada ponto da malha do GCM terá um comportamento diferente. No caso do ponto de malha exemplificado na Figura 9 as três primeiras CTs adicionadas estariam defasadas em 1, 2 e 4 *timesteps* em relação ao preditor original, respectivamente.

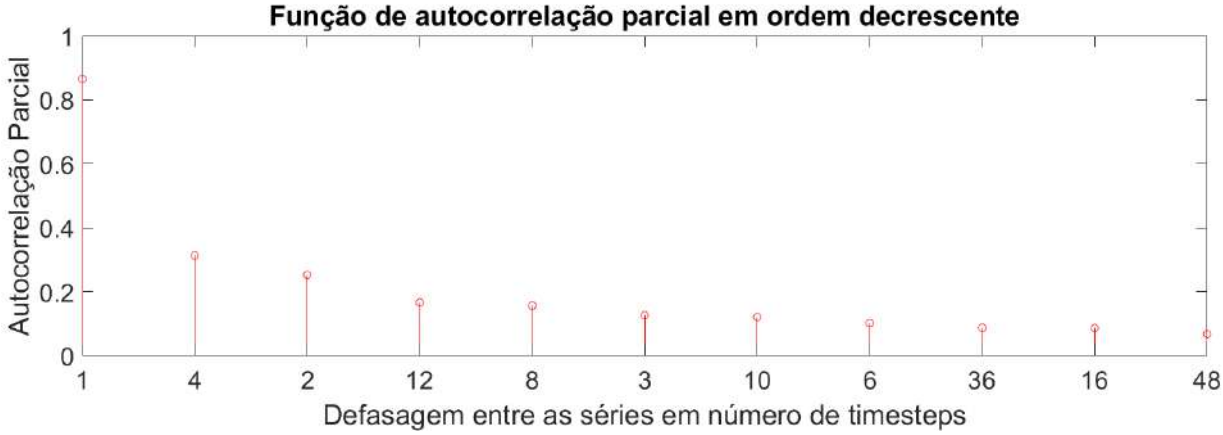
O mesmo raciocínio pode ser utilizado quando se aplica a função de autocorrelação parcial. É importante destacar que as funções de autocorrelação e autocorrelação parcial são

calculadas de formas distintas, podendo gerar um ordenamento diferente das componentes temporais que serão adicionadas.

A diferença evidente entre a função de autocorrelação e a função de autocorrelação parcial, representada na Figura *4* b) é o decaimento mais acelerado desta última. Isso ocorre, pois a autocorrelação parcial leva em consideração apenas a contribuição daquela defasagem retirando a influência das defasagens anteriores, como explicado na seção 2.4 de conceitos preliminares deste trabalho.

Utilizando o mesmo raciocínio da autocorrelação, pode-se fazer o ordenamento das componentes temporais da autocorrelação parcial, como pode ser visto na Figura 10.

Figura 10 - Valor de autocorrelação parcial organizadas em ordem decrescente e as respectivas defasagens



Fonte: O Autor, 2021.

Nota-se que as funções de autocorrelação e autocorrelação parcial geram ordenamentos distintos.

4.2 Modelos

A seção atual apresentará os modelos utilizados neste estudo. Estes são aplicações de uma ou mais das técnicas descritas anteriormente. Para todos os modelos introduzidos nesta seção, exceto os que explicitarem o contrário, utiliza-se o modelo em cada um dos 16 níveis do GCM considerados neste trabalho e seleciona-se o melhor resultado tendo como critério o estatístico SS4 mais elevado.

4.2.1 Interpolação bilinear

A interpolação bilinear (IBL) é o modelo de referência e foi detalhado na seção 4.1.1. No presente trabalho são aplicadas as equações do IBL utilizando-se os 4 pontos mais próximos do local de interesse.

4.2.2 MLR

O MLR foi explicado na seção 4.1.2. Define-se neste trabalho duas regressões lineares múltiplas. A primeira (MLR1) utiliza os 4 pontos mais próximos do local de interesse e a observação para fazer a regressão linear múltipla, enquanto a segunda (MLR2) considera os 36 pontos mais próximos.

4.2.3 MLRTA

O MLRTA é uma regressão linear múltipla com adição de componentes temporais. Utiliza-se a metodologia definida na seção 4.1.5 para seleção das componentes temporais. A equação matricial é dada a seguir:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} & x_{11-\Delta t11} & \cdots & x_{1k-\Delta t1k} \\ 1 & \cdots & x_{22} & \cdots & x_{2k} & x_{12-\Delta t12} & \cdots & x_{2k-\Delta t2k} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{2n} & \cdots & x_{nk} & x_{1n-\Delta t1n} & \cdots & x_{nk-\Delta nk} \end{bmatrix} . \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

Utilizando-se a Figura *3* como exemplo, o valor do Δt é variável para cada um dos preditores, por isso, empregaram-se diferentes Δts na representação matricial. No MLRTA, pode-se utilizar a autocorrelação ou a autocorrelação parcial para seleção das componentes temporais. Ao longo deste estudo, será feita a comparação entre as duas metodologias

Buscando facilitar o entendimento sobre a seleção das CTs quando se utiliza a autocorrelação, pode-se observar a Figura *9*. Na situação ilustrada, as três primeiras componentes temporais associadas a este preditor estão defasadas em 1, 2 e 4 *timesteps* em relação ao preditor original, respectivamente. Ao se empregar a metodologia associada ao uso da função de autocorrelação parcial, observando-se a Figura *10* nota-se que as três primeiras componentes temporais associadas a este preditor estão defasadas em 1, 4 e 2 *timesteps* em

relação ao preditor original, respectivamente. Nota-se que são adicionadas diferentes componentes temporais nas duas diferentes metodologias. Observando a Figura 9 e a Figura 10 nota-se que a diferença entre as defasagens tem uma maior variação com o aumento de componentes temporais utilizadas.

Após a definição da matriz dos preditores, as estimativas são feitas de forma análoga ao MLR. No presente trabalho, optou-se pelo uso do MLRTA com os 4 pontos mais próximos do local de interesse (MLRTA1) e com os 36 pontos mais próximos do local de interesse (MLRTA2).

4.2.4 MLRTAS

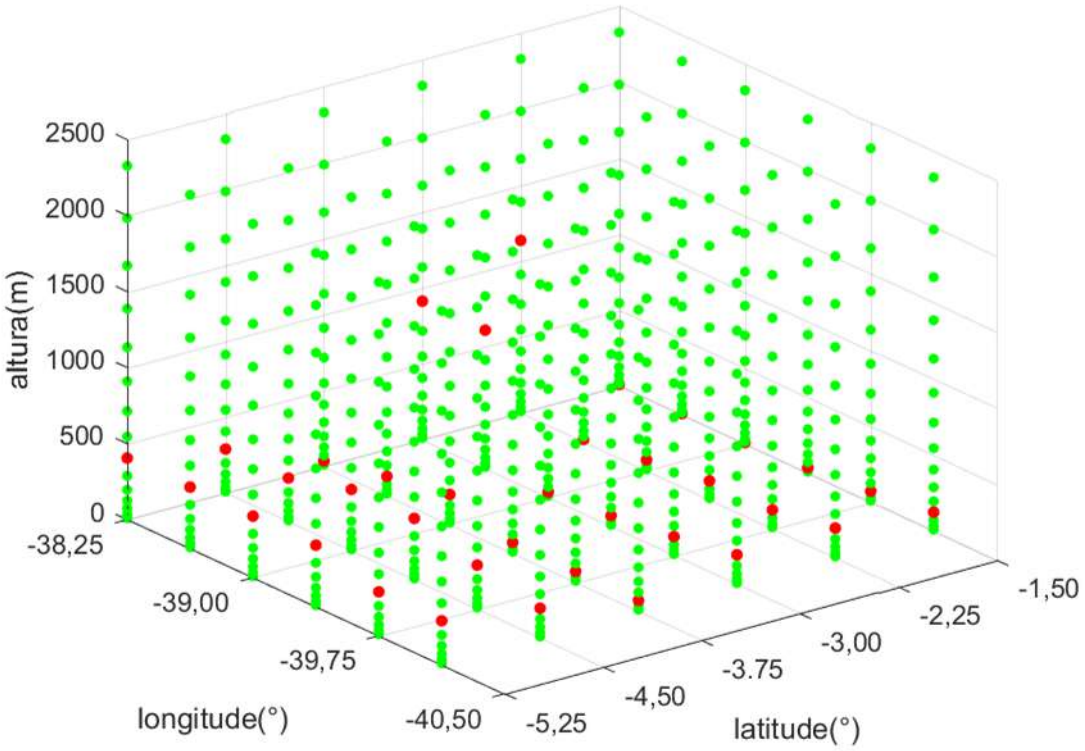
O MLRTAS é um modelo que utiliza a metodologia da regressão linear múltipla com seleção de componentes temporais e um método simples de seleção espacial dos preditores. Esse modelo foi criado para verificar se uma seleção espacial pode melhorar o desempenho da modelagem e para servir como referência para modelos mais complexos que serão explicados neste texto. Durante este estudo, foram considerados diferentes níveis do GCM como explicado na seção 4. Após determinar uma malha com os 36 pontos da malha horizontal mais próximos ao local de interesse e todos os níveis do GCM considerados neste texto, calcula-se o coeficiente de correlação de Pearson, definido na equação 17, entre cada um dos pontos do GCM com os valores observacionais contidos no período de calibração.

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad (17)$$

Onde n é o número de elementos na série, x_i e y_i são os elementos das séries que estão sendo correlacionadas, $\bar{x}$ e $\bar{y}$ são as médias das séries.

Para cada um dos conjuntos de pontos que possuem mesma latitude e longitude, seleciona-se o que possui melhor correlação para fazer parte do conjunto final dos preditores, resultando em um conjunto de 4 ou 36 pontos, dependendo do tamanho da malha inicial. A Figura 11 ilustra o domínio selecionado para um dos locais que forneceram dados observacionais para este trabalho.

Figura 11- Exemplo de seleção de domínio para uma torre anemométrica na costa do nordeste brasileiro



Fonte: O Autor, 2021.

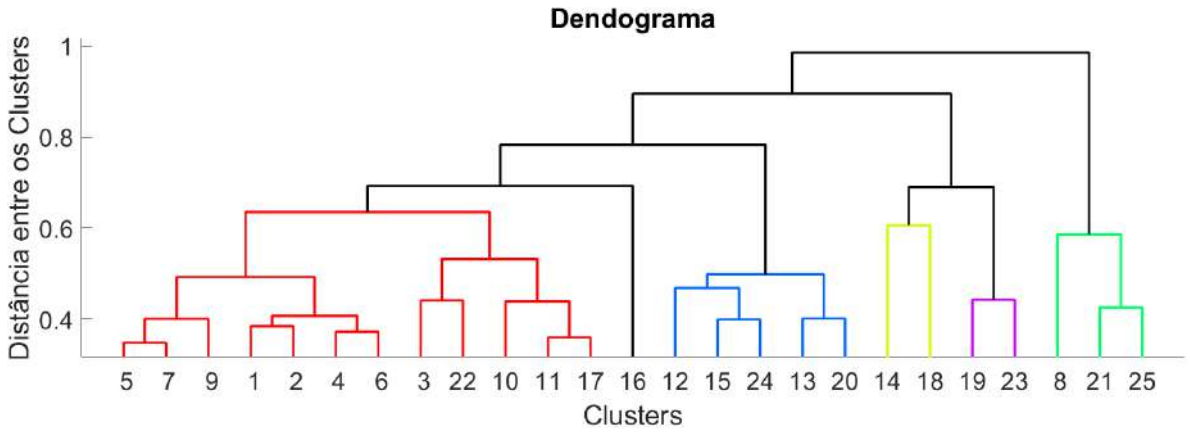
Após a seleção de domínio, o modelo é análogo ao MLRTA.

4.2.5 MLRTACLUSTER

Na seção atual, será determinada a metodologia utilizada para escolher as componentes temporais e para determinar o número de *clusters* hierárquicos que serão utilizados. As componentes temporais são selecionadas a partir da metodologia definida na seção 4.1.5.2. Após a adição das CTs à matriz dos preditores, aplica-se a metodologia descrita na seção 4.1.4. Após se definir um critério de distância e um de dissimilaridade, entre os clusters, estabelece-se os *clusters*. Calcula-se a variação da distância entre os *clusters* quando se diminui o número de agrupamentos. Após uma análise de sensibilidade, estabeleceu-se como critério de parada a variação da distância entre os clusters, normalizado pela maior distância, inferior a 1%. Avaliou-se um número máximo de 100 *clusters*. A Figura 12 ilustra um dendrograma para auxiliar no entendimento da metodologia. A cada fusão de dois *clusters*, por exemplo, a fusão dos *clusters* 14 e 18, diminui-se o número de *clusters* e aumenta-se a distância máxima entre os *clusters* restantes. Quando se tem 100 clusters é calculada a maior

distância possível, a qual ocorre quando existem apenas dois *clusters* e esta é utilizada como referência. As fusões são interrompidas quando após uma fusão que reduz o número de *clusters* para a distância máxima entre os agrupamentos variar menos de 1% em relação a distância máxima quando havia c+1 *clusters* e a partir deste momento as novas fusões gerando c -1, c -2, c -n *clusters*, onde n é um número natural, causam variações na distância máxima entre os *clusters* superior a 1% da distância máxima anteriormente calculada. O processo é repetido cada vez que se adiciona mais uma CT e se escolhe a configuração com o maior SS4.

Figura 12 - Dendograma



Fonte: O Autor, 2021.

Na situação ilustrada na Figura *12* determinou-se que seriam utilizados 6 agrupamentos. Após a determinação do número de agrupamentos pela metodologia hierárquica utiliza-se a metodologia não hierárquica (NHC, *Non-Hierarquical Clustering*) baseado no algoritmo *k-means* (e.g., GUTIÉRREZ *et al*., 2004). Conforme descrito por PERRUCI (2018), no AHC, o número de clusters C é estabelecido à posteriori, em função de um critério de parada de fusões. No NHC, por sua vez, C é decidido à priori. No método utilizado neste trabalho, o número C de *clusters* utilizados no *k-means* são obtidos a partir da configuração determinada pelo método hierárquico. Após a determinação dos agrupamentos utiliza-se uma regressão linear múltipla para cada um destes e a estimativa final é a composição das estimativas de cada MLR.

4.2.6 PCAMLR

Utilizar a análise de componentes principais antes de fazer o MLR é uma modelagem comum e permite utilizar a informação proveniente de um grande volume de dados, reduzindo o número de preditores de forma significativa, perdendo apenas uma pequena parcela da

informação total. É importante destacar que as componentes principais devem ser ortogonais entre si, ou seja, são independentes entre si, o que evita problemas como a formação de uma matriz singular quando se aplica o MLR aos preditores.

