Model Settings in Imputation Experiments

"Matrix and Tensor Model for Spatiotemporal Traffic Data Imputation and Forecasting"

We consider a new mechanism for our missing traffic data imputation tasks (see the illustration below): 1) Suppose the traffic data can be separated into 10 fractions (e.g., 4-week data); 2) Separate the first 5 fractions (**subset #1**, e.g., the data in the first two weeks) into the training set and the validation set, helping tune hyperparameters; 3) Separate the last 5 fractions (**subset #2**, e.g., the data in the last two weeks) into the training set and the testing set, allowing one to report imputation results on this subset. In the experiment setting, both subsets preserve the same missing rate level.

Subset #1 (tuning hyperparameters) Subset #2 (reporting results)

How to Select Hyperparameters in LATC and Baseline Imputation Methods?

Table 1 shows that the best hyperparameters for LATC on the Seattle freeway traffic speed dataset are:

- 30%, RM: $\gamma = 10\lambda$ and $r = 25$;
- 70%, RM: $\gamma = 10\lambda$ and $r = 25$;
- 90%, RM: $\gamma = 10\lambda$ and $r = 20$;
- 30%, NM: $\gamma = 5\lambda$ and $r = 15$;
- 70%, RM: $\gamma = 10\lambda$ and $r = 10$;
- 30%, BM-12: $\gamma = 10\lambda$ and $r = 25$.

Table 2 shows that the best hyperparameters for LRTC-TNN on the Seattle freeway traffic speed dataset are:

- 30%/70%, RM: $r = 25$;
- 90%, RM: $r = 10$;
- 30%/70%, NM: $r = 5$;
- 30%, BM-12: $r = 5$.

Table 4 shows that the best hyperparameters for LATC on the Portland traffic volume dataset are:

- 30%, RM: $\gamma = \lambda$ and $r = 10$;
- 70%, RM: $\gamma = \lambda$ and $r = 25$;

Table 1: Imputation performance (in MAPE/RMSE) of LATC with different truncation values and different ratios γ/λ on the Seattle freeway traffic speed dataset (on the first two-week subset for tuning hyperparameters). Note that the best results are highlighted in bold fonts. The number next to the BM denotes the window length.

Missing rate	γ/λ	Truncation parameter				
		$r = 5$	$r = 10$	$r = 15$	$r = 20$	$r = 25$
30%, RM	$1/10$	5.19/3.33	4.93/3.22	4.77/3.16	4.69/3.13	4.63/3.11
	$1/5$	5.19/3.32	4.93/3.22	4.77/3.16	4.69/3.13	4.63/3.10
	1	5.18/3.32	4.92/3.22	4.77/3.16	4.68/3.12	4.62/3.10
	5	5.13/3.30	4.88/3.20	4.74/3.14	4.65/3.11	4.59/3.09
	10	5.09/3.27	4.85/3.19	4.71/3.13	4.63/3.10	**4.57**/**3.08**
70%, RM	$1/10$	6.64/4.11	6.14/3.88	5.87/3.77	5.77/3.73	5.73/3.72
	$1/5$	6.64/4.11	6.13/3.88	5.87/3.77	5.77/3.73	5.73/3.72
	1	6.62/4.10	6.12/3.87	5.86/3.76	5.75/3.72	5.71/3.71
	5	6.55/4.06	6.07/3.84	5.80/3.73	5.69/3.69	5.63/3.67
	10	6.48/4.02	6.01/3.81	5.75/3.71	5.63/3.66	**5.56**/**3.64**
90%, RM	$1/10$	8.33/4.99	7.60/4.69	7.45/4.65	7.54/4.72	7.68/4.81
	$1/5$	8.33/4.99	7.60/4.69	7.44/4.64	7.54/4.71	7.67/4.81
	1	8.34/4.99	7.57/4.67	7.39/4.62	7.46/4.67	7.57/4.76
	5	8.26/4.95	7.47/4.62	7.22/4.53	7.22/4.55	7.29/4.61
	10	8.22/4.92	7.41/4.58	7.12/**4.48**	**7.08**/**4.48**	7.10/4.51
30%, NM	$1/10$	7.69/4.65	7.30/4.49	7.16/4.46	7.11/4.51	7.09/4.55
	$1/5$	7.69/4.65	7.30/4.49	7.16/4.46	7.10/4.50	7.09/4.57
	1	7.69/4.65	7.29/4.49	7.17/4.46	7.10/4.50	7.07/4.53
	5	7.71/4.65	7.32/4.49	7.15/**4.45**	7.09/4.48	7.05/4.50
	10	7.73/4.66	7.32/4.50	7.18/**4.45**	7.09/4.46	**7.04**/4.48
70%, NM	$1/10$	12.54/**11.35**	12.49/11.38	12.55/11.41	12.64/11.44	12.71/11.48
	$1/5$	12.54/**11.35**	12.52/11.40	12.56/11.41	12.65/11.44	12.68/11.47
	1	12.54/**11.35**	12.48/11.38	12.55/11.41	12.63/11.44	12.66/11.46
	5	12.55/**11.35**	12.50/11.38	12.53/11.38	12.59/11.40	12.63/11.43
	10	12.56/**11.35**	**12.41**/**11.35**	12.50/11.37	12.56/11.40	12.58/11.41
30%, BM-12	$1/10$	8.35/5.08	8.29/5.05	8.32/5.09	8.36/5.11	8.41/5.15
	$1/5$	8.35/5.08	8.28/5.05	8.28/5.07	8.33/5.10	8.37/5.13
	1	8.34/5.07	8.27/5.04	8.25/5.05	8.28/5.08	8.34/5.12
	5	8.28/5.03	8.22/5.01	8.20/5.00	8.18/5.00	8.20/5.01
	10	8.23/5.00	8.17/4.97	8.15/**4.96**	8.16/4.97	**8.14**/**4.96**

