



Universidade
Federal de Juiz
de Fora

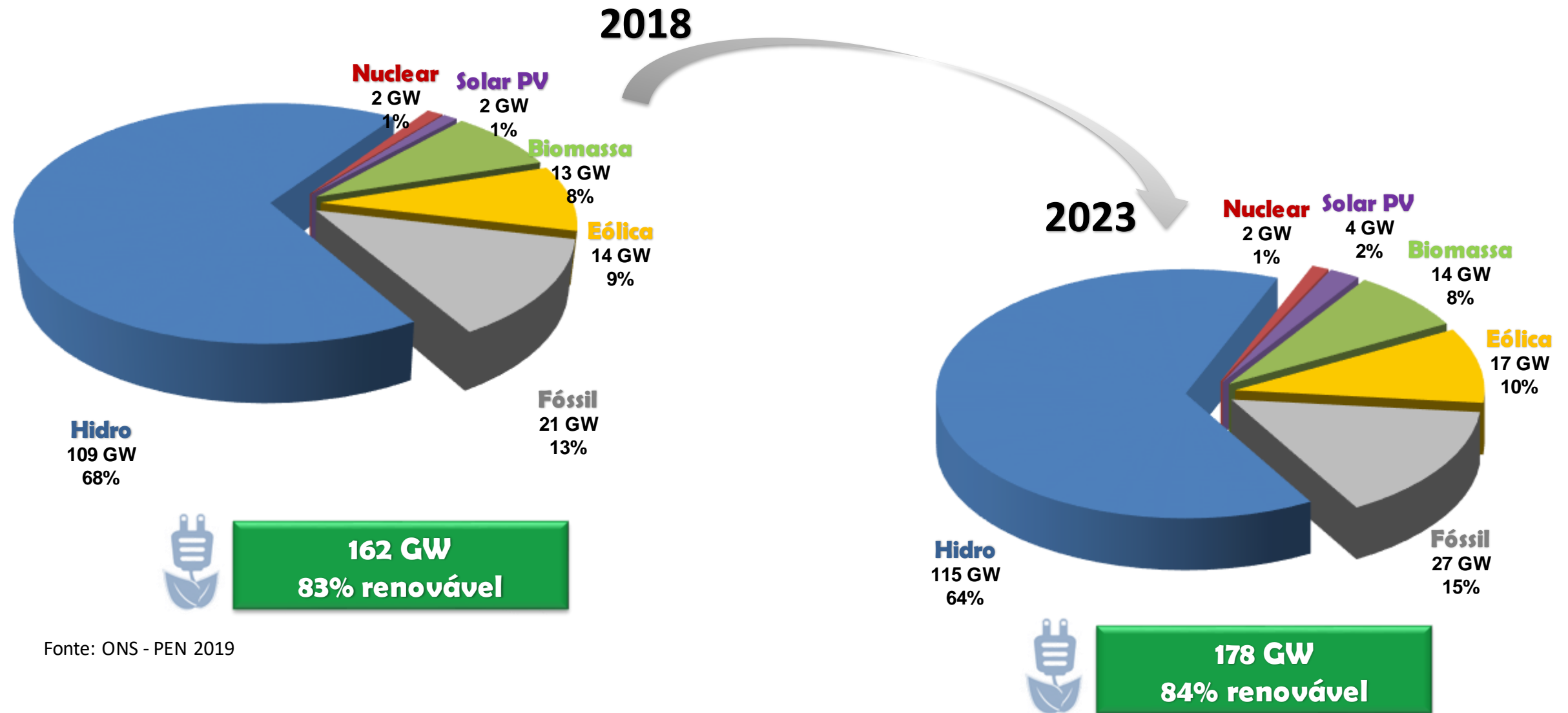
Modelo estatístico para tratamento de *missing e outliers*