No presente trabalho, aplicou-se o PCA, descrito na seção 4.1.3, a uma malha horizontal dos 36 pontos mais próximos e em todos os níveis de modelo. Em seguida, aplica-se o MLR, o qual é descrito na seção 4.1.2. É importante destacar que esse modelo não utiliza componentes temporais.

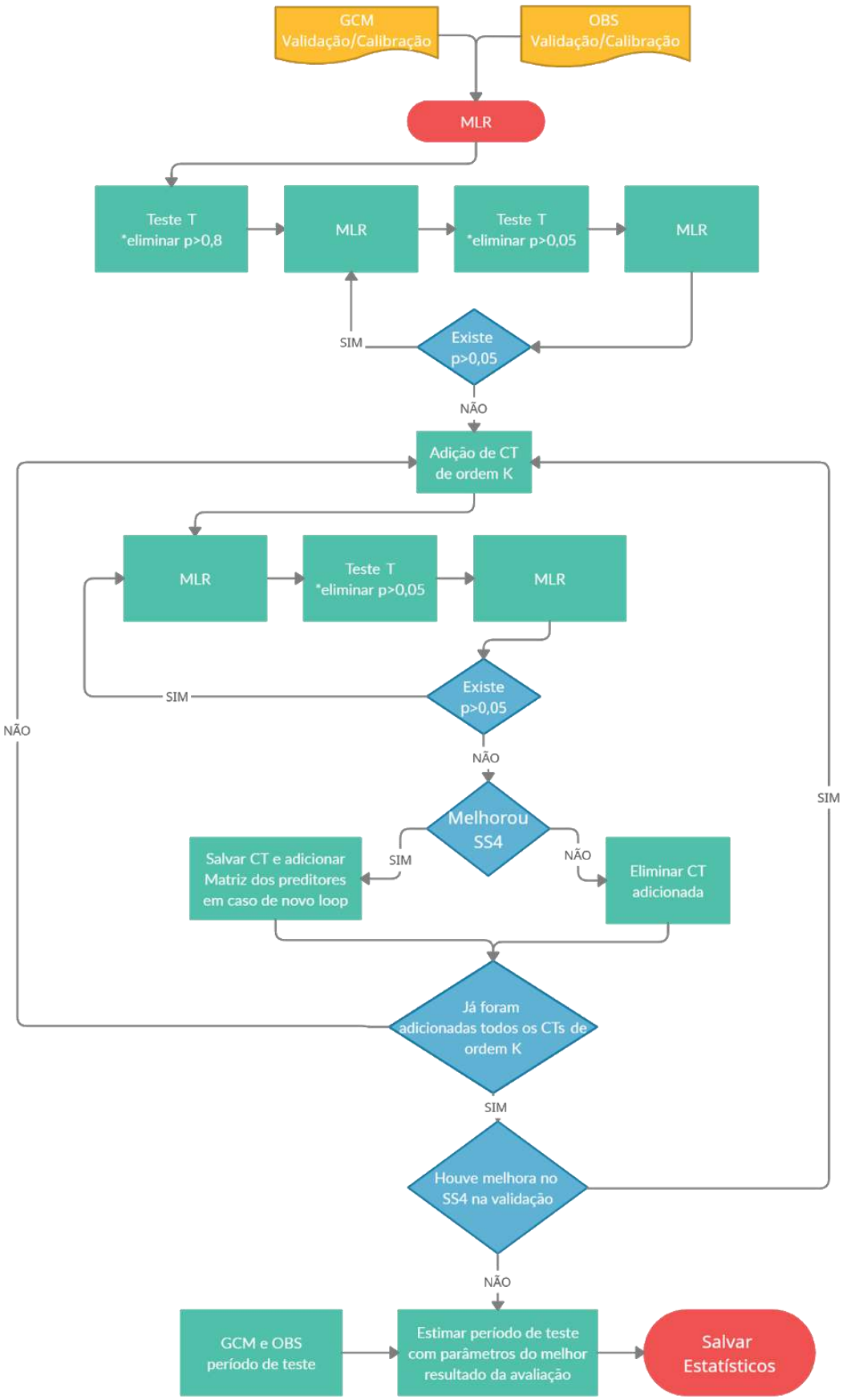
4.2.7 PCASMLR

O PCASMLR inicialmente aplica a seleção de domínio como descrito na seção 4.2.4. Depois aplica-se a metodologia da seção 4.2.6 e armazena-se o resultado. Em seguida, adiciona-se as componentes temporais utilizando o método descrito na seção 4.1.5 de forma cumulativa e aplica-se novamente o PCA e o MLR. Armazena-se os resultados e seleciona-se o número de componentes temporais que proporcione o maior SS4.

4.2.8 RTCT

O modelo chama-se RTCT, pois faz alusão a técnica de Regressão Linear Múltipla (MLR), aos testes de hipótese e as técnicas de *clustering* e uma segunda vez aos testes de hipótese. O fluxograma na Figura 13 ilustra os processos do modelo.

Figura 13 - Fluxograma do modelo RTCT



Fonte: O Autor, 2021.

Inicialmente é feita uma regressão linear múltipla como descrito na seção 4.1.2. Após esses procedimentos, é realizado um teste T para cada um dos coeficientes em que a hipótese nula considera o valor do coeficiente igual a zero. Neste primeiro teste, o p-valor é 0,8, considerado elevado, ou seja, pouco restritivo. O objetivo dessa etapa é filtrar os dados que contribuem apenas como ruído para etapa seguinte.

A etapa seguinte é utilizar o MLR com os preditores restantes. Posteriormente a regressão, faz-se um novo teste de hipótese mais restritivo, utilizando a mesma hipótese e para manter o preditor, o coeficiente deve ter um p-valor menor a 0,05. Em sequência, aplica-se novamente o MLR e verifica-se o p-valor associado a cada um dos coeficientes. Esse procedimento é repetido até que todos os coeficientes tenham um p-valor associado menor que 0,05. Então, estima-se a série de validação e o valor do SS4 é calculado e o resultado armazenado.

Adiciona-se a primeira componente temporal associada ao primeiro preditor do grupo dos preditores que restaram após as etapas anteriores. A componente temporal é adicionada a partir da metodologia descrita na seção 4.1.5.2. Repete-se a etapa 2 e mantem-se a componente temporal na matriz dos preditores, caso o valor do SS4 aumente durante o período de validação, após a adição da componente temporal avaliada, caso contrário, ela é retirada da matriz dos preditores. A adição de componentes é repetida de forma análoga para todos os outros preditores.

Posteriormente, adiciona-se a segunda componente temporal dos preditores que tiveram a primeira componente associada mantidas e utiliza-se novamente o SS4 como referência para manter ou não estas componentes na matriz dos preditores.

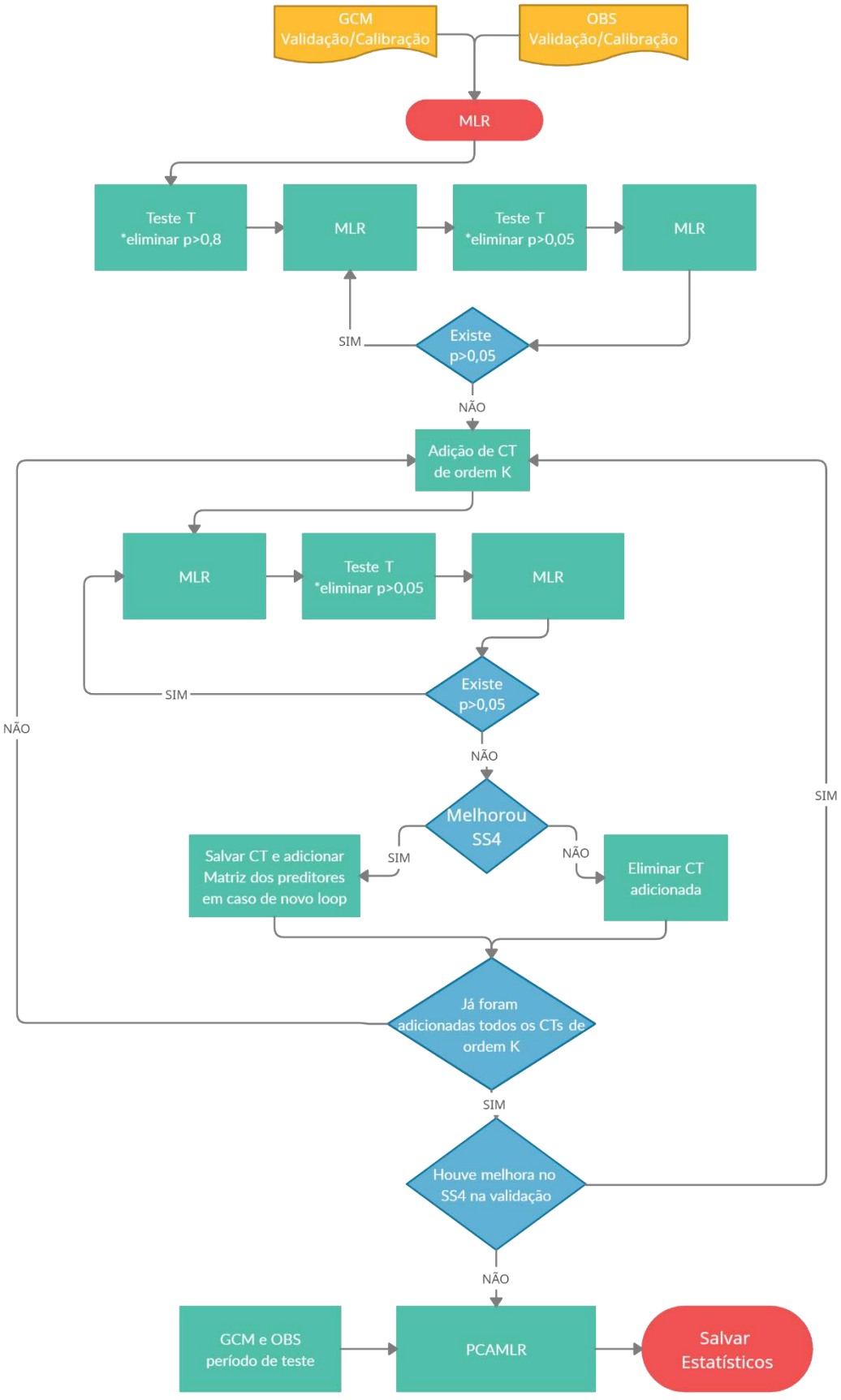
A adição de componentes temporais de ordem k, onde k é um número natural, continua até que não exista melhora na estimativa do período de validação. Após esgotar as possibilidades de adição de componentes temporais utilizando os critérios descritos anteriormente, utiliza-se os parâmetros utilizados na melhor estimativa do período de validação para estimar o período de teste e o resultado é salvo.

4.2.9 RTCTP

De forma análoga ao modelo RTCT o modelo é chamado de RTCTP, pois faz alusão a técnica de Regressão Linear Múltipla (MLR), aos testes de hipótese e as técnicas de *clustering* e a técnica de análise de componentes principais

O fluxograma na Figura *14* ilustra os processos do modelo. A principal diferença entre os modelos é que após a adição do último grupo de componentes temporais os preditores selecionados são utilizados como dados de entrada do modelo descrito na seção 4.2.6.

Figura 14 - Fluxograma do modelo RTCTP



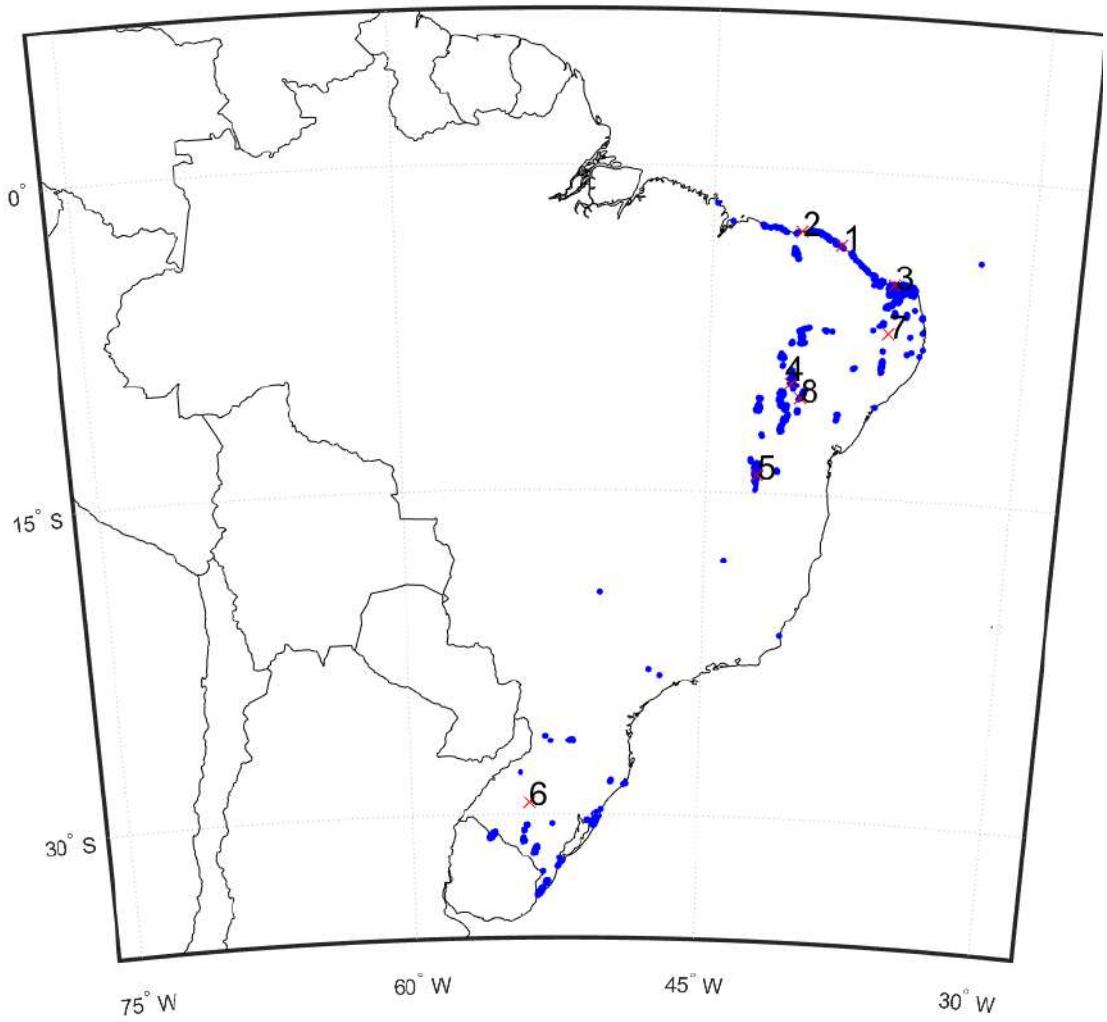
Fonte: O Autor, 2021.

