Atmospheric Chemistry and Physics

Supporting Information for

Rethinking Machine Learning Weather Normalisation: A Refined Strategy for Short-term Air Pollution Policies

Yuqing Dai[1], Bowen Liu[2], Chengxu Tong[1], David C. Carslaw[3], A. Rob MacKenzie[1], Zongbo Shi[1]

[1]School of Geography, Earth and Environmental Science, University of Birmingham, Birmingham, B15 2TT, U.K.

[2]Department of Management, Business School, University of Birmingham, Birmingham, B15 2TT, U.K.

[3]Wolfson Atmospheric Chemistry Laboratories, University of York, York, YO10 5DD, U.K.

Correspondence to: A. Rob MacKenzie (a.r.mackenzie@bham.ac.uk), Zongbo Shi (z.shi@bham.ac.uk)

Contents of this file

Model performance metrics

The following formulas are the model performance metrics used in our study, including faction of predictions within a factor of two (FAC2), fractional bias (FB), coefficient of correlation (r), root mean square error (RMSE), systematic RMSE ($RMSE_s$), unsystematic RMSE ($RMSE_\mu$), index of agreement (IOA):

$$0.5 \le \frac{M_i}{O_i}(FAC2) \le 2.0 \tag{1}$$

$$FB = \frac{\overline{M}-\overline{O}}{0.5(\overline{M}+\overline{O})} \tag{2}$$

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{M_i-\overline{M}}{\delta_M}\right)\left(\frac{O_i-\overline{O}}{\delta_O}\right) \tag{3}$$

$$RMSD = \sqrt{\frac{\sum_{i=1}^{n}(M_i-O_i)^2}{n}} \tag{4}$$

$$RMSD_u = \sqrt{\frac{\sum_{i=1}^{n}(M_i-\widehat{M}_i)^2}{n}} \tag{5}$$

$$RMSD_s = \sqrt{\frac{\sum_{i=1}^{n}(\widehat{M}_i-O_i)^2}{n}} \tag{6}$$

$$IOA = \begin{cases} 1-\frac{\sum_{i=1}^{n}|M_i-O_i|}{2\sum_{i=1}^{n}|O_i-\overline{O}|}, \sum_{i=1}^{n}|M_i-O_i| \le 2\sum_{i=1}^{n}|O_i-\overline{O}| \\ \frac{2\sum_{i=1}^{n}|O_i-\overline{O}|}{\sum_{i=1}^{n}|M_i-O_i|}-1, \sum_{i=1}^{n}|M_i-O_i| > 2\sum_{i=1}^{n}|O_i-\overline{O}| \end{cases} \tag{7}$$

where M_i and O_i denote the i^{th} rank of values from the modelling and observations, respectively; n is the total number of a continuous dataset; δ represents the standard deviation; a bar represents the mean value (e.g., $\overline{M}$); $\widehat{M}_i$ is derived from:

$$\widehat{M}_i = a + b \times O_i \tag{8}$$

a is intercept and b is the slope. Then, we have:

$$RMSD^2 = {RMSD_u}^2 + {RMSD_s}^2 \tag{9}$$

The IOA serves as a normalized metric to quantify the accuracy of model predictions, with its values ranging from -1 to 1. An IOA value of 1 denotes an ideal match between model predictions and observed data. Conversely, a value of -1 represents a complete discordance, indicating an absence of any predictive agreement (Willmott et al., 2012).

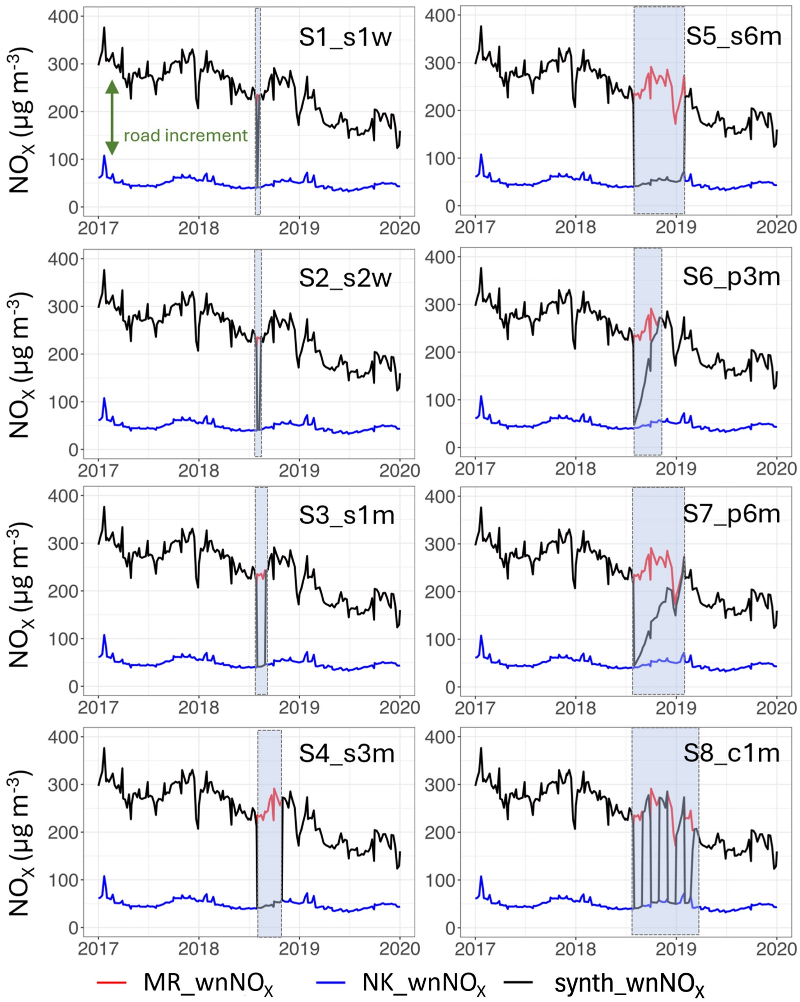
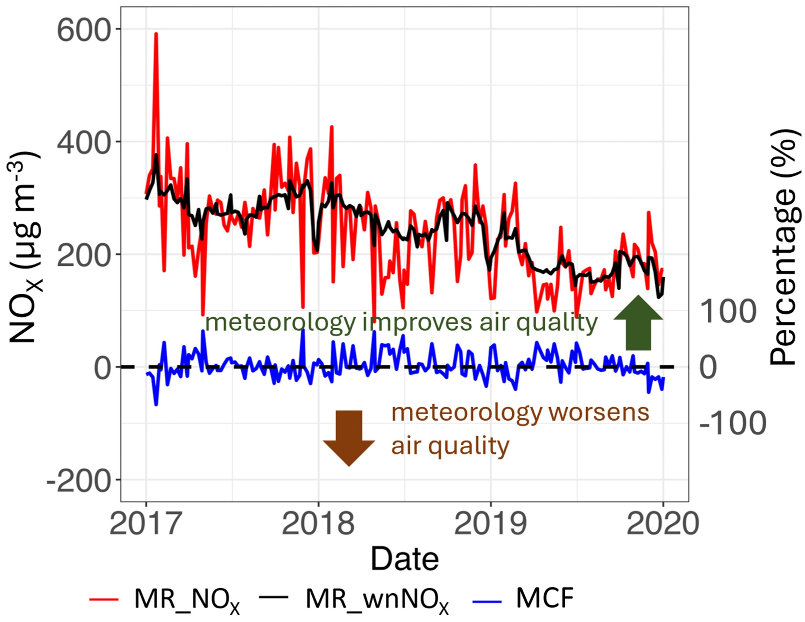