Séries temporais de energia eólica

Juiz de Fora, 13 de julho de 2021

Contextualização

Evolução da capacidade instalada por fontes



Fonte: ONS - PEN 2019

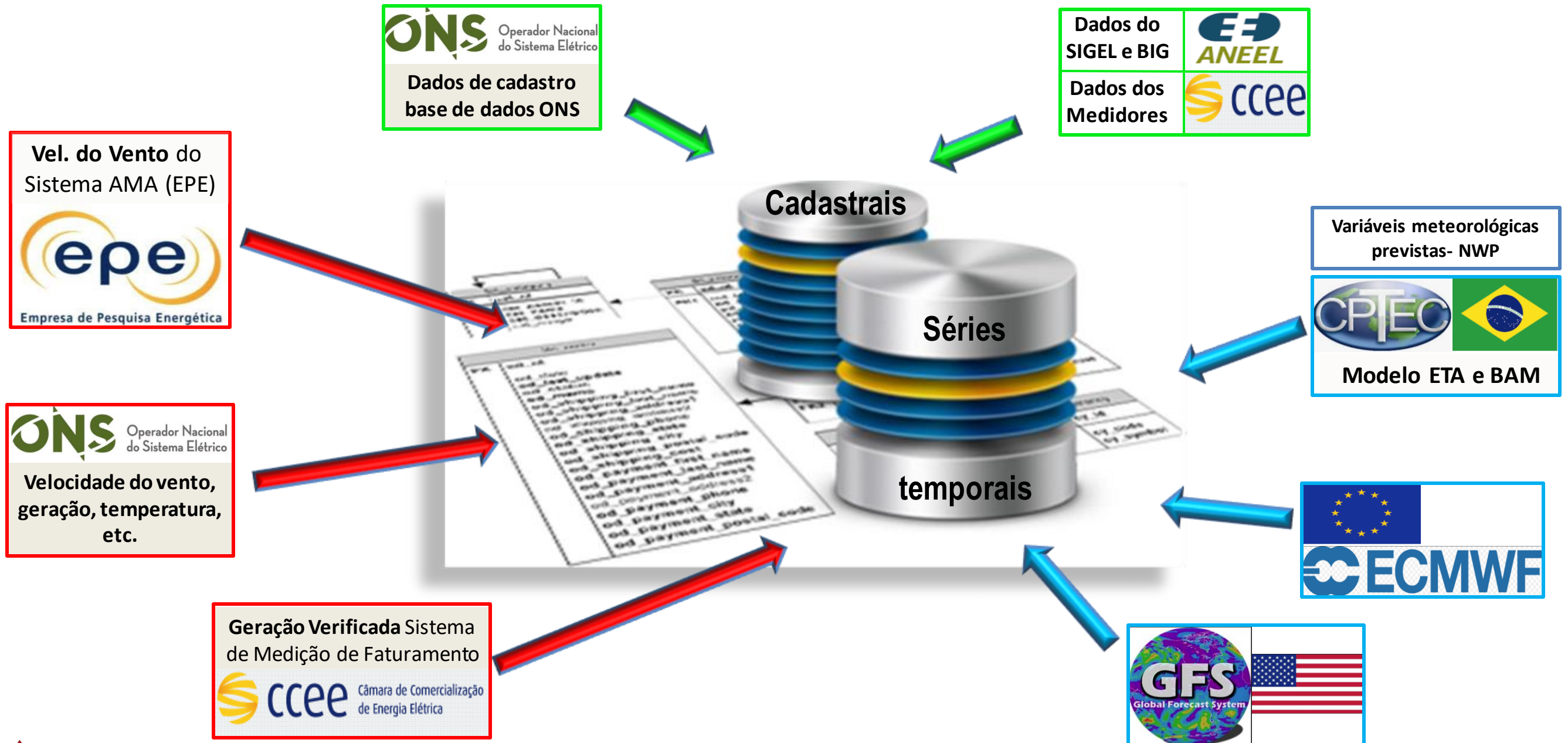
Fonte: ONS - PEN 2019



Universidade
Federal de Juiz
de Fora

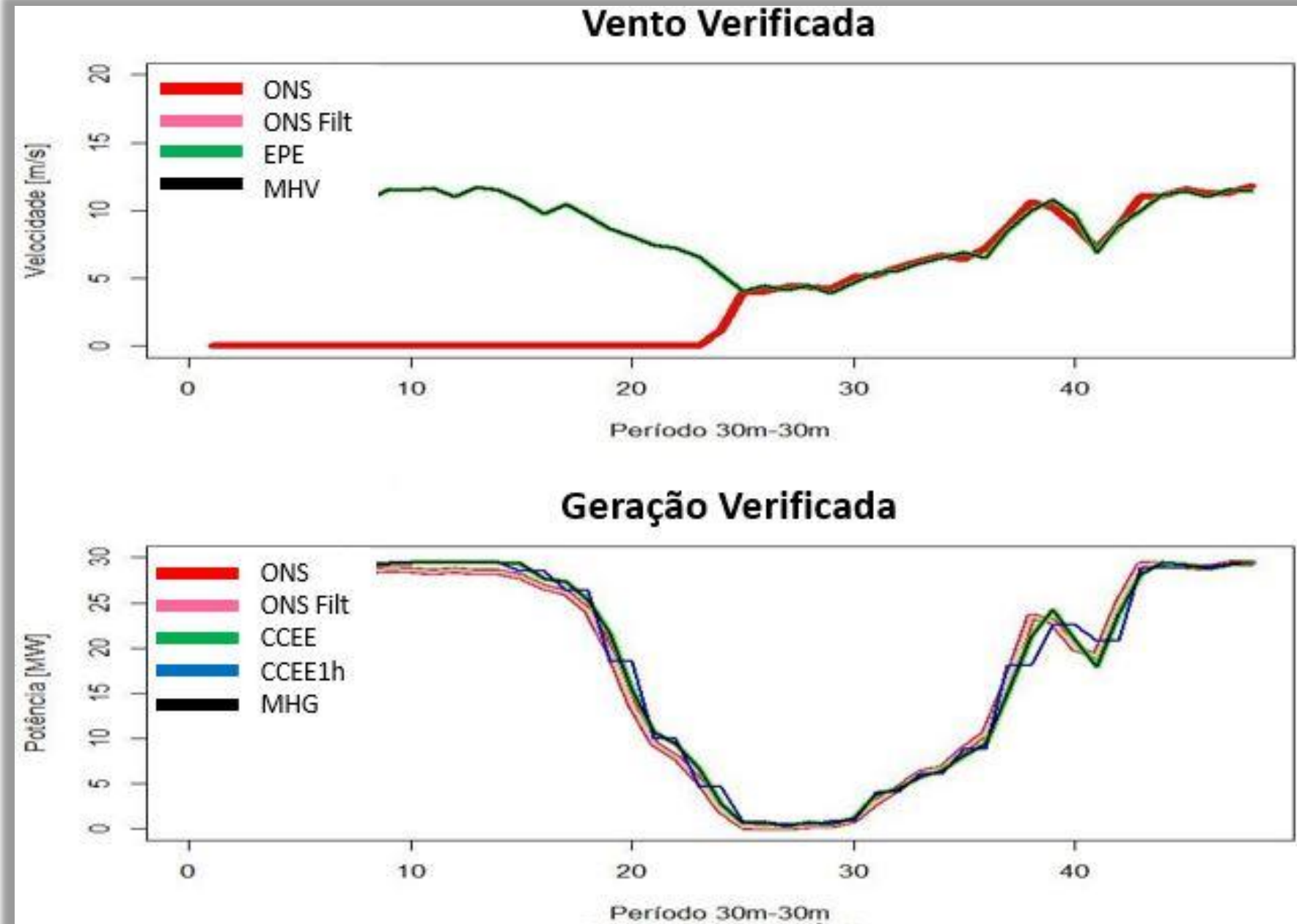
Base de dados e tratamentos iniciais

Base de dados utilizada

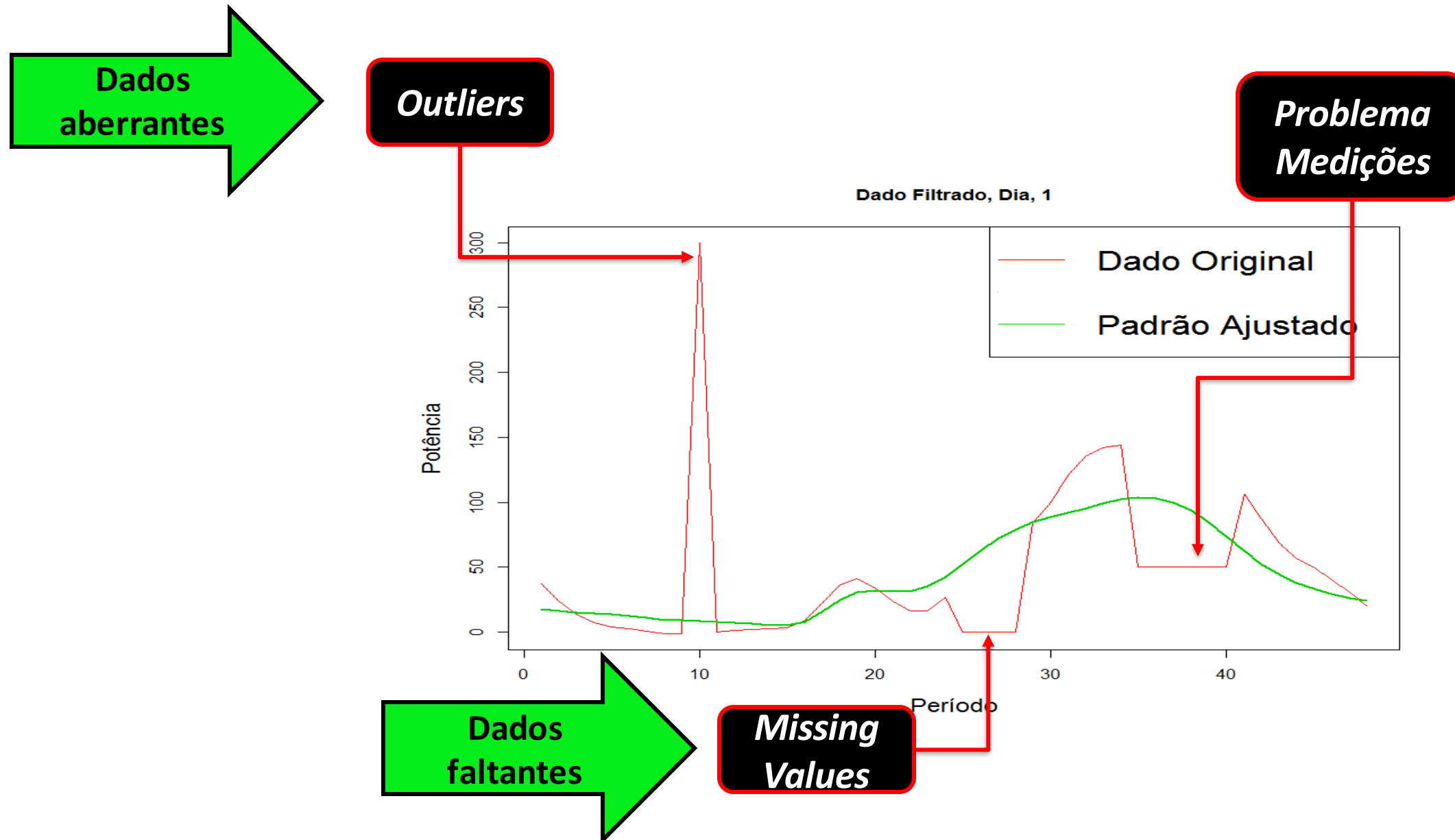


Identificação de falhas de medição

- ✓ Perfil de geração pode ser usado para correção do vento e vice-versa