5 RESULTADOS E DISCUSSÃO

5.1 Base de dados

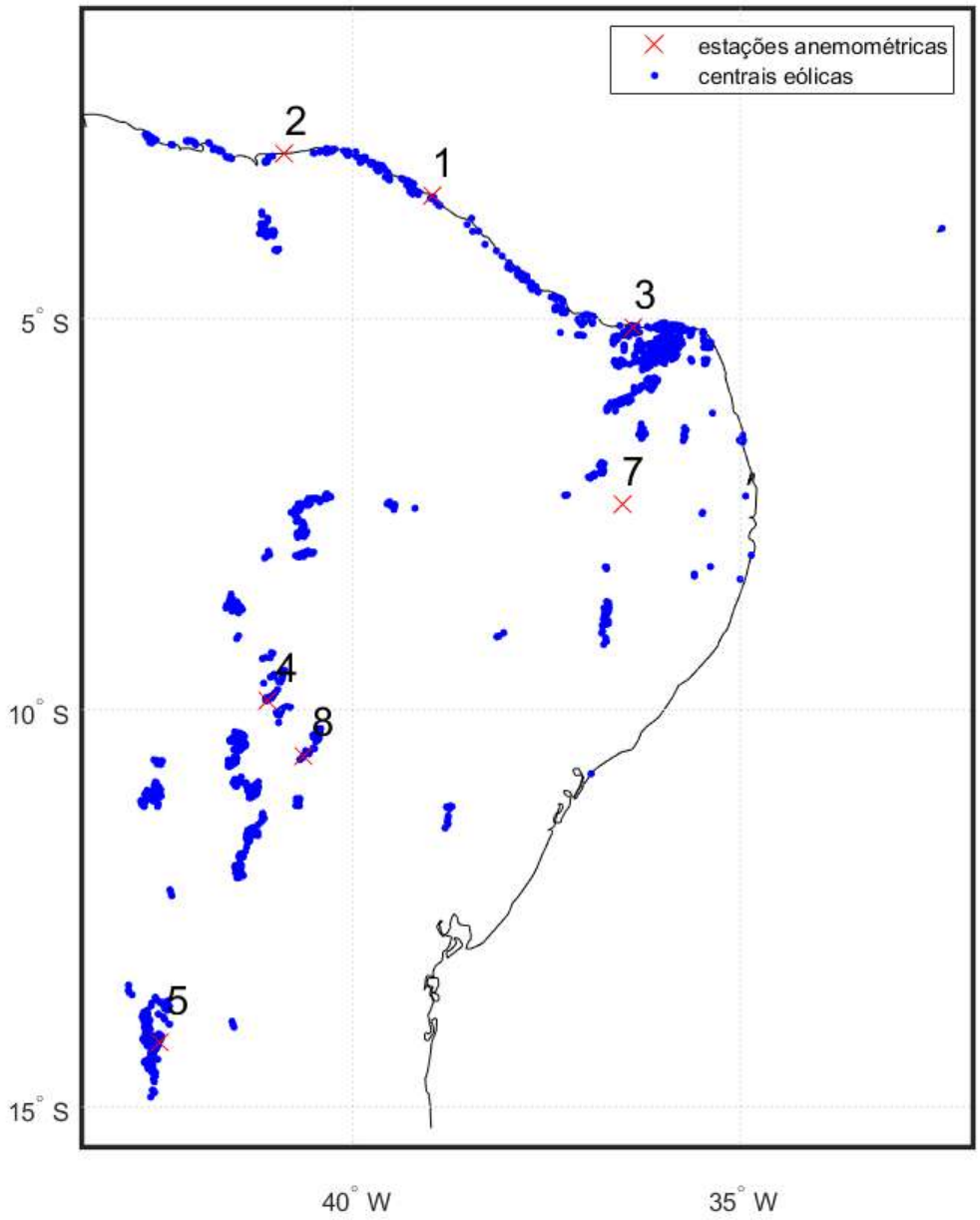
Neste trabalho foram utilizadas 8 torres anemométricas na avaliação de desempenho dos modelos empregados no estudo, localizadas conforme pode ser visto na Figura 15, deste total 7 estão localizadas na região nordeste do Brasil como pode ser visto na Figura 16.

Figura 15 - Estações anemométricas utilizadas no trabalho e as centrais eólicas construídas até 2018



Fonte: Adaptado de PERRUCI, 2018

Figura 16 - Estações anemométricas utilizadas no trabalho localizadas no Nordeste brasileiro e as centrais eólicas construídas até 2018



Fonte: Adaptado de PERRUCI, 2018.

O comportamento do vento foi avaliado na altura dos dados disponibilizados por três instituições distintas: Secretária de Infraestrutura do Estado do Ceará (SEINFRA-CE) e Sistema de Organização Nacional de Dados Ambientais (SONDA). As duas fontes supracitadas disponibilizaram os dados publicamente e, por fim, foram utilizados dados do Operador Nacional do Sistema Elétrico (ONS), estes foram disponibilizados durante o projeto HPC4E. A Tabela 3 sintetiza várias destas informações.

Tabela 3 - Informações sobre as estações anemométricas utilizadas neste trabalho

ÍNDICE DA ESTAÇÃO ANEMOMÉTRICA	ALTURA DE MEDIÇÃO (M)	FONTE DE DADOS
LOCAL 1	60,4	SEINFRA-CE
LOCAL 2	60,4	SEINFRA-CE
LOCAL 3	80,0	ONS
LOCAL 4	50,0	ONS
LOCAL 5	80,0	ONS
LOCAL 6	50,0	SONDA
LOCAL 7	80,0	SONDA
LOCAL 8	80,0	ONS

Fonte: O Autor, 2021.

Os preditores são dados provenientes do GCM conforme explicado na seção 4 e no Quadro *1*

Os comparativos feitos ao longo desta seção de Resultados e Discussão utilizam como métrica para comparação, o estatístico SS4 definido na seção 4. Em adição a isso estão disponíveis os diagramas de Taylor (TAYLOR, 2001) com os resultados do período de teste de todos os modelos avaliados em cada uma das estações da Tabela *3*.

5.2 Autocorrelação e Autocorrelação Parcial

Durante a avaliação preliminar deste estudo verificou-se que a adição de componentes temporais de forma cronológica causava uma melhora no MLR de 4 preditandos, entretanto, não se pode perceber qualquer melhora nos modelos quando se aumentou o domínio para 36 pontos do GCM. Tal fato ocorreu, pois o grande de número de novos preditores adicionados de forma indiscriminada no modelo causou a existência de preditores semelhantes o que acarretou um problema ao calcular os parâmetros utilizando a equação 12, dado que só é possível calcular a inversa de uma matriz se as colunas utilizadas forem linearmente independentes.

Utilizou-se como critério para seleção das componentes temporais a função de autocorrelação e autocorrelação parcial conforme descrito na seção 4.1.5.

Avaliando os resultados dos modelos MLRTA1 com MLR1, ou seja, os modelos análogos divergindo apenas pela adição das componentes temporais houve uma melhora na estimativa do período de teste em 6 das 8 torres disponíveis e em nenhum local houve piora dos resultados. Um fato relevante que se pôde perceber é que as três estações em que não houve melhora estavam distantes da costa, o que demonstra o primeiro indício da dificuldade que os modelos com menor domínio possuem maior dificuldade em descrever tais regiões como será devidamente abordado nas próximas seções.

De forma similar fez-se a comparação dos resultados dos modelos MLRTA2 e MLR2 e houve melhora em 3 dos locais avaliados, sem ocorrer piora. A menor quantidade de estações com melhora nos resultados demonstra que houve um ganho ao alterar a forma de selecionar como as componentes temporais são adicionadas em domínios maiores, entretanto, ainda existe o problema em vários casos de os termos não serem linearmente independentes, por esse motivo novos modelos foram utilizados para melhorar a estimativa utilizando domínios maiores conforme será explicado ao longo desta seção.

Os resultados foram semelhantes em relação a quantidade de melhoras utilizando autocorrelação ou autocorrelação parcial, entretanto, a autocorrelação parcial demonstrou-se, em geral, superior como modelo de seleção de componentes temporais, pois proporcionou melhoras mais significativas. O comparativo pode ser visto na Tabela *4*

Tabela 4 - Comparativo de ganho utilizando autocorrelação e autocorrelação parcial em todos os modelos com componentes temporais

MODELOS	SIMILAR	AC	ACP
MLRTA1	5	1	2
MLRTA2	5	2	1
MLRTAS	5	2	1
MLRTAS2	5	3	0
MLRCLUSTER1	3	1	4
MLRCLUSTER2	4	1	3
RTCT	2	0	6
RTCTP	1	1	6

Fonte: O Autor, 2021.

O melhor resultado, em geral, utilizando a autocorrelação parcial como critério para seleção das componentes temporais deve-se ao fato de a autocorrelação parcial isolar a

influência de uma determinada componente temporal subtraindo o efeito das outras no sinal original (BUENO, 2012). Desta forma evita-se a redundância na informação e isto fica mais evidente nos modelos com maior grau de especialização, o MLRCLUSTER1, MLRCLUSTER2, RTCT e RTCTP. Estes modelos serão discutidos com mais detalhes nas subseções posteriores.

Os modelos MLRTA apresentaram desempenho semelhante para ambos os casos e optou-se por seguir com a autocorrelação parcial para o MLRTA1 e MLRTA2 para a estimativa dos resultados mantendo a concordância com a maioria dos demais modelos. O MLRTAS e MLRTAS2 apresentaram um resultado anômalo sendo mais significativo o ganho utilizando a autocorrelação, por isso para estes modelos foi utilizado, como critério para o ordenamento das componentes temporais, a autocorrelação.

5.3 Seleção espacial e temporal

As técnicas aplicadas neste trabalho, em geral, buscam avaliar a melhora das estimativas de modelos lineares utilizando preditores de instantes passados para melhorar as estimativas do vento local.

Os modelos MLRTAS e MLRTAS2 utilizam uma forma simples de seleção espacial das variáveis aliada a adição de componentes temporais. Buscando avaliar a melhora do uso da seleção espacial em conjunto com adição das componentes temporais foi realizada a comparação entre os resultados dos modelos MLRTA e MLRTAS tendo como referência o estatístico SS4. Este resultado pode ser visto na Tabela *5*

Tabela 5 - Comparação do desempenho de diferentes configurações dos modelos MLRTAS X MLRTA

MODELOS	SIMILAR	SUPERIOR	INFERIOR
MLRTAS - 4 PONTOS	2	5	1
MLRTAS - 36 PONTOS	4	2	2

Fonte: O Autor, 2021.

A Tabela *5* evidencia que para modelos com domínios menores existe um claro ganho ao utilizar a seleção de domínio em conjunto com as componentes temporais, entretanto, não houve um ganho significativo para o modelo com 36 pontos. É importante destacar que a forma de seleção foi simples apenas para ser utilizada como uma referência para desenvolvimentos futuros. Apesar disso, os dois locais em que houve melhora para o domínio

com 36 pontos estão afastados da costa, por isso possuem uma maior quantidade de fatores relevantes que impactam no comportamento do vento local, por exemplo, orografia mais complexa, efeitos térmicos não tão evidentes, o que faz notar que a seleção de domínio em conjunto com as componentes temporais contribuem para descrever regimes de ventos mais complexos, entretanto é necessário avaliações mais robustas na seleção de domínio como feito em (DANTAS, 2020) em conjunto com a adição de componentes temporais para verificar de forma mais acurada os ganhos com a união da seleção espacial e temporal de preditores.

Outro ponto que evidencia que a seleção de domínio do modelo MLRTAS não foi ideal é o comparativo entre os modelos PCAMLR e PCASMLR. O modelo PCAMLR utiliza uma metodologia difundida para filtrar os preditores, a análise de componentes principais sobre o conjunto com 576 pontos e o PCASMLR utilizou a metodologia já detalhada na seção 4.2.6, que é análoga a seleção de domínio feita no modelo MLRTAS, o desempenho do PCAMLR foi superior ao PCASMLR, isto evidencia que a seleção de preditores não foi a ideal, pois o conjunto inicial dos preditores é a única diferença entre os modelos. Dito isto, nota-se que ainda existe possibilidades de aperfeiçoamento da união da seleção espacial e temporal de componentes, especialmente modelos com domínios maiores.

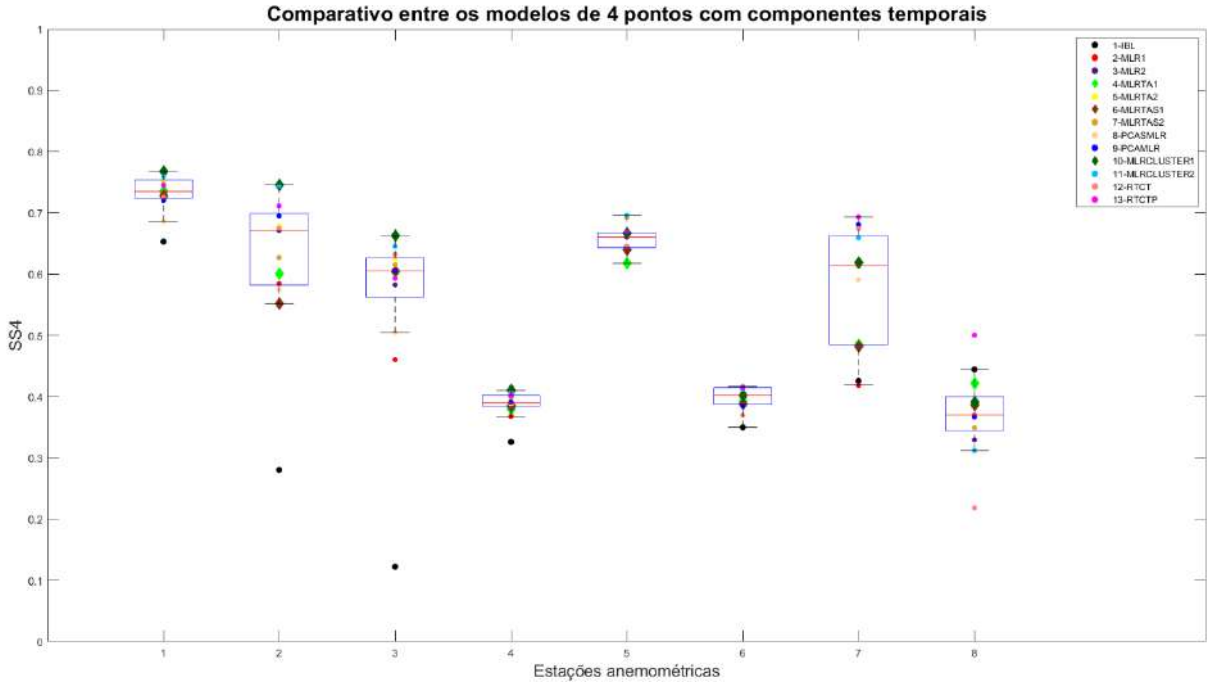
5.4 Adição do comportamento local aos modelos

Os modelos explicados nos tópicos anteriores demonstraram uma dificuldade de melhora ao adicionar as componentes, considerando isto notou-se a necessidade de uma avaliação de comportamentos locais ao longo da série temporal, por isso foram criados os modelos MLRTACLUSTER1 E MLRTACLUSTER2, os quais utilizam as técnicas de agrupamento (GUTIÉRREZ *et al*., 2004) conforme descrito na seção 4.2.5.

