Origin-Destination matrix estimation in the presence of faulty measurements using the cell transmission model

Yiolanda Englezou, **Stelios Timotheou** and **Christos Panayiotou**

englezou.yiolanda@ucy.ac.cy, timotheou.stelios@ucy.ac.cy and christop@ucy.ac.cy

Motivation

Efficient estimation of the origin-destination (O-D) matrix of a particular network of interest is a crucial requirement for transportation planning. The O-D matrix contains information on the amount of vehicles that commute towards specific locations. The main objective is to calculate an O-D matrix based on available traffic counts to reproduce field data as accurately as possible.

A significant challenge when using information obtained from traffic sensors, is that such sensors are subject to considerable disruptions due to system errors, that affect the quality and reliability of the information they deliver (Zhou, et. al., 2017). Common sensor failures include bias, drifting or complete failure that decrease the accuracy and reliability of the sensor measurements and provide incorrect information. This affects the quality of the estimation, hence sensor failures have to be considered explicitly in order to make efficient and reliable O-D matrix estimation.

Contributions

We aim to estimate a static O-D matrix for a pre-specified time period T.

- ▶ We utilise a path-based CTM for O-D matrix estimation in the presence of faulty measurements, where link densities (measurements) are associated with per path densities (state vector) and the path demand.
- ▶ We develop a novel methodology for O-D matrix estimation assuming sensor failures that (i) achieves O-D matrix estimation and (ii) identifies faulty sensors and their fault magnitude, at the same time.

Cell Transmission model (CTM)

- Cell transmission model (CTM): is a macroscopic flow model (Daganzo, 2005) where each cell i is characterised by: the free flow speed, v_i^f (km/h), the backward propagation speed w_i (km/h), the maximum flow, φ_i^{max} (veh/h), the maximum density, ρ_i^{max} (veh/km), the cell length l_i, the inflow, $\bar{\varphi}_i^{in}(t)$, outflow, $\bar{\varphi}_i^{out}(t)$, and density, $\bar{\rho}_i(t)$, of vehicles.
- Upstream connections (i^-): ordinary, $\mathcal{O}$, entering, $\mathcal{E}$, merging, $\mathcal{M}$.
- Downstream connections (i^+): ordinary, $\mathcal{O}$, diverging, $\mathcal{D}$, or exiting, $\mathcal{G}$.
- $\mathcal{N}^-$: the set of upstream neighbours, $\mathcal{N}^+$: the set of downstream neighbours.
- Demand: flow of vehicles that want to exit i at time t, $D_i(t) = \min\{v_i^f \bar{\rho}_i(t), \varphi_i^{max}\}$.
- Supply: flow of vehicles i can receive according to its capacity, $S_i(t) = \min\{\varphi_i^{max}, w_i[\rho_i^{max} - \bar{\rho}_i(t)]\}$.

Objective: estimate path demand $\lambda = [\lambda_1, \ldots, \lambda_Q]^T$ in the presence of faulty sensors and hence estimate the O-D matrix. The demand of each O-D pair $w \in \mathcal{W}$, $d_w(t)$, is related to the path demand pattern, $\lambda_p(t)$ on each path $p \in \mathcal{S}_w$ and time $t = 1, \ldots, K$, by

$$\sum_{p \in \mathcal{S}_w} \lambda_p(t) = d_w(t), \quad \forall w \in \mathcal{W}. \tag{1}$$