- ✓ Identificação de *outliers*
- ✓ Algumas falhas podem ser identificáveis
- ✓ Varias fontes de dados pode ser usada

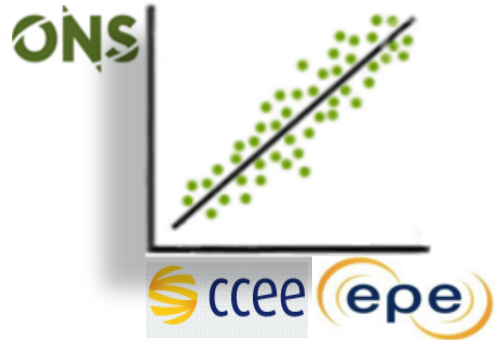


Identificação de falhas de medição



Filtros iniciais dos dados

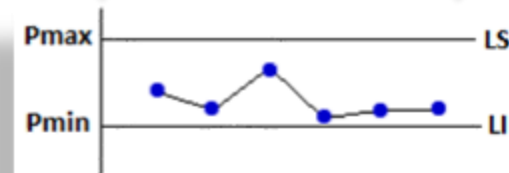
1º Passo
Correlacionar os dados



2º Passo
Eliminar resultados ruins
do Estimador de Estados



3º Passo
 $0\text{MW} < \text{Ger} < 1.1 \cdot P_{\text{inst}}^{\text{MW}}$
 $0\text{m/s} < \text{Ven} < 30\text{m/s}$