Figure S1: Temporal variation of weekly averaged NO_X concentrations after weather normalisation at Marylebone Road (MR_wnNO_X, red line) and North Kensington (NK_wnNO_X, blue line), 2017-2020. The synthetic weather normalised NO_X concentrations (synth_wnNO_X) are denoted by the black line, closely aligning with MR_wnNO_X except during the light blue shaded interval indicative of a policy intervention. The differences between synth_wnNO_X and MR_wnNO_X only attribute to emission changes.



50 **Figure S2: Temporal variation of original (red line) and weather normalised (black line) weekly averaged NO_X concentrations at Marylebone Road, London, 2017-2020 using the weather normalisation approach (ML-WN, denoted as "wn"). The blue line with a unit of percentage (%) represents the extent of "meteorological contribution" (MCF) on NO_X levels; positive values indicate the meteorology contributes for NO_X mitigation, while negative values represent the meteorology led to adverse conditions.**

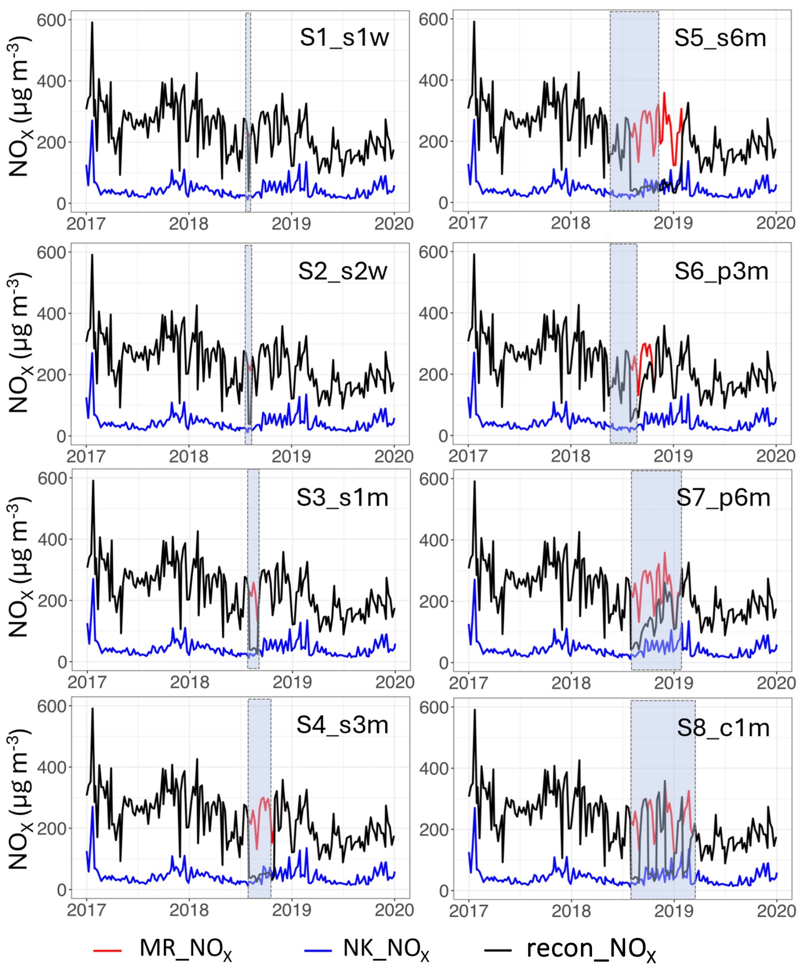
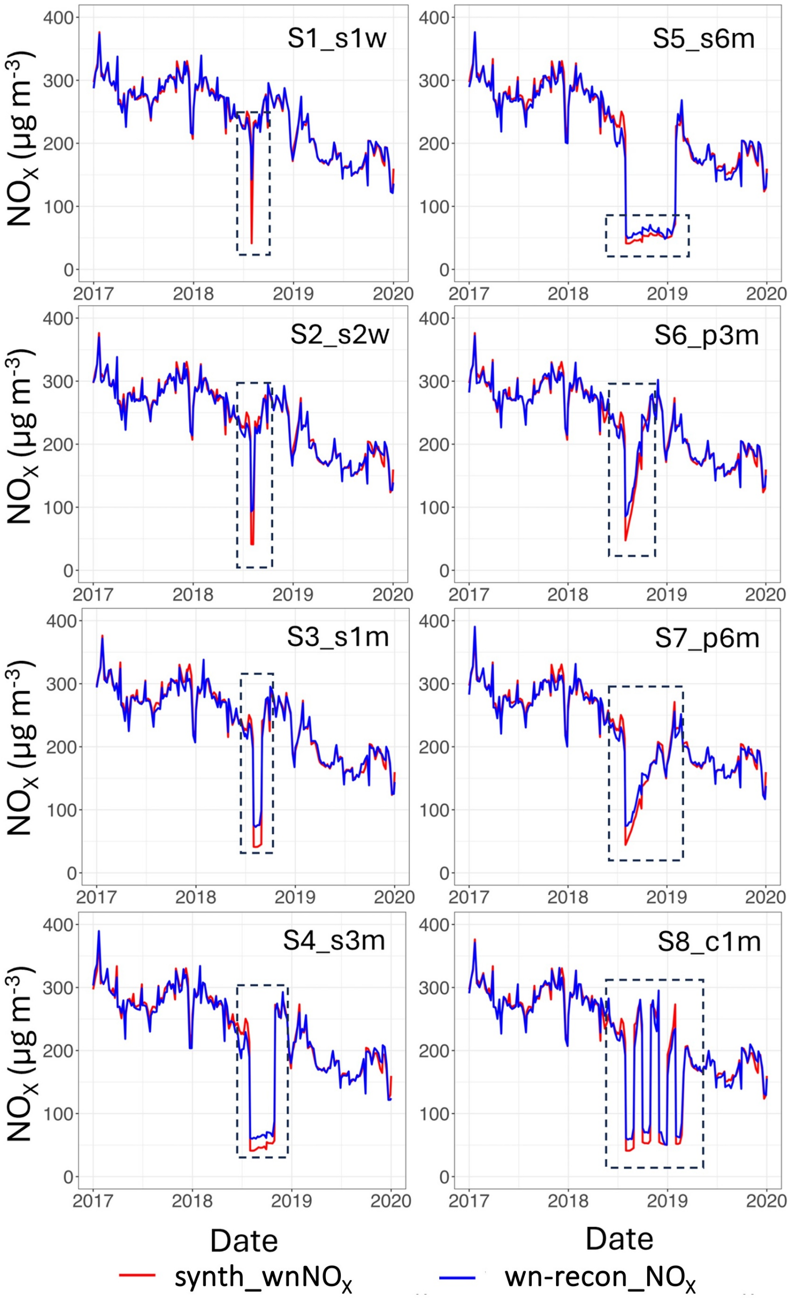