Path-based CTM

- Path-based CTM (Ukkusuri, et al., 2012): an extension of the CTM that allows us to keep track of path-based densities and path-based flows given by

$$\begin{aligned}
\rho_{i,p}(t+1) &= \rho_{i,p}(t) + \frac{t_s}{l_i}\left[\varphi_{i,p}^{in}(t) - \varphi_{i,p}^{out}(t)\right],\\
\varphi_{i,p}^{in}(t) &= \begin{cases} \dfrac{\rho_{j,p}(t)}{\sum_{b\in\mathcal{P}_i\cap\mathcal{P}_j}\rho_{j,b}(t)} D_{ji}(t) \min\left\{1, \dfrac{S_i(t)}{\sum_{r'=1}^{|\mathcal{N}_i^-|} D_{j_{r'}i}(t)}\right\} & \text{if } i^- \in \{\mathcal{O}, \mathcal{M}\}\\ \dfrac{u_p(t)}{\sum_{p\in\mathcal{P}_i} u_p(t)} \min\left\{\sum_{p\in\mathcal{P}_i} u_p(t), S_i(t)\right\}, & \text{if } i^- \in \{\mathcal{E}\}\end{cases}\\
\varphi_{i,p}^{out}(t) &= \begin{cases} \dfrac{\rho_{i,p}(t)}{\sum_{b\in\mathcal{P}_i\cap\mathcal{P}_j}\rho_{i,b}(t)} D_{ij}(t) \min\left\{1, \left\{\dfrac{S_{j'}(t)}{D_{ij'}(t)}\right\}_{j'\in\mathcal{N}_i^+}\right\} & \text{if } i^+ \in \{\mathcal{O}, \mathcal{D}\}\\ v_i^f \rho_{i,p}(t) \; i^+ \in \{\mathcal{G}\} \end{cases}
\end{aligned} \tag{2}$$

where $\varphi_{i,p}^{in}(t)$, $\varphi_{i,p}^{out}(t)$ the per path inflow and outflow.
- $\mathcal{P}_i \subseteq \mathcal{P}$ ($\mathcal{P} = \cup_{i\in\mathcal{R}}\mathcal{P}_i$) the set of all paths passing through cell i.
- $u_p(t)$ the vehicles entering the network at t through the entering connections, $u_p(t)|\lambda_p(t) \sim \text{Pois}[\lambda_p(t)]$.

We assume $\lambda_p(t)$ remains fixed for all $t = 1, \ldots, K$ and hence we drop subscript t

State-space model

Assume a mathematical description of the physical system with state vector $\mathbf{x}_t$ and inputs $\mathbf{u}_t$:

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{A}_t\mathbf{x}_t + \mathbf{B}_t\mathbf{u}_t + \boldsymbol{\epsilon}_t,\\ \mathbf{y}_{t+1} &= \mathbf{H}_{t+1}\mathbf{x}_{t+1} + \boldsymbol{\omega}_{t+1}. \end{aligned} \tag{3}$$

- $\mathbf{x}_0$: the initial conditions of the stochastic process.
- the $(LQ+L)$-state vector $\mathbf{x}_t = [\boldsymbol{\rho}(t)^T, \bar{\boldsymbol{\rho}}(t)^T]^T$.
- $\boldsymbol{\rho}(t) = [\rho_{1,1}(t), \rho_{1,2}(t), \ldots, \rho_{1,Q}(t), \ldots, \rho_{L,1}(t), \rho_{L,2}(t), \ldots, \rho_{L,Q}(t)]^T$ and $\bar{\boldsymbol{\rho}}(t) = [\bar{\rho}_1(t), \ldots, \bar{\rho}_L(t)]^T$.
- the C-vector of observations $\mathbf{y}_t = [\bar{\rho}_{b_1}(t), \ldots, \bar{\rho}_{b_C}(t)]^T$, $b_1, \ldots, b_C \in \mathcal{C}$.
- the Q input vector $\mathbf{u}_t$, denotes the inflow of cell i at time t, with $i^- \in \mathcal{E}$ and $p \in \mathcal{P}_i$.
- $\mathbf{A}_t$ he evolution of the unknown states as time progresses.
- $\mathbf{H}_t$ the matrix of explanatory variables.

  These matrices do not vary with time as a result of the path-based CTM under free-flow conditions, hence drop subscript t.
- Independent Gaussian errors $\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_t^{\epsilon})$ and $\boldsymbol{\omega}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_t^{\omega})$.
- $\boldsymbol{\Sigma}_t^{\epsilon}$ and $\boldsymbol{\Sigma}_t^{\omega}$ model and measurement error matrices, respectively.

