



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

**APLICAÇÃO DA LINGUAGEM DE PROGRAMAÇÃO
PYTHON NO ESTUDO DO DESEMPENHO DE
MODELOS NUMÉRICOS DE TEMPO E CLIMA A
PARTIR DA APLICAÇÃO DO SOFTWARE SCAMTEC**

Theodoro Ilha Görck

Relatório Final de Bolsa de Iniciação Científica, orientado por João Gerd Zell de Mattos

URL do documento original:
<<http://urlib.net/xx/yy>>

INPE
São José dos Campos
15 de Setembro de 2020

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3945-6923/6921

Fax: (012) 3945-6919

E-mail: <pubtc@sid.inpe.br>

**COMISSÃO DO CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO
DA PRODUÇÃO INTELECTUAL DO INPE (DE/DIR-544):****Presidente:**

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Membros:

Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)

Dr. Amauri Silva Montes - Coordenação Engenharia e Tecnologia Espaciais (ETE)

Dr. André de Castro Milone - Coordenação Ciências Espaciais e Atmosféricas
(CEA)

Dr. Joaquim José Barroso de Castro - Centro de Tecnologias Espaciais (CTE)

Dr. Manoel Alonso Gan - Centro de Previsão de Tempo e Estudos Climáticos
(CPT)

Dr^a Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

Clayton Martins Pereira - Serviço de Informação e Documentação (SID)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Simone Angélica Del Duca Barbedo - Serviço de Informação e Documentação
(SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

EDITORAÇÃO ELETRÔNICA:

Marcelo de Castro Pazos - Serviço de Informação e Documentação (SID)

André Luis Dias Fernandes - Serviço de Informação e Documentação (SID)



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

**APLICAÇÃO DA LINGUAGEM DE PROGRAMAÇÃO
PYTHON NO ESTUDO DO DESEMPENHO DE
MODELOS NUMÉRICOS DE TEMPO E CLIMA A
PARTIR DA APLICAÇÃO DO SOFTWARE SCAMTEC**

Theodoro Ilha Görck

Relatório Final de Bolsa de Iniciação Científica, orientado por João Gerd Zell de Mattos

URL do documento original:
<<http://urlib.net/xx/yy>>

INPE
São José dos Campos
15 de Setembro de 2020



Esta obra foi licenciada sob uma Licença Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada.

This work is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported License.

LISTA DE FIGURAS

	<u>Pág.</u>
3.1 Gráfico mostrando a diferença entre a previsão e dados observados	6
4.1 Gráfico da Correlação de Anomalia para dois experimentos com modelos atmosféricos, v2.0.0 e v2.1.0	8
4.2 Idêntico à Fig. 4.1, mas com modificações nas opções gráficas.	9
4.3 Interface gráfica teste desenvolvida em Python com o programa Jupyter Notebook	10
4.4 Tutorial para plotagem dos resultados do SCANTEC feito em Jupyter Notebook	11
4.5 Tutorial para plotagem dos resultados do SCANTEC feito em Jupyter Notebook	11
4.6 Tutorial para plotagem dos resultados do SCANTEC feito em Jupyter Notebook	12

SUMÁRIO

	<u>Pág.</u>
1 Introdução	2
2 Objetivo	2
3 Atividades Desenvolvidas durante o período da bolsa	2
3.1 Softwares, linguagens e bibliotecas	3
3.1.1 Python	4
3.1.2 TKinter	4
3.1.3 Pandas	4
3.1.4 Matplotlib	4
3.1.5 Jupyter Notebook	4
3.2 Modelos de Previsão Numérica	4
3.3 Análise e métricas estatísticas	5
3.3.1 Skill Score (SS)	5
3.3.2 Raíz do Erro Quadrático Médio (RMSE)	6
3.3.3 Erro Absoluto Médio (MAE)	7
3.3.4 Coeficiente de Correlação de Anomalia (ACC)	7
4 Resultados obtidos em função do trabalho proposto	8
5 Conclusões Gerais	12
REFERÊNCIAS BIBLIOGRÁFICAS	13

RESUMO

O Sistema Comunitário de Avaliação de Modelos Numéricos de Tempo e Clima (SCANTEC) é um sistema desenvolvido para comparar os resultados de diversos tipos de modelos de Previsão Numérica de Tempo e Clima (PNTC) com diferentes resoluções, através de valores adequados de referência, para o cálculo de métricas estatísticas. Esse cálculo é feito de forma padronizada, visando identificar deficiências e ressaltar os aspectos significativos da destreza dos modelos numéricos avaliados.

