1 Forecasting Australian Domestics Tourism

Domestic tourism flows in Australia exhibit a natural hierarchical and grouped structure, driven both by geography and by purpose of travel. At the top of this hierarchy lies the national total, which splits into the seven states and territories. Each state is further subdivided into tourism zones, which in turn break down into 77 regions. A complete illustration of this geographic hierarchy appears in Appendix Section 2.1. Intersecting this geographic hierarchy is a second dimension–travel motive–partitioning tourism flows into four categories: holiday, business, visiting friends and relatives, and other. Altogether, this yields a grouped system of 560 series, from the most disaggregated regional-purpose cells up to the full national aggregate. Table 1 depicts this structure.

Table 1: Hierarchical and grouped structure of Australian domestic tourism flows

Geographical division	Number of series per geographical division	Number of series per purpose	Total number of series
Australia	1	4	5
States	7	28	35
Zones	27	108	135
Regions	77	308	385
Total	112	448	560

We quantify tourism demand via "visitor nights", the total number of nights spent by Australians away from home. The data is collected via the National Visitor Survey, managed by Tourism Research Australia, using computer assisted telephone interviews from nearly 120,000 Australian residents aged 15 years and over (*Tourism Research Australia*, 2024).

The data are monthly time series spanning from January 1998 to December 2016, resulting in 228 observations per series, producing a canonical "$n \ll p$" setting which is ideal for evaluating reconciliation approaches that rely on high-dimensional covariance estimation. The extreme dimensionality over sample size mirrors many contemporary business problems, for instance, Starbucks drink sales. Tourism demand is also economically vital yet highly volatile, with geographical and purpose-specific patterns create a realistic stress-test for reconciliation algorithms.

Wickramasuriya et al. (2019) also argued that modelling spatial autocorrelations directly from the start would be challenging as in this case of a large collection of time series. Post-processing reconciliation approaches have the advantage to implicitly model this spatial autocorrelation structure, especially true for MinT.

To assess forecasting performance, we adopt a rolling-window cross-validation scheme. Beginning with the first 120 monthly observations (January 1998-December 2005) as the initial training set, we obtain the best-fitted ARIMA model for each of the 560 series via the automatic algorithm by minimising AICc from Hyndman & Khandakar (2008). The 1- to 12-step-ahead base forecasts are then generated by these ARIMA models, and then reconciled using multiple approaches. We then roll the training window forward by one month and refit all models, rebuild reconciliations, and produce another batch of 1- to 12-step-ahead forecasts, repeating until the training set reaches December 2015. In total, this results in 133 out-of-sample windows.