Figure S3: Temporal variation of original weekly averaged NO_x concentrations at Marylebone Road (MR_NO_X, red line) and North Kensington (NK_NO_X, blue line), 2017-2020. The recon_NO_X concentrations are denoted by the black line and are indistinguishable from the original MR data except during the light blue shaded interval indicative of a policy intervention. The differences between recon_NO_X and MR_NO_X concentrations can be attributed to both emission changes and meteorological effects.



60 **Figure S4: Temporal variation of weekly averaged concentrations for synthetic weather normalised NO_X (synth_wnNO_X, red line) and weather normalised reconstitute NO_X (wn-recon_NO_X, blue line) at Marylebone Road, 2017-2020. The dashed frame indicates the period of interventions occurred.**

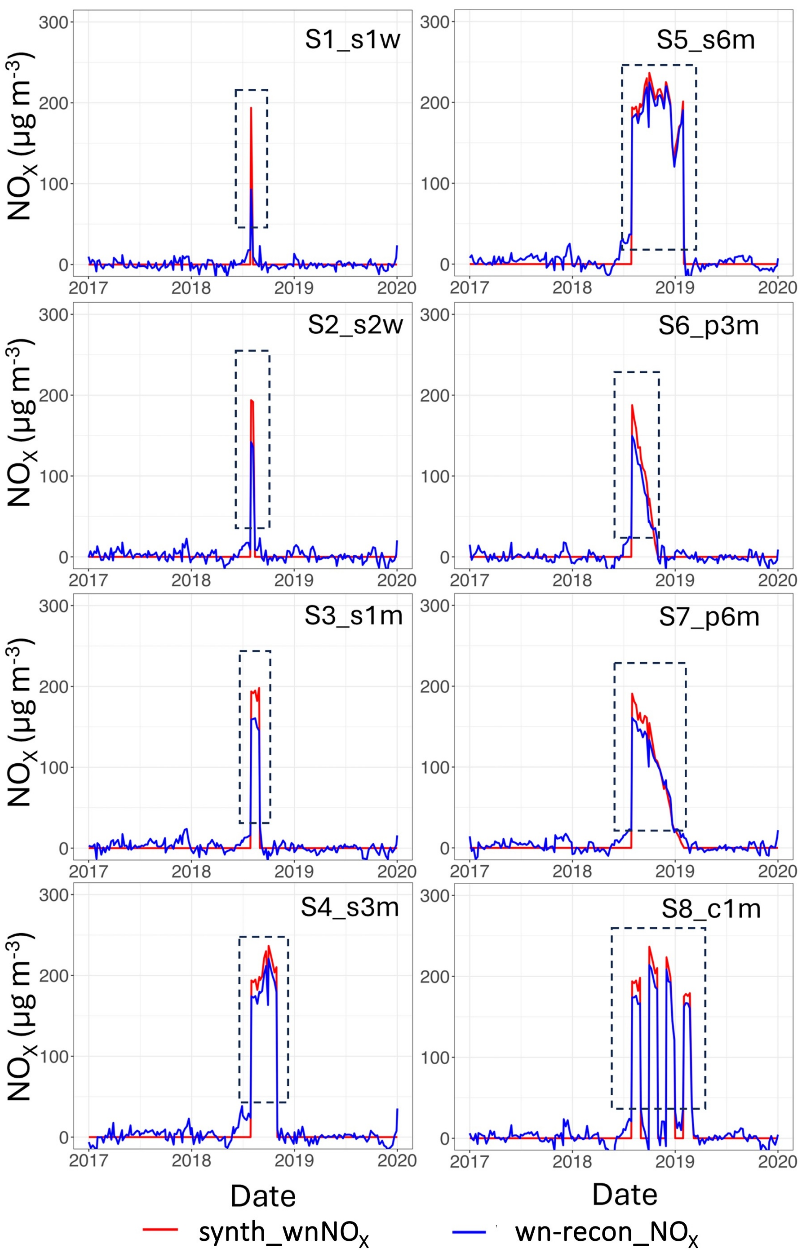


Figure S5: Temporal variation of weekly policy effects on synthetic weather normalised NO_X (synth_wnNO_X, red line) and weather normalised reconstitute NO_X concentrations (wn-recon_NOX, blue line) at Marylebone Road, 2017-2020. The dashed frame indicates the period of interventions occurred.