- 90%, RM: $\gamma = \lambda$ and $r = 10$;
- 30%, NM: $\gamma = \lambda/5$ and $r = 10$;
- 70%, NM: $\gamma = \lambda/10$ and $r = 15$;
- 30%, BM-4: $\gamma = \lambda$ and $r = 5$.

Table 2: Imputation performance (in MAPE/RMSE) of LRTC-TNN with different truncation values on the Seattle freeway traffic speed dataset (on the first two-week subset for tuning hyperparameters). Note that the best results are highlighted in bold fonts. The number next to the BM denotes the window length.

Missing rate	Truncation parameter				
	$r = 5$	$r = 10$	$r = 15$	$r = 20$	$r = 25$
30%, RM	5.24/3.35	4.97/3.25	4.80/3.18	4.71/3.14	4.63/3.11
70%, RM	6.70/4.14	6.16/3.90	5.86/3.77	5.73/3.71	5.65/3.68
90%, RM	8.41/5.04	7.66/4.73	7.60/4.74	7.83/4.91	8.12/5.09
30%, NM	7.70/4.64	15.40/17.04	19.29/20.32	-/-	-/-
70%, NM	24.44/21.76	79.33/49.93	-/-	-/-	-/-
30%, BM-12	8.38/5.08	46.05/35.61	79.33/50.27	-/-	-/-

Table 3: Imputation performance comparison (in MAPE/RMSE) of LATC and baseline models on the Seattle freeway traffic speed dataset. Note that the best results are highlighted in bold fonts.

Missing rate	LATC	LAMC	LRTC-TNN	BTMF	SPC
30%, RM	5.18/3.23				
70%, RM	6.40/3.83				
90%, RM	8.32/4.80				
30%, NM	8.29/4.81				
70%, NM	12.52/8.04				
30%, BM-12	11.77/6.20				

Table 4: Imputation performance (in MAPE/RMSE) of LATC with different truncation values and different ratios γ/λ on the Portland traffic volume dataset (on the first two-week subset for tuning hyperparameters). Note that the best results are highlighted in bold fonts. The number next to the BM denotes the window length.

