CHEMICAL ENGINEERING TRANSACTIONS

VOL. 48, 2016

A publication of
AIDIC
The Italian Association
of Chemical Engineering
Online at www.aidic.it/cet

Guest Editors: Eddy de Rademaeker, Peter Schmelzer
Copyright © 2016, AIDIC Servizi S.r.l.,
ISBN 978-88-95608-39-6; **ISSN** 2283-9216

DOI: 10.3303/CET1648016

Atmospheric Powder Dispersion in an Urban Area

Pierre Lauret[*a], Marcia Perrin[b], Frederic Heymes[a], Laurent Aprin[a], Serge Forestier[c], Pierre Slangen[a], Alexis Pey[c], Marc Steinkrauss[c]

[a]Institute of Risk Science (ISR), Ecole des Mines d'Alès, Alès, France
[b]Novartis Pharma AG, Lichtstrasse 35, CH-4002 Basel
[c]Swissi Process Safety GmbH, Mattenstrasse 24, CH-4002 Basel
pierre.lauret@mines-ales.fr

Dust dispersion in an urban area becomes a major concern in several fields: global safety, pollution tracking, accidental release of highly active substances in powder form.

In Switzerland, this last point became a relevant scenario of major accident in the middle of 2015. Every company producing or working with highly active substances in powder form (e.g. pharmaceutical and chemical firms) must thus assess the consequences of the dispersion of this kind of product and based on the results of the simulations, the authorities grant their approval for the production of this powder.

The main background of dust dispersion modeling relies on heavy gas dispersion modeling. Indeed, air loaded with dust has an apparent density higher than the ambient one and behaves globally as a heavy gas. But other phenomena such as sedimentation, agglomeration have to be considered. Furthermore, in complex configurations such as urban areas, the accuracy of heavy gas models is low.

This paper aims to evaluate the efficiency of an Artificial Neural Networks model to predict the dust dispersion in an urban area. Dust concentration data were collected at different places in a city.

The wind velocity, direction, and atmospheric temperature were measured at nearest Meteoswiss station. This one year long data acquisition is thus a very rare data set that can be really useful to calibrate dust dispersion models in such areas.

Results about the comparison between the experimental concentrations found and the results of Artificial Neural Networks based model of dust dispersion are presented. The results are discussed to explain the trends of the experimental values and the variation of accuracy of the tested models.

1. Introduction

Dust dispersion modeling in urban areas is of major interest for in the fields of health and safety. That is why the Swiss department for the protection against major accidents (OPAM) requires that the accidental release of a highly active powder should be studied. The authorization of production depends on the expected consequences. However, the lack of official emphasized method may lead to a different conclusion of the same scenario. This is why either a unified a defined model should be defined at a global scale, either a refined and more robust model could be defined by production site.

The first step to build a model is to acquire a data set of the desired phenomenon while registering the different influent parameters, i.e, the meteorological data. In that sense, the dust monitoring carried out by Novartis around their production sites since years is of huge interest since it could allow comparing a model with real data points. Any kind of model could be built with this set but in frame of an OPAM risk assessment the most accurate and complete meteorological data set is defined at the nearest MeteoSwiss station. That is why the ideal model should be able to estimate the concentrations of the emitted powder with an important uncertainty of the wind characteristics at the exhaust point. This paper answers the question whether or not is possible to anticipate dust concentrations using artificial neural networks. A dispersion model is thus developed to forecast mean day concentrations of particles at a station 300 m away from the emission area.

Please cite this article as: Lauret P., Perrin M., Aprin L., Heymes F., Forestier S., Slangen P., Pey A., 2016, Powder dispersion in an urban area, Chemical Engineering Transactions, 48, 91-96 DOI:10.3303/CET1648016

1.1 Modeling of dust dispersion

Atmospheric dispersion modeling is an important topic and several models of different complexity exist (Meroney, 1999). Firstly, flow of the wind has to be determined. Depending of the configuration of the site to be modeled, the scale and the fineness required, several strategies might be used. Dealing with complex geometries, the impact of buildings are of primary importance. While Gaussian models use a constant wind hypothesis, Computational Fluid Dynamics (CFD) models are solving Navier-Stokes equations using Finite Elements methods. It results in large differences on the flow but also on the computation time. Once the wind flow is obtained, dispersion can be evaluated from CFD models by using lagrangian models (particle tracking), eulerian models (solving the advection diffusion equation) or empirical models like Gaussian (statistical correlations from reference experimentations). This whole process has to be adapted depending on each situation. Pharmaceutical and chemical firms face this problem when they have to assess the consequences of the dispersion of powder. Moreover, it is often difficult to assess the initial emission source, like it is in our case. In this work, the aim is to create a model that should be able to forecast the mean particle concentration values of the following day in order to prevent extreme values by acting on the process.

