This is a supplementary material for the paper: “A Dynamic Ridesplitting Method with Potential Pick-up Probability Based on GPS Trajectories”. In this document, we provide the details about the Kalman filter model in the pick-up probability prediction. We first give the real-world data analysis to support the white noise assumption in the Kalman filter model, then we provide the details about the matrix settings associated with the Kalman filter model.

1. The White Noise Assumption Analysis

For each grid, the proposed Kalman filter model assumes the number of ride requests in the same time interval to be stable between two consecutive days. Also, the number of ride requests is stable between consecutive time intervals in the same day. Therefore, there are two noise assumptions in the model. For a time interval t in day m, first, the ride requests number difference from the time interval t of the previous day m-1 (i.e., the process noise ω_m) is assumed to be white noise; second, the ride requests number difference from previous time interval $t-1$ of day m (i.e., the measurement noise v_m) is assumed to be the white noise. These two assumptions are supported by our analysis of ride requests number distribution on two real-world datasets.

ω_m and v_m are white noise if the sequence of ride requests number differences is identically distributed with a mean of zero and are not autocorrelated. We checked these two conditions for both datasets.

On the San Francisco dataset, the distribution of process noise ω_m is shown as the following figure 1, the mean is 0.01 and the standard deviation is 1.34:

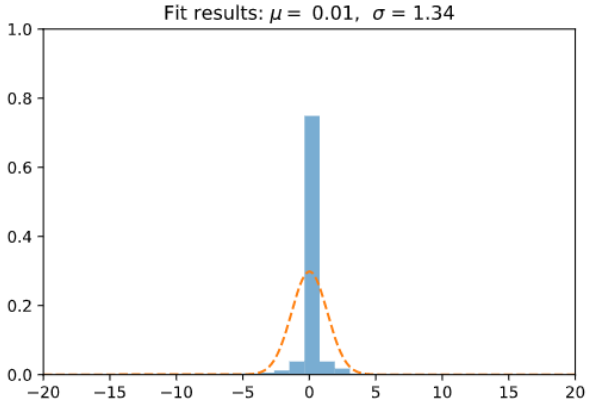

Figure 1. Process noise ω_m distribution on SF

The correlogram of ω_m is shown in the following figure 2, where all spikes are within the 95% confidence interval. The correlogram does not show any obvious autocorrelation pattern of ω_m. Therefore, ω_m is assumed to be the white noise on San Francisco dataset.

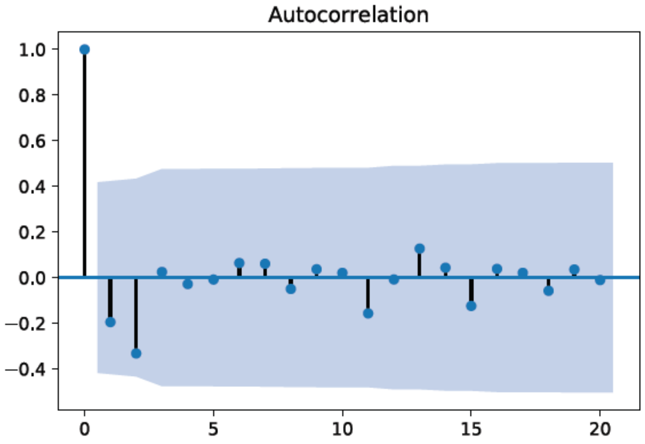

Figure 2. Process noise ω_m autocorrelation on SF

Similarly, we checked v_m on the San Francisco dataset. The distribution and correlogram are shown in Figure 3.

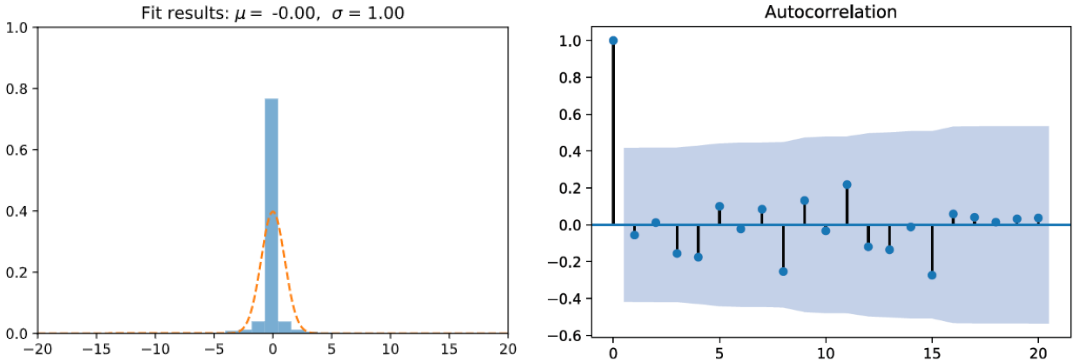

Figure 3. measurement noise v_m distribution and autocorrelation on SF

Since v_m is also identically distributed with a mean of zero and is not autocorrelated, it is assumed to be the white noise on San Francisco dataset.

The distribution and autocorrelation check of ω_m and v_m on Wuhan dataset is shown in Figure 4 and Figure 5, respectively.