Dessa maneira, o SCANTEC apresenta uma interface gráfica, extremamente necessária para que usuários não desenvolvedores possam ter um controle maior sobre a utilização do *software*, escrita em *Java*. No entanto, *Java* não é uma linguagem muito utilizada na comunidade acadêmica e científica o que torna a manutenção e aperfeiçoamento do *software* uma tarefa difícil.

Dessa forma, se faz necessário a criação de uma interface mais moderna e que possa ser melhor utilizada pelos desenvolvedores do CPTEC. Sendo assim, a proposta do atual projeto é desenvolver uma nova interface utilizando a linguagem *Python*, que já está bastante consolidada na comunidade acadêmica e científica.

1 Introdução

O SCANTEC é um sistema baseado em ferramentas livres, podendo ser executado em diversos sistemas operacionais tais como *UNIX*, *Linux*, *Windows* e *MacOS*. O *framework* do sistema é composto por um *kernel* (um núcleo central) totalmente modular para facilitar a implementação de novas funcionalidades (SAPUCCI et al.,). Essa estrutura é desenvolvida em linguagem de programação Fortran 90, seguindo o padrão ANSI, embora haja diversos processos (principalmente ligados à programação orientada à objeto) que exploram a linguagem de programação C (compiladores GCC) (SAPUCCI et al.,). Esse *software* é aplicado para eliminar ou minimizar metodologias de avaliação não padronizadas que impossibilitam a comparação dos indicadores de desempenho e que muitas vezes geram conclusões equivocadas. O sistema oferece aos usuários uma ferramenta completa, integrada, otimizada e consistente para a avaliação de modelos de PNTC, a qual possa ser utilizada livremente por iniciantes, pesquisadores e profissionais nas áreas de meteorologia e ciências afins.

2 Objetivo

Tendo em vista que a linguagem de programação *Java* não está consolidada no meio acadêmico e científico, este trabalho propõe o uso de uma linguagem de programação moderna e de uso global para a criação de uma nova interface gráfica para o SCANTEC. Com isso, o objetivo principal do presente projeto é acoplar o *software* em uma interface desenvolvida em *Python*, tornando esta ferramenta mais atrativa para novos desenvolvedores do meio acadêmico e científico.

3 Atividades Desenvolvidas durante o período da bolsa

O projeto iniciou-se em agosto de 2019 com previsão de término em agosto de 2020. Os primeiros meses foram utilizados para o aprendizado da linguagem de programação *Python* e para o entendimento de assuntos englobados pelo SCANTEC, mais especificamente modelos numéricos para previsão de tempo, além de métricas e análises estatísticas.

Em seguida, o bolsista passou a trabalhar com as tabelas geradas pelo SCANTEC, analisando os dados e plotando gráficos para entender o funcionamento do programa. Junto à isso, começou-se o estudo da biblioteca *TKinter* que seria utilizada para realizar a interface gráfica do SCANTEC.

Perto do fim do projeto, foi decidido que o bolsista iria trabalhar em conjunto com o desenvolvimento do *software* SCANPLOT (específico para plotagem das tabelas geradas pelo SCANTEC), fazendo uso do programa *Jupyter Notebook*, para a criação de uma interface que, depois de pronta, seria transferida para fora do *Jupyter* e desenvolvida com *TKinter*.

Nas próximas seções serão descritos os *softwares* utilizados, além de uma pequena apresentação sobre modelos numéricos para a previsão de tempo e as métricas estatísticas comumente utilizadas para a avaliação objetiva de tais modelos.