O resultado foi uma melhora em todas as estações quando comparado ao MLRTA e MLRTAS, exceto na estação 8, onde houve uma leve piora. O comparativo pode ser visto na Figura *17* com o resultado de todos os modelos desenvolvidos neste trabalho para cada uma das 8 estações anemométricas avaliadas, destacando o MLRTACLUSTER, MLRTAS e MLRTA para os modelos com 4 pontos. O similar é feito para os domínios de 36 pontos na Figura *18*.

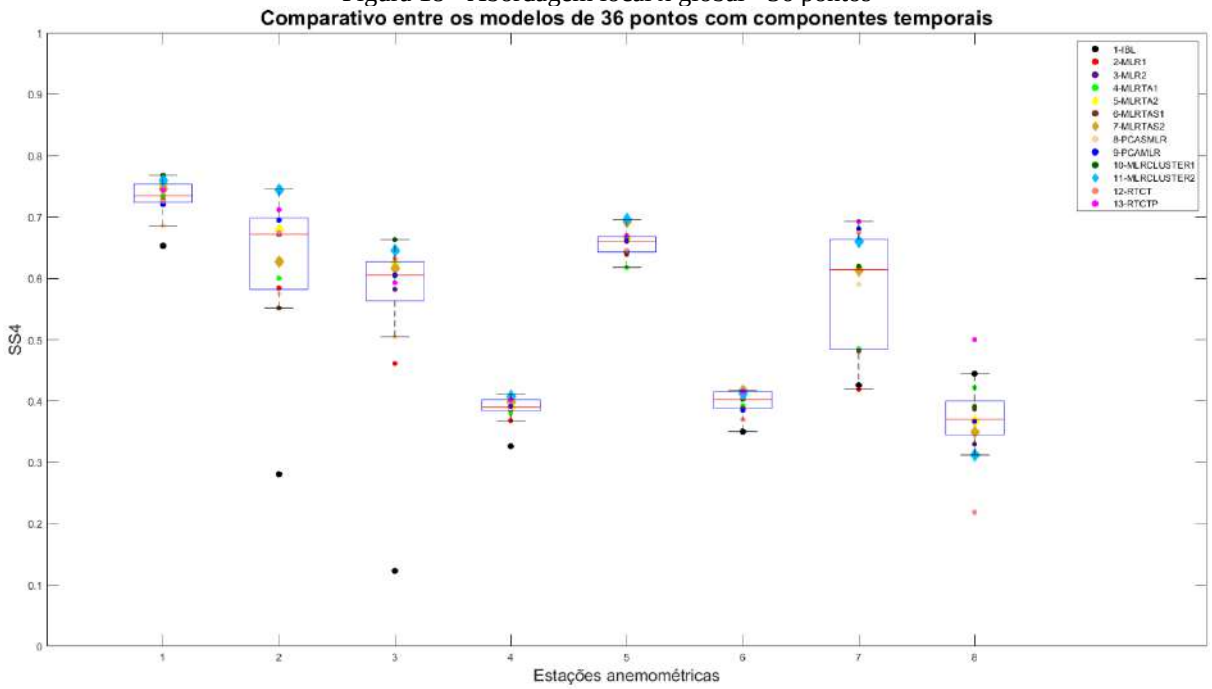
Os resultados do local 8 são anômalos como poderá ser visto nas subseções posteriores.

Figura 17 - Abordagem local x global - 4 pontos



Fonte: O Autor, 2021.

Figura 18 - Abordagem local x global - 36 pontos



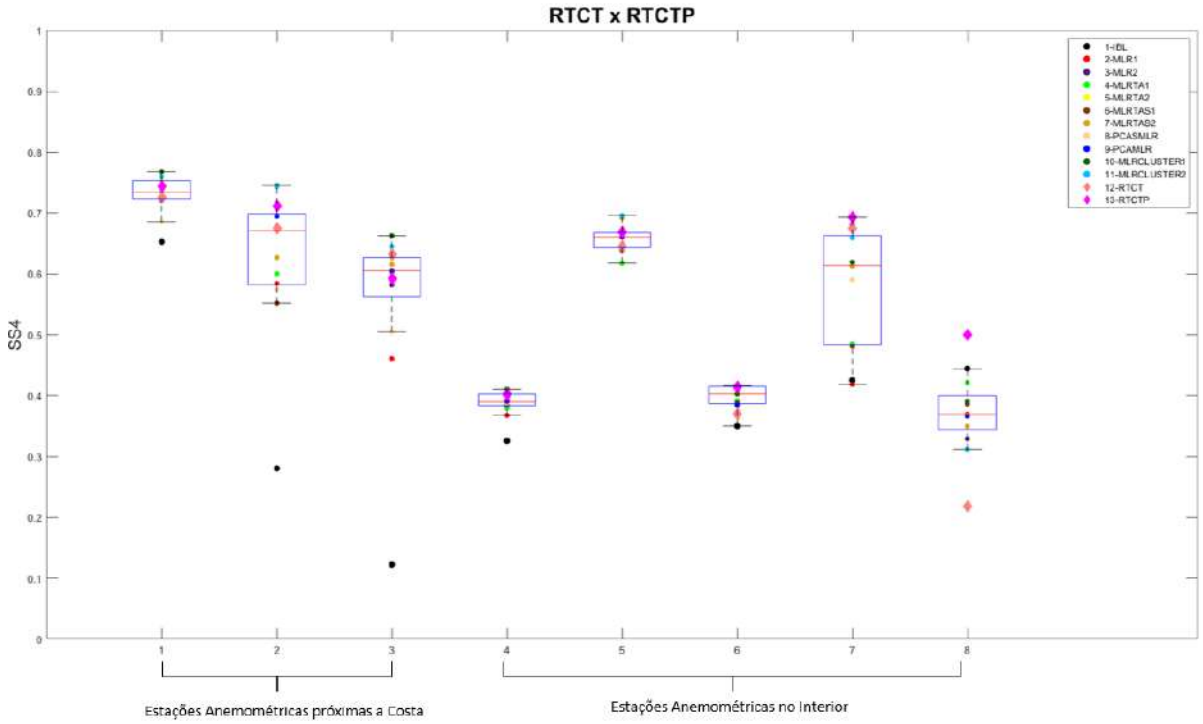
Fonte: O Autor, 2021.

5.5 RTCT e RTCTP

Os modelos discutidos anteriormente apresentaram bons resultados, entretanto, não determinam de forma objetiva o nível que será utilizado nem a quantidade de componentes temporais que será inserida na matriz dos preditores.

RTCT e RTCTP são modelos criados com intuito de fazer a seleção de forma automática de todos os preditores que serão utilizados para estimar o comportamento do vento local, incluindo as componentes temporais. A descrição detalhada destes é encontrada na seção 4.2.8 e 4.2.9, respectivamente. Os resultados destes modelos podem ser vistos na Figura *19*.

Figura 19 - Comparativo entre os modelos RTCT e RTCTP



Fonte: O Autor, 2021.

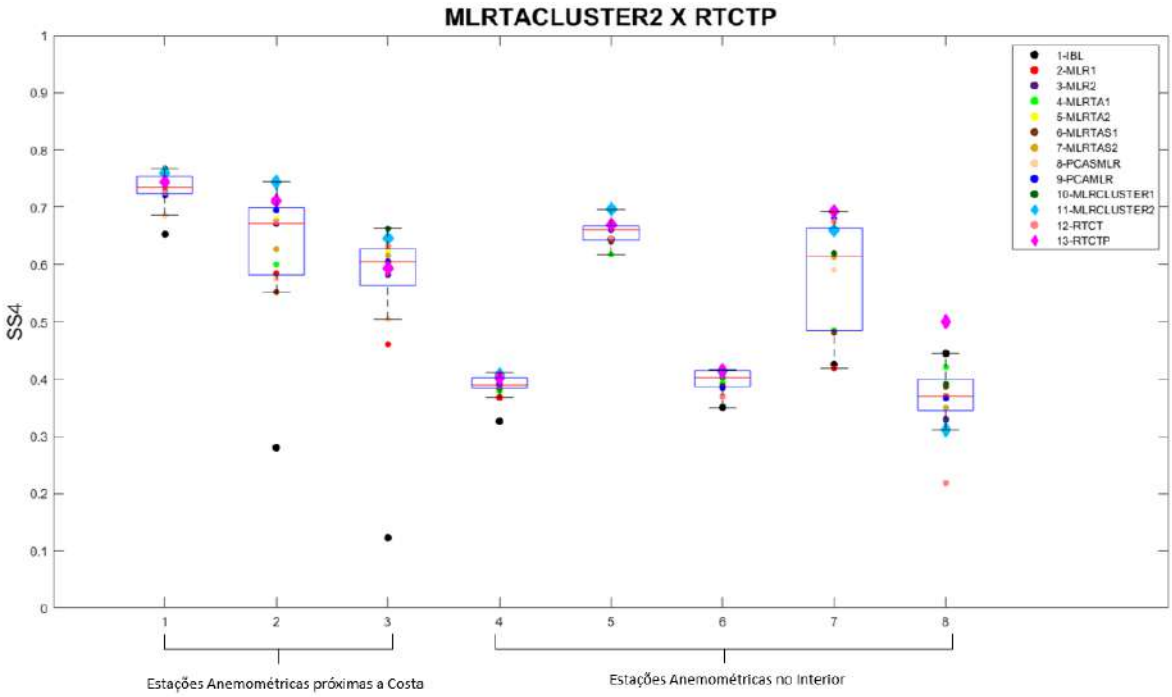
Nota-se que de forma global o modelo RTCTP (linha cheia) tem um desempenho superior ao RTCT, com exceção do local 3. O local apresenta um desempenho muito semelhante para todos os modelos, trata-se de um local de orografia bastante complexa, o que dificulta descrição utilizando apenas um modelo linear, de forma geral, os modelos obtiveram desempenho semelhantes neste local.

O modelo RTCTP foi criado ao se notar que o RTCT ainda apresentava um grau de similaridade considerável entre os preditores e por isso foi utilizada a análise de componentes principais no RTCTP para tornar os preditores ortogonais entre si, evitando a dependência linear.

5.6 RTCTP x MLRTACLUSTER

Dentre os modelos com adição de componentes temporais, os modelos RTCTP e MLRTACLUSTER apresentaram os melhores desempenhos. A Figura *20* realiza o comparativo entre os modelos.

Figura 20 - Comparativo entre os modelos MLRTACLUSTER2 e RTCTP



Fonte: O Autor, 2021.

Dado a existência de padrões distintos nos grupos de Estações anemométricas próximas a costa e as localizadas no interior como pode ser visto em (DANTAS, 2020) foi feita uma análise análoga que pode ser identificada na figura anterior.

De forma geral, o MLRTACLUSTER com 36 pontos apresenta um desempenho superior nas estações próximas a costa. Dado a menor complexidade dos fenômenos próximos a costa a contribuição de diversos níveis utilizado pelo RTCTP não foi o suficiente para o modelo ter um desempenho superior ao MLRTACLUSTER2, entretanto, ao observarmos as estações no interior existe uma inversão de padrões, sendo o RTCTP ou tendo comportamento similar o MLRTACLUSTER2, exceto no local 5, que é uma estação próxima a um grande lago, o que pode causar fenômenos similares ao que ocorre na costa como o efeito da brisa descrito em (CROSMAN, 2010).

O local 8 apresenta resultados anômalos, pois o único modelo a superar o IBL, modelo de referência, é o RTCTP, além da grande discrepância entre este e o MLRTCLUSTER2 que não ocorre em nenhuma das outras estações. Aproximadamente 60% dos dados foram descartados no processo de garantia de qualidade, evidenciando uma campanha de medições ruins. Os resultados das medições do local 8 podem indicar uma grande contribuição da utilização de diferentes níveis de forma simultânea em conjunto com as componentes temporais ou uma série de dados muito inconsistente tornando os resultados desta estação suspeitos.

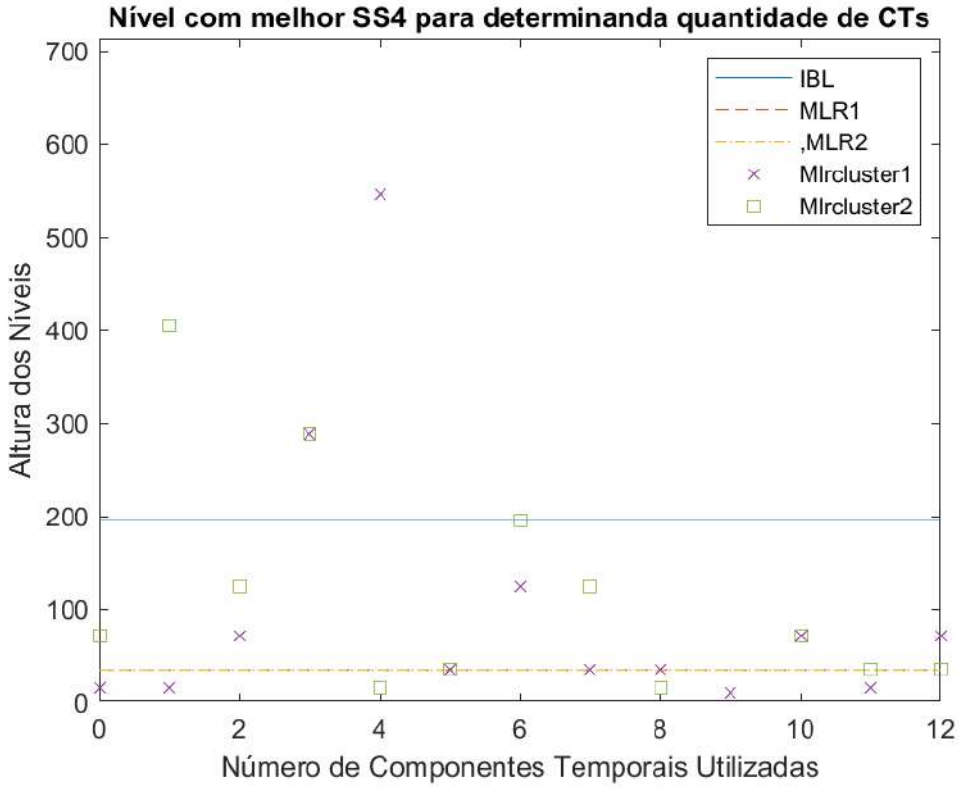
Dado a anomalia do local 8 e a estação 4 ser um local muito complexo e ter uma altura de medição inferior as demais torres anemométricas, o número de estações foi um pouco reduzido, todavia, foi utilizada a base de dados disponível.

