

3 MATERIAL E MÉTODOS

3.1 Material

3.1.1 Dados

Foram utilizados, nesta pesquisa, dados mensais de temperaturas mínimas médias obtidos em 40 estações meteorológicas no Rio Grande do Sul, no período de 1913 a 2002, perfazendo um total de 90 anos de dados, abrangendo três normais climatológicas, o que permite antever resultados satisfatórios pelo período abrangido na matriz de dados. Esses dados foram obtidos junto ao Instituto Nacional de Meteorologia – INMET (8^o Distrito de Meteorologia – Porto Alegre) do Ministério da Agricultura, Pecuária e Abastecimento (MAPA), e junto à Fundação Estadual de Pesquisas Agropecuárias (FEPAGRO), da Secretaria de Ciência e Tecnologia (SCT) do Estado do Rio Grande do Sul.

A tabela 1 mostra as 40 estações meteorológicas das quais os dados originais foram obtidos. Nela estão incluídos os dados de latitude, longitude e altitude de cada estação, bem como o período em que os dados estiveram disponíveis.

TABELA 1. Estações meteorológicas do Estado do Rio Grande do Sul, coordenadas geográficas, período e instituições a que pertencem.

Estação	Latitude	Longitude	Altitude (m)	Período	Fonte
1-Alegrete	29°41'	55°31'	124	1922 – 1998	8º DISME
2-Bagé	31°21'	54°06'	215	1913 – 1998	8º DISME
3-Bento Gonçalves	29°15'	51°31'	619	1922 – 1998	8º DISME
4-Bom Jesus	28°40'	50°26'	1047	1948 – 1998	8º DISME
5-Caçapava do Sul	30°30'	53°29'	450	1915 – 1974	8º DISME
6-Cachoeira do Sul	30°02'	52°53'	72	1913 – 1986	8º DISME
7-Caxias do Sul	29°10'	51°12'	785	1931 – 1998	8º DISME
8-Cruz Alta	28°38'	53°36'	472	1913 – 1998	8º DISME
9-Encruzilhada do Sul	30°32'	52°31'	427	1914 – 1998	8º DISME
10-Farroupilha	29°14'	51°26'	702	1963 – 1998	FEPAGRO
11-Guaporé	28°55'	51°54'	772	1913 – 1981	8º DISME
12-Ijuí	28°23'	53°54'	448	1963 – 1998	FEPAGRO
13-Iraí	27°11'	53°14'	222	1934 – 1998	8º DISME
14-Júlio de Castilhos	29°13'	53°40'	516	1914 – 1998	FEPAGRO
15-Lagoa Vermelha	28°25'	51°35'	836	1931 – 1996	8º DISME
16-Marcelino Ramos	27°27'	51°54'	414	1916 – 1972	8º DISME
17-Osório	29°41'	50°13'	32	1956 – 1998	FEPAGRO
18-Palmeira das Missões	27°53'	53°26'	634	1915 – 1986	8º DISME
19-Passo Fundo	28°15'	52°24'	676	1913 – 1998	8º DISME
20-Pelotas	31°52'	52°21'	13	1913 – 1998	8º DISME
21-Porto Alegre	30°01'	51°13'	46	1916 – 1998	8º DISME
22-Rio Grande	32°01'	52°05'	5	1913 – 1998	8º DISME
23-Santa Cruz do Sul	29°43'	52°25'	56	1915 – 1967	8º DISME
24-Santa Maria	29°42'	53°42'	95	1913 – 1998	8º DISME
25-Santa Rosa	27°51'	54°25'	360	1922 – 1998	8º DISME
26-Santa Vitória do Palmar	33°31'	53°21'	6	1913 – 1998	8º DISME
27-Santana do Livramento	30°53'	55°32'	210	1913 – 1994	8º DISME
28-Santiago	29°11'	54°53'	426	1917 – 1984	8º DISME
29-Santo Ângelo	28°18'	54°15'	289	1914 – 1994	8º DISME
30-São Borja	28°39'	56°00'	96	1914 – 1998	FEPAGRO
31-São Francisco de Paula	29°20'	51°30'	912	1913 – 1960	8º DISME
32-São Gabriel	30°20'	54°19'	124	1913 – 1998	8º DISME
33-São Luiz Gonzaga	28°23'	54°58'	254	1913 – 1998	8º DISME
34-Soledade	29°03'	52°26'	720	1931 – 1963	8º DISME
35-Tapes	30°52'	51°21'	5	1924 – 1973	8º DISME
36-Taquari	29°48'	51°49'	76	1913 – 1998	8º DISME
37-Torres	29°20'	49°43'	43	1931 – 1998	8º DISME
38-Uruguaiana	29°45'	57°05'	74	1931 – 1998	8º DISME
39-Vacaria	28°33'	50°42'	960	1931 – 1998	8º DISME
40-Veranópolis	28°56'	51°33'	705	1913 – 1998	FEPAGRO