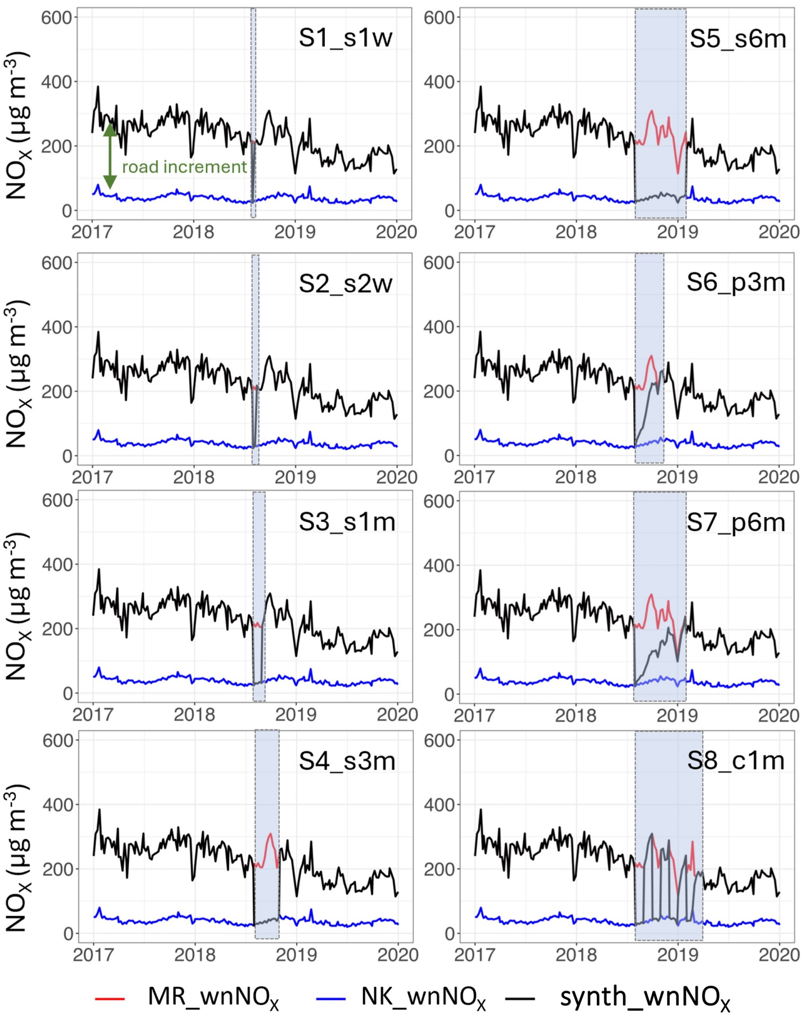


Figure S6: Temporal variation of weekly averaged NO_X concentrations at Marylebone Road (MR_wnNO_X, red line) and North Kensington (NK_wnNO_X, blue line) using the MacLeWN approach, 2017-2020. The synthetic 'updated' weather normalised NO_X concentrations (synth_wnNO_X) are denoted by the black line, closely aligning with the MR data except during the light blue shaded interval indicative of a policy intervention. The differences between synth_wnNO_X and MR_wnNO_X only attribute to emission changes.

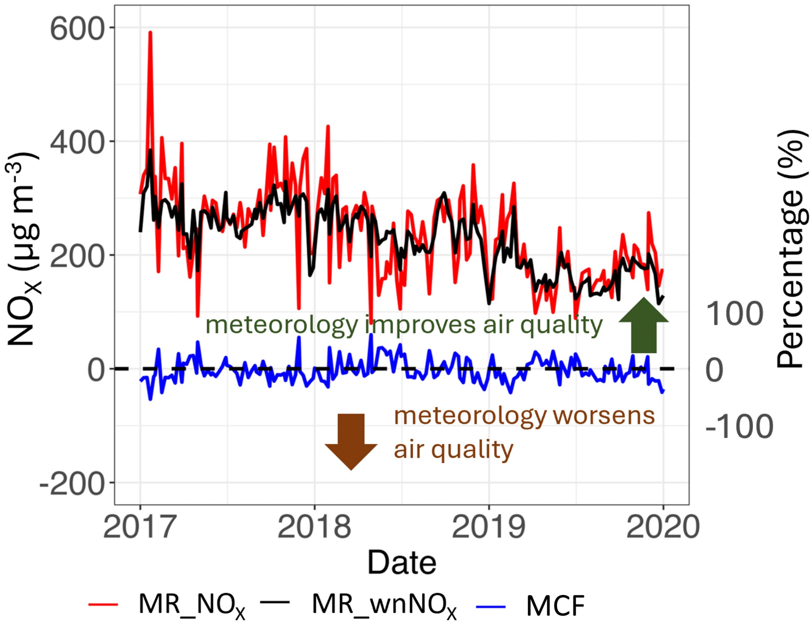


Figure S7: Temporal variation of original (red line) and weather normalised (black line) weekly averaged NO_X concentrations at Marylebone Road from 2017 to 2020 using the updated weather normalisation approach (MacLeWN, denoted as "wn"). The blue line with a unit of percentage (%) represents the extent of "meteorological contribution" (MCF) on NO_X levels; positive values indicate the meteorology contributes for NO_X mitigation, while negative values represent the meteorology led to adverse conditions.