Os resultados mostram um indício que o RTCTP apresenta melhor desempenho no interior e o MLRTACLUSTER na costa, sendo necessário estudos posteriores com uma quantidade maior de torres anemométricas para analisar melhor este comportamento.

Uma vantagem do modelo RTCTP é a maior robustez na seleção dos preditores, pois é determinado um único conjunto analisando de forma conjunta todos os pontos do GCM e as componentes temporais, o que é diferente do MLRTACLUSTER. O nível do GCM selecionado variou de acordo com a quantidade de componentes temporais utilizadas no modelo.

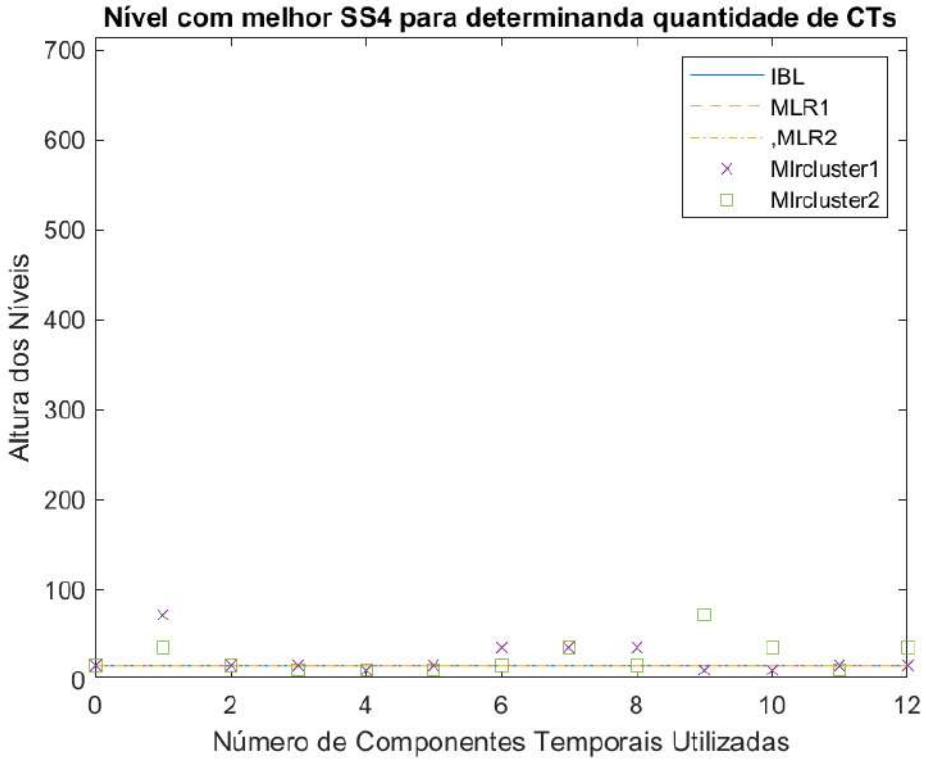
Neste trabalho está sendo analisado o período de reanálise. Foi inserido nas figuras o melhor resultado de cada modelo independentemente do número de componentes temporais e o nível utilizado do GCM. Contudo, esta variação pode significar uma maior instabilidade do modelo durante a fase operacional para realizar previsões, pois como não existe uma avaliação global de todos os preditores disponíveis, como no caso do RTCTP, é mais fácil a adição de um erro no período operacional, sendo necessário testes para avaliar esta hipótese. Esta variação do nível com melhor resultado em relação a quantidade de componentes temporais utilizadas pode ser visto na Figura 21 e Figura 22.

Figura 21 - Melhor nível para diferentes quantidades de CTs adicionadas no MLRTACLSUTER no local 1



Fonte: O Autor, 2021.

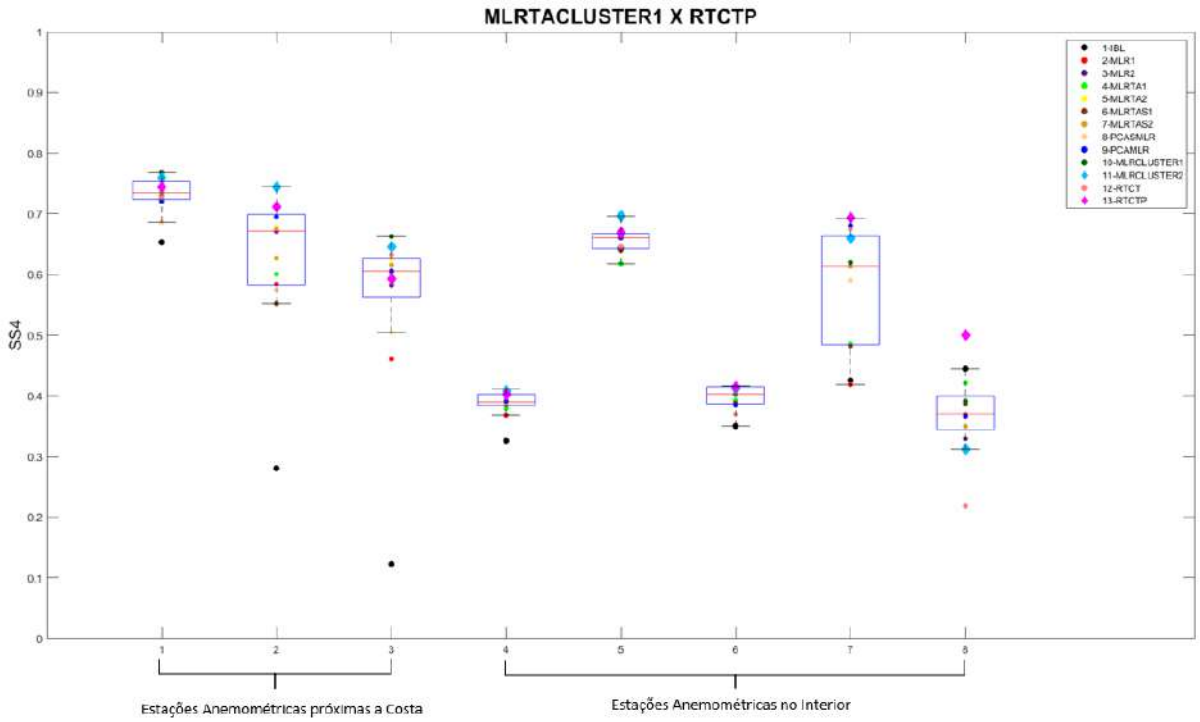
Figura 22 - Melhor nível para diferentes quantidades de CTs adicionadas no MLRTACLSUTER no local 7



Fonte: O Autor, 2021.

De forma análoga ao feito entre RTCTP e MLRTACLUSTER2 pode ser feito a comparação entre o MLRTACLUSTER1 e RTCTP. Os resultados podem ser vistos na Figura 23

Figura 23 - Comparativo entre os modelos MLRTACLUSTER1 e RTCTP

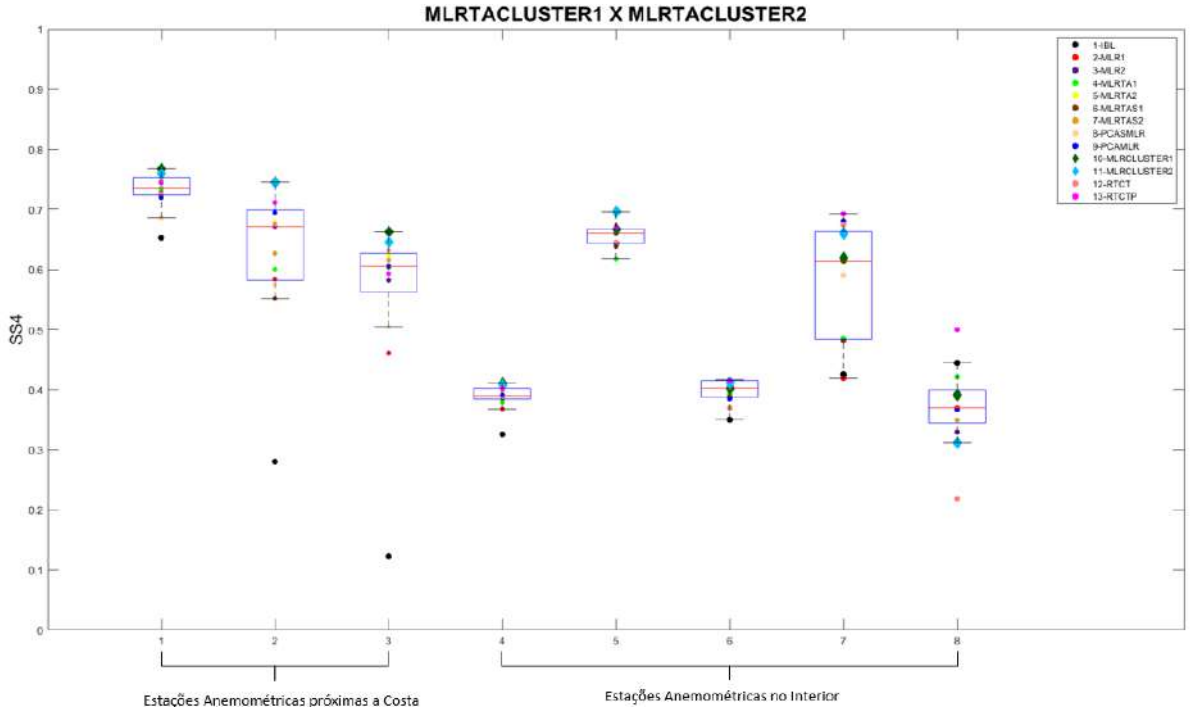


Fonte: O Autor, 2021.

Os resultados são similares ao comparativo com o MLRTACLUSTER2, tendo como principal diferença o desempenho nas estações anemométricas localizadas no interior, pois o MLRTACLUSTER1 não apresentou desempenho melhor em nenhuma das estações do interior, mesmo o local 5, evidenciando a necessidade de um domínio maior para descrever a maior complexidade dos fenômenos destas regiões.

A Figura 24 mostra o comparativo entre o MLRTACLUSTER1 e MLRTACLUSTER2.

Figura 24 - Comparativo entre os modelos MLRTACLUSTER1 e MLRTACLUSTER2



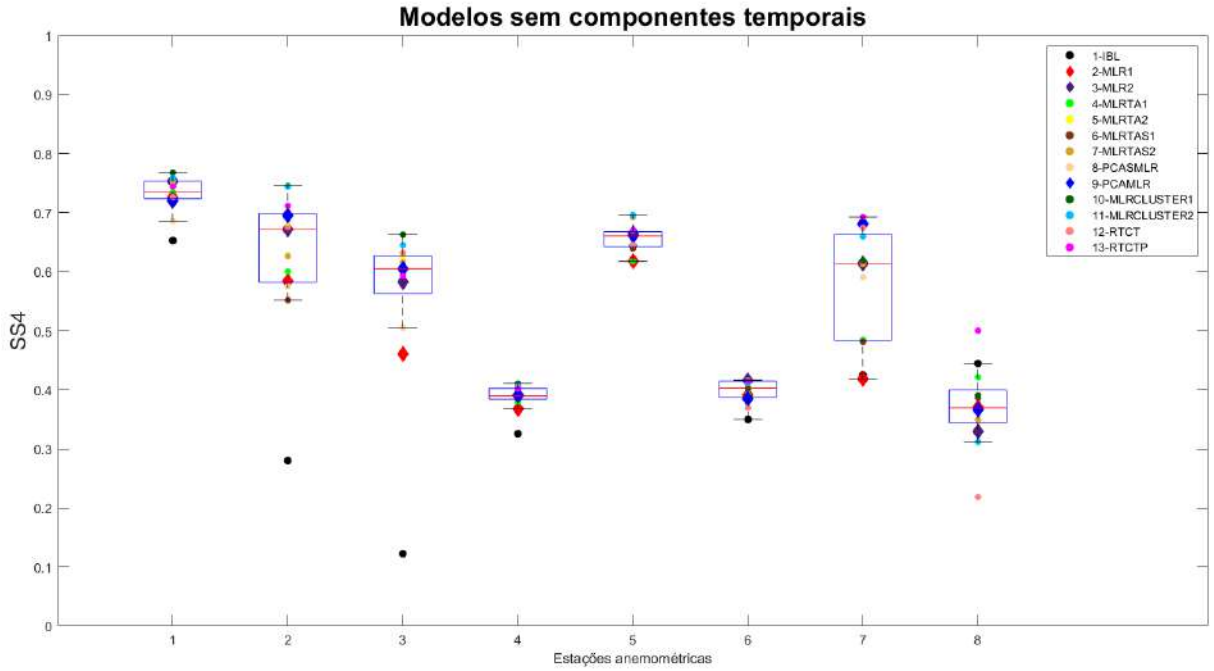
Fonte: O Autor, 2021.

Os resultados nas estações próximas a costa são similares havendo um desempenho levemente superior do MLRTACLUSTER1. Ao analisar as estações no interior, exceto a estação 8, que apresenta um comportamento anômalo, o MLRCLUSTER2 apresenta um desempenho superior, o que está alinhado com a necessidade de mais informações para descrever os fenômenos destas regiões.

5.7 Comparativo entre os modelos adicionados ou não de componentes temporais

A adição das componentes temporais visa a melhor capacidade de descrever os fenômenos na microescala a partir da escala sinóptica. É importante avaliar se os modelos sem componentes temporais têm desempenho melhor que os modelos com componentes temporais. Na Figura 25, de forma global, notamos que o modelo PCAMLR apresenta os melhores resultados entre os modelos que não utilizam componentes temporais.