We can re-write Equations (3) as

$$\begin{aligned} \mathbf{CX} &= \mathbf{AX} + \mathbf{Bu} + \mathbf{b}_\star + \boldsymbol{\epsilon},\\ \mathbf{Y} &= \mathbf{HX} + \boldsymbol{\omega}, \quad t = 1, \ldots, K. \end{aligned} \tag{4}$$

Solution Approach

The inputs, $\mathbf{u}$, as well as the state vector, $\mathbf{X}$, are unknown and hence we define $\mathbf{w} = [\mathbf{X}^T, \mathbf{u}^T]^T$ hence

$$\begin{aligned} \mathbf{C}_\star\mathbf{w} + \mathbf{b}_\star + \boldsymbol{\epsilon} &= \mathbf{0},\\ \mathbf{Y} &= \mathbf{H}_\star\mathbf{w} + \boldsymbol{\omega}, \end{aligned} \tag{5}$$

$$\begin{aligned} \mathbf{C}_\star &= [\mathbf{A} - \mathbf{C}, \mathbf{0}_{(LQ+L)K\times m}] + [\mathbf{0}_{(LQ+L)K\times(LQ+L)K}, \mathbf{B}]\\ \mathbf{H}_\star &= [\mathbf{H}, \mathbf{0}_{C\times Q}]. \end{aligned}$$

Path-based CTM Static O-D matrix estimation

We formulate the path-based CTM static O-D matrix estimation (pCTM-O-D) in an optimisation context:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \Psi(\mathbf{w})\\ \text{s.t.} \quad & \mathbf{w} \geqslant \mathbf{0}, \end{aligned} \tag{6}$$

where the objective function to be minimised is:

$$\Psi(\mathbf{w}) = \tfrac{1}{2}\left[||\mathbf{Y} - \mathbf{Hw}||^2_{[\boldsymbol{\Sigma}_\omega]^{-1}} + ||\mathbf{C}_\star\mathbf{w} + \mathbf{b}||^2_{[\boldsymbol{\Sigma}_\epsilon]^{-1}}\right], \tag{7}$$

where $\boldsymbol{\Sigma}_\omega = \text{blkdiag}(\boldsymbol{\Sigma}_1^\omega, \ldots, \boldsymbol{\Sigma}_K^\omega)$ and $\boldsymbol{\Sigma}_\epsilon = \text{blkdiag}(\boldsymbol{\Sigma}_1^\epsilon, \ldots, \boldsymbol{\Sigma}_K^\epsilon)$.

Fault-Tolerant Path-based CTM O-D matrix estimation

When faulty sensors are present in the network under study, a more appropriate model to describe the evolution of traffic density of the path-based CTM is

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{A}_t\mathbf{x}_t + \mathbf{B}_t\mathbf{u}_t + \boldsymbol{\epsilon}_t,\\ \mathbf{y}_{t+1} &= \mathbf{H}_{t+1}\mathbf{x}_{t+1} + \boldsymbol{\omega}_{t+1} + \mathbf{o}_t, \end{aligned} \tag{8}$$

with $\mathbf{o}_t = [o_{1,t}, \ldots, o_{C,t}]^T$ the *sensor fault residuals*, with $o_{i,t}$ having a zero value if sensor $i \in \mathcal{C}$ is not faulty and non-zero value if sensor $i \in \mathcal{C}$ is faulty at time t (Timotheou, et. al., 2015).