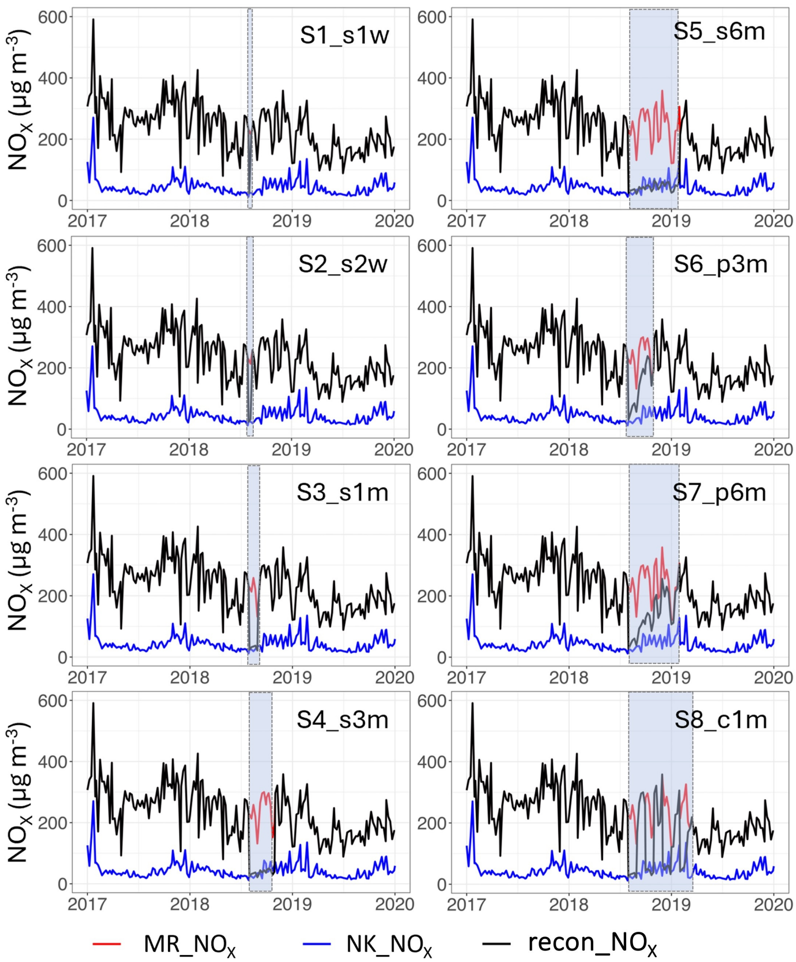
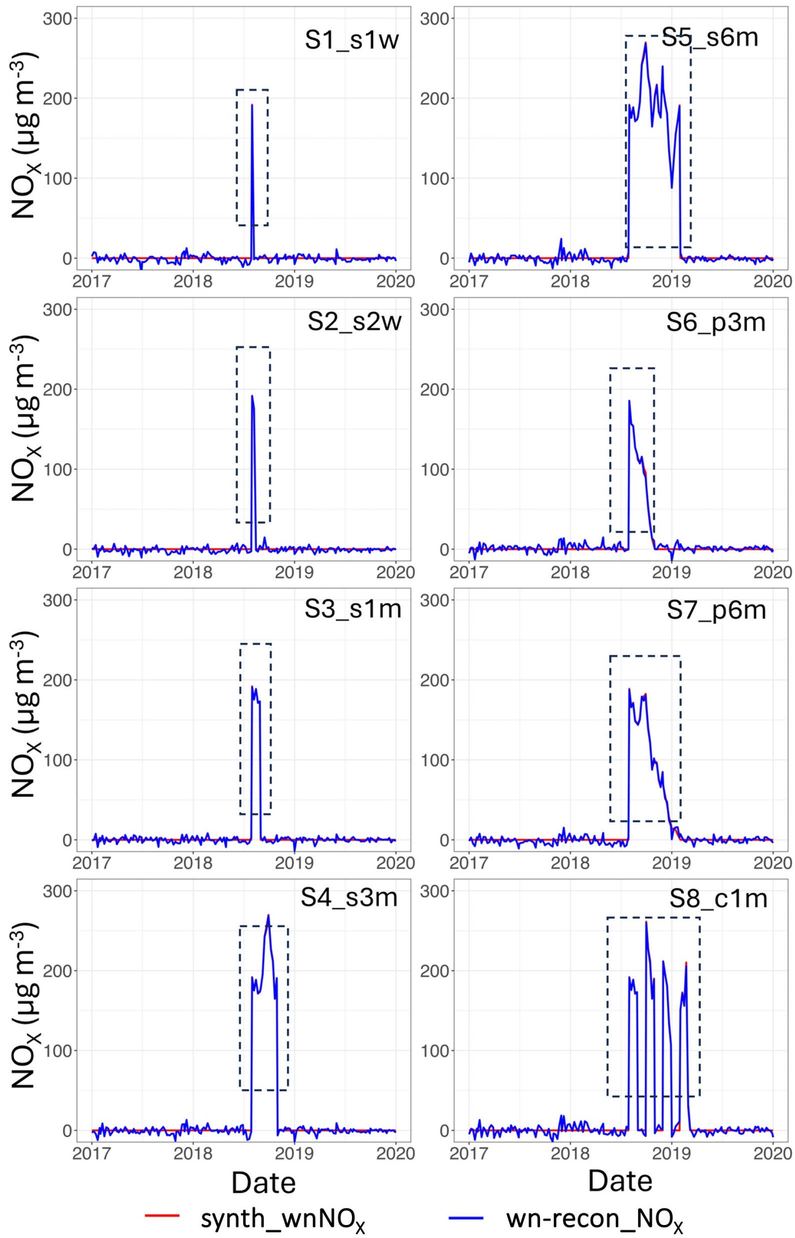
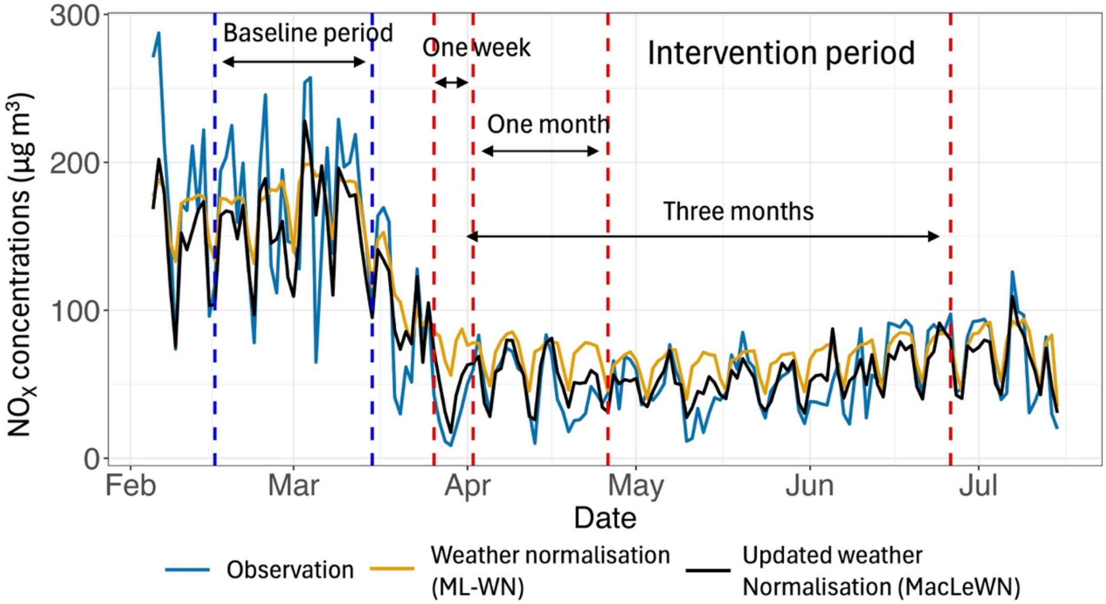