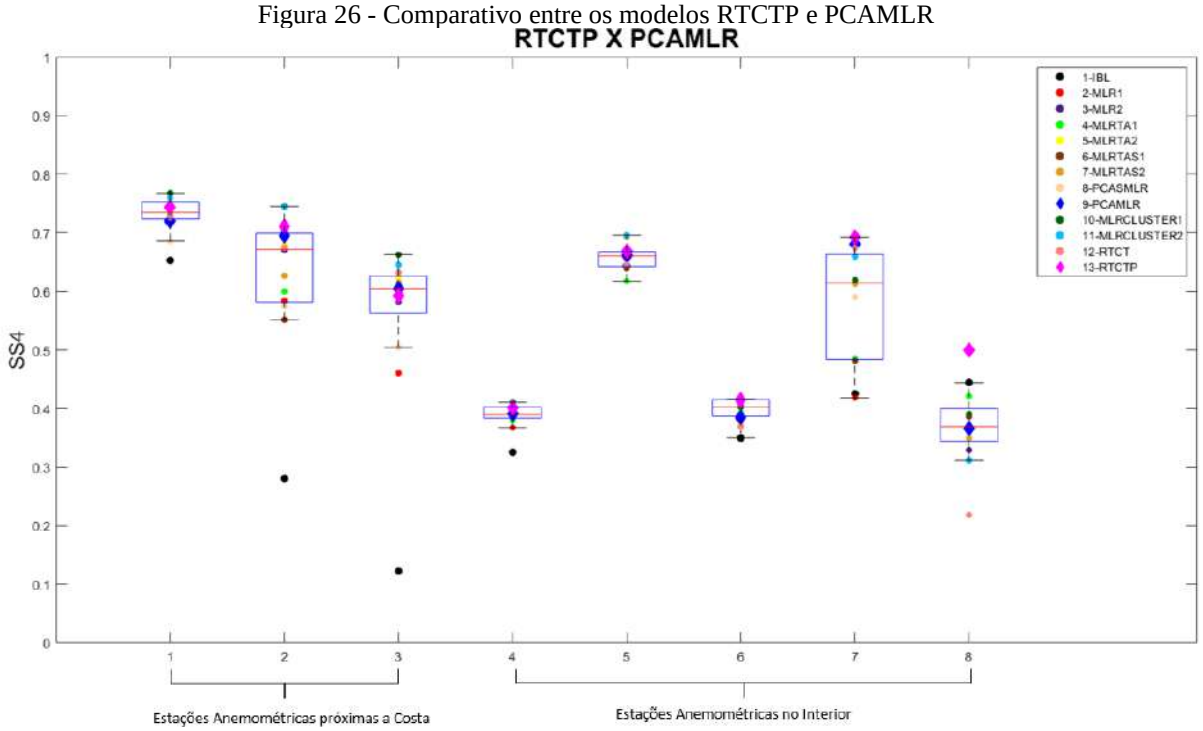
Figura 25 - Comparativo entre os modelos sem componentes temporais



Fonte: O Autor, 2021.

Dado que o PCAMLR foi o melhor modelo sem utilização de componentes temporais e utilizou a informação dos 576 pontos do GCM utilizaremos este para comparar com o RTCTP e MLRTACLUSTER. O RTCTP tem um desempenho superior ao PCAMLR de forma global como pode ser visto na Figura 26.

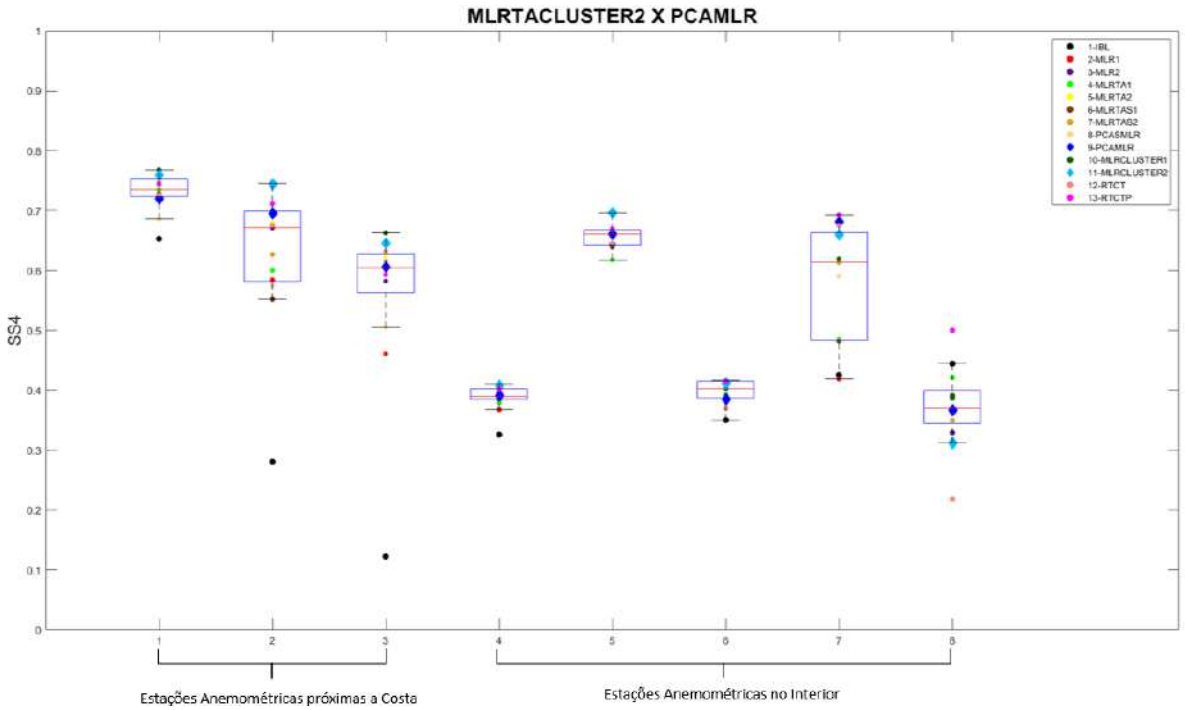
Figura 26 - Comparativo entre os modelos RTCTP e PCAMLR



Fonte: O Autor, 2021.

A Figura 27 exibe uma melhora significativa MLRTACLUSTER2 em relação ao PCAMLR nas torres próximas a costa, enquanto nas torres localizadas no interior, em uma análise global, o MLRTACLUSTER2 é levemente superior.

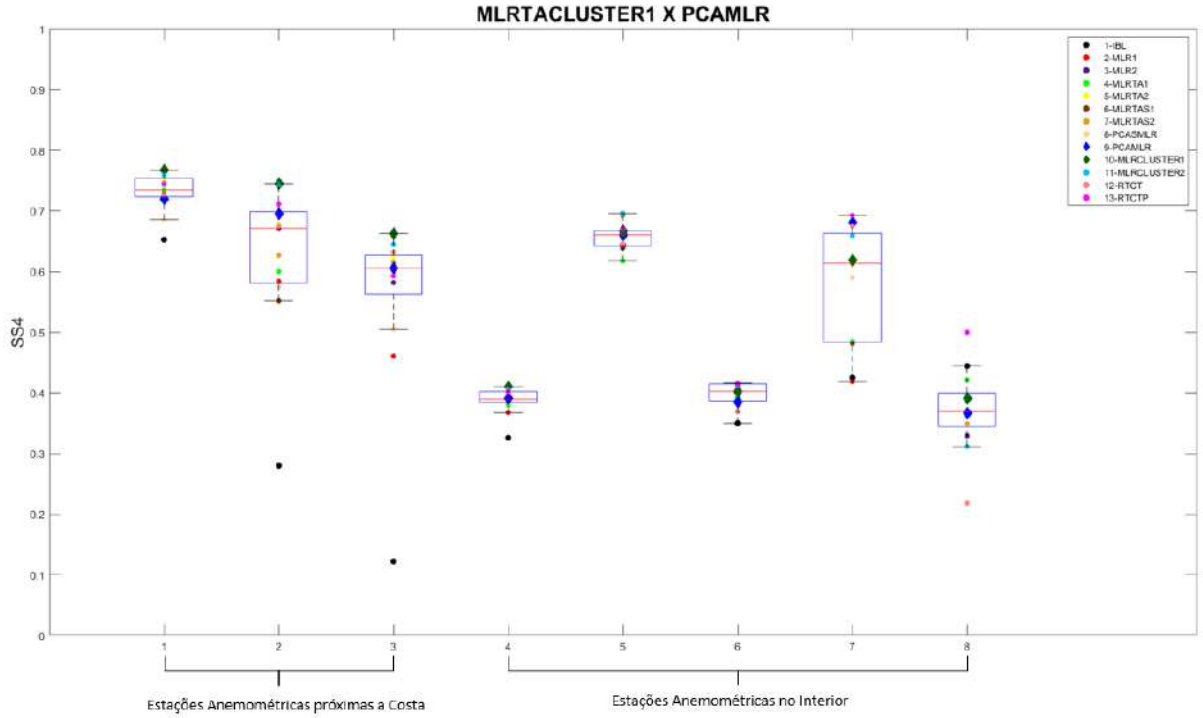
Figura 27 - Comparativo entre os modelos MLRTACLUSTER*2 e PCAMLR*



Fonte: O Autor, 2021.

Na Figura 28 notamos um desempenho superior do MLRTACLUSTER1 nas estações próximas a costa, entretanto, nas estações próximas ao interior, existe um desempenho semelhante entre os dois modelos, com exceção ao local 7, onde o PCAMLR exibe uma melhora significativa. O resultado é esperado como citado em outros momentos desta discussão devido ao melhor desempenho de modelos com maior domínio para as estações localizadas no interior.

Figura 28 - Comparativo entre os modelos MLRTACLUSTER1 e PCAMLR



Fonte: O Autor, 2021.

Em síntese as componentes temporais trouxeram um ganho para a avaliação do comportamento do vento no local de interesse.

6 CONCLUSÕES E PERSPECTIVAS

Os resultados da modelagem do vento local deste estudo demonstram que os preditores anteriores ao instante estimado contribuem para a descrição deste último, o que foi evidenciado pela melhora dos resultados entre modelos que se distinguem apenas pelo emprego, ou não, das componentes temporais, pois o último grupo teve melhores resultados para os locais avaliados neste trabalho. Além disso, os modelos empregados descreveram de maneira satisfatória o comportamento do vento na maioria das estações utilizadas neste trabalho, entretanto, as estações 4 e 6, apesar de terem as melhores estimativas feitas com modelos que utilizam componentes temporais, tratam-se de locais de alta complexidade orográfica e necessitam de refinamento nas modelagens, possivelmente utilizando modelos não lineares. A metodologia desenvolvida pode ser utilizada para avaliação do recurso eólico em avaliações de longo prazo, o que é uma etapa fundamental para decisão da implantação de um novo projeto eólico.

Outro importante fator que pode ser percebido é a diferença de complexidade no comportamento do vento no local de interesse quando este está localizado próximo a costa e quando se localiza no interior. O vento no interior do continente, a princípio, necessita de uma maior quantidade de informações para descrever de forma mais acurada o comportamento deste, por isso, acredita-se que é necessário a utilização de uma maior quantidade de preditores para descrevê-lo no ponto de interesse.

Em adição ao exposto acima, ao se comparar os modelos MLRTA e MLRTAS, também se notou que a seleção de domínio adequada pode contribuir de forma significativa para melhoria do desempenho dos modelos, sendo possível unir a seleção espacial e temporal para melhorar a estimativa do comportamento do vento local. Entretanto, ainda existe ampla possibilidade de melhora na combinação da seleção espacial e temporal de preditores, além das possibilidades de melhora das técnicas de seleção temporal empregadas neste trabalho.

Por fim, acredita-se que dentre os modelos empregados neste trabalho, o MLRTACLUSTER tem o maior potencial para descrever de forma mais acurada o comportamento do vento em regiões próxima a costa e o RTCTP tem maior potencial para descrição de locais no interior e ambos têm desempenho muito superior ao modelo de referência e até mesmo a modelos consolidados empregados na literatura como a análise de componentes principais combinada com regressões lineares múltiplas.

Dado o exposto, sugere-se para novos estudos os pontos listados abaixo:

a) Em relação aos padrões sugeridos no texto entre costa e interior: Avaliação dos modelos listados neste trabalho em uma base de dados maior para verificar se os padrões das estações na costa e no interior continuam se repetindo.
b) Em relação ao modelo RTCTP: Utilizar a adição de componentes temporais em ordem cronológica e avaliar se a metodologia de seleção do próprio modelo gera resultados superiores a realização de uma seleção prévia utilizando como critério a autocorrelação ou autocorrelação parcial.
c) No tocante as técnicas de *clustering*: Avaliar novas metodologias para seleção do número de clusters que será utilizado e validar a partir de uma análise de sensibilidade testando a maior quantidade de possibilidades possíveis.
d) No que faz referência a robustez dos modelos: Alterar os tamanhos dos conjuntos de calibração e validação e verificar o impacto no desempenho dos modelos.
e) No que concerne as técnicas de seleção temporal: Avaliar os preditores no domínio da frequência e determinar padrões para auxiliar na seleção das componentes temporais que serão consideradas como preditores.
f) No que diz respeito a seleção espacial: Combinar técnicas mais robustas de seleção espacial em conjunto com as técnicas de seleção temporal para melhorar o desempenho das estimativas do vento local.

REFERÊNCIAS

Accadia, C. M. (2003). Sensitivity of precipitation forecast skill scores to bilinear interpolation and a simple nearest-neighbor average method on high-resolution verification grids. . *Weather and forecasting*, pp. 918-932.

AMS. (2012). *Meteorology Glossary. American Meteorology Society*.

Bernhardt, M. L. (2010). High resolution modelling of snow transport in complex terrain using downscaled MM5 wind fields. pp. 99-113.

BORDONI, S., & STEVENS, B. (2006). Principal component analysis of the summertime winds over the Gulf of California: A gulf surge index. . *Monthly weather review,*, pp. 3395-3414.

Box, G. E., Jenkins, G. M., & Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*. Wiley.

Bueno, R. d. (2012). *Econometria de Séries Temporais*. CENGAGE.

BUSUIOC, A., TOMOZEIU, R., & CACCIAMANI, C. (2008). Statistical downscaling model based on canonical correlation analysis for winter extreme precipitation events in the Emilia-Romagna region. *International Journal of Climatology*, pp. 449-464.

COSTA, A., CRESPO, A., NAVARRO, J., LIZCANO, G., MADSEN, H., & FEITOSA, E. (2008). A review on the young history of the wind power short-term prediction. *Renewable and Sustainable Energy Reviews*.

CURRY, C. L., VAN DER KAMP, D., & MONAHAN. (2012). A. H. Statistical downscaling of historical monthly mean winds over a coastal region of complex terrain I. Predicting wind speed. *Climate dynamics*, pp. 1281-1299.

Dee, D. P. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*.

Dibike, Y. B., & Coulibaly, P. (2006). Temporal Neural Networks for Downscaling. *Temporal Neural Networks for Downscaling*.

EMPRESA DE PESQUISA ENERGÉTICA. *Balanço Energético Nacional*. Rio de Janeiro, EPE, 2018. Disponível em < https://ben.epe.gov.br. Acesso em:

F., B., BOSSY, M., CHAUVIN, C., DROBINSKI, P., ROUSSEAU, A., & SALAMEH, T. (2009). Stochastic downscaling method: application to wind refinement. *Stochastic Environmental Research and Risk Assessment*, pp. 851-859.

Feng, C., Cui, M., Hodge, B.-M., & JieZhanga. (2017). A data-driven multi-model methodology with deep feature selection for short-term wind forecasting. *Applied Energy*.

George E. P. Box, G. M. (1994). *Time Series Analysis.: Forecasting and Control*. John Wiley & Sons, 2015.

GUTIÉRREZ, J. M., S., C. A., R., C., & M., R. (2004). Clustering methods for statistical downscaling in short-range weather forecasts. *Monthly Weather Review*, pp. 2169-2183.