A figura 1 mostra a localização espacial das estações meteorológicas. Como pode ser visto, elas estão distribuídas de modo a cobrir todo o território do Rio Grande do Sul. Embora sua distribuição não seja a ideal, concentrando-se mais em algumas regiões do que em outras, pode-se considerar os dados como representativos de todos as sub-regiões ecoclimáticas do Estado.

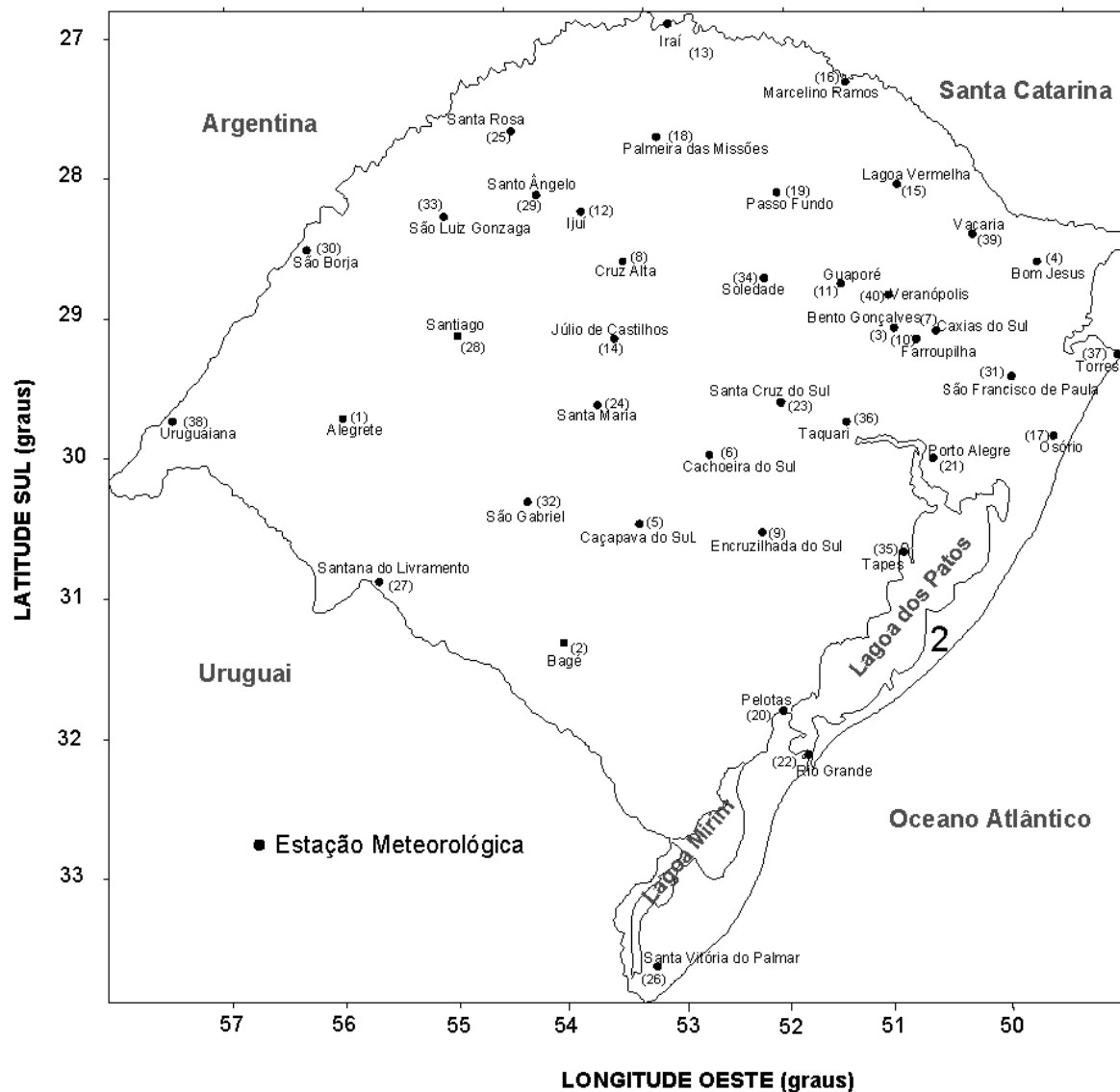


Figura 1: Distribuição espacial das estações meteorológicas do Estado do Rio Grande do Sul utilizadas neste trabalho.

3.1.2 Preenchimento de Falhas

As séries já tiveram suas falhas preenchidas através do “método das correlações” aplicado por Diniz (2002). Esta técnica consiste em formar, inicialmente, uma matriz de correlação entre todas as estações do conjunto de dados. As estações,

com alto grau de coeficiente de correlação entre elas (igual ou superior a $|0,8|$), foram agrupadas. Diniz verificou, a seguir, quais tinham altitudes semelhantes, que formaram subgrupos dos primeiros. Por último, foi levada em consideração a proximidade física entre as estações, de forma a garantir mais homogeneidade entre elas. Obtidos os grupos definitivos de estações correlacionadas, as falhas foram preenchidas dentro do grupo pela média dos valores das temperaturas de cada estação.

Além deste método, foram testados pelo pesquisador os métodos da análise de fatores em componentes principais e o método da regressão linear múltipla. A escolha do método mais adequado foi fundamentada na comparação entre eles, usando como parâmetro de comparação o erro relativo médio do grupo de estações meteorológicas que tiveram suas falhas preenchidas. A técnica das correlações foi a que apresentou o menor erro relativo.

3.1.3 Organização dos Dados