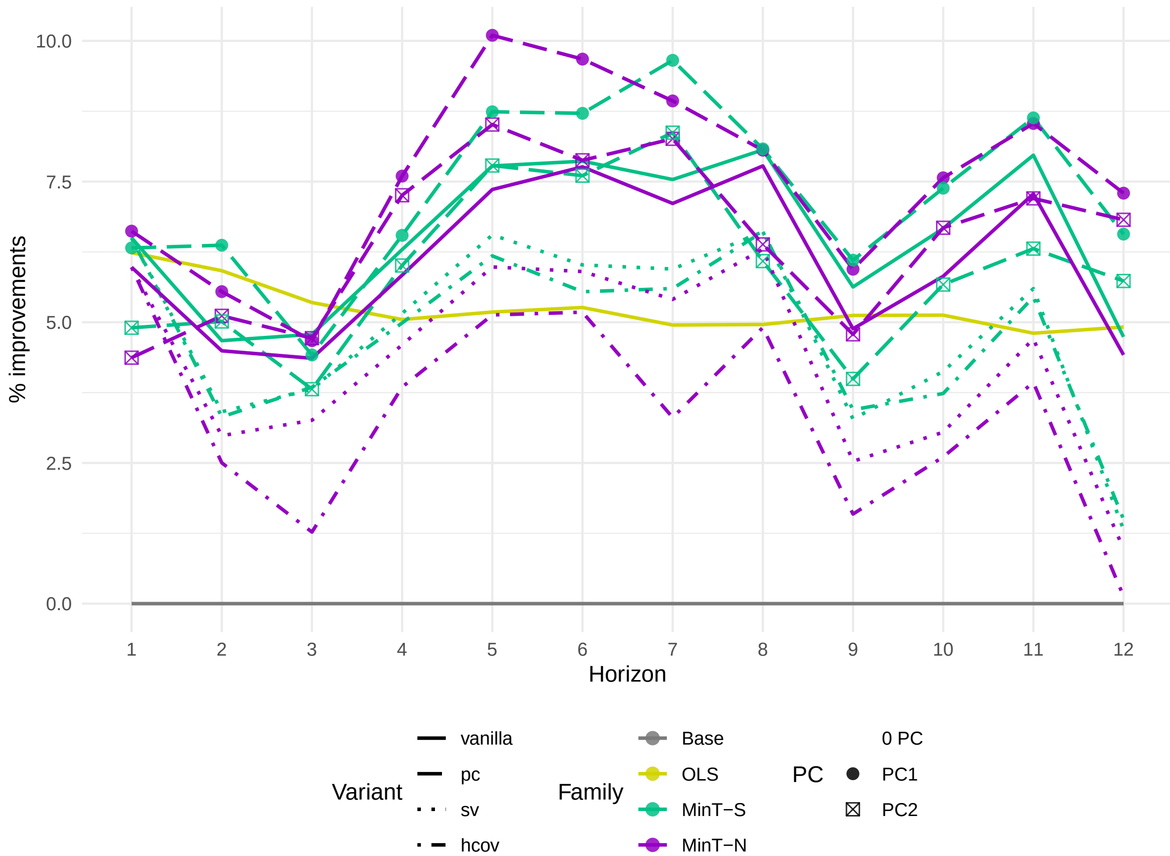

Figure 1: Percentage relative improvement in the mean squared error (MSE) of different reconciled forecasts over the base forecasts for the Australian domestic tourism data, for 1 to 12 steps ahead forecasts. The positive entries indicate an decrease in MSE.

Figure 1 and Figure 2 show that reconciliation is beneficial for point forecasts: across horizons $h = 1, 2, \ldots, 12$, all reconciled methods improve MSE over incoherent base ARIMA on average. Among vanilla MinT variants, the shrinkage estimator (*MinT-S*) marginally outperforms NOVELIST (*MinT-N*) at most horizons, though the gap is small. Adjusting for a single dominant factor via PC decomposition tightens performance further: both *MinT-S(PC1)* and *MinT-N(PC1)* beat their unadjusted counterparts, with *MinT-N(PC1)* narrowly ahead. Adding more than one PC brings no additional benefit, likely due to injecting estima-

