HYBRID ARTIFICIAL NEURAL NETWORK COUPLED WITH KALMAN FILTERS FOR AIR QUALITY FORECASTING IN GUADALAJARA, MEXICO

C. González-Figueredo, E. A. Egurrola-Hernández, R. L. Ramírez-Briseño, A. De los Reyes-Corona, H. De Alba-Martínez.
Technological and Industrial Processes Department, Western Institute of Technology and Higher Education (ITESO),
Tlaquepaque, Jalisco, 45604, México;
Presenting author email: figueredo@iteso.mx

Summary
This study aims to develop a novel hybrid scheme of Artificial Neural Networks (ARN) coupled to a non-linear Kalman filter for air quality forecasting in Guadalajara Metropolitan Area, in Mexico. ARN's are widely used for air quality forecasting, however these schemes need large amounts of data regarding the pollutants concentration levels and meteorological data in order to manage reliable forecasting. To address this issue, we present a scheme consisting of Neural Network models assisted by nonlinear Kalman filter that manage to considerably improve the forecasting performance, adding robustness in case of lack of data, and reducing the need of retraining over time.

Introduction
As in many megacities, air pollution is a growing concern, and poor air quality episodes have become more frequent. In Guadalajara city, in Mexico, government agencies have the Atmospheric Monitoring System for the Guadalajara Metropolitan Area, SIMAJ (http://siga.jalisco.gob.mx/), that presents real time air pollution measurements, but fails to provide specific forecasting of air quality. ARN models can be used to predict air quality behavior (Russo, *et. al.* 2013), but their performance is severely hindered by the quality of the available information. Coupling an ARN prediction model with a non-linear Kalman filter scheme, introduces robustness to the forecasting process in case of the lack of data. In addition, this hybrid-forecasting scheme does not require complex ARN architectures and is able to predict air quality levels for a 48-hour horizon, within reasonable tolerances.

Methodology and Results
The hybrid ARN + Kalman forecasting system was constructed using air quality and meteorological data from years 2013 and 2014, obtained from SIMAJ. We treated these data sets, first by removing the spurious data, then by filling in the empty spaces with sliding window averages, and finally normalizing the data taking into account the limit values defined by the respective standards. We determined the variables with statistical significance using principal component analysis. We constructed the respective training sets for a Feed-forward ARN with two hidden layers of neurons; the pollutant concentration values for a 48-hour horizon were defined as the target values. The neural networks were trained using the error back-propagation scheme; we tried from 10 to 15 neurons on both hidden layers, and then selected the one with the best correlation coefficient. We used 50% of the data for the training process, and 50% for validation and testing.

We used a novel scheme on coupled ARN with non-lineal Kalman filter that corrects the quantitative errors of the ARN, as described in Ramírez-Álvarez *et. al*., 2013. The prediction error covariance was defined as the measurement error covariance, to reduce the prediction error. The hybrid ARN + Kalman scheme achieved better correlation coefficient values, in comparison to the simple ARN, as shown on table 1. An example of the hybrid scheme performance is shown in figure 1.

Horizon (hours)	Hidden layer neurons	RNA correlation R	RNA + Kalman correlation R
+1	10	0.9226	0.9990
+2	11	0.8701	0.9963
+3	11	0.8356	0.9931
+4	13	0.8218	0.9915
+8	13	0.7990	0.9980
+12	11	0.7897	0.9265
+24	13	0.6944	0.9830

Table 1. Forecasting correlation coefficients for CO forecasts in "Las Águilas" station location.

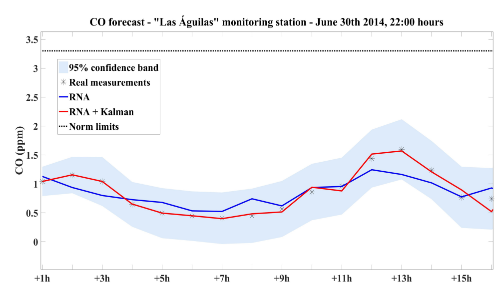

Fig. 1 CO forecast with hybrid model.

Conclusions
The hybrid ARN model coupled to a Kalman filter, manages to correct the neural networks predictions, significantly improving its performance. Kalman parameter tuning is a key issue; the measurement error covariance is set to equal the ARN prediction error covariance. Kalman provides the model robustness, and allows its use online without the need for periodic updates. Nonetheless, the correct ARN architecture design plays a major role in the forecasting process.

References
Russo, A., Raischel, F., & Lind, P. G. (2013). Air quality prediction using optimal neural networks with stochastic variables. Atmospheric Environment, 79, 822–830. https://doi.org/10.1016/j.atmosenv.2013.07.072 .

Ramírez-Álvarez, E., Calderón Ramírez, M., Rico-Martínez, R., González-Figueredo, C., & Parmananda, P. (2013). Detecting bifurcations in an electrochemical cell employing an assisted reference model strategy. Journal of Physical Chemistry A, 117(3), 535–540. https://doi.org/10.1021/jp310127a .