

## MÉTODO EMPÍRICO PARA DETERMINAÇÃO DE *OUTLIERS* EM SÉRIES DE FLUXOS DE DADOS MICROMETEOROLÓGICOS PÓS-PROCESSADOS

Dornelles Vissotto Junior, Nelson Luís Dias

Programa de Pós-Graduação em Métodos Numéricos em Engenharia/UFPR

### RESUMO

Os *outliers* são registros de dados que apresentam valores fisicamente inconsistentes. A eliminação destes dados é de fundamental importância para a obtenção de uma série confiável de dados de fluxos. É necessário o uso de técnicas empíricas, pois não é possível o cálculo de parâmetros físicos para dados pós-processados. O objetivo é apresentar uma técnica para eliminação de *outliers* através da estatística média em blocos de dados de mesmo horário de medição. A técnica se mostrou eficiente na eliminação da maioria dos *outliers*.

### ABSTRACT

The outliers are data records that have physically inconsistent values. The elimination of these data is crucial for obtaining a reliable fluxes of data series. It is necessary to use empirical techniques because there is no possible determination of physical standard in post processing data. We present a technique for eliminating outliers through the statistics of blocks of the same measurement time. The technique proved effective in elimination of most outliers.

### INTRODUÇÃO

O monitoramento micrometeorológico tem sido amplamente utilizado como uma importante ferramenta na análise dos dados ambientais, em especial a medição de fluxos superficiais através do método *eddy covariance*, ou Covariâncias Turbulentas (DIAS *et al.*, 2002; AUBINET *et al.*, 2012).

Estudos ao longo dos anos vêm buscando melhorar a qualidade das medições de *eddy covariance*, eliminando os efeitos devido a sensibilidade dos sensores, em especial na análise dos dados pós-processamento (LEE *et al.*, 2004).

Após a aplicação de técnicas de filtragem, ainda restam nos dados medidos, valores que não são fisicamente aceitáveis, os *outliers*. Este trabalho propõe uma abordagem simples para a filtragem dos dados, analisando a estatística de distribuição dos fluxos em cada meia hora de medição.

### MATERIAL E MÉTODOS

Os dados foram obtidos pelo LEMMA-UFPR (Laboratório de Monitoramento e Modelagem Ambientais) durante o período de 22/07/2004 a 14/09/2004 em uma campanha realizada sobre o Lago de Furnas. Foram utilizados os sensores anemômetro sônico Young 81000, um higrômetro capacitivo CSI CS500, 3 termopares CSI FW03 e um sensor infravermelho LICOR LI7500. Os dados foram obtidos a uma frequência de 20Hz, analisados, processados e armazenados em médias de flutuações a cada 30 min.

Para estabelecer uma metodologia de detecção para os *outliers* das séries temporais, foi separado o conjunto de dados por blocos, utilizando a relação temporal, onde os blocos de dados foram subdivididos pela meia hora correspondente ao seu período de medição.

Desta maneira, os blocos divididos em cada meia hora, estabeleceram 48 blocos de controle para cada variável, onde as características dos fluxos devem assumir um comportamento fisicamente semelhante.

Foi aplicada a estatística padrão, considerando que a distribuição dos dados em cada bloco de meia hora obedecem uma distribuição normal. O critério de seleção de dados correspondeu ao intervalo de confiança de 99%, ou seja, dada a distribuição normal, são considerados *outliers* os dados que estão abaixo de 0,5% e acima de 99,5% da distribuição acumulada ( $\pm 2,576\sigma$ ).

## RESULTADOS E DISCUSSÃO

A aplicação do filtro com intervalo de confiança de 99% sobre a média em cada bloco de 30 min. é capaz de identificar *outliers*, como podemos observar na figura 1.

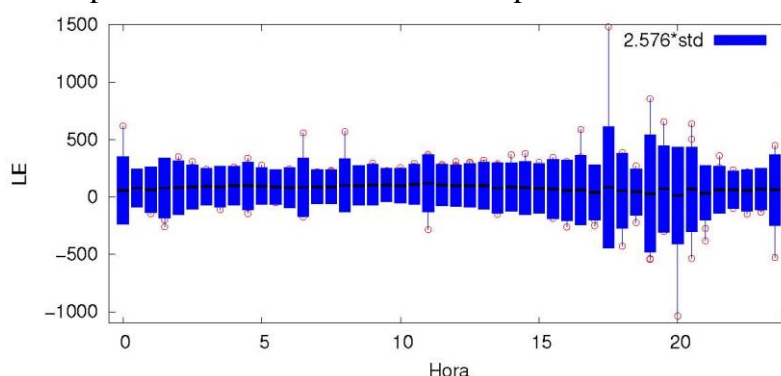


Figura 1: Distribuição dos Dados com intervalo de confiança ( $\pm 2,576\sigma$ ).