3.1 Softwares, linguagens e bibliotecas

Criada na década de noventa por Guido Van Rossum, a linguagem de programação *Python* (BORGES, 2004), tem como seus principais objetivos a produtividade e a facilidade na programação, destacando o fato dela ser de livre distribuição (sem custo financeiro) e por existir versões funcionais para *Windows*, *Linux* e *Mac* (sendo possível a execução pelo terminal nesses dois últimos). O *Python* contém uma vasta biblioteca com inúmeras subrotinas de diversas aplicações como, por exemplo, *Numpy* e *Scipy* com enfoque matemático, *PyQt* para desenvolvimento de interfaces gráficas, *Matplotlib* e *Plotly* na criação de gráficos e entre milhares de outras. Por esse motivo, a linguagem é uma facilitadora no desenvolvimento de *softwares* em geral.

Apesar de ter se originado nos anos noventa, o *Python* começou a ganhar espaço nos últimos anos, tendo uma comunidade de usuários extremamente ativa e que vem auxiliando o desenvolvimento em diversas áreas de pesquisa, se destacando tanto no meio acadêmico quanto no profissional, sendo tudo isso possível devido às inúmeras vantagens descritas acima.

A escolha das bibliotecas *Python* utilizadas neste trabalho foram em razão do fácil manuseio e a grande quantidade de materiais didáticos, sendo suas rotinas simples e vastamente utilizadas para trabalhos não muito complexos.

Abaixo podem ser vistos os *softwares*, linguagens e bibliotecas utilizadas pelo bolsista durante o projeto.

3.1.1 Python

Python é uma linguagem de programação de alto nível muito utilizada no meio acadêmico e científico e foi aplicada para analisar dados e gerar gráficos, além de que, com o auxílio de bibliotecas, será utilizada para criar a nova interface do SCANTEC.

3.1.2 TKinter

TKinter é uma biblioteca padrão da linguagem *Python* para a criação de interfaces gráficas. Será empregada na criação da interface final do SCANTEC.

3.1.3 Pandas

Pandas é uma biblioteca do *Python* para análise e manipulação de dados. Ela foi utilizada para analisar as tabelas geradas pelo SCANTEC.

3.1.4 Matplotlib

Assim como *pandas*, a *Matplotlib* é uma biblioteca do *Python*. Com ela foi possível gerar gráficos através dos valores das tabelas geradas pelo SCANTEC, facilitando a visualização dos dados.

3.1.5 Jupyter Notebook

O *Jupyter* é uma ferramenta para desenvolvimento de *software* que está sendo utilizada para a criação do SCANPLOT e para a criação dos primeiros protótipos da interface gráfica do SCANTEC.

3.2 Modelos de Previsão Numérica

A principal ferramenta utilizada para as previsões de tempo e de clima são os modelos de previsão numérica conhecidos como **Modelos Numéricos de Circulação Geral** – MCG, que são constituídos por um conjunto de equações físicas descritas em forma numérica e resolvidas com o auxílio de computadores. As equações nos modelos de previsão de tempo e clima são resolvidas em pontos de grade, isto é,

a superfície do planeta é dividida em latitudes e longitudes e a intersecção delas fornece os pontos de grade.

Para uma previsão do futuro, o modelo é iniciado com as condições iniciais observadas e a previsão para aquele momento. Também é informado o intervalo de tempo (resolução temporal previamente determinada), em que será feita a integração das equações até chegar ao final do tempo desejado (por exemplo, 24 h, 48 h, 72 h). As condições da atmosfera observadas são utilizadas apenas na primeira integração, pois nas demais o modelo utilizará a previsão feita anteriormente como condição inicial. Com isso, é elaborado um conjunto de previsões: para algumas horas, um dia ou alguns dias e etc.

3.3 Análise e métricas estatísticas

A medida da exatidão e da habilidade dos Modelos de Previsão Numérica é realizada por meio de avaliações objetivas. Essas avaliações podem ser utilizadas na comparação da performance relativa de diferentes modelos. Segundo (FONSECA, 1998), pode-se classificar as avaliações em dois tipos básicos: aquelas que medem a habilidade das previsões e aquelas que medem o grau de realismo com que uma previsão simulou um dado fenômeno atmosférico.