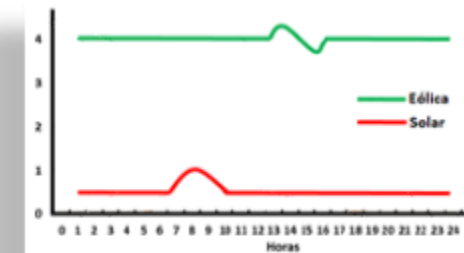


Se a quantidade de dados considerados
inválidos for maior do que 80% do histórico,
todo os dados são descartados

4º Passo
Eliminar dados congelados (Ao
longo dos dias e das horas)



5º Passo
Eliminar dados dos dias com
menos de 2,5 horas válidos



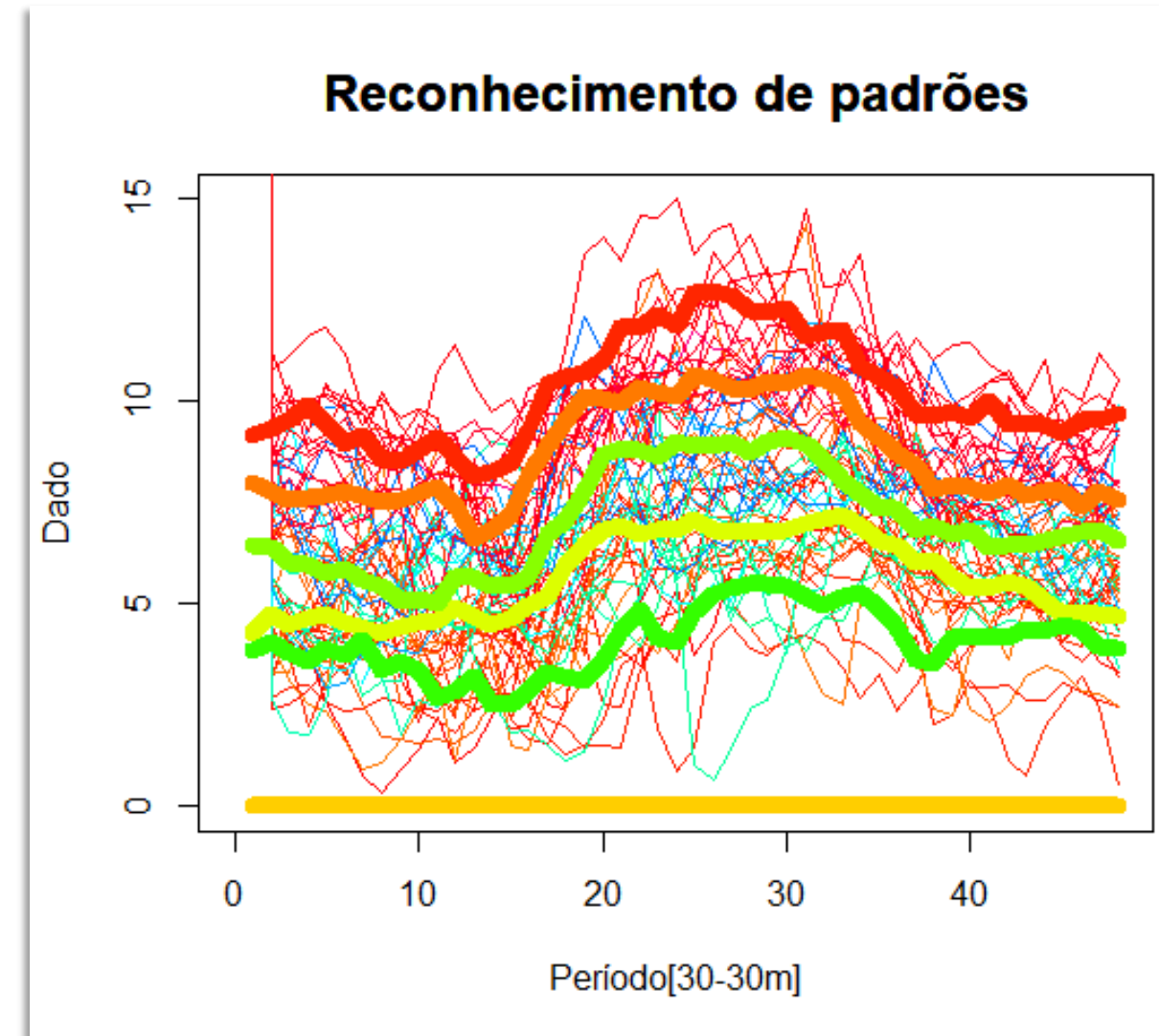


Universidade
Federal de Juiz
de Fora

Modelo de preenchimento de dados faltantes

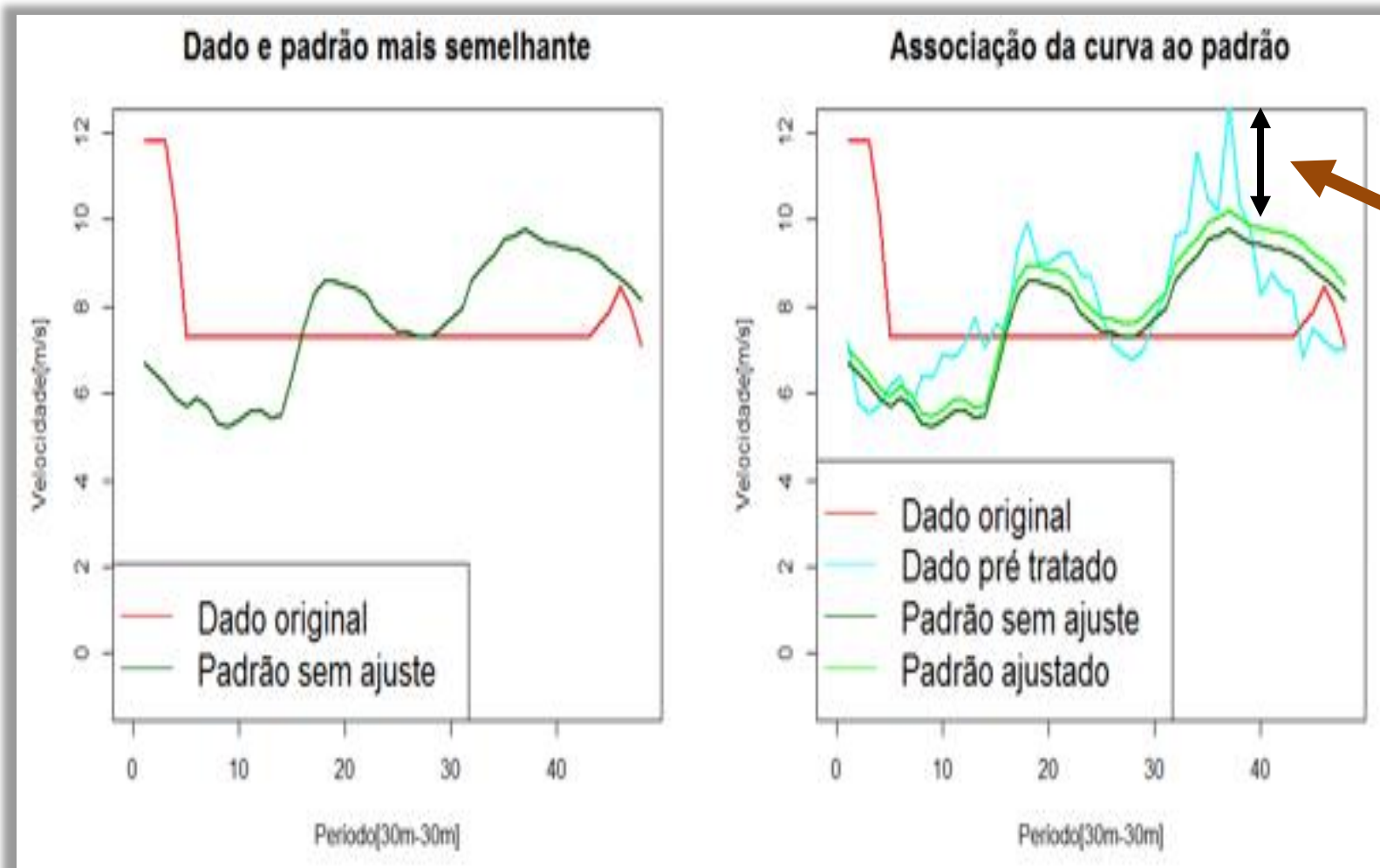
Classificar as curvas de geração e/ou vento e identificar os padrões característicos

- Objetivo: classificar, identificar padrões e estimar variâncias
- Metodologia *K-means*
