By communicating its local summary to every other sensor, each mobile sensor can then construct a globally consistent summary from the received local summaries:

**Definition 3 (Global Summary)** Given a common support set \( U \subset V \) known to all \( K \) mobile sensors and the local summary \((\hat{z}_U^k, \hat{\Sigma}_{UU}^k)\) of every mobile sensor \( k = 1, \ldots, K \), the global summary is defined as a tuple \((\hat{z}_U, \hat{\Sigma}_{UU})\) where

\[
\hat{z}_U^k \triangleq \sum_{k=1}^{K} \hat{z}_U^k \tag{8}
\]

\[
\hat{\Sigma}_{UU} \triangleq \hat{\Sigma}_{UU} + \sum_{k=1}^{K} \hat{\Sigma}_{UU}^k . \tag{9}
\]

**Remark.** In this paper, we assume all-to-all communication between the \( K \) mobile sensors. Supposing this is not possible and each sensor can only communicate locally with its neighbors, the summation structure of the global summary (specifically, (8) and (9)) makes it amenable to be constructed using distributed consensus filters (Olfati-Saber, 2005). We omit these details since they are beyond the scope of this paper.

Finally, the global summary is exploited by each mobile sensor to compute a globally consistent predictive Gaussian distribution, as detailed in Theorem 1A below, as well as to perform decentralized active sensing (Section 4):

**Theorem 1** Let a common support set \( U \subset V \) be known to all \( K \) mobile sensors.

A. Given the global summary \((\hat{z}_U, \hat{\Sigma}_{UU})\), each mobile sensor computes a globally consistent predictive Gaussian distribution \( \mathcal{N}(\tilde{\mu}_U, \tilde{\Sigma}_{UU}) \) of the measurements at any set \( Y \) of unobserved road segments where

\[
\tilde{\mu}_U \triangleq \mu_Y + \hat{\Sigma}_{UU}^{-1} \hat{z}_U \tag{10}
\]

\[
\tilde{\Sigma}_{YY} \triangleq \Sigma_{YY} - \hat{\Sigma}_{UU}^{-1} \hat{\Sigma}_{UY} \hat{\Sigma}_{UU}^{-1} \Sigma_{UY} . \tag{11}
\]

B. Let \( \mathcal{N}(\mu_{\text{PITC}}_{Y|D}, \Sigma_{\text{PITC}}_{Y|D}) \) be the predictive Gaussian distribution computed by the centralized sparse partial independent training conditional (PITC) approximation of GP model (Quiñonero-Candela and Rasmussen, 2005) where

\[
\mu_{\text{PITC}}_{Y|D} \triangleq \mu_Y + \Gamma_Y (\Gamma_D + \Lambda)^{-1}(\hat{z}_U - \mu_D) \tag{12}
\]

\[
\Sigma_{\text{PITC}}_{Y|D} \triangleq \Sigma_{YY} - \Gamma_Y (\Gamma_D + \Lambda)^{-1} \Gamma_D \tag{13}
\]

such that

\[
\Gamma_{BB} \triangleq \Sigma_{BU} \Sigma_{UU}^{-1} \Sigma_{UB} . \tag{14}
\]

and \( \Lambda \) is a block-diagonal matrix constructed from the \( K \) diagonal blocks of \( \Sigma_{DD|U} \), each of which is a matrix \( \Sigma_{D_kD_k|U} \) for \( k = 1, \ldots, K \) where \( D = \bigcup_{k=1}^{K} D_k \). Then, \( \tilde{\mu}_U = \mu_{\text{PITC}}_{Y|D} \) and \( \hat{\Sigma}_{YY} = \Sigma_{\text{PITC}}_{Y|D} \).

Its proof is given in (Chen et al., 2012). The equivalence result of Theorem 1B bears two implications:

**Remark 1.** The computation of PITC can be parallelized and distributed among the mobile sensors in a Google-like MapReduce paradigm (Chu et al., 2007), thereby improving the time efficiency of prediction: Each of the \( K \) mappers (sensors) is tasked to compute its local summary while the reducer (any sensor) sums these local summaries into a global summary, which is then used to compute the predictive Gaussian distribution. Supposing \(|Y| \leq |U|\) for simplicity, the \( O(|D|(|D|/K)^2 + |U| \log K) \) time incurred by PITC can be reduced to \( O((|D|/K)^2 + |U| \log K) \) time of running our decentralized algorithm on each of the \( K \) sensors, the latter of which scales better with increasing number \(|D|\) of observations.

**Remark 2.** We can draw insights from PITC to elucidate an underlying property of our decentralized algorithm: It is assumed that \( Z_{D_k}, \ldots, Z_{D_k}, Z_Y \) are conditionally independent given the measurements at the support set \( U \) of road segments. To potentially reduce the degree of violation of this assumption, an informative support set \( U \) is actively selected, as described earlier in this section. Furthermore, the experimental results on real-world urban road network data\(^2\) (Section 6) show that \( D^2\)FAS can achieve predictive performance comparable to that of the full GP model while enjoying significantly lower computational cost, thus demonstrating the practicality of such an assumption for predicting traffic phenomena. The predictive performance of \( D^2\)FAS can be improved by increasing the size of \( U \) at the expense of greater time and communication overhead.

### 4 Decentralized Active Sensing

The problem of active sensing with \( K \) mobile sensors is formulated as follows: Given the set \( D_k \subset V \) of observed road segments and the currently traversed road segment \( s_k \in V \) of every mobile sensor \( k = 1, \ldots, K \), the mobile sensors have to coordinate to select the most informative walks \( w_1^k, \ldots, w_K^k \) of length (i.e., number of road segments) \( L \) each and with respective origins \( s_1, \ldots, s_K \) in the road network \( G \):

\[
(w_1^k, \ldots, w_K^k) = \arg \max_{(w_1^k, \ldots, w_K^k)} \mathbb{E} \left[ Z_{\bigcup_{k=1}^{K} Y_{w_k}^k} \right] \tag{15}
\]

where \( Y_{w_k} \) denotes the set of unobserved road segments induced by the walk \( w_k \). To simplify notation, let a joint walk be denoted by \( w \triangleq (w_1, \ldots, w_K) \) (similarly, for \( w^* \)) and its induced set of unobserved road segments be \( Y_{w^*} \triangleq \bigcup_{k=1}^{K} Y_{w^k} \) from now on. Interestingly, it can be shown using the chain rule for entropy that these maximum-entropy walks \( w^* \) minimize the posterior joint entropy (i.e., \( \mathbb{H}(Z_{Y \setminus (D \cup Y_{w^*})}) \)) of the measurements at the remaining unobserved segments (i.e., \( V \setminus (D \cup Y_{w^*}) \)) in the road network. After executing the walk \( w_k^* \), each mobile sensor \( k \) observes the set \( Y_{w_k^*} \) of road segments and updates its local information:

\[
D_k \leftarrow D_k \cup Y_{w_k^*}^k, \quad z_{D_k} \leftarrow z_{D_k} \cup Y_{w_k^*}^k, \quad s_k \leftarrow \text{terminus of } w_k^* . \tag{16}
\]