



# INSTITUTO NACIONAL DE PESQUISA ESPACIAL – INPE

## PROVA DISCURSIVA

### TG35

#### DESENVOLVIMENTO E ANÁLISE DE PRODUTOS DE PREVISÃO NUMÉRICA DE TEMPO E DADOS AMBIENTAIS



#### SUA PROVA

- Além deste caderno contendo **5 (cinco)** questões discursivas **com as respectivas folhas de rascunho**, você receberá do fiscal de prova as folhas de textos definitivos;



#### TEMPO

- Você dispõe de **4 (quatro) horas** para a realização da prova;
- **2 (duas) horas** após o início da prova, é possível retirar-se da sala, sem levar o caderno de questões;
- A partir dos **30 (trinta) minutos** anteriores ao término da prova é possível retirar-se da sala **levando o caderno de questões**.



#### NÃO SERÁ PERMITIDO

- Qualquer tipo de comunicação entre os candidatos durante a aplicação da prova;
- Anotar informações relativas às respostas em qualquer outro meio que não seja no caderno de questões e nas folhas de textos definitivos;
- Levantar da cadeira sem autorização do fiscal de sala;
- Usar o sanitário ao término da prova, após deixar a sala.



#### INFORMAÇÕES GERAIS

- Verifique se seu caderno de questões está completo, sem repetição de questões ou falhas. Caso contrário, **notifique imediatamente o fiscal da sala**, para que sejam tomadas as devidas providências;
- Confira seus dados pessoais, especialmente nome, número de inscrição e documento de identidade e leia atentamente as instruções para preencher as folhas de textos definitivos;
- Para o preenchimento das folhas de textos definitivos, use somente caneta esferográfica, fabricada em material transparente, com tinta preta ou azul;
- Assine seu nome apenas no(s) espaço(s) reservado(s) no cartão de respostas;
- Caso você tenha recebido caderno de cargo **diferente** do impresso em suas folhas de textos definitivos, o fiscal deve ser **obrigatoriamente** informado para o devido registro na ata da sala;
- O preenchimento das folhas de textos definitivos é de sua responsabilidade e **não será permitida a troca de folha de texto definitivo em caso de erro cometido pelo candidato**;
- Para fins de avaliação, serão levadas em consideração apenas os textos das folhas de textos definitivos;
- A FGV coletará as impressões digitais dos candidatos na lista de presença;
- Os candidatos serão submetidos ao sistema de detecção de metais quando do ingresso e da saída de sanitários durante a realização das provas.
- **Boa prova!**

### Questão 1

A precisão nas previsões de temperatura é vital para a gestão de recursos energéticos, planejamento agrícola e resposta a eventos climáticos extremos. Métricas específicas, como o Viés, o Erro Médio Absoluto (MAE), a Raiz do Erro Quadrático Médio (RMSE) e o Coeficiente de Correlação, são fundamentais para avaliar o desempenho de modelos de previsões numéricas meteorológicas. Essas métricas estatísticas podem ser utilizadas para avaliar a precisão da previsão de temperatura e o papel das condições meteorológicas regionais na interpretação desses resultados.