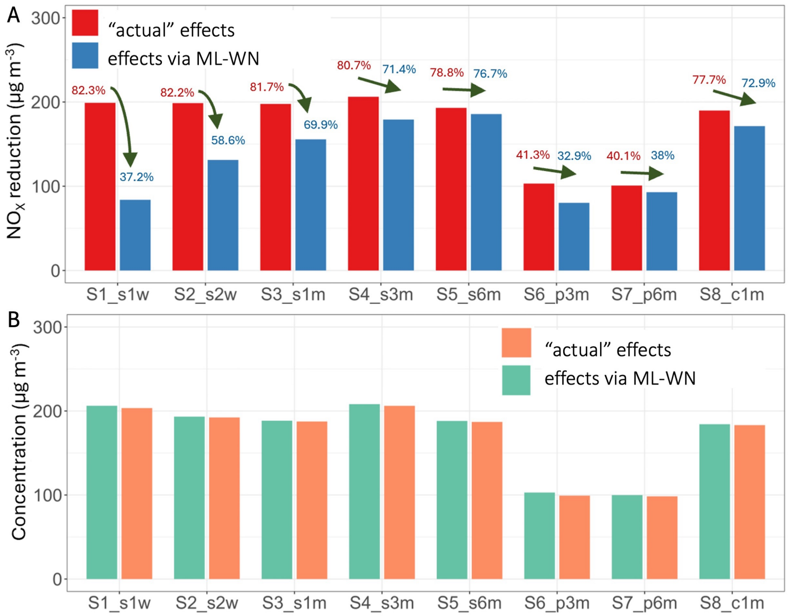
Figure S8: Temporal variation of original weekly averaged NO_X concentrations at Marylebone Road (MR_NO_X, red line) and North Kensington (NK_NO_X, blue line), 2017-2020. The reconstructed NO_X concentrations are denoted by the black line and are indistinguishable from the original MR data except during the light blue shaded interval indicative of a policy intervention. The differences between $recon\_NO_X$ and MR_NO_X concentrations can be attributed to both emission changes and meteorological effects.



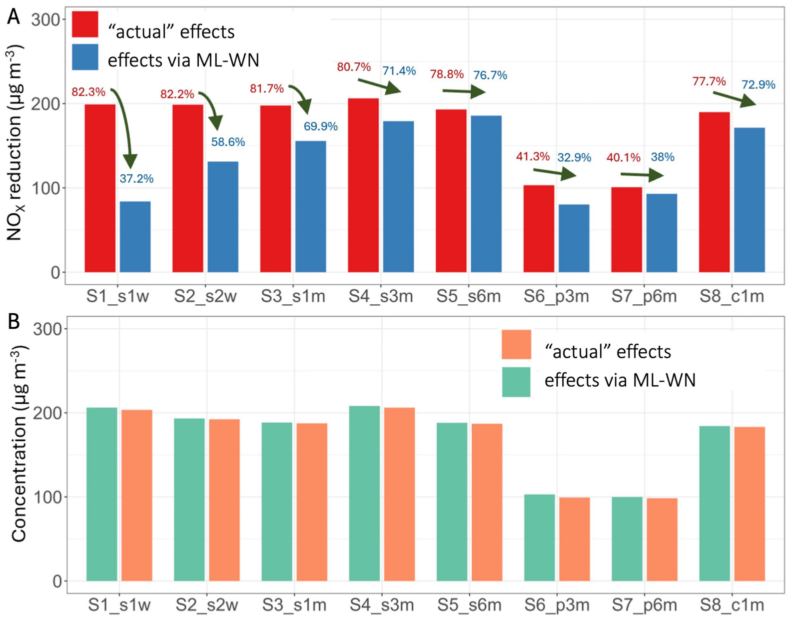
90 **Figure S9: Temporal variation of weekly policy effects on synthetic weather normalised NO_X (synth_wnNO_X, red line) and weather normalised reconstitute (wn-recon_NO_X, blue line) concentrations at Marylebone, 2017-2020, using the updated weather normalisation approach (MacLeWN). The dashed frame indicates the period of interventions occurred.**



95 **Figure S10. Comparison of daily NO_X concentrations at London Marylebone Road using the direct observations, the machine learning weather normalisation approach (ML-WN), and the updated machine learning weather normalisation approach (MacLeWN). Time period starts from 15th Feb to 15th April 2020. The area between blue dashed line indicates the 'baseline period', and the area between red dashed line indicates the implementation of the COVID-19 lockdown measures.**



100 **Figure S11: Average intervention effects under difference scenarios using the XGBoost (eXtreme Gradient Boosting) model for weather normalisation.**



105 **Figure S12: Average intervention effects under difference scenarios using the DRF (distributed random forest) model for weather normalisation.**

Table S1: GBM (gradient boosting machine) model hyper-parameters from the H2O.ai's AutoML for the machine learning weather normalisation (ML-WN).