Situation presented here is specific because of the complexity of the site (buildings, Rhin river, car circulation …), the difficulty to determine source term and the need for fast modeling. In this context, machine learning tools like Artificial Neural Networks (ANN) are of particular interest. Indeed, these types of models are capable to forecast a phenomenon based on past data, through a learning phase. In our case, such a database exists: each station records daily concentration and information of the weather is provided by the MeteoSwiss.

1.2 Atmospheric dispersion modeling by Artificial Neural Networks

Artificial Neural Networks (ANNs) came from the original idea to emulate the structure and behavior of the brain (Minsky and Papert, 1969). It consists of several mathematical functions, termed as neurons, which are linked in a network. ANN are powerfull non-linear statistical data modeling tools. They are generally used when the process to model is not fully known thanks to two essential properties: first the universal approximation (Hornik et al., 1989), and second the parsimony (Barron, 1993). Thanks to these properties ANNs are able to predict efficiently future behaviors on never encountered situations. ANNs can be used in classification, in text recognition for example (Dreyfus, 2004). They can also be used to forecast physical phenomenon, presenting powerful models (Lauret et al., 2015). The information about the non-linear phenomenon to forecast must be provided using a database. As previously presented, ANNs act generally like a black-box: the physics cannot be extracted from the results. A neuron is a nonlinear, parameterized, bounded function. Variables are assigned to the inputs of the neuron. Output of a neuron is the result of nonlinear combination of the inputs, weighted by the parameters and using an s-shaped function like a sigmoid. A neural network is the composition of several neurons. Parameters calibration is done through application of an algorithm using the training database and designed to decrease the model error. In this work, the Levenberg-Marquardt method is used (Hagan and Menhaj, 1994). The function realized by the ANN is continuously tested on a disjoined set of examples, namely the validation set. This set is employed to avoid overtraining using early stopping (Sjöberg et al., 1995). Lastly, performances of the model must be measured on another set, never used during training or stopping: the test set. ANN have already been used in atmospheric dispersion. Prevision of concentrations of tracers in complex terrain have been made by training an ANN using databases of values coming from various sensors spatially distributed (Podnar, 2002). Predicted concentrations were the output variable of the ANN at a specific point. ANN were compared to other statistical methods and revealed better performance with ANN, especially for long-term prediction. Boznar et al., (1993, 2004) use meteorological values (air temperature, global solar radiation, wind speed, wind direction, maximal air temperature) and previous pollutant concentrations (NO, NO2, NOx, CO, O3) to perform a 12-hours forecasting of Ozone concentrations. Previous work of the author have been focused on atmospheric dispersion in the near field of an industrial site (Lauret et al., 2013, 2014). In this work, ANN are good candidates because of the complexity of the flow, the need for fast forecasting and the lack of hourly data for concentrations that make 24 hours prediction a long term prediction.

2. Methodology

2.1 Inputs selection

First of all, it is important to define variables used as inputs of the ANN. The inputs available are meteorological data computed from values of the six intervals of ten minutes from the previous hour: minimum, maximum and mean temperature at two meters of altitude, mean relative humidity of air at two meters of altitude, sum of the precipitations, duration of shining sun, mean radiation, mean wind direction, mean wind speed, maximum wind speed and mean atmospheric pressure (respectively noted here $T_{mean,}$ T_{min},

T_{max}, MRH, Rain, Ray_m, Ray, Dir_w, w, w_{max}, P, std_{dw}, C). Other available data are given by measurements on the different locations around the emission source. These data are the outputs of the model. They are measured on a daily basis. Meteorological data have to be processed to fit day duration. In order to avoid the gap between 0° and 360° in the direction of the wind, values of the cosinus of this variable is used. Characteristics of the variables are given on table 1. To select data the most correlated to the concentration, a correlation matrix is computed between each of these variables. The more the value is close to one means the variables are correlated. Negative values imply opposite trend. If the value is close to zero, the two variables are not correlated: modification of the first variable value only does not impact the second variable value.