Missing rate	γ/λ	Truncation parameter				
		$r = 5$	$r = 10$	$r = 15$	$r = 20$	$r = 25$
30%, RM	$1/10$	19.50/17.43	19.13/16.97	19.04/17.11	18.97/17.23	18.85/17.41
	$1/5$	19.49/17.41	19.12/16.96	19.03/17.05	18.95/17.16	18.83/17.26
	1	19.40/17.28	19.03/**16.88**	18.93/16.97	18.84/16.97	18.73/16.99
	5	19.25/17.56	18.96/17.41	18.83/17.44	18.76/17.40	**18.65**/17.35
	10	19.28/18.23	19.06/18.22	18.94/18.26	18.88/18.23	18.79/18.20
70%, RM	$1/10$	21.79/20.34	21.15/20.08	20.99/19.75	20.89/19.59	20.91/19.58
	$1/5$	21.77/20.31	21.10/19.87	20.95/19.65	20.85/19.51	20.84/19.51
	1	21.66/20.21	20.99/19.76	20.84/19.49	20.71/19.38	20.68/**19.36**
	5	21.48/20.43	20.97/20.28	20.83/20.36	20.76/20.28	**20.67**/20.35
	10	21.40/20.83	21.01/20.92	20.93/21.09	20.87/21.07	20.81/21.15
90%, RM	$1/10$	25.19/24.49	24.71/24.39	24.83/24.50	25.03/24.74	25.41/24.95
	$1/5$	25.16/24.46	24.62/24.17	24.65/24.12	24.89/24.35	25.25/24.53
	1	25.17/24.58	24.35/**24.00**	24.48/24.07	24.56/24.20	24.81/24.35
	5	25.01/24.86	24.22/24.90	24.26/25.11	24.31/25.26	24.46/25.42
	10	24.91/25.50	24.29/26.80	**24.12**/27.53	24.18/27.84	24.44/28.24
30%, NM	$1/10$	22.01/22.61	21.62/22.65	21.67/23.52	21.81/23.75	21.82/23.87
	$1/5$	22.01/22.65	**21.55**/**22.60**	21.69/23.70	21.84/23.87	21.82/23.72
	1	22.05/22.75	21.64/23.45	25.24/52.86	25.46/52.94	25.39/52.78
	5	22.37/23.56	26.75/55.29	51.87/124.84	52.52/125.87	52.89/125.99
	10	22.88/24.44	38.10/89.49	64.03/148.00	64.31/148.53	64.86/148.72
70%, NM	$1/10$	52.17/108.21	57.34/112.18	**48.04**/**97.43**	48.29/97.47	48.51/97.46
	$1/5$	53.33/110.79	59.89/116.05	54.61/109.94	54.75/109.95	55.02/109.99
	1	51.04/105.28	72.48/138.43	73.73/143.10	73.78/143.10	74.08/143.09
	5	55.95/114.62	82.36/157.25	86.18/161.35	85.83/161.71	85.39/161.38
	10	53.63/109.04	80.04/156.88	90.15/168.07	89.45/168.49	89.16/168.30
30%, BM-4	$1/10$	31.18/57.15	33.34/75.29	34.16/79.52	34.39/79.76	34.49/79.87
	$1/5$	26.96/40.67	31.16/65.41	31.74/69.04	31.99/69.27	31.98/69.40
	1	24.31/**23.07**	26.25/42.89	26.19/44.60	26.37/44.79	26.33/44.87
	5	23.55/23.08	23.44/23.73	23.52/25.98	23.53/26.10	23.70/26.28
	10	23.29/23.35	**23.15**/24.31	23.41/25.77	23.50/25.89	23.76/26.05

Table 5: Imputation performance comparison (in MAPE/RMSE) of LATC and baseline models on the Portland highway traffic volume dataset. Note that the best results are highlighted in bold fonts.

Missing rate	LATC	LAMC	LRTC-TNN	BTMF	SPC
30%, RM	17.57/15.81				
70%, RM	19.60/19.20				
90%, RM	23.84/23.78				
30%, NM	20.77/21.42				
70%, NM	50.98/108.83				
30%, BM-12	24.80/24.51				