Harpham, C., & L.Wilby, R. (2005). Multi-site downscaling of heavy daily precipitation occurrence and amounts. *Journal of Hydrology*.

HART, N. C., GRAY, S. L., & CLARK, P. A. (2015). Detection of Coherent Airstreams Using Cluster Analysis: Application to an Extratropical Cyclone. *Monthly Weather Review*, pp. 3518-3531.

Heap, J. L. (2008). *A Review of Spatial Interpolation Methods for Environmental Scientists* . GEOSCIENCE AUSTRALIA.

Hessami, M., Gachon, P., Ouarda, T. B., & St-Hilaire, A. (2008). Automated regression-based statistical downscaling tool. *Environmental Modelling & Software*.

Hewitson, B., & Crane, R. (1996). Climate downscaling: techniques and application. *Climate Research*.

HUTH, R. (1999.). Statistical downscaling in central Europe: evaluation of methods and potential predictors. *Climate Research*, pp. 91-101.

JOLLIFFE, I. T. (1986). *Principle component analysis.* Springer Verlag.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., . . . Joseph, D. (1996). The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society*.

LANDBERG, L., & WATSON, S. J. (1994). Short-term prediction of local wind conditions. *Boundary-Layer Meteorology*, 171-195.

Mendes, D. &. (2010). Temporal downscaling: a comparison between artificial neural network and autocorrelation techniques over the Amazon Basin in present and future climate change scenarios. *Theoretical and Applied Climatology*, pp. 413-421.

MURPHY, J. (1999). An evaluation of statistical and dynamical techniques for downscaling local. *Journal of Climate*, pp. 2256-2284.

PALOMINO, I., & MARTIN, F. (1995). A simple method for spatial interpolation of the wind in complex. *Journal of Applied Meteorology*, pp. p. 1678-1693.

Perruci, V. P. (2018). *Análise de complementariedade entre diferentes técnicas estatísticas para aumento na resolução espacial do comportamento do vento local.* Recife: PROTEN.

SCHOOF, J. T., & PRYOR, S. C. (2001). Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks. *International Journal of climatology*, pp. 773-790.

SEMENOV, M. A., & BARROW, E. M. (1997). Use of a stochastic weather generator in the development of climate change scenarios. *Climatic change*, pp. 397-414.

VAN DEN DOOL, H. M. (1994). Searching for analogues, how long must we wait? *Tellus A,*, pp. 314-324.

VRAC, M., STEIN, M., & HAYHOE, K. (2007). Statistical downscaling of precipitation through nonhomogeneous stochastic weather typing. *Climate Research*, pp. 169-184.

Wilby, R. L. (2008). Constructing climate change scenarios of urban. *Environment and Planning B: Planning and Design*.

Wilby, R. L. (2013). The statistical downscaling model: insights from one decade of application. International. *Journal of Climatology*, 1707-1719.

WILBY, R. L., CHARLES, S. P., ZORITA, E., TIMBAL, B., WHETTON, P., & L.O., M. (2004). *Guidelines for use of climate scenarios developed from statistical downscaling methods.*

Wilby, R., & Wigley, T. (1997). Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography*.

WILKS, D. S. (2010). Use of stochastic weather generators for precipitation downscaling. *Wiley Interdisciplinary Reviews: Climate Change*, pp. 898-907.

WILKS, D. S. (2011). *Statistical methods in the atmospheric sciences.* Academic press.

Xu, C.-y. (1999). From GCMs to river flow: a review of downscaling methods and hydrologic modelling approaches. *Progress in Physical Geography: Earth and Environment*.

ZORITA, E., & VON STORCH, H. (1999). The analog method as a simple statistical downscaling technique: comparison with more complicated methods. *Journal of Climate*, pp. 2474-.

APÊNDICE A – DIAGRAMAS DE TAYLOR

As figuras contidas neste apêndice trazem os diagramas de Taylor para as estações de 1 a 8 avaliadas neste trabalho. O diagrama tem as informações dos estatísticos: desvio padrão, correlação e a raiz do erro quadrático das anomalias (RMSD).

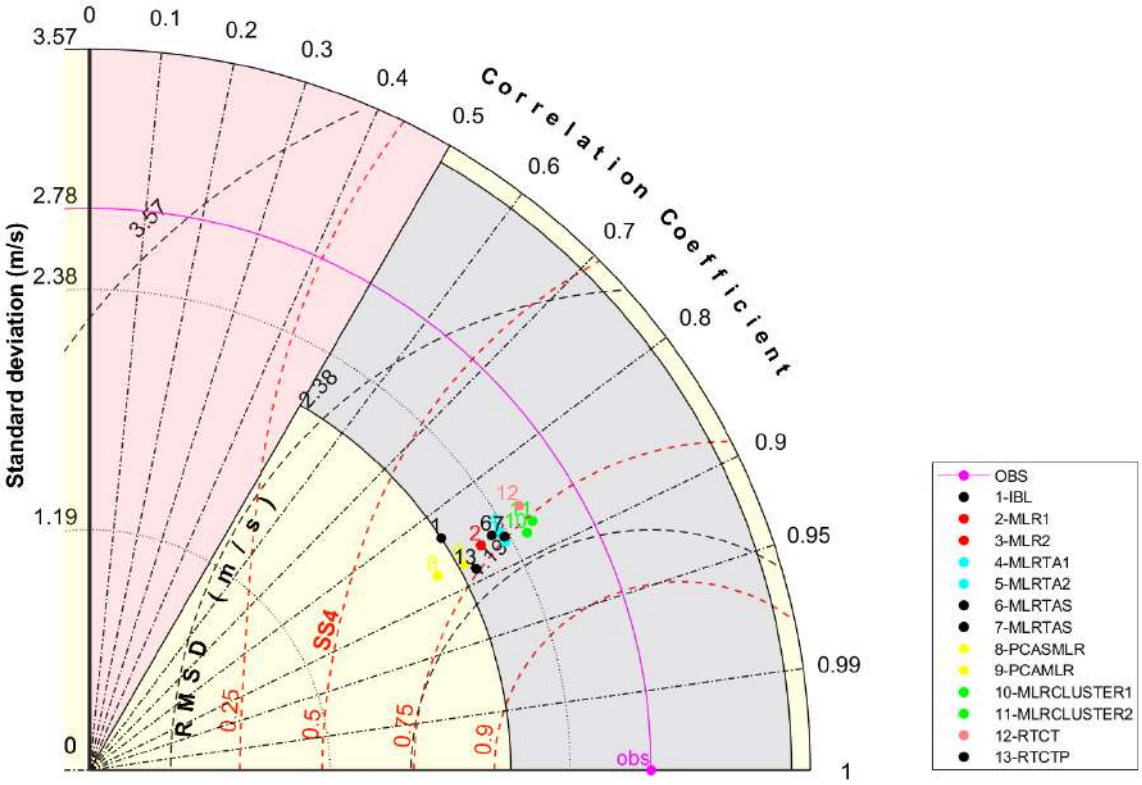
Os pontos enumerados no plano polar da figura representam os resultados estimados na posição da estação anemométrica avaliada por cada um dos modelos, que podem ser vistos na legenda do diagrama. A posição dos pontos no diagrama dar-se em função dos estatísticos avaliados para as séries temporais estimadas por cada um dos modelos relativas à série temporal observada (ponto nomeado como “obs”).

Podemos avaliar os estatísticos citados anteriormente observando diferentes aspectos do diagrama. As distâncias radiais com respeito à origem são proporcionais ao desvio padrão das séries temporais. A posição azimutal indica o coeficiente de correlação entre a estimativa do modelo avaliado e as observações, indicado no gráfico pelo cosseno do ângulo formado entre uma reta contendo a origem e a abcissa. O RMSD é proporcional às distâncias radiais centradas na série temporal medida no local avaliado, a leitura do SS4 é feita de forma análoga. Semelhantemente ao RMSD, o SS4 é um parâmetro para avaliar o desempenho global dos modelos e foi definido na seção 4.

As Figura A1, A2, A3, A4, A5, A6, A7, A8 trazem o diagrama de Taylor para as estações de 1 a 8 avaliadas neste trabalho.

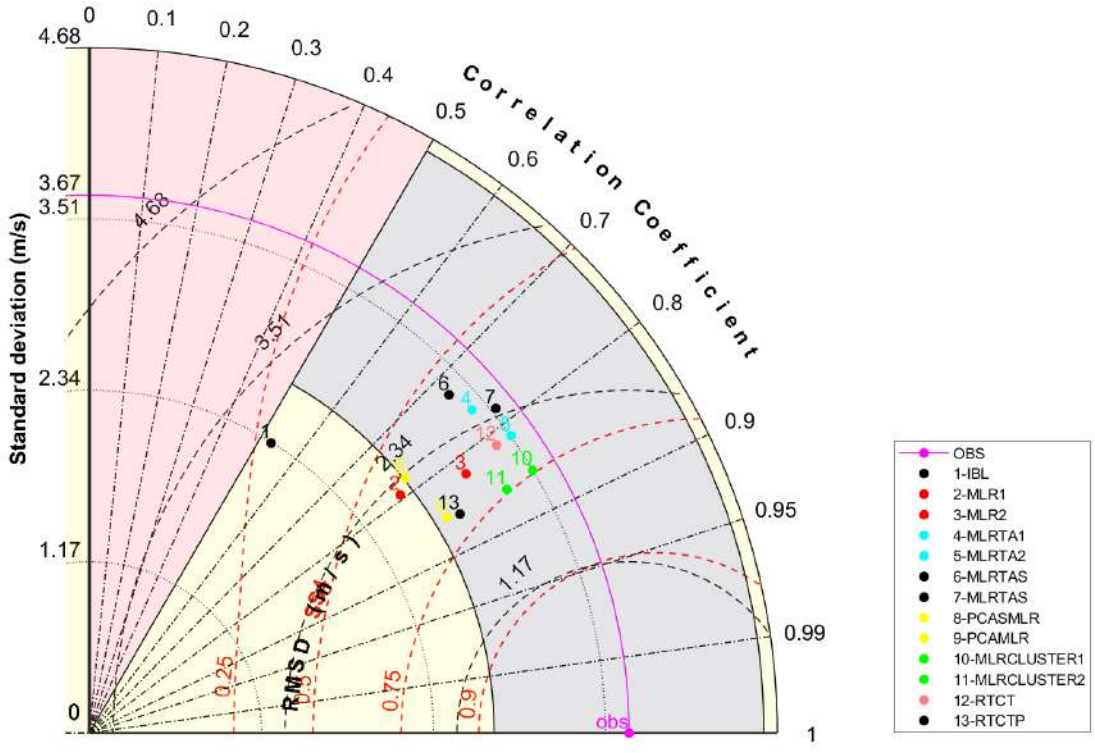
Os resultados vistos nos diagramas de Taylor das figuras deste apêndice confirmam os pontos já discutidos na seção 5 quando se observa os valores do SS4. Além disto, ao observar o comportamento global dos modelos que tem adição de componentes temporais também se pode perceber uma tendência de melhora da descrição da variabilidade das séries temporais avaliadas. Devido a capacidade de condensar diversas informações em um único diagrama, os diagramas de Taylor deste apêndice poderão ser utilizados em análises futuras para aprofundar o entendimento do comportamento dos modelos desenvolvidos neste trabalho.

Figura A1 - Diagrama de Taylor com os resultados para o local 1



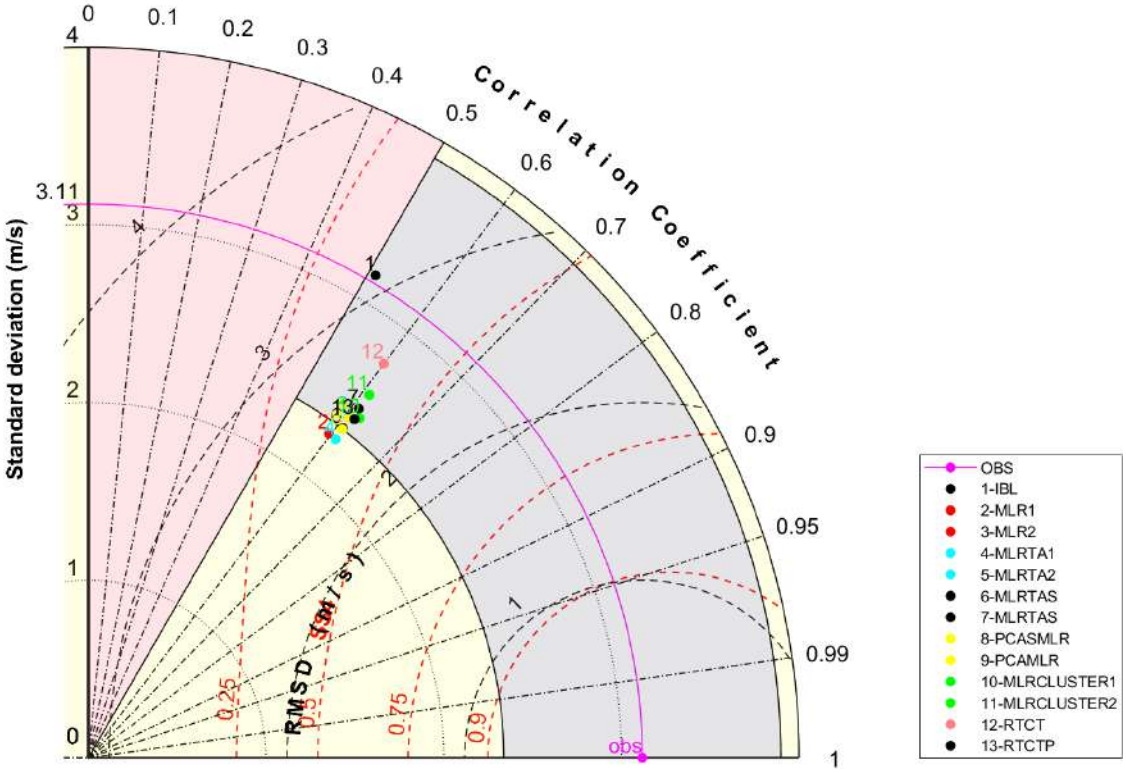
Fonte: O Autor, 2021.