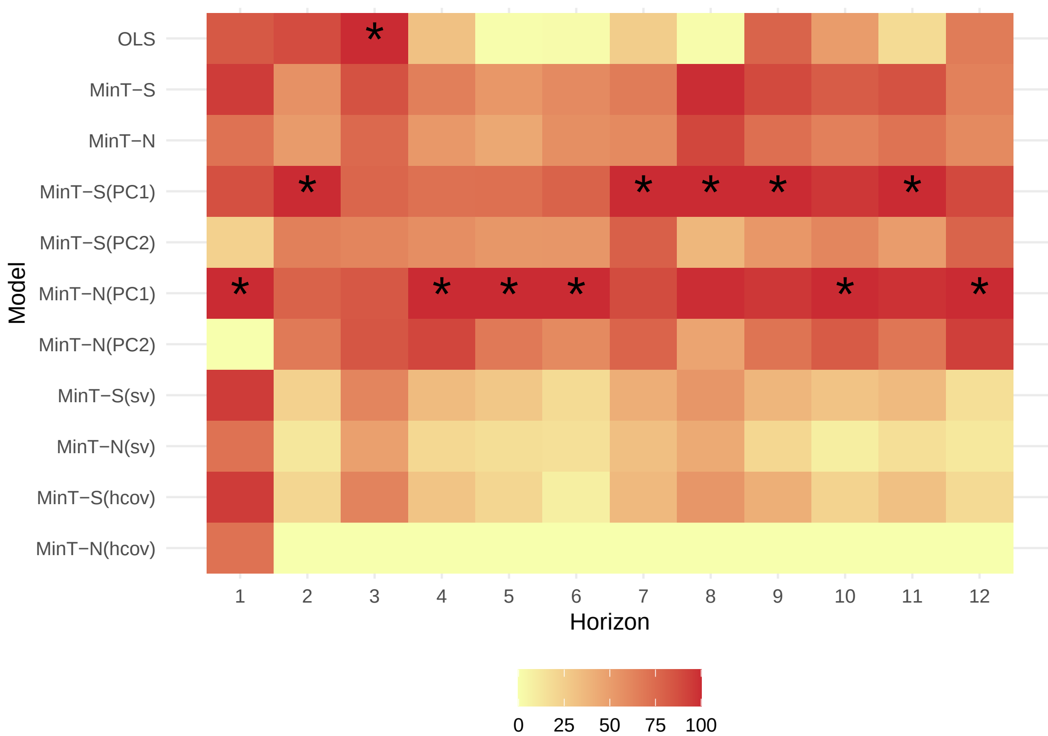

Figure 2: Heatmap of relative improvement in the mean squared error (MSE) of different reconciled forecasts over the base forecasts for the Australian domestic tourism data, for 1 to 12 steps ahead forecasts. The values are scaled to the range of 0 to 100 for better visualisation, with darker colors indicating greater improvement and best performance is noted by a star.

tion noise from weaker components. Variants that modify the multi-step covariance, either via scaled-variance or direct h-step residual covariances, underperform standard MinT, suggesting that extra estimation at horizon $h > 1$ is not rewarded in this setting.

Turning to probabilistic forecasts (1-step-ahead forecasts), Figure 3 shows that *MinT-N* consistently outperforms *MinT-S* across CRPS and both Winkler scores, and PC1-adjustment again yields the best improvements. Results are aligned across probabilistic scoring rules, but the 95% Winkler reveals a notable pattern, as all MinT variants lose to the base while *OLS* performs best. In the multivariate evaluation, the Energy score places *OLS* close to the PC1-adjusted MinT methods, indicating that simple coherent projection might be able to capture a substantial share of cross-series dependence benefits even without sophisticated covariance regularisation.

Discuss the 95% empirical coverage

The summary radar graph in Figure 4 consolidates these findings: among the selected MinT models by probabilistic criteria, *MinT-N(PC1)* clearly leads across CRPS, W80, W95, and Energy, extending the improvements seen in the single-metric panels.

Taken together, the evidence supports the use of reconciliation for both point and probabilistic forecasts in this high-dimensional setting. For point forecasts, MinT with shrinkage is a solid default; for probabilistic forecasts, NOVELIST performs more reliably. PC adjustment with a single dominant factor consistently enhances performance and should be considered when a dominant latent factor is evident. More complex adjustments, such as multiple PCs or horizon-specific covariances, add estimation noise without clear benefits and can be omitted for parsimony.