A matriz dos dados originais foi organizada calculando-se a média, por trimestre, das temperaturas mínimas médias de janeiro, fevereiro e março, designado “trimestre 1”; de abril, maio e junho, “trimestre 2”; de julho, agosto e setembro, “trimestre 3”; de outubro, novembro e dezembro, “trimestre 4”. Optou-se por trabalhar com trimestres do ano civil, ao invés do astronômico, pela facilidade de se trabalhar com dados do mesmo ano e por essa escolha não alterar substancialmente os objetivos da pesquisa.

Para cada trimestre foi organizada uma matriz de dados com “n” linhas, representando as estações meteorológicas da rede, e em “p” colunas, indicando os 90 anos de observações, incluindo-se nestas as coordenadas de posição (altitude, latitude e longitude). As linhas variam, portanto, de 1 a 40 e as colunas de 1 a 93.

$$X_{(n \times p)} = \begin{pmatrix} X_{11} & X_{12} & . & . & . & X_{1p} \\ X_{21} & X_{22} & . & . & . & X_{2p} \\ . & & & & & . \\ . & & & & & . \\ X_{n1} & X_{n2} & . & . & . & X_{np} \end{pmatrix} \quad (54)$$

Como foram incluídas as coordenadas de posição, os dados da matriz ficaram com unidades diferentes. Para se obter a uniformidade dos mesmos, foi necessário padronizá-los. A padronização é obtida pelo cociente entre a diferença, de cada dado e a média da série, e o seu desvio padrão, ou seja

$$y = \frac{X - \bar{X}}{\sigma} \quad (55)$$

onde: x : dado

$\bar{X}$: média da série

σ : desvio padrão

y : dado padronizado

3.2 Metodologia

3.2.1 Testes de Normalidade e Homogeneidade das Variâncias

Na aplicação de muitas técnicas estatísticas a um conjunto de dados, como na análise de séries temporais, é importante verificar, segundo Jenkins (1979), citado por Ferraz (1999), (ou Baptista da Silva, 2001; Amaral, 1975, Assis, 1996), a normalidade e a homogeneidade das variâncias dos mesmos.

Foram aplicados, com essa finalidade, testes estatísticos adequados, mais precisamente, o teste de Shapiro-Wilk (Teste W) para a normalidade e o de Cochran para a homogeneidade . Caso não se verificassem essas propriedades, buscar-se-ia uma transformação adequada dos dados.