Model parameters	Marylebone Road	North Kensington
training_frame	Automls	Automl
nfolds	5	5
keep_cross_validation_models	FALSE	FALSE
keep_cross_validation_predictions	TRUE	TRUE
keep_cross_validation_fold_assignment	FALSE	FALSE
score_each_iteration	FALSE	FALSE
score_tree_interval	5	5
fold_assignment	Modulo	Modulo
response_column	y	y
ignore_const_cols	TRUE	TRUE
balance_classes	FALSE	FALSE
max_after_balance_size	5	5
max_confusion_matrix_size	20	20
ntrees	154	110
max_depth	10	11
min_rows	10	5
nbins	20	20
nbins_top_level	1024	1024
nbins_cats	1024	1024
r2_stopping	1.80e+308	1.80e+308
stopping_rounds	0	0
stopping_metric	deviance	Deviance
stopping_tolerance	0.00714	0.00707
max_runtime_secs	0	0
build_tree_one_node	FALSE	FALSE
learn_rate	0.1	0.1
learn_rate_annealing	1	1
distribution	gaussian	gaussian
quantile_alpha	0.5	0.5
tweedie_power	1.5	1.5
huber_alpha	0.9	0.9
sample_rate	0.8	0.9
col_sample_rate	0.8	0.7
col_sample_rate_change_per_level	1	1
col_sample_rate_per_tree	0.8	1
min_split_improvement	1.0e-05	1.0e-04
histogram_type	UniformAdaptive	UniformAdaptive
max_abs_leafnode_pred	1.8 e+308	1.8e+308
pred_noise_bandwidth	0	0
categorical_encoding	Enum	Enum
calibration_method	PlattScaling	PlattScaling
in_training_checkpoints_tree_interval	1	1
gainslift_bins	-1	-1

Table S2: GBM (gradient boosting machine) model performance statistics for the weather normalisation approach (ML-WN).

Metrics	Unit	Marylebone Road	North Kensington
The number of complete pairs of data (n)	-	4908	5004
Fraction of predictions within a factor of two (FAC2)	-	0.97	0.96
Fractional bias (FB)	-	0.00038	0.011
Coefficient of correlation (r)	-	0.93	0.91
Root mean square error (RMSE)	$\mu g\ m^{-3}$	62.26	24.62
Systematic RMSE	$\mu g\ m^{-3}$	25.43	13.38
Unsystematic RMSE	$\mu g\ m^{-3}$	56.83	20.66
Index of agreement (IOA)	-	0.84	0.82

Table S3: Absolute and percentile policy impacts on NO_X concentrations under different scenarios. "actual" represents idealised policy effects, and "wn" represents "weather normalised" effects using the machine learning approach (ML-WN).

Scenario	Units	Type	One week	Two weeks	One month	Three months	Six months
S1_s1w	μg m^{-3}	actual	193.8	-	-	-	-
		wn	92.7	-	-	-	-
	%	actual	81.4	-	-	-	-
		wn	39.2	-	-	-	-
S2_s2w	μg m^{-3}	actual	193.8	192.7	-	-	-
		wn	141.9	138.7	-	-	-
	%	actual	81.4	81.4	-	-	-
		wn	59.1	57.9	-	-	-
S3_s1m	μg m^{-3}	actual	193.8	192.7	191.3	-	-
		wn	159	159.4	156.1	-	-
	%	actual	81.4	81.4	81	-	-
		wn	67.8	68.4	67.2	-	-
S4_s3m	μg m^{-3}	actual	193.8	192.7	191.3	206	-
		wn	174.1	173.3	172.3	188.1	-
	%	actual	81.4	81.4	81	80.5	-
		wn	73.7	73.9	73.7	74.2	-
S5_s6m	μg m^{-3}	actual	193.8	192.7	191.3	206	196.6
		wn	180.4	181.7	181.4	195.7	188.7
	%	actual	81.4	81.4	81	80.5	78
		wn	76.1	77.2	77.3	77	75.6
S6_p3m	μg m^{-3}	actual	187.7	179.3	160.5	101.7	-
		wn	149.6	146.5	132.3	85.1	-
	%	actual	78.8	75.6	67.8	41.1	-
		wn	64.4	63.3	57.5	35.1	-
S7_p6m	μg m^{-3}	actual	190.8	186	175.9	153.9	101.5
		wn	160.9	159.2	153.9	136.7	95.1
	%	actual	80.1	78.5	74.4	60.8	39.9
		wn	68	67.8	65.7	54.5	38
S8_c1m	μg m^{-3}	actual	193.8	192.7	191.3	-	194.3
		wn	174	173.9	171.8	-	178.3
	%	actual	81.4	81.4	81	-	77.7
		wn	74.1	74.5	73.9	-	72.9

Table S4: GBM (gradient boosting machine) model hyper-parameters from the H2O.ai's AutoML for the updated weather normalisation approach (MacLeWN).