Table 1: Characteristics of available variables

Variables	T_{mean}	T_{min}	T_{max}	MRH	Rain	Ray_m	Ray	Dir_w	W	W_{max}	P	Std_{dw}	C
Unit	°C	°C	°C	%	mm	min	W/m^2	-	$km.h^{-1}$	$km.h^{-1}$	hPa	-	$\mu g.m^{-3}$
Minimum	-3	-5.8	0.3	43.3	0	0	4.6	-0.84	2.7	10.1	945	0.04	4e-4
Maximum	26.7	19.3	35.5	95.6	32.4	890	359.4	0.95	20.1	92.9	1000	0.84	0.41

Table 2: Correlation matrix of normalized meteorological data and concentrations

Variables	T_{mean}	T_{min}	T_{max}	MRH	Rain	Ray_m	Ray	Dir_w	W	W_{max}	P	std_{dw}	C
T_{mean}	1,00	0,95	0,97	-0,32	0,14	0,42	0,70	0,06	-0,22	0,04	-0,06	0,28	0,41
T_{min}	0,95	1,00	0,87	-0,09	0,24	0,19	0,51	0,03	-0,20	0,05	-0,14	0,20	0,33
T_{max}	0,97	0,87	1,00	-0,44	0,07	0,56	0,79	0,06	-0,23	0,02	0,01	0,31	0,42
MRH	-0,32	-0,09	-0,44	1,00	0,31	-0,68	-0,71	-0,17	-0,14	-0,19	-0,12	-0,21	-0,35
Rain	0,14	0,24	0,07	0,31	1,00	-0,30	-0,18	-0,21	0,10	0,29	-0,25	-0,08	0,03
Ray_m	0,42	0,19	0,56	-0,68	-0,30	1,00	0,85	0,21	-0,18	-0,14	0,29	0,31	0,31
Ray	0,70	0,51	0,79	-0,71	-0,18	0,85	1,00	0,22	-0,21	-0,04	0,14	0,41	0,47
Dir_w	0,06	0,03	0,06	-0,17	-0,21	0,21	0,22	1,00	-0,17	-0,12	0,18	0,32	0,09
W	-0,22	-0,20	-0,23	-0,14	0,10	-0,18	-0,21	-0,17	1,00	0,76	-0,13	-0,55	-0,04
W_{max}	0,04	0,05	0,02	-0,19	0,29	-0,14	-0,04	-0,12	0,76	1,00	-0,19	-0,30	0,12
P	-0,06	-0,14	0,01	-0,12	-0,25	0,29	0,14	0,18	-0,13	-0,19	1,00	0,13	0,01
std_{dw}	0,28	0,20	0,31	-0,21	-0,08	0,31	0,41	0,32	-0,55	-0,30	0,13	1,00	0,16
C	0,41	0,33	0,42	-0,35	0,03	0,31	0,47	0,09	-0,04	0,12	0,01	0,16	1,00

From table 2, one can see that all variables relative to air temperature and to solar radiation are correlated to the concentration. The relative humidity is negatively correlated to the concentration. Rain and atmospheric pressure seems not to affect the concentration. Wind parameters are not well correlated with the concentration. This point may be explained by the urban configuration where the global wind data can be strongly modified by the local configuration. Variables with a correlation value above 0.1 with the concentration are selected as inputs. To avoid over-influence of one specific variable, they are normalized (centered and reduced) before training the neural networks.

2.2 ANN architecture and model selection

The structure of the neural network corresponds to a classical two-layer perceptron. Input variables are linked to the neurons of the hidden layer. The output layer contains a unique linear neuron. Complexity selection is done through the use of cross validation (Stone, 1974) with variation of the number of hidden neurons, from 1 to 20. Using the training set of D subsets, each subset one at a time is reserved as the validation set. Training is then performed D times on D subsets. The mean quadratic error is thus calculated D times. To assess the model's generalization capability the cross-validation score (S_{cv}) is compared for the investigated configurations.

$$S_{cv} = \sqrt{\frac{1}{N_e}\sum_{i=1}^{D}\sum_{k\in i}^{D}\left(y_k^p - g(x_k, w_i)\right)^2} \qquad (1)$$

Early stopping is used to avoid overtraining: it consists in dividing the database in three parts: one set is the training set and represents a certain percentage of the database. In this work, several divisions are tested: 60%, 70%, 80% and 90% respectively feed the training set. Left examples are divided equally for the stop and test sets. The stop set is used to avoid overtraining: when the mean squared error stops decreasing on it, the training phase is interrupted. Finally, the test set is used to assess the model quality. Initialization of parameters is known to have influence on results of training phase. Once architecture is determined, 20 initializations are made in order to obtain the best model. At the end of the training process, the mean squared error and the coefficient of determination are computed on the test set assessing the generalization capabilities of the model.