A escolha de uma única medida não é suficiente para se determinar a exatidão ou a habilidade de uma previsão obtida a partir de um modelo numérico. Deve-se escolher um conjunto que avalia as diferentes variáveis produzidas por um modelo (ANTHES et al., 1989). Além disso, torna-se também necessário subdividir as regiões para que se possa determinar onde o modelo apresenta maior deficiência.

A seguir serão apresentados alguns índices estatísticos que medem o grau de acurácia dos modelos de previsão, os quais serão utilizados nesta pesquisa. Para um melhor entendimento desses índices, será utilizada uma variável qualquer α , em que os índices O , A , P e C representarão, respectivamente, as observações, as análises, as previsões e a climatologia. Os índices I e J representarão o número total de pontos de grade nas direções y e x , respectivamente.

3.3.1 Skill Score (SS)

O *Skill Score* mede a precisão de um conjunto de previsões em relação à um conjunto de previsões padrão. Este índice compara o *score* obtido pela nova previsão com o *score* obtido pela previsão padrão, de referência. O resultado mostrará a porcentagem da habilidade de previsão do novo sistema em relação ao sistema de referência.

Uma porcentagem positiva do *skill score* significa que a previsão teve uma melhora em relação à previsão padrão. Um valor negativo significa que a previsão tem uma precisão inferior à previsão padrão.

A fórmula para calcular o *skill score* é dado por:

$$SS = \frac{A - A_{ref}}{A_{pref} - A_{ref}} \times 100 \quad (3.1)$$

Em que A é o valor da nova previsão, Aref é o valor da previsão padrão e Apref é o valor do que seria uma previsão perfeita. O *score* perfeito está relacionado com previsões perfeitas.

3.3.2 Raíz do Erro Quadrático Médio (RMSE)

Para calcular a Raíz do Erro Quadrático Médio pegamos um modelo e dados de referência (dados observados). A partir de cada ponto de dado (pontos pretos no gráfico) conseguimos calcular a diferença entre o Y do dado e o Y do modelo (pontos brancos no gráfico). A diferença (linha em azul no gráfico) é chamada de erro (*residuals*) e mede o quão distante os pontos de dados estão dos pontos da linha de previsão, ou seja, a magnitude dos erros.

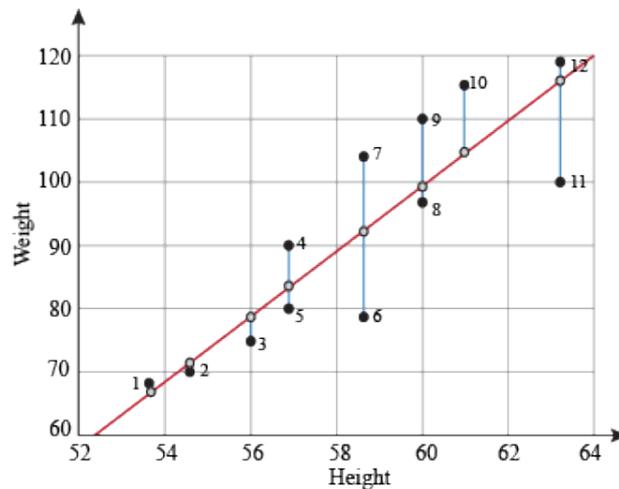


Figura 3.1 - Gráfico mostrando a diferença entre a previsão e dados observados

Para calcular o erro, fazemos a diferença entre o Y do dado e o Y da previsão e elevamos ao quadrado. Somamos todos os erros e fazemos a média. Por fim, tiramos

a raiz quadrada. A fórmula da RMSE pode ser vista abaixo (BARNSTON, 1992):

$$\mathbf{RMSE}(\alpha) = \frac{1}{N} \sum_{n=1}^N \left[\frac{1}{I \cdot J} \sum_{i=1}^I \sum_{j=1}^J (\alpha_{i,j,n}^P - \alpha_{i,j,n}^O)^2 \right]^{\frac{1}{2}} \quad (3.2)$$

Um score perfeito seria RMSE igual à 0. Isto indicaria que todos os pontos de dado estão perfeitamente alinhados sobre a linha de previsão.