Model parameters	Marylebone Road	North Kensington
training_frame	Automls	Automl
nfolds	5	5
keep_cross_validation_models	FALSE	FALSE
keep_cross_validation_predictions	TRUE	TRUE
keep_cross_validation_fold_assignment	FALSE	FALSE
score_each_iteration	FALSE	FALSE
score_tree_interval	5	5
fold_assignment	Modulo	Modulo
response_column	y	y
ignore_const_cols	TRUE	TRUE
balance_classes	FALSE	FALSE
max_after_balance_size	5	5
max_confusion_matrix_size	20	20
ntrees	154	110
max_depth	10	11
min_rows	10	5
nbins	20	20
nbins_top_level	1024	1024
nbins_cats	1024	1024
r2_stopping	1.79e+308	1.80 e+308
stopping_rounds	0	0
stopping_metric	deviance	deviance
stopping_tolerance	7.12e-03	7.07e-03
max_runtime_secs	0	0
build_tree_one_node	FALSE	FALSE
learn_rate	0.1	0.1
learn_rate_annealing	1	1
distribution	gaussian	Gaussian
quantile_alpha	0.5	0.5
tweedie_power	1.5	1.5
huber_alpha	0.9	0.9
sample_rate	0.8	0.9
col_sample_rate	0.8	0.7
col_sample_rate_change_per_level	1	1
col_sample_rate_per_tree	0.8	1
min_split_improvement	1.0e-05	1.0e-05
histogram_type	UniformAdaptive	UniformAdaptive
max_abs_leafnode_pred	1.80e+308	1.79 e+308
pred_noise_bandwidth	0	0
categorical_encoding	Enum	Enum
calibration_method	PlattScaling	PlattScaling
in_training_checkpoints_tree_interval	1	1
gainslift_bins	-1	-1

Table S5: GBM (gradient boosting machine) performance statistics for the updated weather normalisation approach (MacLeWN).

Metrics	Unit	Marylebone Road	North Kensington
The number of complete pairs of data (n)	-	4908	5004
Fraction of predictions within a factor of two (FAC2)	-	0.97	0.96
Fractional bias (FB)	-	0.00038	0.011
Coefficient of correlation (r)	-	0.93	0.91
Root mean square error (RMSE)	$\mu g\ m^{-3}$	62.26	24.62
Systematic RMSE	$\mu g\ m^{-3}$	25.43	13.38
Unsystematic RMSE	$\mu g\ m^{-3}$	56.83	20.66
Index of agreement (IOA)	-	0.84	0.82

Table S6: Absolute and percentile policy impacts on NO_X concentrations under different scenarios. "actual" represents idealised policy effects, and "wn" represents "weather normalised" effects using the updated machine learning approach (MacLeWN).

Scenario	Units	Type	One week	Two weeks	One month	Three months	Six months
S1_s1w	μg m^{-3}	actual	191.9	-	-	-	-
		wn	190.8	-	-	-	-
	%	actual	87.2	-	-	-	-
		wn	86.7	-	-	-	-
S2_s2w	μg m^{-3}	actual	191.9	183.8	-	-	-
		wn	191.3	183.4	-	-	-
	%	actual	87.2	85.8	-	-	-
		wn	87	85.6	-	-	-
S3_s1m	μg m^{-3}	actual	191.9	183.8	181.1	-	-
		wn	190.9	183.1	180.7	-	-
	%	actual	87.2	85.8	84.6	-	-
		wn	86.7	85.4	84.4	-	-
S4_s3m	μg m^{-3}	actual	191.9	183.8	181.1	204	-
		wn	191.3	183.2	180.8	203.3	-
	%	actual	87.2	85.8	84.6	84	-
		wn	87	85.5	84.5	83.8	-
S5_s6m	μg m^{-3}	actual	191.9	183.8	181.1	204	187.0
		wn	191.1	183.2	180.8	203.1	186.2
	%	actual	87.2	85.8	84.6	84	81.0
		wn	86.8	85.5	84.4	83.6	80.6
S6_p3m	μg m^{-3}	actual	185.6	171.3	152.3	100.2	-
		wn	185.4	170.9	152.1	99.7	-
	%	actual	84.4	79.8	71	43	-
		wn	84.4	79.7	70.9	42.6	-
S7_p6m	μg m^{-3}	actual	188.7	177.5	166.7	152.1	98.6
		wn	187.6	176.7	165.9	150.9	97.9
	%	actual	85.8	82.8	77.8	63.5	41.4
		wn	85.2	82.4	77.4	63	40.7
S8_c1m	μg m^{-3}	actual	191.9	183.8	181.1	-	185.3
		wn	191.4	183.5	180.9	-	184.3
	%	actual	87.2	85.8	84.6	-	81.1
		wn	86.9	85.6	84.5	-	80.7

Table S7: DRF (distributed random forest) model performance statistics.

Metrics	Unit	Marylebone Road	North Kensington
The number of complete pairs of data (n)	-	4908	5004
Fraction of predictions within a factor of two (FAC2)	-	0.96	0.95
Fractional bias (FB)	-	0.0084	0.015
Coefficient of correlation (r)	-	0.92	0.89
Root mean square error (RMSE)	$\mu g\ m^{-3}$	66.93	27.65
Systematic RMSE	$\mu g\ m^{-3}$	35.14	18.86
Unsystematic RMSE	$\mu g\ m^{-3}$	56.96	20.22
Index of agreement (IOA)	-	0.83	0.80

Table S8: XGBoost (eXtreme Gradient Boosting) model performance statistics.

Metrics	Unit	Marylebone Road	North Kensington
The number of complete pairs of data (n)	-	4908	5004
Fraction of predictions within a factor of two (FAC2)	-	0.96	0.94
Fractional bias (FB)	-	0.0022	-0.00063
Coefficient of correlation (r)	-	0.93	0.91
Root mean square error (RMSE)	$\mu g\ m^{-3}$	62.26	24.70
Systematic RMSE	$\mu g\ m^{-3}$	22.99	13.30
Unsystematic RMSE	$\mu g\ m^{-3}$	57.86	20.81
Index of agreement (IOA)	-	0.84	0.81

References

Willmott, C. J., Robeson, S. M., and Matsuura, K.: A refined index of model performance, Int. J. Clim., 32, 2088-2094, https://doi.org/10.1002/joc.2419, 2012.