2.3 Performance criteria used

To improve the performance of dispersion modeling, several criteria were proposed by Chang and Hanna (2004). The present study uses the following set of criteria: factor of two (*FAC2*), Normalized Mean Squared Error (*NMSE*), and Fractional Bias (*FB*). Because the coefficient of determination R^2 is widely used to evaluate performance in the field of artificial neural networks, it replaces the correlation coefficient in the present study. It is ranging from -∞ to 1. Negative values indicate that the mean of the data gives a better forecasting than the fitted function values. The target values for these criteria are as following: R^2 and *FAC2*=1; and *FB* and *NMSE*=0. *FB* measures systematic errors which lead to mostly underestimate or overestimate measured values. *FB* values ranges between -2 (extreme underprediction) to 2 (extreme overprediction). Therefore, matching perfect target *FB* value does not mean perfect modeling, because of possible cancelling errors. *NMSE* measures systematic and random errors. Acceptable values are within +/- 30% of the mean fractional bias ($|FB| < 0.3$), random scatter is about a factor of two to three of the mean ($NMSE < 1.5$), the factor of two is superior to 0.5. For this reason it is necessary to use simultaneously several criteria.

3. Results

3.1 Training phase evaluation

Database contains 442 values of each variable selected from the 21th of November 2013 to the 1st of April 2015. The model selection has required 1,600 trainings. It is possible to assess the respective influence of the percentage of examples in the training set, the number of neurons and the parameters initialization. Four percentages were tested by computing the value of coefficient of determination. Percentage equal to 90% shows better median (0.44) and maximum values (0.6139) with wider range than results for lower percentage. Training results are evaluated by the coefficient of determination for different numbers of neurons in hidden layer on Figure 1a. For each number of neurons, the top, middle and bottom of the rectangle respectively represent the 3rd quartile, median and 1st quartile. On this graph is also plotted the maximum value for the coefficient of determination. It is not obvious that the mean quality of the training phase is improved by adding neurons (increasing the model complexity). On figure 1b, values of R2 are plotted for 20 different initializations while keeping the percentage of examples in the training set to 90% and the number of neurons in hidden layer at 7. This graph demonstrates that it is important to try several initializations because results can vary a lot from one to another. In this case, if only positive values are considered, the deviation reaches 45%. Finally, the best model selected corresponds to 7 neurons in hidden layer, for a percentage in the training set of 90%. The ANN is fed with normalized data and the output is a normalized concentration too. In order to avoid unphysical values, negative concentrations outputs are set to zero.