Os dados mais extremos de cada bloco de 30 min. que se afastam muito da média ( $\pm 2,576\sigma$ ) são eliminados (figura 2).

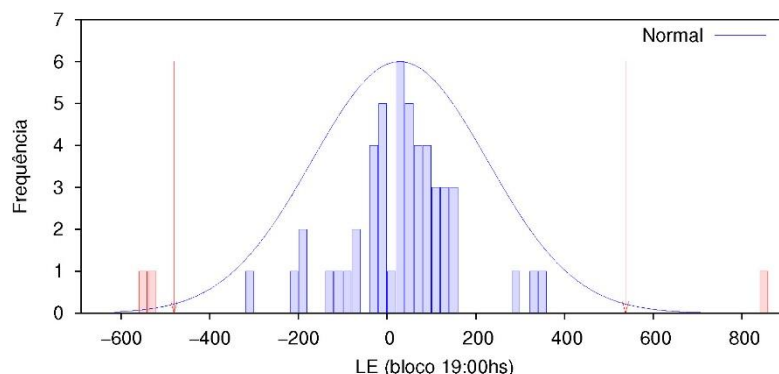
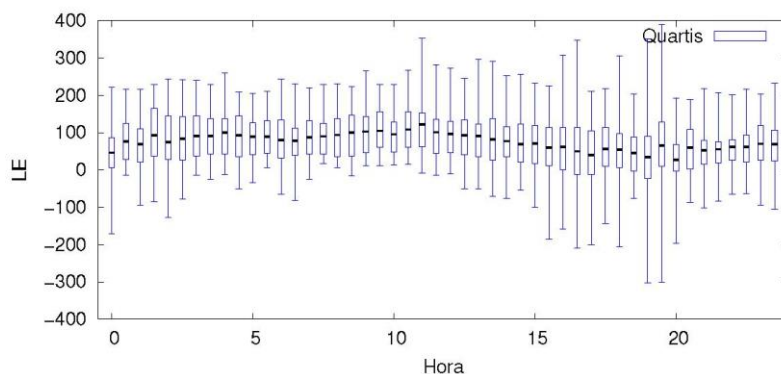


Figura 2: Histograma para o bloco 19:00hs, distribuição normal e dados "bons" (em azul) e intervalo de confiança e outliers em vermelho.

O comportamento de distribuição do bloco de 30 min. após a eliminação dos *outliers* ficou mais homogêneo (figura 3) se comparado com os blocos com *outliers* (figura 1).

## CONCLUSÃO

A determinação de *outliers* é uma questão ainda em desenvolvimento no tratamento de dados micrometeorológicos. Como não é possível a determinação de equações físicas em dados pós- processados, o estabelecimento de critérios empíricos tem sido uma saída muito versátil.



**Figura 3: Distribuição dos Dados “bons” (sem outliers).**

O critério de determinação pelo intervalo de confiança de 99% sobre a média de cada bloco de 30 min. é uma boa ferramenta para a eliminação da maioria dos *outliers*. Dos 2587 dados válidos no período, 2489 foram preservados. A maioria dos *outliers* foi encontrada no período de final de tarde e início de noite. Embora esta técnica elimine alguns dados fisicamente inconsistentes, outros critérios ainda devem ser aplicados para validação dos dados de fluxos.

#### **BIBLIOGRAFIA**

AUBINET, M.; VESALA, T.; PAPALE, D., editors. **Eddy Covariance: A Practical Guide to Measurement and Data Analysis**. Springer, Dordrecht, Heidelberg, London, New York, 2012.

DIAS, N. L.; KAN, A.; GRODZKI, L.; SANCHEZ, S. D.; VISSOTTO Jr., D. **O Método de Covariâncias Turbulentas Atenuadas (MCTA) para medição de fluxos de calor sensível e latente**: aplicação ao lago de Itaipu e seu redor. *Revista Brasileira de Recursos Hídricos*, 7(1): 143-160, 2002.

LEE, X; MASSMAN, W. J. e LAW; B. E., editors. **Hanbook of Micrometeorology. A Guide for Surface Flux Measurements**. Kluwer, Dordrecht, 2004.