3.2.2 Análise de Agrupamentos

À cada matriz de dados padronizados foram aplicados métodos hierárquicos aglomerativos de agrupamentos, mais precisamente: o da ligação simples, o da ligação completa, o da centróide e o de Ward, tendo como função de agrupamento a distância euclidiana, com exceção do método de Ward, para o qual a função de agrupamento foi a distância quadrática, por ser mais adequada, segundo recomendação contida no programa utilizado. A aplicação de vários métodos se justifica pelo fato de que cada um gera diferentes estruturas, podendo-se optar por aquela que melhor reflita as características climáticas do Estado do Rio Grande do Sul.

O número de grupos e sua composição foi determinado de forma subjetiva através de um corte no dendrograma. A escolha foi feita de modo a garantir um número razoável de estações em cada grupo, sem excessiva fragmentação da superfície estudada. Cada grupo formado foi designado “região homogênea”.

3.2.3 Análise Espectral

Obtidas as regiões homogêneas do Estado do Rio Grande do Sul, foi feito um estudo do comportamento da temperatura mínima média, por trimestre, numa dessas regiões. Optou-se por aquela que contém a estação meteorológica de Pelotas, onde se situa a Universidade Federal de Pelotas, e por ser a que cobriu uma maior área do

Estado. Foi determinada a média das temperaturas mínimas médias mensais das estações que compõe essa região, obtendo-se séries temporais regionais por trimestre. Cada série, “trimestre”, está composta por noventa dados, um para cada ano de observação.

A cada série trimestral da região homogênea escolhida foi aplicada a técnica da Análise Espectral, baseada na Transformada Discreta de Fourier, para identificação dos possíveis harmônicos existentes e de sua importância na composição do modelo de temperaturas mínimas médias para a região. Foram utilizados dados de 84 anos, de 1913 a 1996, para a Análise Espectral e para a composição dos modelos e 6 anos, de 1997 a 2002, para previsões.

Este modelo, baseado na representação periódica, apresenta a forma

$$\hat{X}_t = \frac{A_0}{2} + \sum_n [A_n \cos \omega_n t + B_n \sin \omega_n t] + e_t \quad (56)$$

onde: $\hat{X}_t$: valor estimado da variável temporal.

ω_n : frequência angular da n-ésima onda.

A_n e B_n : parâmetros a serem estimados.

e_t : erros

A seleção das ondas utilizadas na composição dos modelos foi baseada, conforme procedeu Baptista da Silva (2001), na magnitude dos picos no periodograma e na observação dos valores estimados da função de densidade espectral. Foram incluídas, em cada modelo, as componentes senoidais ($\sin \omega_n$ e $\cos \omega_n$) que tiveram coeficientes significativos, estimados segundo a técnica da regressão linear múltipla. Cada modelo está composto por um número de ondas adequado, de modo a garantir uma boa representatividade dos dados, sem que se torne excessivamente complexo.

O periodograma apresenta no eixo das abscissas a frequência e nas ordenadas o valor do periodograma, calculado pela soma dos quadrados dos coeficientes do seno e do cosseno, em cada frequência, multiplicada por $N/2$, ou seja

$$I_n = \frac{N}{2} (A_n^2 + B_n^2) \quad (57)$$

O periodograma pode, igualmente, ser expresso na forma complexa, conforme equação (52).

O periodograma mostra como as variâncias da série, constituídas da mistura de senos e cossenos, estão distribuídas entre as várias frequências harmônicas distintas, assim como o espectro mostra como as variâncias de um processo estocástico estão distribuídas entre os intervalos contínuos de frequências.

Para Box, Jenkins e Reinsel (1994), algumas vezes é conveniente basear a definição do espectro nas autocorrelações em lugar das autocovariâncias. A função resultante é chamada “função de densidade espectral”. Esta tem as mesmas propriedades de uma função de distribuição de probabilidade comum.

As estimativas da densidade espectral são mais bem calculadas pela suavização dos valores do periodograma usando as especificações na janela de dados (Data Windows) para densidade espectral. Pela suavização do periodograma podem-se identificar as regiões das frequências que contribuem significativamente no comportamento cíclico das séries. Os pesos usados na suavização são padronizados, tal que somam 1,0. No início e no final das séries, a suavização é feita por reflexão.

