**INTRODUCCIÓN**

La importancia del análisis de valores atípicos en series temporales radica en que a partir de éste se puede tener certeza de la calidad de los datos y de la viabilidad para tenerlos en cuenta en la toma decisiones o para entrenamiento de modelos de previsión basados en series temporales.

La presencia de valores atípicos en conjuntos de datos registrados periódicamente puede corresponder principalmente al comportamiento estocástico de las variables físicas estudiadas y a fallas en los sensores utilizados o en los nodos dispuestos para almacenar o transmitir los datos registrados.

Existen diferentes tipos de datos atípicos que pueden presentarse en series temporales y que se clasifican por su efecto sobre los datos, los principales son los atípicos aditivos y de cambio de nivel, los cuales pueden afectar la serie en instantes de tiempo determinados o persistir sobre los datos siguientes a su ocurrencia; También pueden ser clasificados a partir de la descripción estadística de los datos, como extremos que corresponden a valores muy lejanos a la media del conjunto de los datos o atípicos enmascarados que corresponden a variaciones cercanas a la media de los datos pero que no corresponden al comportamiento de la variable. Por otro lado, puede darse el caso de que valores correctos sean considerados atípicos empantanados debido a la presencia de valores atípicos cercanos.

A partir del análisis de los datos atípicos se decide si estos afectan al conjunto de datos y deben ser eliminados o ajustados. Teniendo en cuenta que la eliminación de los mismos puede afectar la integridad de los datos incluso más que su presencia. Debe escogerse con cuidado los datos que se van a eliminar. Siendo una buena alternativa el ajuste de los mismos.

Los métodos de ajuste o suavizado de curvas por lo general implican la reducción o eliminación de la cantidad de ruido en los registros y la consecución de una curva en la que se pueden identificar más fácilmente la tendencia, los patrones y estacionalidad de una serie temporal.

Sin embargo. La aplicación de un suavizado al igual que la eliminación de los valores considerados atípicos, puede causar pérdida de información importante para el análisis. Por tal motivo se propone un algoritmo que **permite aplicar un suavizado a los datos y sí es necesario corrige los valores afectados por atípicos de cambio de nivel. *Aún no he cuadrado lo de corregir los valores afectados por el cambio de nivel, eso lo haría comparando valores entre nodos para registros simultáneos.***

**El algoritmo**

Se aplica un filtro para los valores que pueden presentarse en los registros de cada variable. El filtro corresponde aún rango de medición que ofrecen los fabricantes de los sensores utilizados y el rango de Tukey para los valores registrados. A partir de un procedimiento iterativo en el que se desplazan los límites superiores e inferiores del filtro hasta determinar el rango que garantiza la mejor correlación entre la variable filtrada y la variable que presenta la mayor relación con ésta

Con la aplicación de este filtro se consigue la eliminación de valores atípicos extremos, que pueden corresponder a descalificación de los sensores o una falla en el funcionamiento del mismo por defectos de fabricación. Se sospecha que también pudieron ser causados por la presencia de insectos dentro de los nodos, dado que éstos se encontraban a la intemperie y en las visitas realizadas al escenario de medición se encontraron chinches y hormigas viviendo dentro de los nodos.

Posteriormente se aplican ventanas móviles y se determinan los registros que presentan mayores desviaciones estándar, estos son ajustados teniendo en cuenta la mediana móvil de los datos.....

De manera que se eliminan los atípicos enmascarados dentro del rango de datos

***Sin embargo al algoritmo le falta reducir el impacto que pueden causar los atípicos en los valores siguientes su ocurrencia***

Para esto se realiza un ajuste basado en la relación entre los nodos