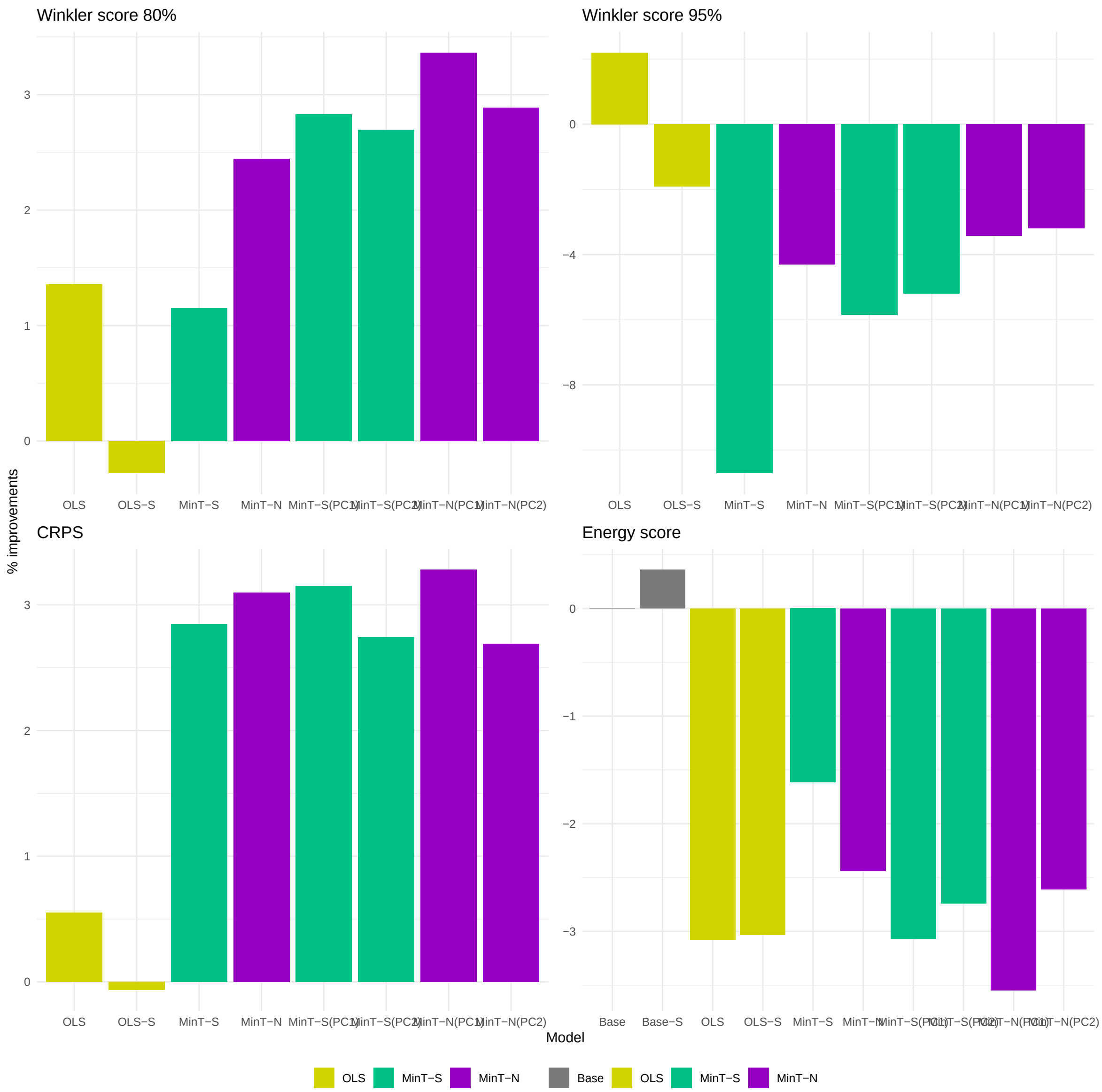

Figure 3: Percentage relative improvement in the Winkler score at 80% and 95% nominal coverage, CRPS, and Energy score of multiple reconciled forecasts over the base forecasts for the Australian domestic tourism data, for 1-step-ahead forecasts. The positive entries indicate an decrease in the probabilistic scores.

Figure 4: Radar plot of relative improvements in probabilistic scores (Winkler 80, Winkler 95, CRPS, Energy) over the base forecasts. The scores are scaled to a range of 0 to 100, with larger values indicating better performance. Only the top 5 models are shown.