 - ✓ Alta capacidade de processamento em problemas de grandes dimensões
 - ✓ Algoritmo *machine learning* não supervisionado
 - ✓ Realizam inferências sem necessidades de resultados conhecidos
 - ✓ Atribui dados a um cluster de modo que a soma da distância quadrada seja mínima



Identificação de *outliers*

Associar curva a ser filtrada à classe correspondente



$$dist_k = \sqrt{\sum_{j=1}^{48} (dado_j - Padrao_{j_{ajust}}^k)^2}$$

Identificação dos dados aberrantes e transformação em faltantes

✓ Erro da regressão

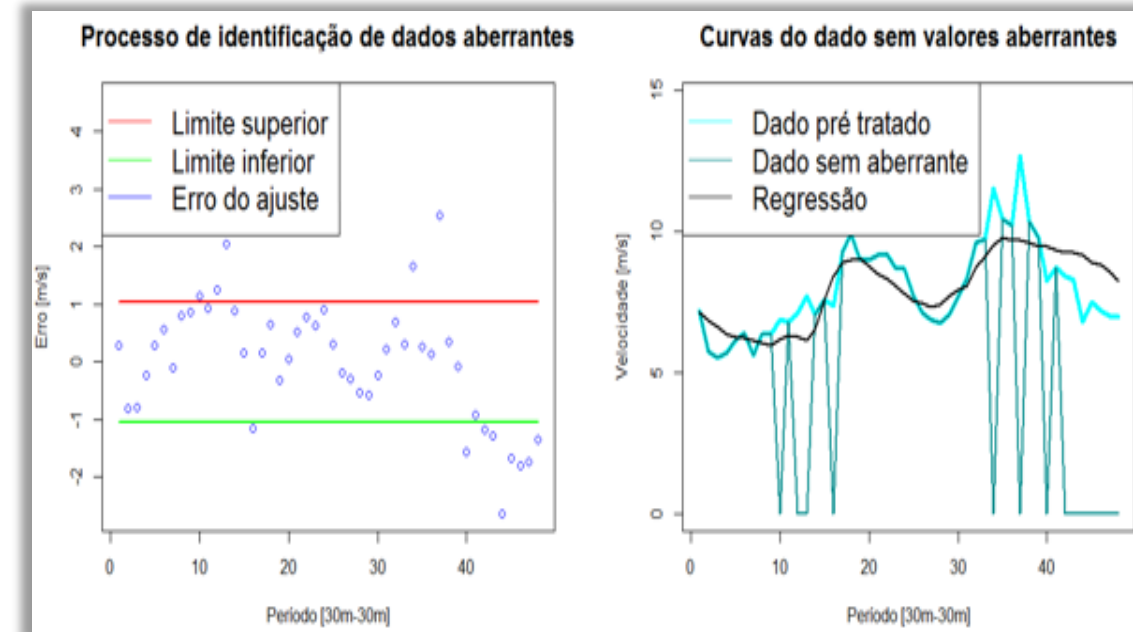
$$Erro_{reta_i} = Dado_{filtrar_i} - Classe_{escolhida_i}$$

