

## Description about the smoothing methods in BlueTOAD data modeling

All smoothing algorithms are based on an adaptive rolling window. For each timestamp, we choose the latest N speeds or all available measurements in a certain time period of T to estimate the smoothed speed. N and T are two control parameters used for performance optimization. The timestamp of the latest valid measurement is used to mark the age of the smoothed speed. The details about these methods are described below.

- 1) Two stage median smoothing method: This smoothing algorithm is divided into two stages in calculating the smoothed speed. In the first stage, we pick the latest N=15 pair speeds or all available speeds in T=8 hours. Of these speeds, the 11 largest speeds are used to form an initial data set. In the second stage, we pick the latest 7 speeds from this initial data set to form the data set, from which the median speed is estimated as the smoothed speed. The purpose of the first stage is to eliminate the possible low-speed outliers in a larger window. We use the median speed as the smoothed speed in the second stage to minimize the impact of the high speed outliers. This method is designed to handle the measurements with significant outliers in both high- and low-speed sides.
- 2) Two stage 85<sup>th</sup> percentile smoothing method: All are the same as for two stage median smoothing method except that we calculate the 85<sup>th</sup> percentile speed as the smoothed speed in the last step. This method is designed to handle the measurements with significant outliers in the low-speed sides and should be the first choice for the long distance pairs on the freeway.
- 3) Two stage mean smoothing method: The algorithm consists of two stages also. In the first stage we filter out outliers in a larger window. In the second stage we estimate the mean speed from these filtered speeds in a smaller window. In order to remove extremely high or low speeds, but to keep low speeds resulted from actual congestion, we estimate the upper and lower cut-off speeds based on the mean speeds of raw data in the first stage. This method is designed for scattering data set from BlueToad units with LR radio.
- 4) Multiple speeds smoothing method: The algorithm is modified from method 1). The latest 25 points are used and both the 5 highest and 5 lowest speeds are removed in the first stage. Similar to method 3), the lower cut-off speed is estimated dynamically from the mean of raw speeds. This algorithm is designed to suppress impacts from HOV lane and ramp traffic.

The time window T used in all the methods is up to 8 hours, so if there is a data gap, the latest available data points will be used in calculating the smoothed speeds. The age of the smoothed speeds will be indicated by the last match time.

“Staleness” field is used to specify the upper limit of the age of the smoothed speeds in XML feeds. If the last match time is earlier than the current time minus the staleness value, the smoothed speed will show “no data”.