



Universidade Estadual de Campinas

ME607

Séries Temporais

Análise de Modelos para Séries Temporais

Gabriel Alvaro Batista RA 171822

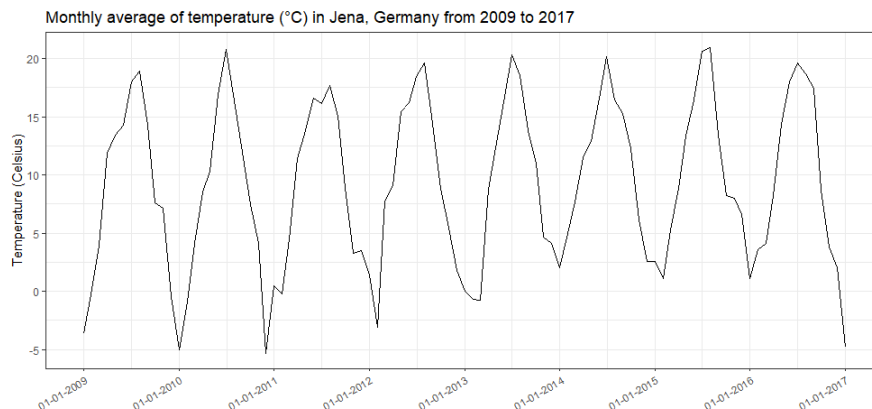
Prof. Carlos Trucios Maza

Campinas - SP
6 de abril de 2022

Análise Exploratória

O conjunto de dados escolhido para a análise dos modelos nos traz observações de variáveis climáticas coletadas todos os dias, com recorrência de 10 minutos, pelo Instituto Max Planck de Biogeoquímica da cidade de Jena, na Alemanha, entre os anos de 2009 e 2016. Para facilitar o tratamento e a análise dos dados e modelos, agrupamos essas observações por mês utilizando a média, e selecionamos somente a variável referente à temperatura.

Os dados originais podem ser acessados pelo [Kaggle](#). O código usado na análise e os dados tratados estão disponíveis no [GitHub](#).



A característica mais marcante dessa série é sua rigorosa sazonalidade, com temperaturas altas próximas ao meio do ano (Verão), e temperaturas amenas no fim (Inverno). Por conta disso, entre os modelos básicos, espera-se que o **Naive Sazonal** demonstre maior eficiência, já que as temperaturas máximas são praticamente constantes e as temperatura mínimas apresentam um leve aumento com o passar dos anos, o que faz com que todos os anos apresentem comportamento muito parecido.

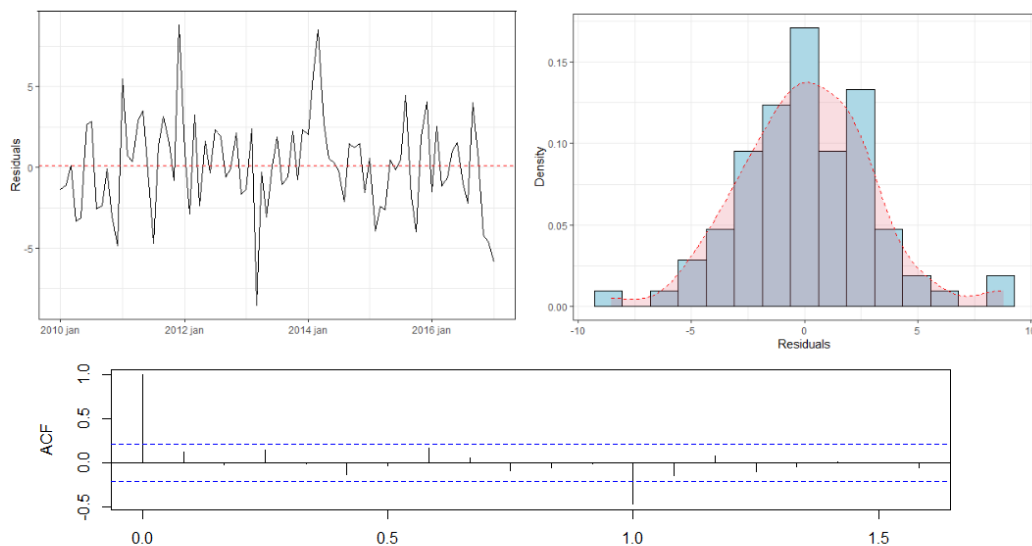
Utilizando cross-validation, podemos avaliar a performance dos modelos para o conjunto de dados escolhido. Na tabela abaixo, temos o **RMSE** (Root Mean Squared Error, métrica utilizada para a avaliação) para previsões h tempos (meses) à frente. A cross-validation foi feita com a técnica *stretching windows*, utilizando tamanho inicial igual a 12. Os modelos avaliados foram: Naive Sazonal (Snaive), Suavização Exponencial com Holt-Winters (ExpHW), Naive, Drift, Média e Regressão Linear (LRM).

h	Snaive	ExpHW	Naive	Drift	Mean	LRM
1	2.92	3.08	4.27	4.33	7.13	7.44
2	2.93	4.20	7.13	7.32	7.16	7.89
3	2.95	5.34	9.73	10.1	7.21	8.25
4	2.97	6.34	11.9	12.5	7.25	8.41
5	2.96	7.02	13.2	14.1	7.22	8.28

Como esperado, o melhor modelo foi o Naive Sazonal para todos os horizontes de previsão. O modelo de suavização exponencial apresentou performance próxima para $h = 1$, porém sua performance decai rapidamente conforme aumentamos os passos.

Diagnóstico do Modelo

O modelo Naive Sazonal foi o que apresentou melhores resultados dentre os modelos testados, porém não sabemos se ele é de fato um bom modelo. Para isso, realizaremos o diagnóstico do modelo, olhando para os resíduos da aplicação no nosso conjunto de dados com $h = 5$.



Pelas duas primeiras figuras, com o gráfico dos resíduos e o histograma, vemos que eles estão bem comportados - a média está bem próxima de zero (0.096) e aparentemente seguindo distribuição Normal. Realizando o teste de Shapiro-Wilk, obtemos $p\text{-valor} = 0.4623$, corroborando a normalidade dos resíduos.

Já no gráfico de baixo, após o lag 0, vemos autocorrelação alta somente no lag = 1, o que nos leva a crer que não existe autocorrelação entre os resíduos e, portanto, possuem independência. Aplicando o teste de Box-Pierce e o teste de Ljung-Box, temos:

Teste	p-valor
Box-Pierce	0.09784
Ljung-Box	0.02493

A análise dos gráficos nos indica que o modelo se adequa bem aos dados, porém, a um nível de significância $\alpha = 0.05$, rejeitamos a hipótese nula no teste de Ljung-Box, o que diz que os resíduos não são independentes.

Como todo o resto da análise nos indica uma boa aderência, é provável que o Naive Sazonal seja um bom modelo para os nossos dados, mas a rejeição no teste de Ljung-Box e a dubiedade do gráfico de autocorrelação faz com que seja necessário realizarmos ajustes nos parâmetros do modelo e talvez até testar outros modelos para garantirmos uma aderência verdadeiramente boa.