✓ Com base na estatística R^2

$$R^2 = 1 - \frac{\sum_{i=1}^n (Dado_{filtrar_i} - \widehat{Dado_{estimado_i}})^2}{\sum_{i=1}^n (Dado_{filtrar_i} - \overline{Dado_{filtrar}})^2}$$

✓ Estima-se limite inferior e superior

$$Lim = \pm desvio_{padr\tilde{a}o}(Erro_{reta}) \cdot Qte_{desvio}$$



Modelo de correção de dados faltantes – filtro de Kalman

Modelo de nível e tendência local

$$\begin{aligned} Y_t &= \mu_t + v_t & v_t &\sim N(0, V_t) \\ \mu_t &= \mu_{t-1} + \beta_t + \omega_{1t} & \omega_{1t} &\sim N(0, W_{1t}) \\ \beta_t &= \beta_{t-1} + \omega_{2t} & \omega_{2t} &\sim N(0, W_{2t}) \end{aligned}$$

Priori em t:

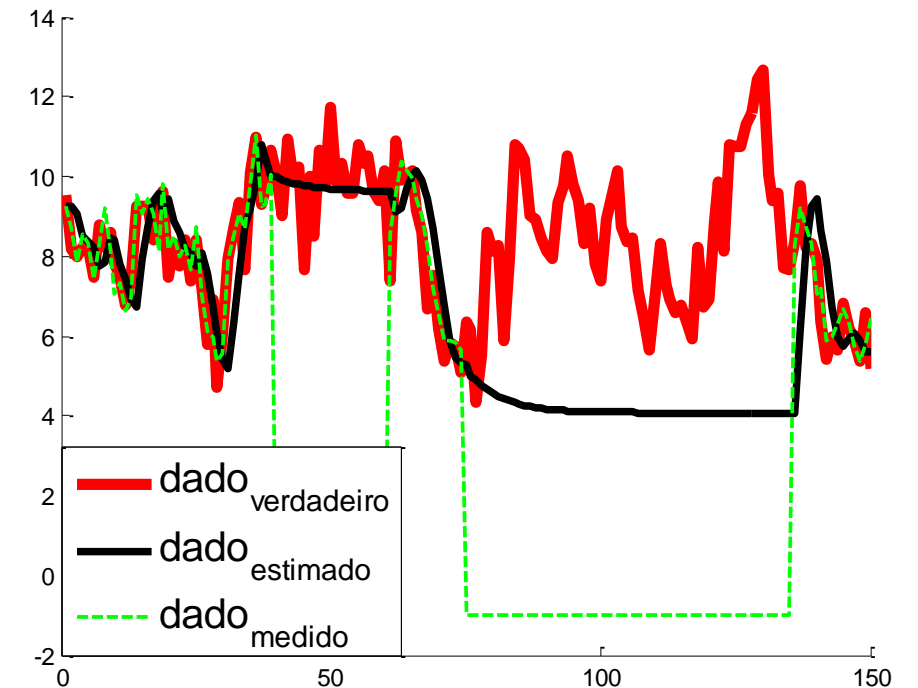
$$\begin{aligned} \theta_t | D_{t-1} &\sim N[a_t, R_t] \\ a_t &= G_t m_{t-1} \\ R_t &= G_t C_{t-1} G'_t + W_t \end{aligned}$$

Posteriori em t:

$$\begin{aligned} Y_t | D_{t-1} &\sim N[f_t, Q_t] \\ f_t &= F'_t a_t \\ Q_t &= F'_t R_t F_t + V_t \end{aligned}$$

Previsão 1-passo à frente:

$$\begin{aligned} \theta_t | D_t &\sim N[m_t, C_t] \\ m_t &= a_t + A_t e_t \\ A_t &= R_t F_t Q_t^{-1} \\ e_t &= Y_t - f_t \\ C_t &= R_t - A_t A'_t Q_t \end{aligned}$$



**Estimação inconsistente
quando não há inovação**

Modelo de correção de dados faltantes

Previsão 1-passo à frente:

$$\theta_t | D_t \sim N[m_t, C_t], \text{ onde:}$$

$$m_t = a_t + A_t e_t$$

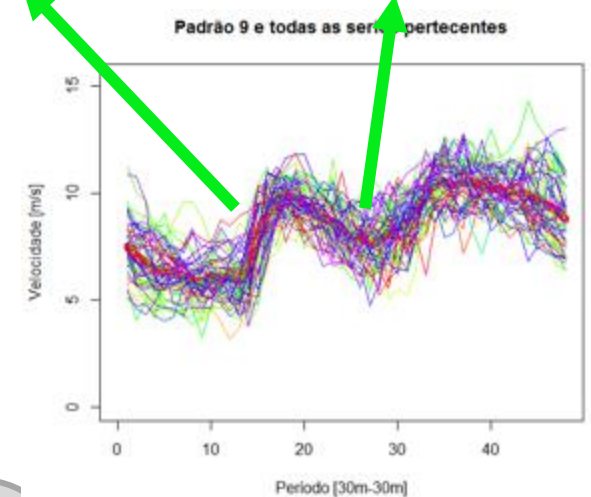
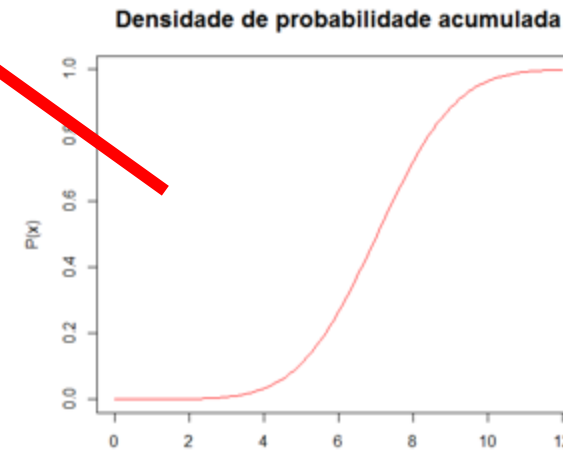
$$A_t = R_t F_t Q_t^{-1}$$

$$e_t = Y_t - f_t$$

$$C_t = R_t - A_t A'_t Q_t$$

A cada iteração calcula-se:

$$Prob = N[Dado_{filtrar}; media = media_{padrao}; \sigma^2 = \sigma_{padrao}^2]$$



