

## Supplement

**Table S1:** Low-cost sensor specifications. Superscript 1 indicates the lowest concentration difference that can be distinguished by the sensor while 2 the lowest detectable reading that can be reliably measured and displayed. ISB stands for individual sensor board.

	<b>CO-B41</b>	<b>NO<sub>2</sub>-B43F</b>	<b>OX-B431</b>	<b>SO<sub>2</sub>-B4</b>
<b>Detection range (ppm)</b>	1000	20	20	100
<b>Temperature range (°C)</b>	-30 to 50	-30 to 40	-30 to 40	-30 to 50
<b>Relative Humidity Range (%)</b>	15 to 90	15 to 85	15 to 85	15 to 19
<b>Response time (s)</b>	< 25 (From 0 to 10 ppm)	<60 (From 0 to 2 ppm)	< 60 (From 0 to 1ppm)	< 60 (From 0 to 2ppm)
<b>Resolution<sup>1</sup> ± 2 standard deviation (ppb)</b>	4	15	15	5
<b>Major cross-sensitivity gases</b>	H <sub>2</sub> S	O <sub>3</sub> , Cl <sub>2</sub>	NO, NO <sub>2</sub> , Cl <sub>2</sub>	NO, CO, H <sub>2</sub> S
<b>Sensitivity<sup>2</sup> from 20 °C to 50 °C / zero current (nA)</b>	-50 to -200	-10 to 50	5 to 100	10 to 30
<b>Operation until the loss of 50% of the original signal (months)</b>	>36 (24 warranted)	>24 (24 warranted)	>24 (24 warranted)	>36 (24 warranted)
<b>Dimensions (mm)</b>	32 x 16.5	32 x 16.5	32 x 16.5	32 x 16.5
<b>Weight (g)</b>	<13	<13	<13	<13
<b>Cost with ISB (USD)</b>	153	158	164	153

5 **Table S2:** Specifications of reference instruments used by Cyprus Department of Labour and Inspection for air quality monitoring.

	<b>CO</b>	<b>NO<sub>2</sub></b>	<b>O<sub>3</sub></b>	<b>SO<sub>2</sub></b>
<b>Model</b>	Ecotech Serinus 30	Ecotech Serinus 40	Thermo Scientific 49i	Ecotech Serinus 50
<b>ISO_EN Standard</b>	CYS EN	CYS EN	CYS EN	CYS EN
<b>Number</b>	14626:2012	14211:2012	14625:2012	14212:2012
<b>Method</b>	Infrared Spectroscopy	Chemiluminescence	Ultraviolet Photometry	Ultraviolet Fluorescence
<b>Sample flow rate</b>	1.0 slpm	0.3 slpm (0.6 slpm total flow for the NO and NOX flow path)	1–3 lpm	0.750 slpm
<b>Precision</b>	20 ppb or 0.1 % of the reading, whichever is greater	0.4 ppb or 0.5% of the reading, whichever is greater		0.5 ppb or 0.15 % of reading, whichever
<b>Response time</b>	60 seconds to 95%	15 seconds to 90%	20 seconds (10 seconds lag time)	60 seconds to 95 %
<b>Lower detectable limit (ppb)</b>	40	0.4	1	0.3

**Table S3:** Optimal values of the main hyper-parameters for SVR, RF, ANN and XGBoost ML algorithms.

<b>SVR-CO</b>	<b>RF-CO</b>	<b>ANN-CO</b>	<b>XGBoost-CO</b>
kernel='rbf' C=1 $\epsilon=1$	n_estimators=236 max_leaf_nodes=744 max_features=0.51	number of hidden layers=3 number of neurons in the 1st and 2nd hidden layers=128 number of neurons in the 3rd hidden layer=100 activation function in the input and output layers='linear' activation function in the in the hidden layers='ReLU' optimizer='Adam', learning rate=0.01 batch size=100 , number of epochs=200	n_estimators=970 max_depth=0 eta=0.092 subsample=0.76 colsample_bytree=1 alpha=0.091
<b>SVR-NO<sub>2</sub></b>	<b>RF-NO<sub>2</sub></b>	<b>ANN-NO<sub>2</sub></b>	<b>XGBoost-NO<sub>2</sub></b>
kernel='rbf' C=1 $\epsilon=1$	n_estimators=194 max_leaf_nodes=1564 max_features=0.31	number of hidden layers=3 number of neurons in the 1st and 2nd hidden layers=128 number of neurons in the 3rd hidden layer=100 activation function in the input and output layers='linear' activation function in the in the hidden layers='ReLU' optimizer='Adam', learning rate=0.01 batch size=100 , number of epochs=200	n_estimators=499 max_depth=0 eta=0.038 subsample=0.515 colsample_bytree=0.85 alpha=0.0017
<b>SVR-O<sub>3</sub></b>	<b>RF-O<sub>3</sub></b>	<b>ANN-O<sub>3</sub></b>	<b>XGBoost-O<sub>3</sub></b>
kernel='rbf' C=1 $\epsilon=1$	n_estimators=418 max_leaf_nodes=1836 max_features=0.72	number of hidden layers=3 number of neurons in the 1st and 2nd hidden layers=128 number of neurons in the 3rd hidden layer=100 activation function in the input and output layers='linear' activation function in the in the hidden layers='ReLU' optimizer='Adam', learning rate=0.01 batch size=100 , number of epochs=200	n_estimators=1260 max_depth=0 eta=0.099 subsample=0.9 colsample_bytree=0.38 alpha=0.9
<b>SVR-SO<sub>2</sub></b>	<b>RF-SO<sub>2</sub></b>	<b>ANN-SO<sub>2</sub></b>	<b>XGBoost-SO<sub>2</sub></b>
kernel='rbf' C=1 $\epsilon=1$	n_estimators=474 max_leaf_nodes=557 max_features=0.32	number of hidden layers=3 number of neurons in the 1st and 2nd hidden layers=128 number of neurons in the 3rd hidden layer=100 activation function in the input and output layers='linear' activation function in the in the hidden layers='ReLU' optimizer='Adam', learning rate=0.01 batch size=100 , number of epochs=200	n_estimators=924 max_depth=0 eta=0.024 subsample=0.56 colsample_bytree=0.89 alpha=0.045



O=October

N=November

D=December

J=January

F=February

M=March

**Figure S1:** Data splitting into training and testing set. The training set was derived from the first 80% of the data of each month, whereas the testing set comprises of the remaining 20% data of each month.