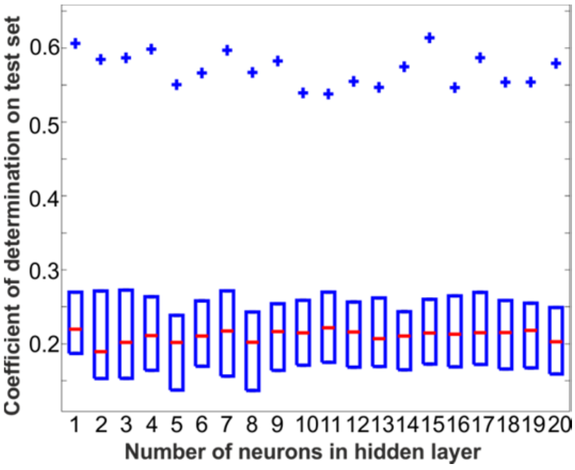

Figure 1a – influence of number of hidden neurons

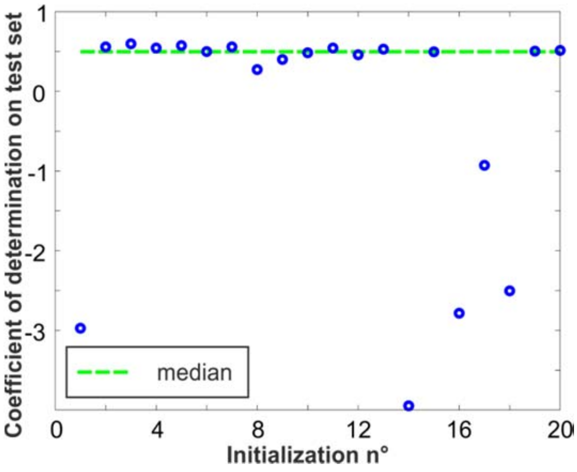

Figure 1b – influence of number of initialization

3.2 Model evaluation

ANN model is evaluated with atmospheric dispersion performance criteria. Factor of two for forecasted concentrations on the period from November the 13th 2013 and the 1st of April 2015 reach 0.48. This criterion can be observed on Figure 2. It corresponds to examples located between the red lines. Two examples of high concentrations are underestimated by the ANN model. Other predictions over the limit value of 0.15 µg.m^{-3} are between factor two over/underestimation. Value of fractional bias on this set (0.01) indicates a light underprediction. Normalized Mean Square Error reach a value of 1.4 which is under the recommendation of Hanna and Chang (2012) for urban dispersion models. On the linear time graph on Figure 3, most of the concentration values are well predicted. Two peaks of high concentrations appear to be difficult to model. Low values at the beginning on the 150 first days are well predicted even if small peak at day 80 is 79% overpredicted. Between day 200 to 300, increasing in observed concentrations is well correlated to modeled concentrations.