Following Timotheou, et. al. (2015) the *maximum sensor fault-residuals* $y_j(t)$, $j = 1, \ldots, C$ are

$$y_j(t) = \max_{t=1,\ldots,T}\{|o_{j,t}|\}, \tag{9}$$

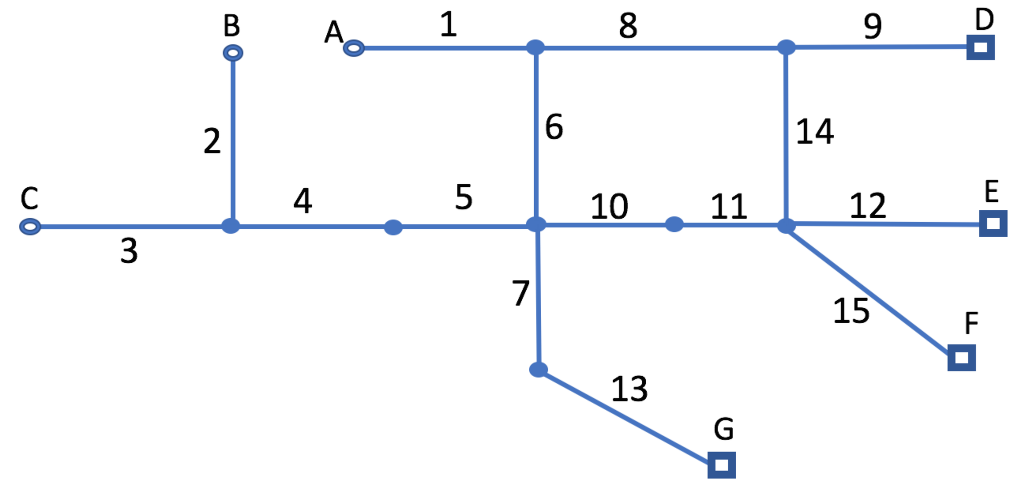
relating a value of sensor j equal to the maximum residual $o_{j,t}$. This results in the below convex formulation:

$$\begin{aligned} \min_{\mathbf{w},\mathbf{o}} \quad & \Psi(\mathbf{w}, \mathbf{O}) + \mu \textstyle\sum_{j\in\mathcal{C}} y_j\\ \text{s.t.} \quad & o_{j,t} \leqslant y_j(t), t = 1, \ldots, K, \quad j \in \mathcal{C}\\ & -o_{j,t} \leqslant y_j(t), t = 1, \ldots, K, \quad j \in \mathcal{C}, \end{aligned} \tag{10}$$

