Supplementary Material for Forecasting Surface O_3 in Texas Urban Areas Using Random Forest and Generalized Additive Models

SUPPLEMENTAL DATA

Supplement Table 1. Significance of different predictors in the quantitative forecast models. Significance codes are: '***', significant at the α = 0.001 level; '**', significant at the α = 0.01 level; '*', significant at the α = 0.05 level; '.', significant at the α = 0.1 level. The top table is the forecasting model as described in Table 2, and each sub-table underneath it represents a different sensitivity study with additional predictors. The four sensitivity studies were replacing ρ_{WV} with H_R and T_{DP}, replacing ρ_{WV} with T_{DP} only, replacing ρ_{WV} with H_R only, and the Table 2 model with ΔT_{NCEP}.

Predictor	ARR	BPA	DFW	HGB	SA	TLM
D_w	***	***	***	***	***	***
D_i	***	***	***	***	***	***
T_{2100}	***	***	***	***	*	***
ΔT	.	.	***	*	***	*
ρ_{wv}	***	***	***	***	***	***
W_S	***	***	***	***	***	***
W_D	*	***	.	***	.	
$\hat{O}_{3,\ MDA8}$	***	***	***	***	***	***
H_R	**		***	.	***	***
T_{DP}	***	***	***	***	***	*
T_{DP}	***	***	**	***	***	***
H_R	***	***	***	***	***	***
ΔT_{NCEP}	***		.		***	

Supplement Table 2. Summary statistics for the probabilistic forecast model training.

Urban Area	Threshold (ppbv)	R^2	Variance Explained (%)
DFW	55	0.549	49.7
	71	0.531	51.7
	86	0.287	41.0
HGB	55	0.615	56.5
	71	0.601	60.6
	86	0.443	54.1
ARR	55	0.626	59.8
	71	0.425	53.4
BPA	55	0.592	55.5
	71	0.487	56.7
SA	55	0.534	49.7
	71	0.521	59.0
TLM	55	0.519	47.7
	71	0.371	45.6

Supplement Table 3. Similar to Supplement Table 1, but for the probabilistic models.

Urban Area	Threshold (ppbv)	D_w	D_i	T_{2100}	ΔT	ρ_{WV}	W_S	W_D	$\hat{O}_{3,\,MDA8}$
DFW	55		***	***	***	**	***		***
	71	***	***		***	*	***	.	***
	86	***			***	*	**	*	***
HGB	55	***	***	***	*	***	***	***	***
	71	***	***	***	*	***	***	.	***
	86	***	***	***		***	***		***
ARR	55	***	***	***	.	***	***	.	***
	71	***	**	**		**	***	.	***
BPA	55	***	***	***		***	***	*	***
	71	***	***	*	.	*	***	*	***
SA	55	***	***		**	**	***		***
	71	***	***	*		.	***		***
TLM	55	***	***	***		**	***		***
	71	***		**		*	***	*	***

Supplement Table 4. Reliability and ROC Area statistics for the probabilistic model of each urban area.

Urban Area	*Threshold (ppb)*	*Reliability*	*ROC Area*
ARR	55	0.01	0.89
BPA	55	0.01	0.78
DFW	55	0.00	0.90
	71	0.01	0.92
HGB	55	0.02	0.82
	71	0.00	0.78
SA	55	0.01	0.91
TLM	55	0.03	0.92
	71	0.01	0.99

Supplement Table 5. Confusion matrix for the Austin/Round Rock O_3 classification model. The columns represent the classifications based on $O_{3,\,MDA8}$ calculations from TAMIS O_3 data, while rows represent the forecasted class.

Austin	***Green Observation***	***Yellow Observation***	***Orange Observation***
Green Forecast	130	12	0
Yellow Forecast	7	4	1
Orange Forecast	0	0	0

Supplement Table 6. Same as Supplement Table 5, but for the Beaumont/Port Arthur urban area.

Beaumont	***Green Observation***	***Yellow Observation***	***Orange Observation***
Green Forecast	130	15	0
Yellow Forecast	5	3	1
Orange Forecast	0	0	0

Supplement Table 7. Same as Supplement Table 5, but for the Dallas/Fort Worth urban area.

Dallas	***Green Observation***	***Yellow Observation***	***Orange Observation***	***Red Observation***
Green Forecast	81	15	0	0
Yellow Forecast	8	35	8	0
Orange Forecast	0	3	3	1
Red Forecast	0	0	0	0

Supplement Table 8. Same as Supplement Table 5, but for the Houston/Galveston/Brazoria urban area.

Houston	***Green Observation***	***Yellow Observation***	***Orange Observation***
Green Forecast	93	35	7
Yellow Forecast	3	10	6
Orange Forecast	0	0	0

Supplement Table 9. Same as Supplement Table 5, but for the San Antonio urban area.

San Antonio	***Green Observation***	***Yellow Observation***	***Orange Observation***
Green Forecast	126	10	1
Yellow Forecast	4	5	1
Orange Forecast	0	0	0

Supplement Table 10. Same as Supplement Table 5, but for the Tyler/Longview/Marshall urban area.

Tyler	***Green Observation***	***Yellow Observation***
Green Forecast	132	11
Yellow Forecast	4	7

Supplement Table 11. Categories for NAM MOS cloud forecasts and their associated values for quantitative forecasting.

MOS CLD Category	***Definition***	***Model Assignment***
CL	Clear	0
FW	0 to 2 octas of sky cover	1
SC	2-4 octas of sky cover	2
BK	6-8 octas of sky cover	3
OV	8 octas (total) of sky cover	4

Supplement Table 12. Statistics for $O_{3,\ MDA8}$ measurements and predictions for day D_i using day D_{i-2} $O_{3,\ MDA8}$ and MOS cloud forecasts as predictors. Compare with Table 6.

Urban Area	*a*	*b*	*r*	*μ (ppbv)*	*σ (ppbv)*
ARR	0.64	14.21	0.76	-1.35	6.39
BPA	0.59	13.60	0.71	-3.18	8.84
DFW	0.63	21.08	0.75	1.15	8.43
HGB	0.46	21.72	0.62	-6.23	11.01
SA	0.70	15.03	0.75	2.39	7.25
TLM	0.59	19.92	0.76	2.71	7.25

Supplement Table 13. Statistics for $O_{3,\ MDA8}$ measurements and predictions for day D_{i+1} using day D_{i-2} $O_{3,\ MDA8}$ and MOS cloud forecasts as predictors. Compare with Table 6.

Urban Area	*a*	*b*	*r*	*μ (ppbv)*	*σ (ppbv)*
ARR	0.45	22.39	0.69	-1.71	7.51
BPA	0.44	20.32	0.69	-2.45	8.83
DFW	0.55	25.68	0.71	1.74	8.86
HGB	0.34	27.64	0.56	-6.57	11.48
SA	0.60	20.74	0.72	3.46	7.94
TLM	0.45	26.99	0.68	3.86	8.04

Supplement Table 14. Statistics for $O_{3,\ MDA8}$ measurements and predictions for day D_i using day D_{i-1} $O_{3,\ MDA8}$ and MOS cloud forecasts as predictors. Compare with Table 6.

Urban Area	*a*	*b*	*r*	*μ (ppbv)*	*σ (ppbv)*
ARR	0.66	13.69	0.77	-1.25	6.60
BPA	0.62	13.06	0.76	-2.54	8.00
DFW	0.70	16.80	0.79	0.78	7.78
HGB	0.52	20.10	0.68	-4.95	10.21
SA	0.76	12.17	0.79	1.76	7.07
TLM	0.65	17.36	0.76	2.68	7.23