3.2.4 Testes de Ajuste dos Modelos

A metodologia adotada nesta pesquisa foi a de selecionar as ondas mais destacadas em cada periodograma, construir, a partir delas, os modelos e verificar sua validade pela análise dos resíduos. Se os modelos são verdadeiros, os desvios devem ser normais e independentes, ou seja, constituir-se num ruído branco. Se os modelos são adequados, os desvios estimados devem estar próximos dos reais e aproximadamente não correlacionados.

Foram aplicados, com essa finalidade, testes estatísticos adequados, mais precisamente, a análise das “autocorrelações dos resíduos” e o “teste t”.

3.2.4.1 Autocorrelação dos Resíduos

Estimado o modelo, a verificação do seu ajuste pode ser feita pela análise das autocorrelações dos resíduos. Se os r_k indicarem as autocorrelações dos resíduos $\hat{e}_t$, então deve-se ter $r_k \cong 0$.

Tendo-se em vista que r_k tem aproximadamente distribuição normal, com média zero e variância $1/N$, pode-se considerar, de um modo geral, que o modelo é adequado quando r_k está dentro do limite $\pm 2/\sqrt{N}$, aceitando-se, no máximo, 5% dos seus valores fora deste intervalo.

Determina-se, também, a existência de ruído branco a partir da comparação de r_k com $\pm 2\sigma(r_k)$, onde $\sigma(r_k)$, desvio padrão de r_k , é definido como

$$\sigma(r_k) \cong \left[\frac{1}{N} \left(1 + 2 \sum_{v=1}^q \rho_v^2 \right) \right]^{\frac{1}{2}}, \quad k > q \quad (58)$$

onde q é o maior “lag”, além do qual as funções de autocorrelação tendem a zero. (Mais detalhes em Box, Jenkins e Reinsel, 1994).

Considerando-se que a série seja um processo “ruído branco”, isto é, que todas as autocorrelações são iguais a zero, o erro padrão de r_k fica determinado como

$$\sigma(r_k) = \sqrt{\frac{1}{N} \left[\frac{N-k}{N+2} \right]}. \quad (59)$$

Novamente, aceita-se como ruído branco quando, no máximo, 5% das autocorrelações ficam fora do intervalo $\pm 2\sigma(r_k)$.

3.2.4.2 Teste t

O teste t, pode, igualmente, ser utilizado para a verificação do ajuste dos modelos (Togrul & Togrul, 2002).

Embora os indicadores estatísticos MBE (*mean bias error*) e RMSE (*root mean square error*) forneçam um razoável procedimento para comparar modelos, eles não indicam objetivamente se as estimativas do modelo são estatisticamente significativas. A medida estatística baseada no teste t permite que os modelos sejam comparados e, ao mesmo tempo, indica se suas estimativas são estatisticamente significativas.

O valor de t, a ser calculado, é

$$t = \sqrt{\frac{(N-1)MBE^2}{RMSE^2 - MBE^2}}, \quad (60)$$

$$\text{onde: } MBE = \frac{1}{N} \sum_{i=1}^N (X_o - X_e), \quad RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_o - X_e)^2}$$

X_o = valor observado e X_e = valor estimado pelo modelo.

Este valor deve ser comparado com o valor de t da tabela bilateral, para o nível de significância escolhido, com (N-1) graus de liberdade. Quanto menor o valor de t, comparado com o valor limite da tabela, melhor a performance do modelo.

3.2.5 Médias Móveis

Para melhor verificar a qualidade dos ajustes dos modelos em relação aos dados, foi utilizada a técnica da suavização, tanto dos dados originais como dos valores estimados pelo modelo. Foi utilizado o “filtro de médias móveis”, conforme descrito no item 2.2.2.

3.2.6 Modelos Suavizados

Além da descrição no domínio da frequência, pelo modelo espectral, foi utilizado um modelo de alisamento. Os modelos de alisamento fazem parte dos modelos de previsão ditos automáticos. O princípio básico desses modelos consiste em alisar os valores da série temporal em estudo para identificar o padrão básico de comportamento e utilizá-lo para previsão.

Optou-se pelo modelo de “Alisamento Exponencial Linear de Brown” (AELB), descrito na seção 2.2.4, por ser adequado às séries que apresentam tendência.