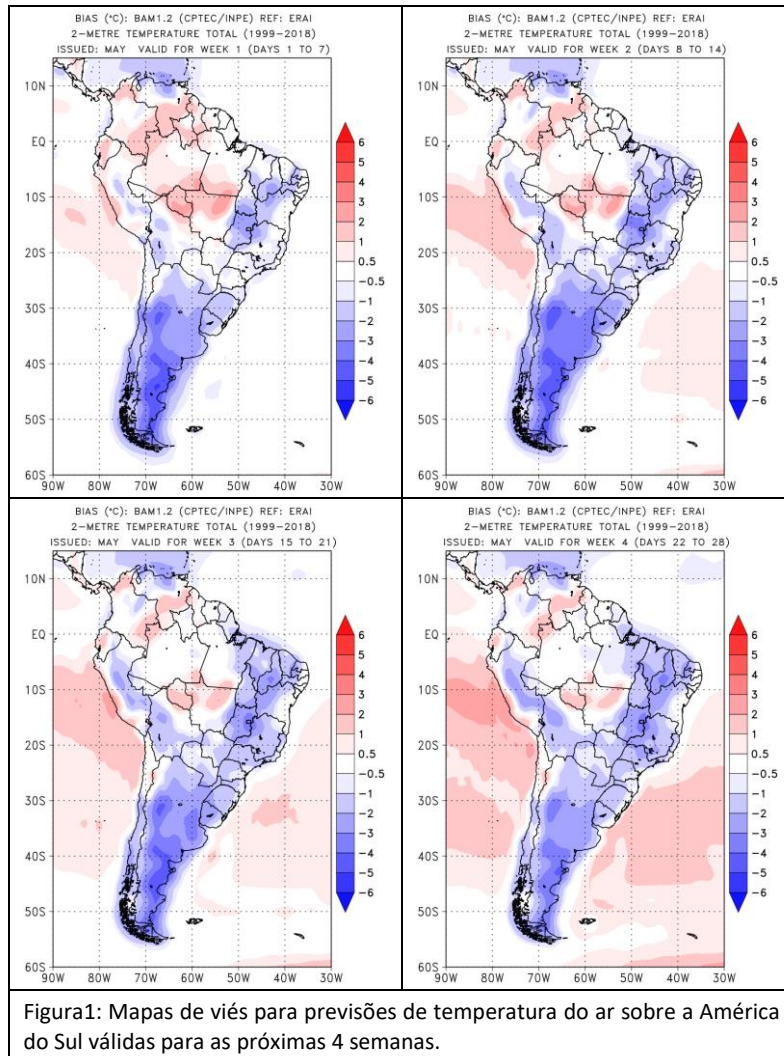


Figura1: Mapas de viés para previsões de temperatura do ar sobre a América do Sul válidas para as próximas 4 semanas.

- A) Explique a aplicação do Viés, do MAE, da RMSE e do Coeficiente de Correlação na avaliação de modelos de previsão numérica e diga como é feita a interpretação conjunta dessas métricas.
- B) Explique como essas métricas podem ser calculadas a partir dos dados fornecidos por previsões numéricas de tempo. Use a tabela abaixo para demonstrar os cálculos:

Previsão (P)	Observação (O)
20.5	21.0
22.0	21.5
19.0	18.5
23.5	22.0
21.0	21.5



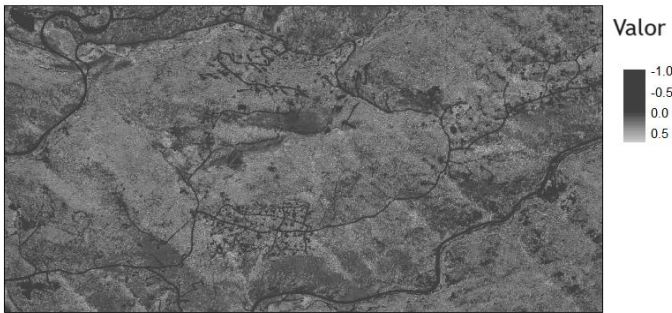
## Questão 2

O índice de vegetação da diferença normalizada (*Normalized Difference Vegetation Index-NDVI*) quantifica a densidade da vegetação capturada por uma imagem de satélite, entre outras coisas.

Ele é definido como segue:  $NDVI = (NIR-RED)/(NIR+RED)$ , em que NIR representa a intensidade da luz refletida na faixa infravermelha e RED a intensidade da luz refletida na faixa vermelha.

O NDVI assume valores no intervalo  $[-1, 1]$  tais que valores próximos de  $-1$  indicam nuvem, água e neve, enquanto que valores próximos de  $1$  representam evidências de florestas tropicais. A Figura 1(a) mostra uma imagem de NDVI em nível de cinza do local de incêndio conhecido como *Cold Springs Fire*, perto do Colorado na Holanda. A Figura 1(b) apresenta (i) um gráfico de dispersão entre 50 valores de NDVI e 50 valores de intensidade da luz na faixa vermelha e (ii) um modelo de regressão linear ajustado.

(a) Imagem NDVI (em nível de cinza)



(b) Ajuste do modelo de regressão linear

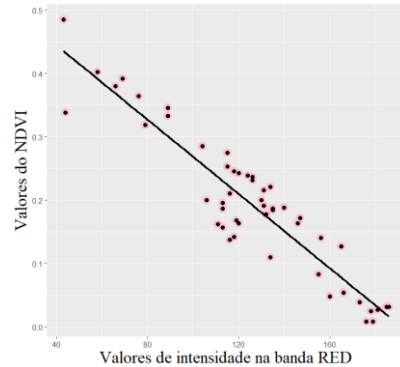


Figura 1. Uso de modelo de regressão linear para NDVI.