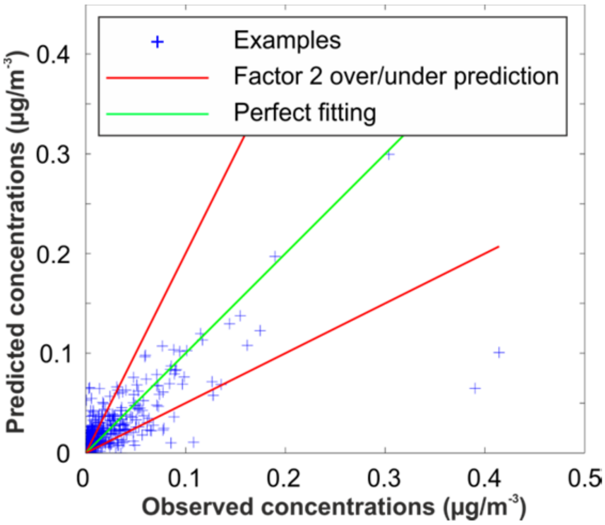

Figure 2 – Scatter plot of observed and predicted concentrations

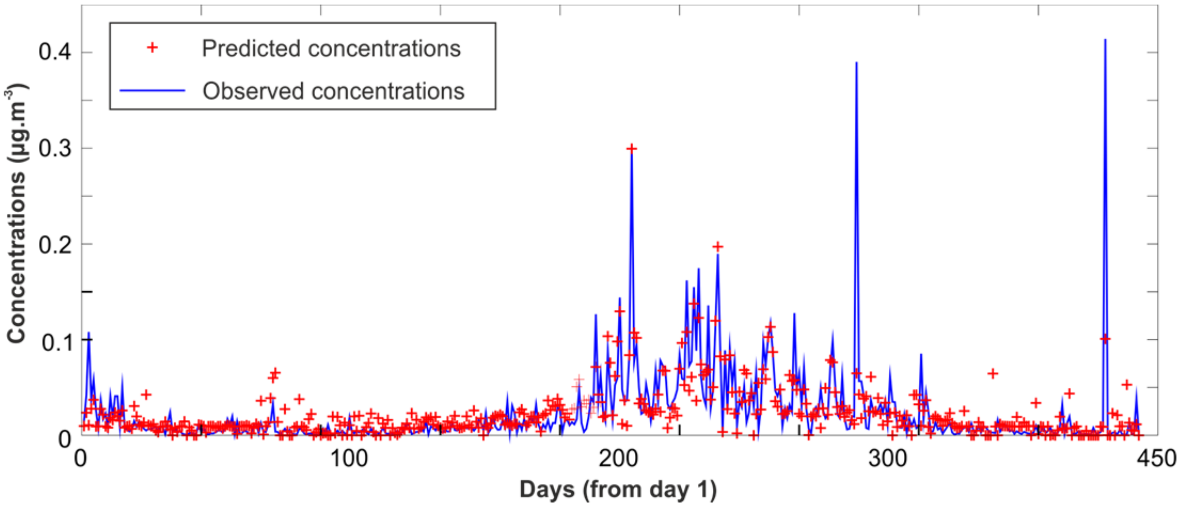

Figure 3 – Observed vs forecasted concentrations in a time line

4. Conclusion

A dispersion model is developed using artificial neural networks to forecasted mean day concentrations of particles at a particles measuring station 300 m away from the emission area. Model is trained through meteorological and concentration database previously acquired. Principal component analysis is applied to select inputs of the model. Architecture of the ANN is realized through complexity selection: best model contains 7 neurons and is selected from 20 different initialization parameters. Results show acceptable forecasting even without flow rate of the emission source. An analysis is realized using quality performance criteria suitable for atmospheric dispersion. ANN model show very low bias. The global error is maintained under reasonable values. Main errors are done on unexpected peaks, difficult to assess with meteorological data only.

5. Improvements

It is important to remember that no information on the emission sources is required in this model. By adding such information, the model might be improved. Moreover, as the sampling is going on, database grows and better trainings can be realized. Authors suggested the implementation of a self-training algorithm, improved each day from data acquired the day before. The final goal of this study is to forecast concentrations for each measurement station and realized a mesh on the entire city.

References

Barron, A.R., 1993. Universal approximation bounds for superpositions of a sigmoidal function. IEEE Trans. Inf. Theory 39, 930–945. doi:10.1109/18.256500

Boznar, M., Lesjak, M., Mlakar, P., 1993. A neural network-based method for short-term predictions of ambient SO2 concentrations in highly polluted industrial areas of complex terrain. Atmos. Environ. Part B. Urban Atmos. 27, 221–230. doi:10.1016/0957-1272(93)90007-S

Boznar, M.Z., Mlakar, P., Grasic, B., 2004. Neural networks based ozone forecasting, in: 9th International Conference on Harmonisation within Atmosphéric Dispersion Modelling for Regulatory Purposes. pp. 356–360.

Chang, J.C., Hanna, S.R., 2004. Air quality model performance evaluation. Meteorol. Atmos. Phys. 87, 167–196. doi:10.1007/s00703-003-0070-7

Dreyfus, G., 2004. Neural Networks, Methodology and Applications, Springer. ed, Neural Networks.

Hagan, M.T., Menhaj, M.B., 1994. Training Feedforward Networks with the Marquardt Algorithm. IEEE Transations Neural Networks 5, 989–993.

Hanna, S., Chang, J., 2012. Acceptance criteria for urban dispersion model evaluation. Meteorol. Atmos. Phys. 116, 133–146. doi:10.1007/s00703-011-0177-1

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer Feedforward Networks are Universal Approximators. Neural Networks 2, 359–366.

Lauret, P., Heymes, F., Aprin, L., Johannet, A., Lapébie, E., Osmont, A., 2014. Atmospheric Turbulent Dispersion Modeling Methods using Machine learning Tools. Chem. Eng. Trans. 36, 517–522.

Lauret, P., Heymes, F., Aprin, L., Johannet, A., Munier, L., Lapébie, E., 2013. Near Field Atmospheric Dispersion Modeling on an Industrial Site Using Neural Networks 31, 151–156.

Lauret, P., Heymes, F., Aprin, L., Johannet, A., Slangen, P., 2015. 2D Modeling of Turbulent Flow around a Cylindrical Storage Tank by Artificial Neural Networks 43, 1621–1626. doi:10.3303/CET1543271

Meroney, R.N., 1999. Perspectives on Air Pollution Aerodynamics, in: 10th Internation Wind Engineering Conference. p. 14.

Minsky, M., Papert, S., 1969. Perceptrons. MIT Press, Cambridge MA.

Podnar, D., Koračin, D., Panorska, A., 2002. Application of artificial neural networks to modeling the transport and dispersion of tracers in complex terrain. Atmos. Environ. 36, 561–570. doi:10.1016/S1352-2310(01)00446-0

Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., Deylon, B., Glorennec, P.Y., 1995. Nonlinear Black-Box Modeling in System Identification: a Unified Overview. Automatica 31, 1691–1724.

Stone, M., 1974. Cross-Validatory Choice and Assessment of Statistical Predictions. J. R. Stat. Soc. Ser. B 36, 111–147.