3.3.3 Erro Absoluto Médio (MAE)

O cálculo do MAE é relativamente simples. Envolve somar a magnitude dos erros para obter o “erro total” e então dividir este total por n (número de pontos de dado) (WILLMOTT; MATSUURA, 2005). A fórmula pode ser vista abaixo:

$$\mathbf{MAE}(\alpha) = \frac{1}{I \cdot J} \sum_{i=1}^I \sum_{j=1}^J (\|\alpha_{i,j}^P - \alpha_{i,j}^O\|) \quad (3.3)$$

O MAE aumenta com o aumento da diferença entre as observações e as previsões. O cálculo do MAE pode ser utilizado para medir a habilidade de um sistema de previsão sobre o outro por meio da equação do *skill score* (MATTOS,).

Quando as duas métricas, MAE E RMSE são calculadas, o resultado do MAE costuma ser bem menor, já que o RMSE penaliza erros grandes, enquanto o MAE atribui o mesmo peso a todos os erros (WILLMOTT; MATSUURA, 2005).

3.3.4 Coeficiente de Correlação de Anomalia (ACC)

A qualidade de uma previsão pode ser avaliado usando o Coeficiente de Correlação de Anomalia (ACC). O ACC é um *score* adimensional e positivamente orientado definido para “premiar” as melhores previsões dos padrões observados, apresentando pouca sensibilidade aos valores corretos da variável (MATTOS,). A fórmula pode ser vista abaixo:

$$\mathbf{ACC} = \frac{\sum_{i=1}^I \sum_{j=1}^J [(\alpha_{i,j}^P - \alpha^C) \cdot (\alpha_{i,j}^A - \alpha^C)]}{\left\{ \left[\sum_{i=1}^I \sum_{j=1}^J (\alpha_{i,j}^P - \alpha^C)^2 \right] \cdot \left[\sum_{i=1}^I \sum_{j=1}^J (\alpha_{i,j}^A - \alpha^C)^2 \right] \right\}} \quad (3.4)$$

Se o valor do ACC for igual a um, ambos os dados estão correlacionados de forma positiva, o que significa uma qualidade perfeita de previsão. Se o valor for igual

A partir deste gráfico o usuário pode realizar uma série de modificações, tais como ter grade ou não, alterar a cor das linhas e do fundo e tirar os marcadores, as quais podem ser vistas na Figura 4.2. Novas modificações ainda serão adicionadas.

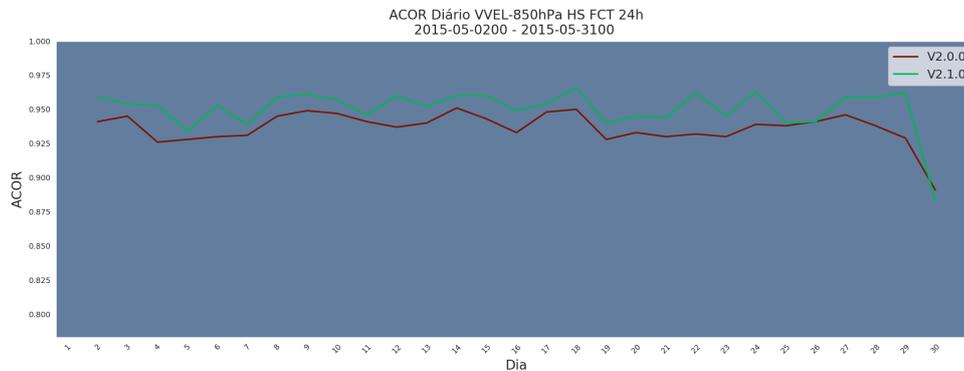


Figura 4.2 - Idêntico à Fig. 4.1, mas com modificações nas opções gráficas.