Considere o seguinte modelo de regressão linear: Para os valores do subíndice  $i=1, \dots, 50$ ,

$$Y_i = \beta_0 + \beta_1 x_i + E_i, \tag{1}$$

em que  $Y_i$  representa uma variável aleatória para descrever o NDVI no  $i$ -ésimo ponto selecionado a partir da Figura 1(a),  $x_i$  indica uma variável fixada representando o  $i$ -ésimo valor na intensidade da luz na faixa vermelha,  $E_i$  é uma variável aleatória representando um erro tal que seu valor esperado é zero e sua variância é  $\sigma^2 > 0$  e  $\beta_0 \in \mathbb{R}$  e  $\beta_1 \in \mathbb{R}$  são os coeficientes de regressão. Assuma para o Modelo (1) que a covariância entre  $E_i$  e  $E_j$  é nula para quaisquer  $i$  e  $j$  distintos.

Com base nessas informações, responda ao que se pede a seguir.

- A) Em Estatística, "parâmetro" pode ser definido como uma quantidade numérica desconhecida na população que descreve um aspecto populacional (por exemplo, dada uma variável aleatória, as expressões para sua média e variância). Quais os parâmetros relacionados ao Modelo (1)?
- B) Quais as expressões para o valor esperado e a variância de  $Y_i$  dado  $x_i$ ?
- C) Em Estatística, "estimador" pode ser definido como uma função de valor real de uma amostra aleatória usada para obter o parâmetro (por exemplo, dada uma amostra aleatória, versões amostrais para média e variância). Mostre que os estimadores de mínimos quadrados ordinários para  $\beta_0$  e  $\beta_1$  são dados por, respectivamente:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x} \quad \text{e} \quad \hat{\beta}_1 = S_{xy}/S_{xx},$$

em que  $S_{xy} = (\sum_{i=1}^{50} x_i Y_i) - \frac{1}{50} (\sum_{i=1}^{50} x_i) (\sum_{i=1}^{50} Y_i)$ ,  $S_{xx} = (\sum_{i=1}^{50} x_i^2) - \frac{1}{50} (\sum_{i=1}^{50} x_i)^2$ ,  $\bar{x} = \frac{1}{50} \sum_{i=1}^{50} x_i$  e  $\bar{Y} = \frac{1}{50} \sum_{i=1}^{50} Y_i$ . Faça isso resolvendo o sistema abaixo:

$$\begin{aligned} \left( \sum_{i=1}^{50} x_i Y_i \right) &= \hat{\beta}_0 \left( \sum_{i=1}^{50} x_i \right) + \hat{\beta}_1 \left( \sum_{i=1}^{50} x_i^2 \right), \\ \sum_{i=1}^{50} Y_i &= 50 \hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{50} x_i. \end{aligned}$$

- D) O modelo ajustado correspondente à Figura 1(b) é dado por  $\hat{Y}_i = 0,5616 - 0,0029x_i$ . Como você interpreta a estimativa para um dos coeficientes de regressão dada por  $\hat{\beta}_1 = -0,0029$ ?



### Questão 3

Um modelo para a previsão meteorológica de temperatura para dois dias foi proposto. As frequências de previsões e observações de uma amostra aleatória das previsões durante dez anos foram registradas. As categorias de resposta para a previsão são: "abaixo do normal", "perto do normal" e "acima do normal" e os resultados estão descritos na Tabela 1 a seguir.

**Tabela 1** – Frequências relativas  $p_{ij}$  em percentual (tamanho total da amostra,  $n=650$ ) para previsões de temperatura de sete dias em um certo país em dez anos.

Previsão	Observação			Distribuição marginal
	abaixo do normal	perto do normal	do acima do normal	
abaixo do normal	15	6	$p_{13}$	$p_1^*$
perto do normal	5	23	16	44
acima do normal	$p_{31}$	13	20	34
Distribuição marginal	$p_1$	42	37	100

Alguns cálculos fornecidos:

$21/22 = 0,95$ ;  $22/21 = 1,05$ ;  $15/21 = 0,71$ ;  $21/15 = 1,4$ ;  $23/42 = 0,55$ ;  $20/37 = 0,54$ ;  $44/42 = 1,05$ ;  $34/37 = 0,92$ ;

A) Calcule  $p_1^*$ ,  $p_1$ ,  $p_{13}$  e  $p_{31}$ .

B) A acurácia de um modelo é a proporção de previsões corretas, e pode ser utilizada para comparar modelos pelas proporções de acertos. A literatura de meteorologia aponta que nos modelos atuais para previsões de temperatura para dois dias é comumente alta. Apresente a expressão para o cálculo da acurácia do modelo em função dos termos  $p_{ij}$ , calcule a acurácia do modelo em estudo e conclua se pode ser considerada alta ou baixa.