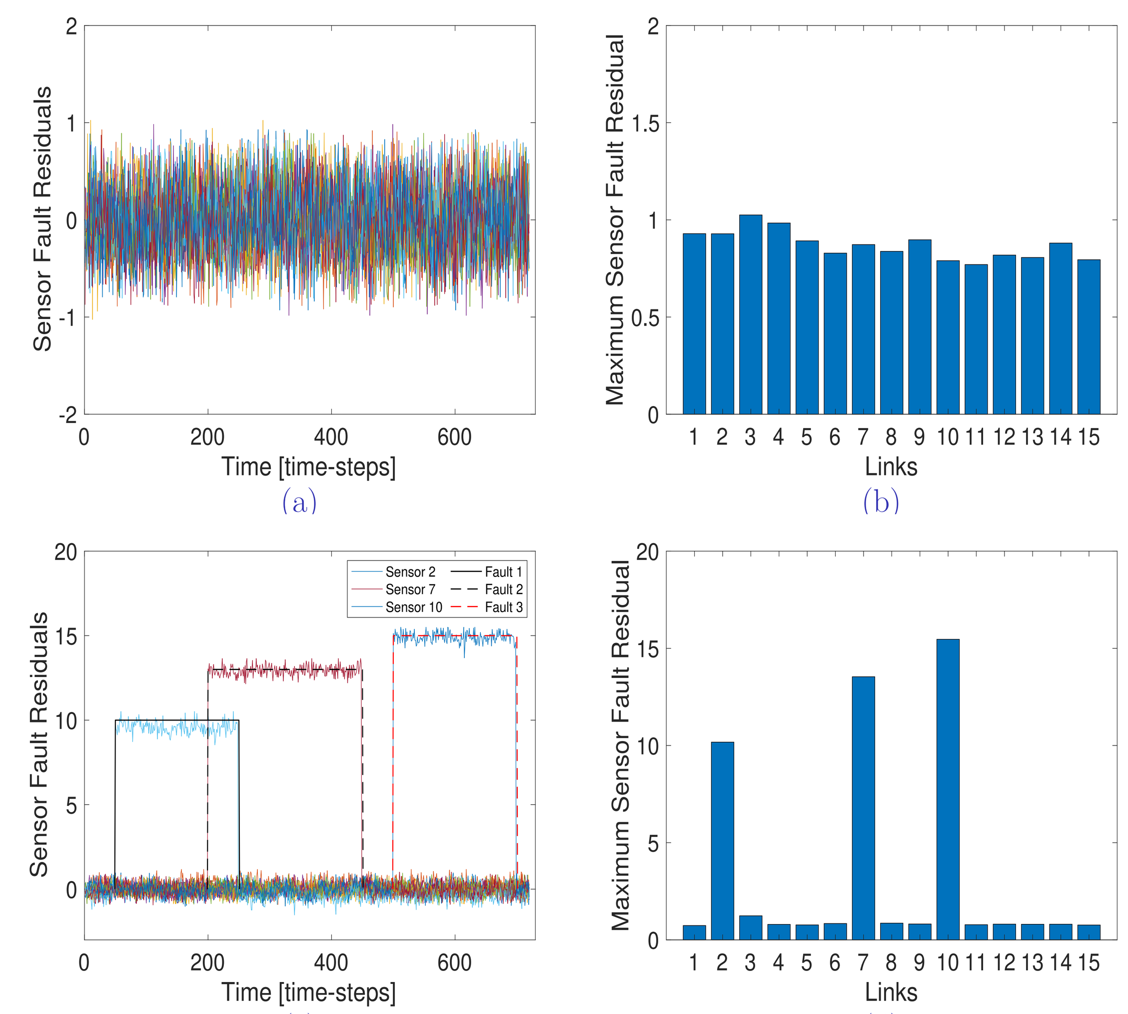
with $\mathbf{O} = [\mathbf{o}_1, \ldots, \mathbf{o}_K]^T$, referred to as pCTM-O-D-F.

Simulation Results

- ▶ Consider the following arterial network which consists of $L = 15$ directed measured links (cells), $Q = 12$ pre-specified paths and $W = 10$ O-D pairs.
- ▶ Traffic enters the network from the upstream boundary of cells 1, 2 and 3 and exits from the downstream boundary of cells 9, 12, 13 and 15.



- ▶ Collect measurements when there are (i) no faulty sensors in the network, $n_f = 0$; (ii) three faulty sensors in the network, $n_f = 3$.
- ▶ Estimate $\boldsymbol{\lambda}$ for both cases using (a) pCTM-O-D and (b) pCTM-O-D-F.



(a) (b) (c) (d)

- ▶ (a) shows that the estimated fault residual parameter values are small around zero, suggesting that no faulty sensors have been detected.
- ▶ (b) shows the estimated maximum fault residual parameter values for each link which are again low and close to zero also supporting that no faulty measurements have been identified.
- ▶ (c) shows that the proposed methodology offers effective fault detection, isolation and identification of the three problematic sensors, through the fault residual values.
- ▶ (d) shows the estimated maximum fault residual parameter values for each link which support the faults at the particular sensors.

	Formulation	SE_λ	
Sensor Errors		$n_f = 0$	$n_f = 3$
Poisson variables $\mathbf{u} \sim \text{Pois}(\boldsymbol{\lambda})$	pCTM-O-D	73.97	684.03
	pCTM-O-D-F	69.62	80.11

Acknowledgements

This work has been supported by the European Union's Horizon 2020 research and innovation programme under grant agreements No 739551 (Project: KIOS CoE) and No 101003435 (Project: BITS), the Government of the Republic of Cyprus through the Directorate General for European Programmes, Coordination and Development and through the Research Promotion Foundation (Project: CULTURE/BR-NE/0517/14).