Durante o desenvolvimento da pesquisa para o uso da interface gráfica do SCAN-TEC, iniciou-se o trabalho de colaboração para o desenvolvimento do SCANPLOT. Este *software* foi uma iniciativa anterior para o desenvolvimento da interface gráfica, porém com o uso do programa *Jupyter Notebook*. Um exemplo deste *software* é visto na Figura 4.3.

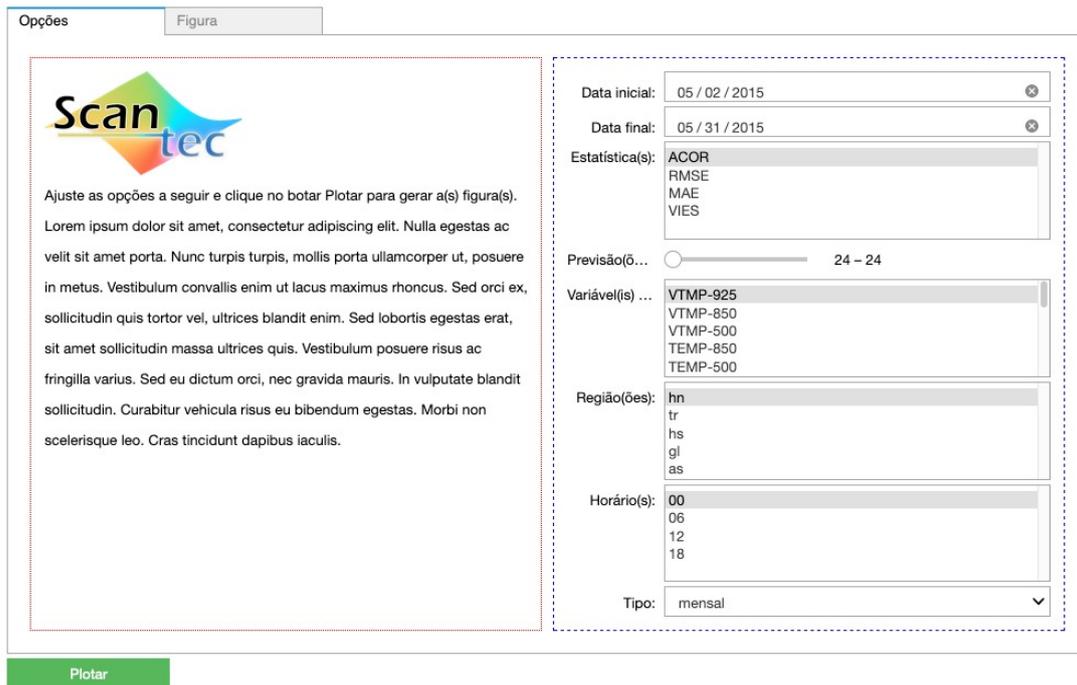


Figura 4.3 - Interface gráfica teste desenvolvida em Python com o programa Jupyter Notebook

O SCANPLOT foi inicialmente concebido para que fosse possível ao usuário do SCANTEC entender como poderiam ser plotados os gráficos estatísticos com o uso da biblioteca *MatPlotLib* do *Python*. Por este motivo, a base do SCANPLOT foi escrita dentro do *Jupyter Notebook*. Assim, o usuário seria capaz de compreender o passo-à-passo para gerar um gráfico estatístico. As Figuras 4.4, 4.5 e 4.6, mostram as páginas iniciais do SCANPLOT. A Figura 4.4 apresenta o texto introdutório que é apresentado ao usuário e indica o início do tutorial para plotagem dos resultados do SCANTEC. Já a Figura 4.5 apresenta ao usuário como alguns *scripts* podem ser carregados em *Python* para a plotagem das estatísticas do SCANTEC. Por fim, na Figura 4.6, é apresentada uma imagem do *Jupyter Notebook* indicando como o usuário pode definir alguns parâmetros para a criação e modificação dos gráficos estatísticos.