C) O viés (bias) e a probabilidade de detecção (POD) de cada categoria de resposta são definidos, respectivamente, por  $bias_i = p_i^*/p_i$  e  $POD_i = p_{ii}/p_i$ .

Estime o viés e a probabilidade de detecção de cada categoria de resposta. Ordene de forma decrescente os vieses e probabilidade de detecção.

D) Cite pelo menos outras duas formas de avaliar o modelo de previsão apresentado.

E) Um intervalo de confiança aproximado para a proporção de previsões corretas é dado por

$$[ p^* - 1,96 \sqrt{p^* (1-p^*) / n} , p^* + 1,96 \sqrt{p^* (1-p^*) / n} ] = [0,542; 0,618]$$

Você diria que, estatisticamente, a probabilidade de previsões corretas poderia ser considerada 60%? Justifique.



## Questão 4

---

O Fortran 90/95/2003 trabalha com variáveis do tipo: REAL, INTEGER, LOGICAL, CHARACTER e COMPLEX. Com relação às variáveis do Fortran 90/95/2003, responda ao que se pede a seguir.

- A) Para que serve o comando **IMPLICIT NONE**?
- B) Algumas linguagens de programação, como o Shell scripts, não possuem a variável **LOGICAL**. Explique o que é essa variável, descrevendo quais são os possíveis valores que ela pode assumir, informando qual a principal vantagem de se utilizar uma variável lógica em vez de inteira ou caracter e dê um exemplo de uma situação em que é adequado a utilização de uma variável lógica.
- C) Informe como se deve declarar, no Fortran 90/95/2003, uma variável chamada “nome\_competo”, que irá armazenar o nome completo de uma pessoa.
- D) Qual função do Fortran 90/95/2003 pega a parte inteira de uma variável real, aproximando para o inteiro mais próximo?
- E) Se definirmos “a” e “b” como sendo variáveis reais e “c” como variável inteira, qual será o valor atribuído para “a”, “b”, e “c” ao executar no Fortran as seguintes expressões, na sequência apresentada? Faça um breve comentário sobre o valor atribuído a cada variável:
- a=3/2  
b=2.0  
b=b\*\*3  
c=sqrt(16.0) + 8.5/10





## Questão 5

---

A questão completa consiste em usar o programa `cdo` para concatenar arquivos de saída de um modelo de previsão tempo, que estão no formato `netCDF4`, escrever um programa em python para ler os arquivos concatenados, gerar uma série temporal e em seguida plotar a série temporal na tela.

Cada arquivo de saída do modelo contém uma variável bidimensional (latitude, longitude) para uma data de previsão. O nome do arquivo está no formato `"var_TYPE_YYYYMMDDHH.nc"`, onde `TYPE` é uma string que identifica a variável e `YYYYMMDDHH` uma string que representa o ano (YYYY), mês (MM), dia (DD) e hora (HH) da previsão (ex: `"var_t2m_2024013123.nc"`).

Não é necessário implementar os programas solicitados nos itens, mas códigos terão pontuação adicional em todos os itens. Considere as versões do Python 3.6 em diante, as versões do `cdo` 1.9.3. em diante e que todos os códigos dos itens serão implementados em um único arquivo Python.

- A) Escreva uma chamada ao programa `"cdo"` para concatenar todos os arquivos do mês de janeiro de 2024, para a variável `"t2m"`, e outra chamada para a variável `"rainc"`. Os dois arquivos resultantes deverão ser nomeados como `"var_t2m.nc"` e `"var_rainc.nc"`. A ordem dos arquivos a serem concatenados pode ser ignorada.
- B) Descreva como você implementaria uma função em Python para ler um arquivo concatenado do Item 1 e retornar o arquivo na saída da função. Parâmetro de entrada: `nome_do_arquivo` (tipo string). Retorno da função: `file (netCDF4.Dataset)`
- C) Descreva como você implementaria uma função em Python para ler o arquivo que seria retornado pela função descrita no Item 2, gerar uma série temporal e plotar a série na tela, usando os parâmetros de entrada: `nome_do_arquivo` (tipo string), `nome_da_variavel` (tipo string), `latitude` (tipo float), `longitude` (tipo float). A série temporal deverá ser criada com base nos parâmetros fornecidos.
- D) Descreva como você implementaria um programa em Python para gerar e plotar uma série temporal para cada uma das variáveis `"t2m"` e `"rainc"`, usando os arquivos `"var_t2m.nc"` e `"var_rainc.nc"` para a `latitude=-22.67` e `longitude=-44.99`, usando a função do Item 3.



Realização