Figura A2 - Diagrama de Taylor com os resultados para o local 2



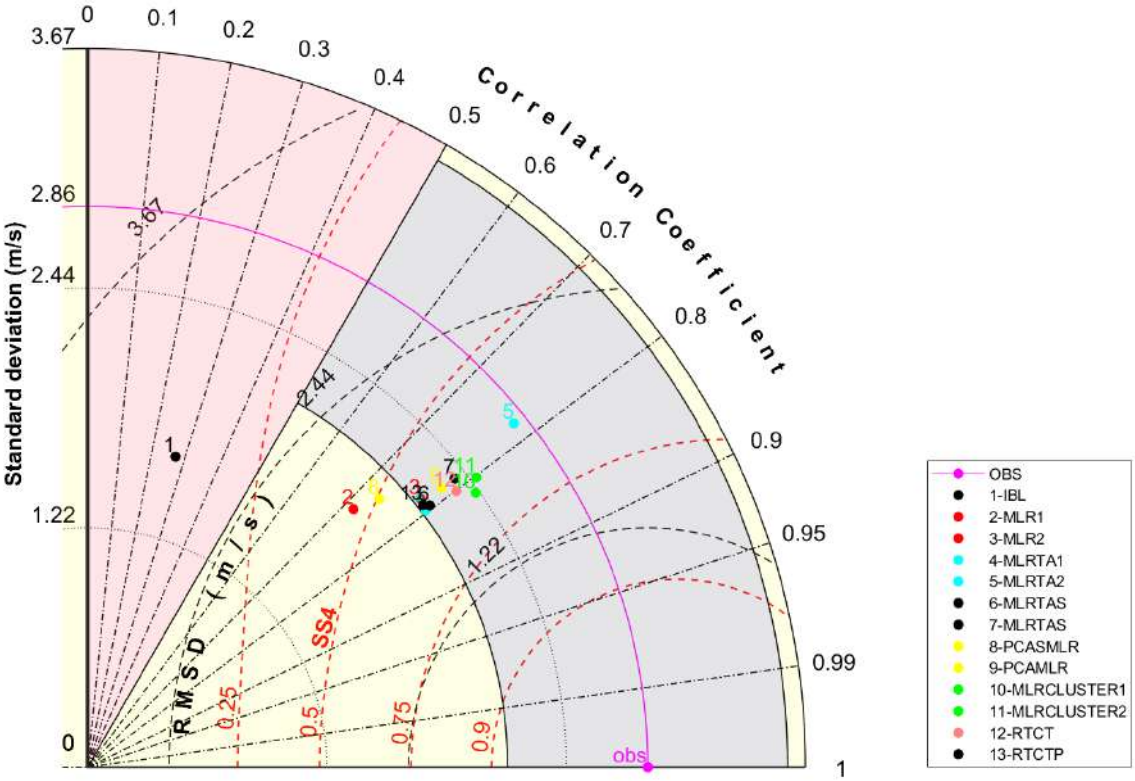
Fonte: O Autor, 2021.

Figura A3 - Diagrama de Taylor com os resultados para o local 3



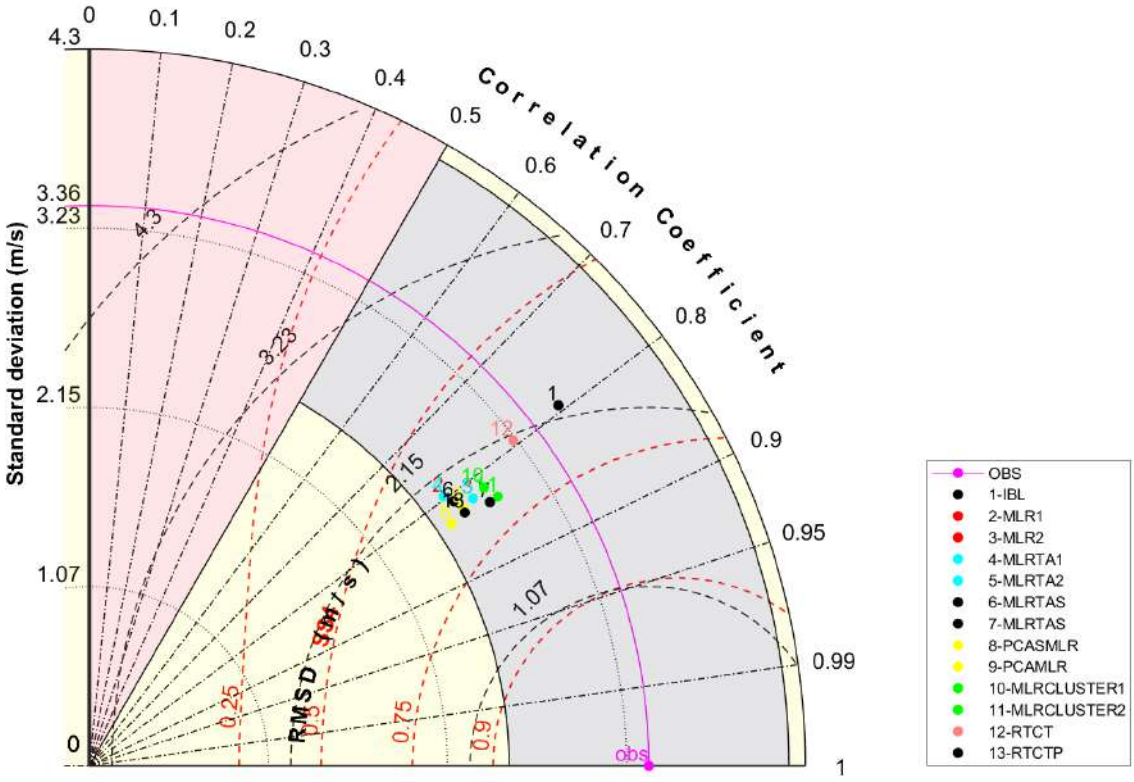
Fonte: O Autor, 2021.

Figura A4 - Diagrama de Taylor com os resultados para o local 4



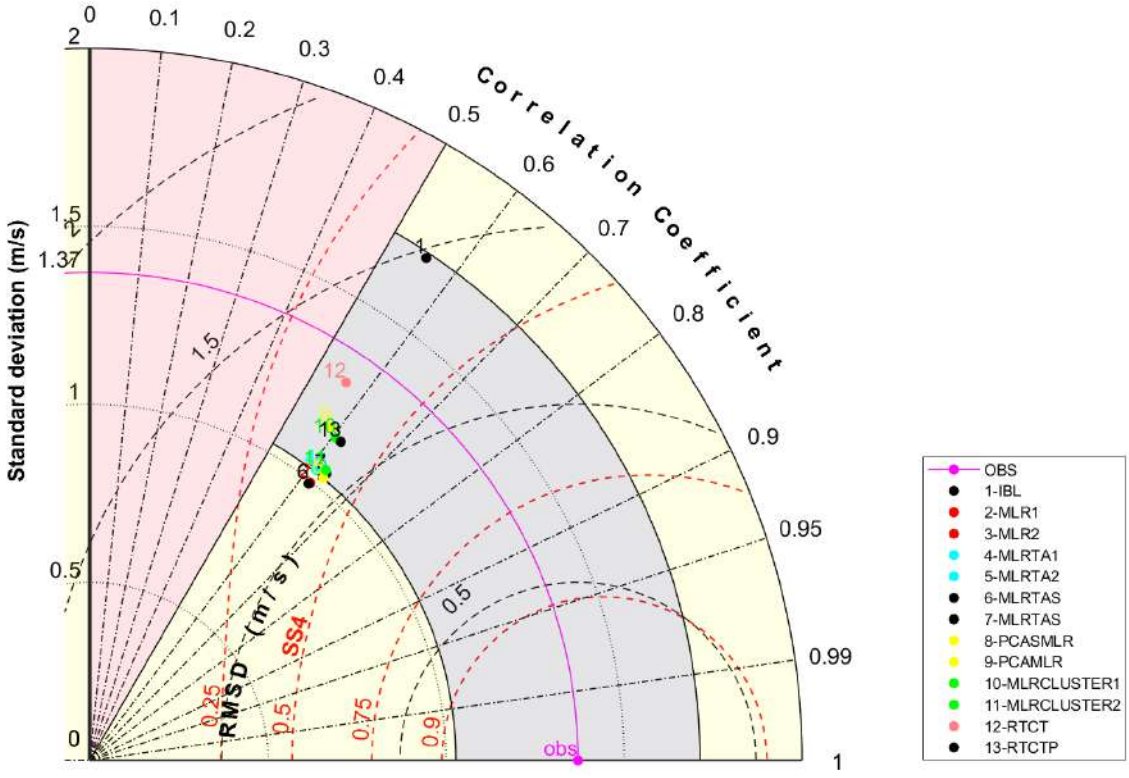
Fonte: O Autor, 2021.

Figura A5 - Diagrama de Taylor com os resultados para o local 5



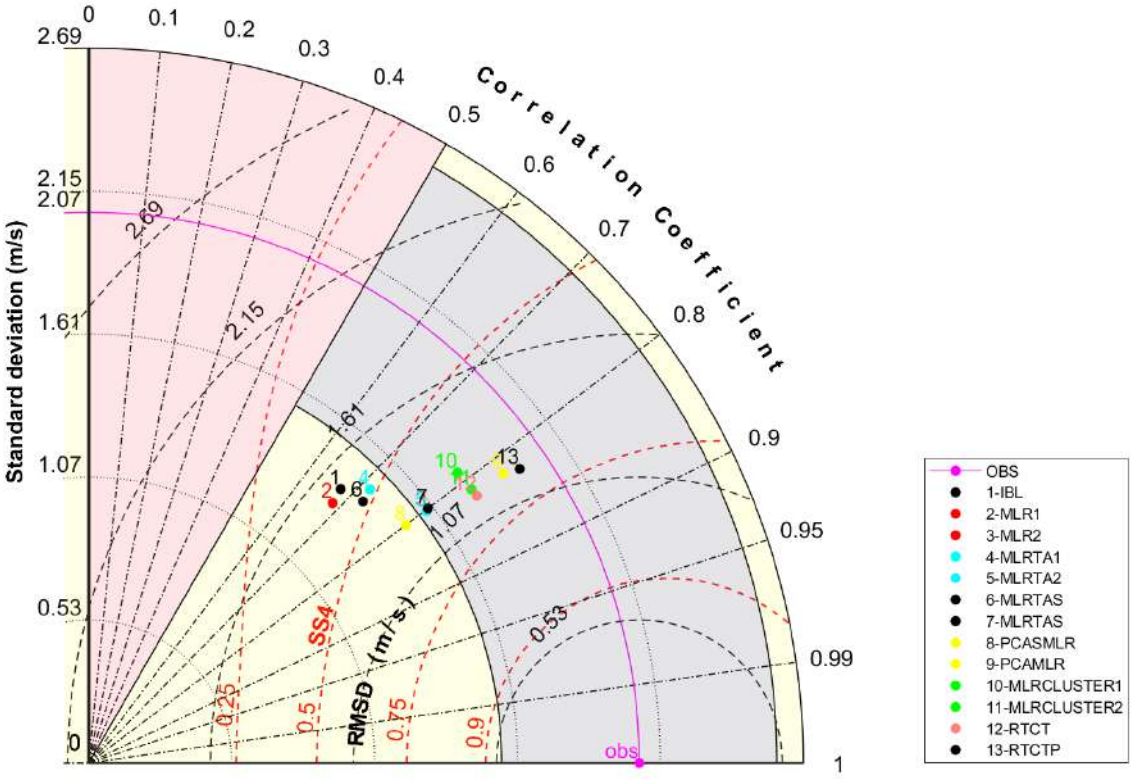
Fonte: O Autor, 2021.

Figura A6 - Diagrama de Taylor com os resultados para o local 6



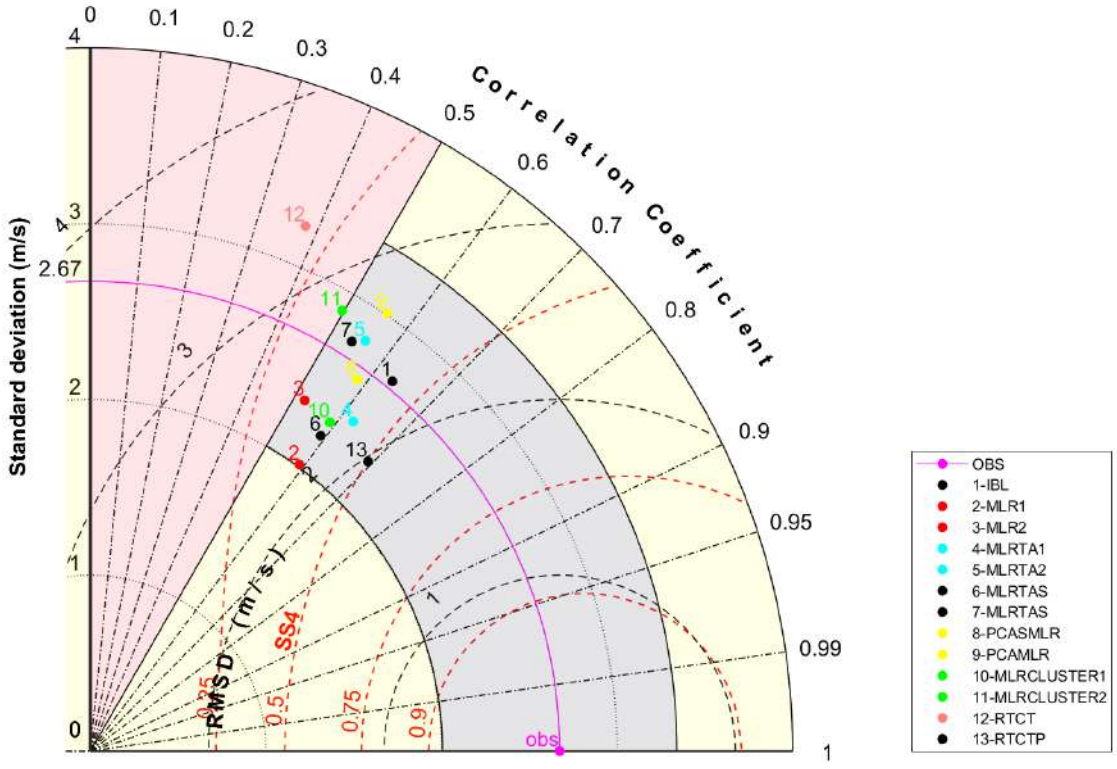
Fonte: O Autor, 2021.

Figura A7 - Diagrama de Taylor com os resultados para o local 7



Fonte: O Autor, 2021.

Figura A8 - Diagrama de Taylor com os resultados para o local 8



Fonte: O Autor, 2021.

APÊNDICE B – COMPARATIVO DE PERFORMACE

A tabela contida neste apêndice traz a síntese de resultados dos principais modelos desenvolvidos neste trabalho, em comparação ao IBL, utilizando como métrica o SS4.

Tabela B1 - Melhora de desempenho em relação ao IBL

ESTAÇÃO	RTCTP	MLRTACLUSTER2	MLRTACLUSTER 1
LOCAL 1	14.0%	16.3%	17.6%
LOCAL 2	153.7%	165.2%	165.8%
LOCAL 3	383.8%	426.4%	440.7%
LOCAL 4	23.4%	25.0%	26.0%
LOCAL 5	3.9%	8.0%	3.6%
LOCAL 6	18.4%	17.6%	15.0%
LOCAL 7	62.9%	55.0%	45.6%
LOCAL 8	12.5%	-29.9%	-12.1%

Fonte: O Autor, 2021.