Caso o dado Y_1 não for outlier

$$e_t = Y_t^{medido} - f_t$$

Se o dado Y_1 for um outlier

$$e_t = VN \left[Prob_{t-1}; media = media_{t_{padrao}}; \sigma^2 = \sigma_{t_{padrao}}^2 \right] - f_t$$



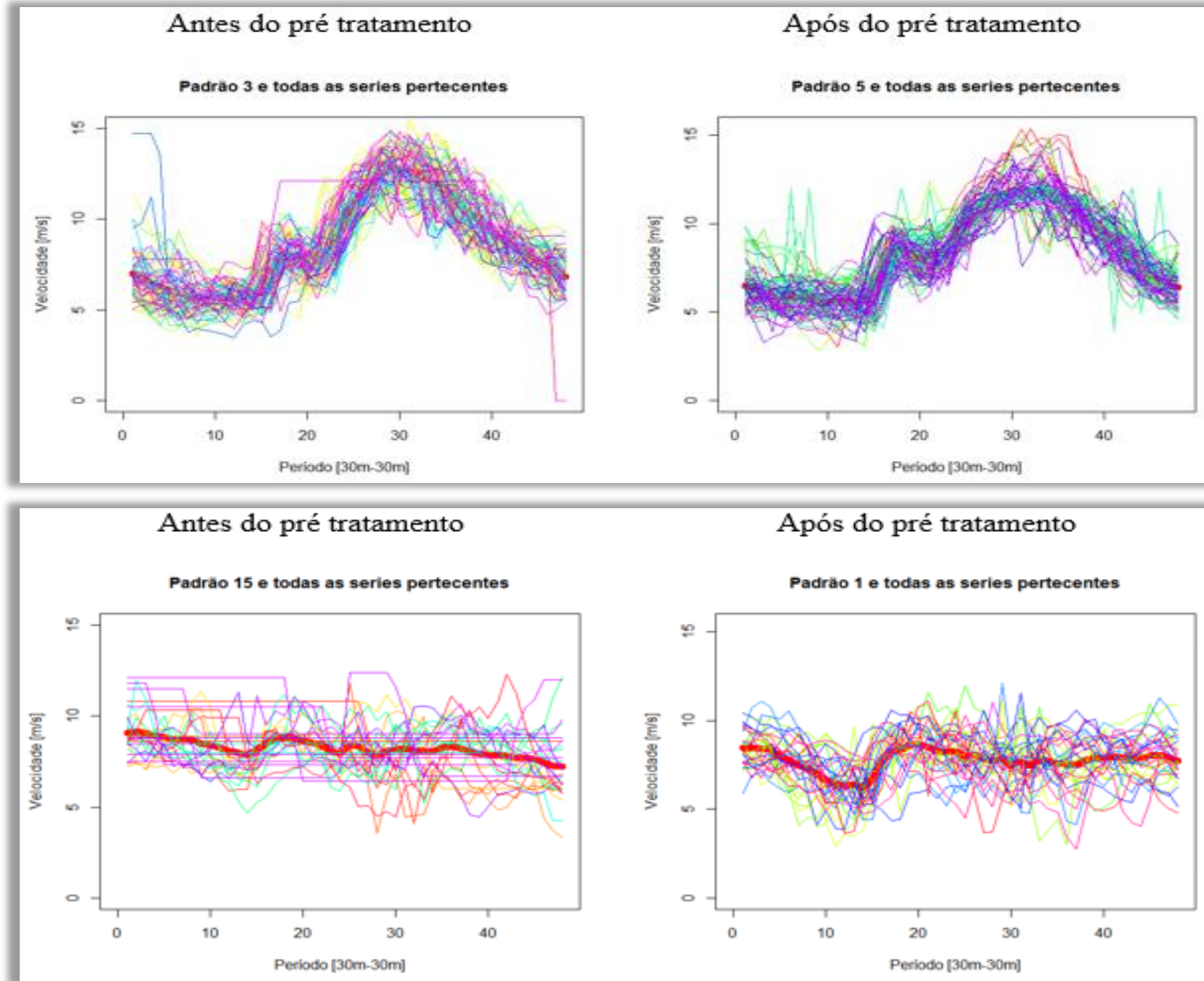
Universidade
Federal de Juiz
de Fora

Resultados do modelo proposto

Apresentação e Discussão dos Resultados

Processo de classificação

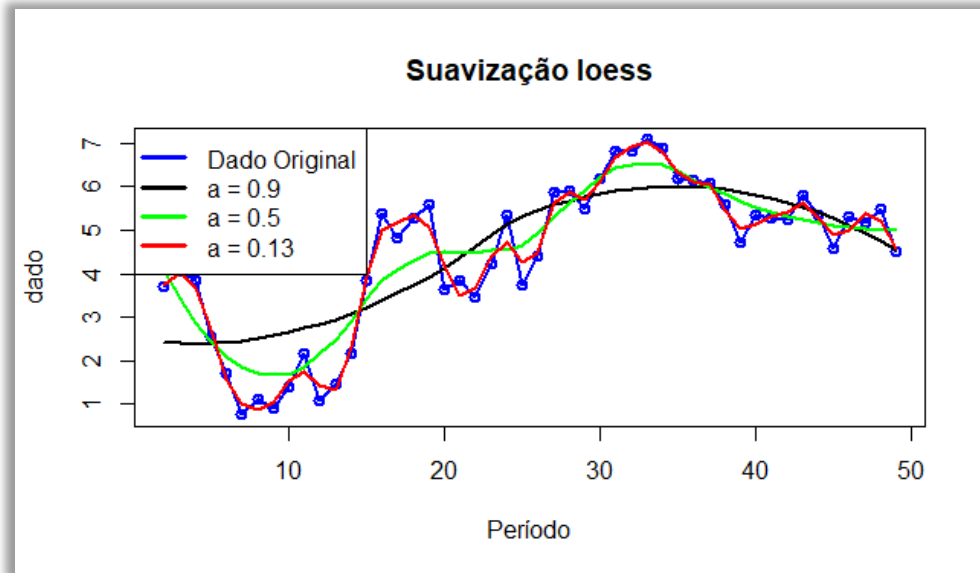
- ✓ Necessário realizar um pré-tratamento antes da *clusterização*
- ✓ Distorção dos centroides por *outliers*
- ✓ *K-means* gerou clusters consistentes



