Neste estágio, estou empenhado em adquirir conhecimentos técnicos para manipular séries temporais, utilizando como base o trabalho de Serpelloni et al. (2018) que abordou séries temporais de GPS. Este relatório detalha os processos e metodologias empregados para refinar as séries temporais, essenciais para a análise e detecção de deslocamentos em uma região de karst, conforme investigado no referido estudo.

Para facilitar a aplicação de modelos e detecção de padrões nas dimensões norte, leste e vertical, optei por escolher estações de GPS próximas entre si. As estações selecionadas foram BRSE, TRIE, PORD, KOPE e MPRA, todas disponíveis no site do Nevada Geodetic Laboratory.

1. **Pré-processamento**

**Avaliação da Estacionariedade**

Inicialmente, utilizei o teste Augmented Dickey-Fuller (ADF) da biblioteca "statsmodels.tsa.stattools" para avaliar estatisticamente a estacionariedade dos dados da série temporal. O objetivo principal é determinar se há evidências estatísticas suficientes para concluir se a série temporal é estacionária ou não. A estacionariedade é crucial para aplicar procedimentos estatísticos, simplificando o processo de modelagem e facilitando a identificação de padrões.

O teste ADF retorna o parâmetro "p-value", utilizado para aceitar ou rejeitar a hipótese nula da estacionariedade, e os "valores críticos" (1%, 5% e 10%) são referências para comparar com o valor do teste estatístico. Ao constatar que a série não é estacionária, rejeitando a hipótese nula, optei por aplicar uma transformação logarítmica nas medidas na tentativa de estabilizar a variância. A escolha por "log1p" (tsLog = np.log1p(instance)) se mostrou mais apropriada, pois lida de forma adequada com valores negativos nas leituras das séries.

**Diferenciação para Estacionarização**

Posteriormente, apliquei a técnica de diferenciação para tornar a série estacionária, subtraindo cada valor da série pelo anterior:

tsLog = tsLog - tsLog.shift()

Outra abordagem seria utilizar a média móvel, subtraindo-a da série. Essa técnica é útil para remover sazonalidade ou padrões de curto prazo. A expressão utilizada para essa abordagem seria:

tsLog = tsLog.rolling(window=7).mean()

O parâmetro "window" refere-se à frequência dos dados.

**Tratamento de Valores Não Nulos (NaN) após Diferenciação**

Ao realizar a diferenciação, é possível que ocorram valores não nulos (NaN). Portanto, é crucial realizar o devido tratamento desses dados, podendo ser removidos ou substituídos pela média. Considerando que essas são as últimas etapas antes do processamento, seria recomendável também a remoção de outliers para garantir que a modelagem seja mais precisa e adequadamente ajustada aos dados.

Esses processos visam preparar as séries temporais para a aplicação eficaz de técnicas analíticas, contribuindo para uma modelagem mais robusta e facilitando a detecção de padrões relevantes.

1. **Modelamento**

Antes de iniciar o modelamento, realizei a decomposição individual da tendência, da sazonalidade e dos resíduos presente na série. Realizar esse procedimento é importante pois a tendência ajuda a entender o movimento dos dados, a sazonalidade ajuda no entendimento dos ciclos, enquanto os resíduos ajudam a capturar padrões não identificados e serve como parâmetro para saber se o meu modelo se adequa ou não aos dados. Portanto, faço uma decomposição inicial para ter como referência para futuras análises.

Ao explorar conceitos de modelagem para séries temporais, deparei-me com o modelo conhecido como Médias Móveis Integradas e Autorregressivas (ARIMA), frequentemente recomendado para análise e previsão em séries temporais. O ARIMA baseia-se na identificação de autocorrelações presentes nos dados em diferentes instantes de tempo, utilizando parâmetros fundamentais: p, d e q para a construção do modelo.

* **Componente Autorregressivo (AR - AutoRegressive):**

O parâmetro *p* no ARIMA está associado ao componente autorregressivo (AR). Esse componente reflete a correlação entre as observações atuais e as observações em períodos temporais anteriores. Em outras palavras, ele modela a relação ou dependência entre diferentes momentos no tempo.

* **Componente de Diferenciação (I - Integrated):**

O parâmetro d*d* refere-se ao componente de diferenciação (I). Ele representa o número de diferenças não sazonais necessárias para tornar a série temporal estacionária. A estacionarização é crucial para garantir propriedades estatísticas constantes ao longo do tempo.

* **Componente de Média Móvel (MA - Moving Average):**

O parâmetro q*q* está associado ao componente de média móvel (MA). Este componente modela a correlação entre as observações atuais e os erros (resíduos) de observações em períodos temporais anteriores.

É importante notar que o processo de modelagem ARIMA envolve a análise de gráficos de Autocorrelação (ACF) e Autocorrelação Parcial (PACF) para identificar os valores apropriados de *p*, *d* e *q*. Sendo, portanto, determinados após à decomposição individual da tendência, sazonalidade e resíduos citados anteriormente.

Essa abordagem oferece uma estrutura robusta para lidar com uma variedade de padrões temporais presentes em séries temporais. Após a produção dos gráficos ACF e PACF, estimei o modelo e obtive os gráficos para a análise da série temporal. Esse procedimento foi aplicado para todas as séries citadas no início do relatório (BRSE, TRIE, PORD, KOPE e MPRA) para a análise nas direções leste, norte e vertical.