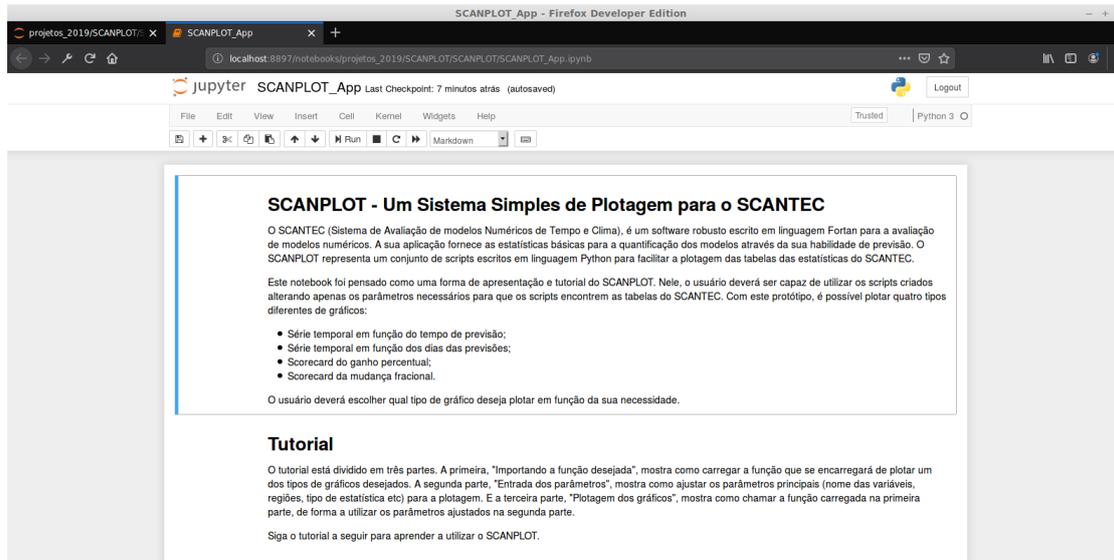


Figura 4.4 - Tutorial para plotagem dos resultados do SCANTEC feito em Jupyter Notebook



Figura 4.5 - Tutorial para plotagem dos resultados do SCANTEC feito em Jupyter Notebook

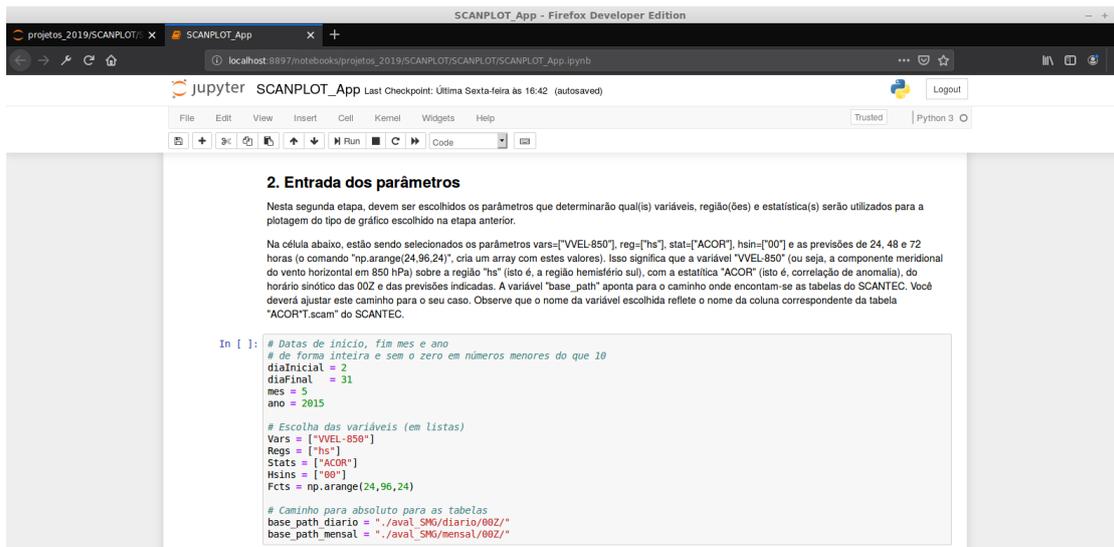


Figura 4.6 - Tutorial para plotagem dos resultados do SCANTEC feito em Jupyter Notebook