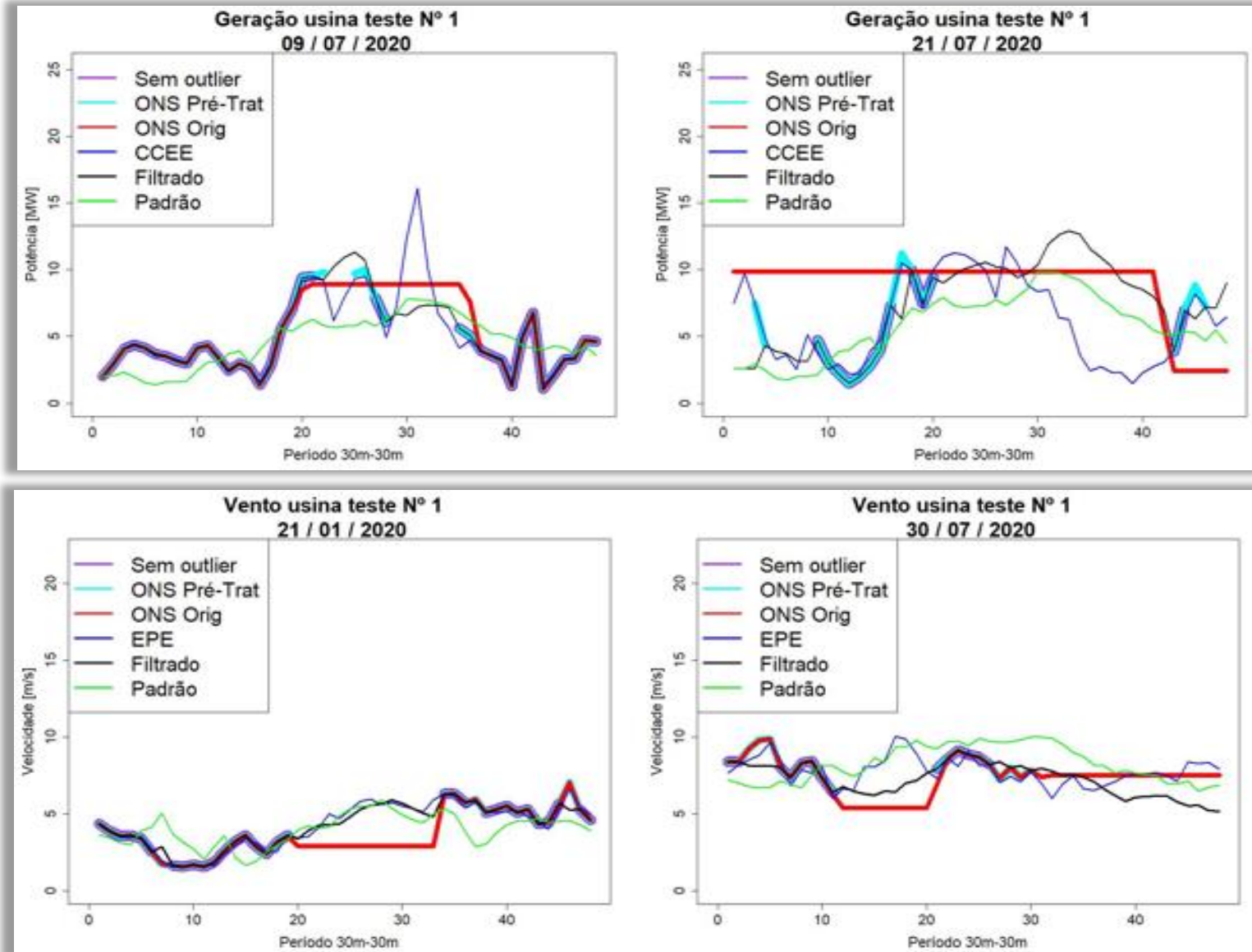
Apresentação e Discussão dos Resultados

Processo de preenchimento de dados faltantes

- ✓ Aplicação do filtro de suavização *loess*



- ✓ Preenchimento robusto em falhas curtas
- ✓ Em falhas longas, pode ter descolamentos em relação a geração verdadeiro





Universidade
Federal de Juiz
de Fora

Considerações finais

Considerações finais



- Metodologia proposta:
 - ✓ Atendeu de forma eficaz o objetivo
 - ✓ Diferentes graus de severidade em termos de falhas de medição
 - ✓ Identificação/substituição de falsos *outliers*
 - ✓ Baixa complexidade e alta velocidade de processamento.
- Como desdobramentos futuros:
 - ✓ Aplicação em dados da fonte solar fotovoltaica
 - ✓ Extrapolação do modelo para casos multivariados.
 - ✓ Adição grandezas meteorológicas, como velocidade e direção do vento, temperatura, irradiância, etc.



Universidade
Federal de Juiz
de Fora



Obrigado!