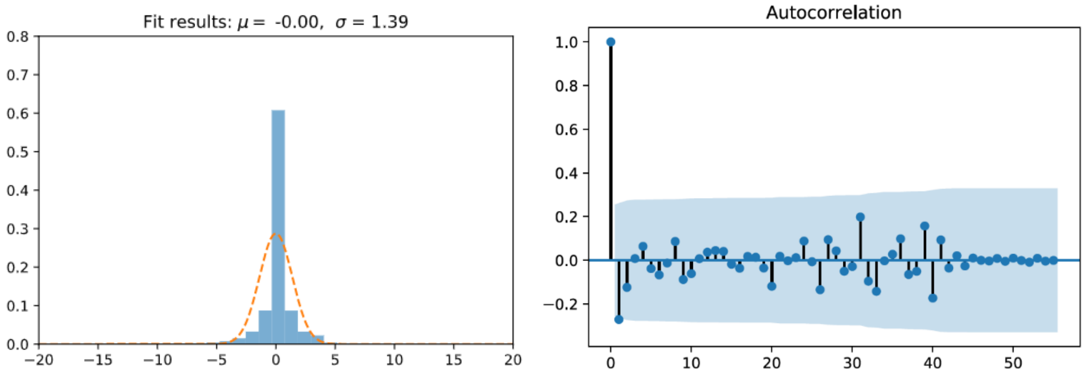

Figure 4. Process noise ω_m distribution and autocorrelation on WH

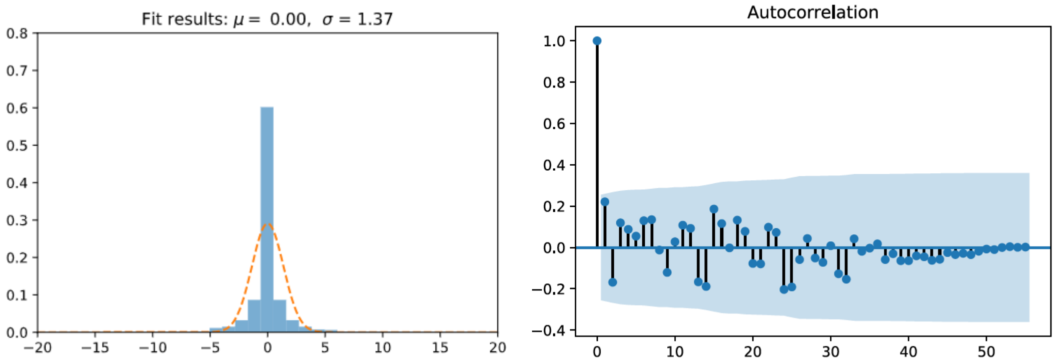

Figure 5. measurement noise v_m distribution and autocorrelation on WH

Based on the distribution and autocorrelation, it is assumed ω_m and v_m are white noise on the Wuhan dataset.

Therefore, we assume ω_m and v_m are white noise.

2. Matrix Settings Associated with the Kalman Filter Model

There are four matrices associated with the Kalman filter model: state transition matrix, measurement matrix, process noise covariance matrix and the measurement noise covariance matrix.

The state transition matrix $\boldsymbol{F}_m$ translates from the pick-up probability of grids $\boldsymbol{x}_{\boldsymbol{t}}^{\boldsymbol{m-1}}$ in time interval t in the previous day $m-1$ to the pick-up probability $\boldsymbol{x}_{\boldsymbol{t}}^{\boldsymbol{m}}$ in time interval t in day m by giving the linear relationship. Because the pick-up probability for a grid in the same time interval is assumed to be stable between two consecutive days, the pick-up state transition matrix $\boldsymbol{F}_m$ is set to an n-order identity matrix in this study, where n is the number of grids, i.e., $\boldsymbol{F}_m = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$.

The measurement matrix $\boldsymbol{H}_m$ translates from pick-up probability x_t^m to the noisy measurement z_t^m, i.e., the pick-up probability of grids in the previous time interval of the same day m. Similarly, according to passenger distribution between consecutive time intervals, matrix $\boldsymbol{H}_m$ is set to n-order identity matrix as well. i.e., $\boldsymbol{H}_m = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$.

For the matrices $\boldsymbol{Q}_{\boldsymbol{m}}$ and $\boldsymbol{R}_{\boldsymbol{m}}$, since the rider dynamics are different between grids, they are set to diagonal matrices. Each element in $\boldsymbol{Q}_{\boldsymbol{m}}$ represents the process noise covariance for the grid. For instance, element $\boldsymbol{Q}_{\boldsymbol{m}}[j,j]$ represents the process noise covariance for the grid g_j. Similarly, each element in $\boldsymbol{R}_{\boldsymbol{m}}[j,j]$ represents the measurement noise covariance for g_j. $\boldsymbol{Q_m}$ is initialized as a small value, i.e., 0.01. $\boldsymbol{R_m}$ is initialized from the historical data. Specifically, we fit the normal distribution to the pick-up probability data for each grid, then $\boldsymbol{R_m}$ is set to the $(3\sigma)^2$ according to the “three-sigma limits”, where σ is the standard deviation of pick-up probability data. $\boldsymbol{Q_m}$ and $\boldsymbol{R_m}$ are then adjusted from the experiment to obtain the filter’s good performance.