2 Appendix

2.1 Appendix: Australian Domestic Tourism Geographical Hierarchy

Table 2: Geographical divisions of Australia.

Series	Name	Label	Series	Name	Label
1	Australia	Total	57	Bundaberg	CAA
2	NSW	A	58	Capricorn	CAB
3	NT	B	59	Fraser Coast	CAC
4	QLD	C	60	Gladstone	CAD
5	SA	D	61	Mackay	CAE
6	TAS	E	62	Southern Queensland Country	CAF
7	VIC	F	63	Outback Queensland	CBA
8	WA	G	64	Brisbane	CCA
9	ACT	AA	65	Gold Coast	CCB
10	Metro NSW	AB	66	Sunshine Coast	CCC
11	Nth Coast NSW	AC	67	Townsville	CDA
12	Nth NSW	AD	68	Tropical North Queensland	CDB
13	Sth Coast NSW	AE	69	Whitsundays	CDC
14	Sth NSW	AF	70	Clare Valley	DAA
15	Central NT	BA	71	Flinders Ranges and Outback	DAB
16	Nth Coast NT	BB	72	Murray River, Lakes and Coorong	DAC
17	Central Coast QLD	CA	73	Riverland	DAD
18	Inland QLD	CB	74	Adelaide	DBA
19	Metro QLD	CC	75	Adelaide Hills	DBB
20	Nth Coast QLD	CD	76	Barossa	DBC
21	Inland SA	DA	77	Fleurieu Peninsula	DCA
22	Metro SA	DB	78	Kangaroo Island	DCB
23	Sth Coast SA	DC	79	Limestone Coast	DCC
24	West Coast SA	DD	80	Eyre Peninsula	DDA
25	Nth East TAS	EA	81	Yorke Peninsula	DDB
26	Nth West TAS	EB	82	East Coast	EAA
27	Sth TAS	EC	83	Launceston and the North	EAB
28	East Coast VIC	FA	84	North West	EBA
29	Metro VIC	FB	85	West Coast	EBB
30	Nth East VIC	FC	86	Hobart and the South	ECA
31	Nth West VIC	FD	87	Gippsland	FAA
32	West Coast VIC	FE	88	Lakes	FAB
33	Nth WA	GA	89	Phillip Island	FAC
34	Sth WA	GB	90	Geelong and the Bellarine	FBA
35	West Coast WA	GC	91	Melbourne	FBB
36	Canberra	AAA	92	Peninsula	FBC
37	Central Coast	ABA	93	Central Murray	FCA
38	Sydney	ABB	94	Goulburn	FCB
39	Hunter	ACA	95	High Country	FCC
40	North Coast NSW	ACB	96	Melbourne East	FCD
41	Blue Mountains	ADA	97	Murray East	FCE
42	Central NSW	ADB	98	Upper Yarra	FCF
43	New England North West	ADC	99	Ballarat	FDA

44	Outback NSW	ADD	100	Bendigo Loddon	FDB
45	South Coast	AEA	101	Central Highlands	FDC
46	Capital Country	AFA	102	Macedon	FDD
47	Riverina	AFB	103	Mallee	FDE
48	Snowy Mountains	AFC	104	Spa Country	FDF
49	The Murray	AFD	105	Western Grampians	FDG
50	Alice Springs	BAA	106	Wimmera	FDH
51	Barkly	BAB	107	Great Ocean Road	FEA
52	Lasseter	BAC	108	Australia's North West	GAA
53	MacDonnell	BAD	109	Australia's Golden Outback	GBA
54	Darwin	BBA	110	Australia's Coral Coast	GCA
55	Katherine Daly	BBB	111	Australia's South West	GCB
56	Litchfield Kakadu Arnhem	BBC	112	Destination Perth	GCC

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, *27*, 1–22. https://doi.org/10.18637/JSS.V027.I03

Tourism research australia. (2024). https://www.tra.gov.au/.

Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J. (2019). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association*, *114*(526), 804–819. https://doi.org/10.1080/01621459.2018.1448825