Embora o tutorial desenvolvido em *Jupyter Notebook* seja bastante interessante, pois mostra ao usuário o passo-a-passo de como podem ser utilizadas as ferramentas em *Python* para a produção de gráficos do SCANTEC, o usuário ainda acaba gastando um tempo desnecessário para a criação dos gráficos, conseqüentemente perdendo o seu foco que são os resultados estatísticos.

Com isso, foi iniciada a criação da interface gráfica, baseando-se na idéia já apresentada na Figura 4.3, em que o usuário somente necessita indicar onde estão os dados de saída gerados pelo SCANTEC e também de qual variável do modelo numérico que deve ser apresentado os resultados estatísticos. Esta interface já está sendo desenvolvida e utiliza os *scripts* de plotagem desenvolvidos no decorrer do último ano.

5 Conclusões Gerais

A linguagem de programação *Python* utilizada no trabalho se mostrou eficiente e estável no desenvolvimento de aplicações para visualização dos resultados estatísticos fornecidos pelo SCANTEC, se mostrando uma boa alternativa dentre as existentes para esse tipo de aplicação. Além disso, *Python* possui uma curva de aprendizagem bastante suave. Suas estruturas são de altíssimo nível e a grande variedade de bibliotecas prontas facilitaram bastante a programação. Foi visto que com um

tempo considerado relativamente pequeno, foi possível desenvolver aplicações com funções consideradas avançadas. Porém, observa-se que algumas classes e funções ainda possuem pouca documentação.

Desta forma, apesar do projeto não ter sido concluído, o período de vigência da bolsa foi de extrema importância para que o bolsista pudesse aprender todas as ferramentas que seriam utilizadas para o desenvolvimento da interface. Além disso, bons resultados foram alcançados apesar das limitações impostas no decorrer do período de vigência da bolsa, como a implementação de funcionalidades gráficas no SCANTEC que hoje só estão disponíveis na interface escrita em *Java*.

Os resultados alcançados demonstram que a continuação do projeto trará benefícios concretos ao uso do SCANTEC pelo meio acadêmico e científico para a avaliação de modelos meteorológicos.

REFERÊNCIAS BIBLIOGRÁFICAS

ANTHES, R. A.; KUO, Y. H.; HSIE, E. H.; LOW-NAN, S.; BETTGE, T. W. Estimation of skill and uncertainty in regional numerical models. **Quarterly Journal of the Royal Meteorological Society**, v. 115, n. 488A, p. 763–806, July 1989. 5

BARNSTON, A. G. Correspondence among the correlation, rmse, and heidke forecast verification measures; refinement of the heidke score. **Weather and Forecasting**, v. 7, n. 4, p. 699–709, 1992. 7

BORGES, L. E. **Python para Desenvolvedores**. 3^o. [S.l.]: Novatec, 2004. ISBN 978-85-7522-405-2. 3

FONSECA, J. F. B. **Modelos regionais de previsão numérica de tempo: um estudo comparativo das destrezas dos modelos FSU e ETA, sobre a América do Sul**. 98 p. Dissertação (Mestrado em Meteorologia) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Brasil, July 1998. 5

MATTOS, J. G. Z. de. Sensibilidade do uso de “pseudo-temps” na assimilação de dados do modelo de circulação geral atmosférica do cptec/cola. 7

SAPUCCI, L. F.; MATTOS, J. G. D.; MATTOS, A. F. D.; GONÇALVES, L. G. G. D.; BASTARZ, C.; SCHUBERT, D. V. Potencialidades do scamtec (sistema de avaliação de modelos numérico de previsão de tempo e clima) na avaliação de previsões de precipitação. 2

WILLMOTT, C. J.; MATSUURA, K. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. **Climate research**, v. 30, n. 1, p. 79–82, 2005. [7](#)