QUANTIFICATION OF UNCERTAINTIES ASSOCIATED WITH AN INTEGRATED GAUSSIAN LINE SOURCE MODEL USING ENSEMBLES

Sam-Erik Walker
Norwegian Institute for Air Research (NILU), Kjeller, Norway

INTRODUCTION

Atmospheric dispersion models will always be uncertain due to the inevitable uncertainties associated with input data and physical formulations. It is important, however, to try to quantify such uncertainties in order to ensure more trust and transparency of accuracy in the modelling result. This paper deals with calculations of uncertainty in association with a newly developed dispersion model for open roads called WORM (Weak Wind Open Road Model).

THE WORM MODEL

The WORM model is an integrated Gaussian puff/plume model for calculating hourly average concentrations from open roads and highways in an arbitrary set of (nearby) receptor points. The model is thus similar to other well-known integrated Gaussian line source models, such as e.g., the CAR-FMI model, CALINE-4 model etc. In contrast to these models, however, the WORM model has been developed specifically with the aim of producing concentration values with some calibrated quantification of their uncertainties. More specifically, the WORM model aims at producing, for each receptor point, not a single concentration value, but rather a range of concentration values in the form of an ensemble, approximating a probability distribution of the conceived underlying, but unknown, true concentration.

The current version of the WORM model consists of the following system components:

- An emission pre-processor
- A pre-processor for meteorological data
- A background concentration pre-processor
- An integrated Gaussian puff/plume type of dispersion model

The emission pre-processor generates hourly emission data (Q in g/ms) for each lane of the roadway based on traffic data (*AirQUIS*, 2005). The background concentration pre-processor generates hourly background concentrations for the road, based on using nearby (upwind) background stations, or urban/regional scale models (*AirQUIS*, 2005). Background concentrations are added to the WORM model concentrations to make them comparable with local (roadside) air quality observations.

A meteorological pre-processor calculates meteorological parameters such as horizontal and vertical diffusivities (σ_u, σ_v, σ_w), mixing height (H_{mix}), Lagrangian time scales (T_{Lx}, T_{Ly}, T_{Lz}), friction velocity (u_*), temperature scale (θ_*) and Monin-Obukhov length scale (L), etc., based on Monin-Obukhov similarity theory, and hourly data for local wind speed and stability (vertical temperature gradient) (*AirQUIS*, 2005; *Walker, S.E. and J. Berger*, 2007). For the current version of the WORM model, a minimum value of horizontal plume diffusivity (σ_u and σ_v) equal to 0.5 m/s is used.

The dispersion model calculates hourly average concentrations in one or more receptor points by integrating a Gaussian puff or plume function along each lane of the road, adding up the contribution from each lane. The Gaussian plume function with dispersion parameters σ_y and

σ_z is used whenever the slanted plume approximation ($\sigma_x/x << 1$) is valid, but the model switches to a puff formulation, with dispersion parameter σ_x along the downwind x-axis, in low wind speed conditions, when the slanted plume approximation is no longer valid. Thus, the concentration in a receptor point $r = (x_r, y_r, z_r)$ is calculated by:

$$C_r = \int_{s=0}^{S}\int_{t=0}^{T} \frac{Q}{(2\pi)^{3/2}\cdot\sigma_x(t)\cdot\sigma_y(t)\cdot\sigma_z(t)} \cdot \exp\left(-\frac{(x_r(s) - U_{eff}\cdot t)^2}{2\sigma_x^2(t)}\right)\cdot\exp\left(-\frac{y_r^2(s)}{2\sigma_y^2(t)}\right) \cdot \left\{\exp\left(-\frac{(z_r - H_{eff})^2}{2\sigma_z^2(t)}\right) + \exp\left(-\frac{(z_r + H_{eff})^2}{2\sigma_z^2(t)}\right)\right\} dsdt \quad (1)$$

where Q is the emission intensity (g/ms), U_{eff} is the plume effective wind speed (m/s), H_{eff} is the plume effective height above ground (m), and where the coordinates of the receptor point and dispersion parameters in the integrand generally depends on the position s on the road, and time t since release. The concentration is obtained by integrating all the infinitesimal puffs over the length S (m) of the road, during the current hour (T = 3600 s). The integral is calculated numerically by using a highly accurate Gaussian quadrature routine.

In the model, growth of dispersion parameters $\sigma_x = \sigma_y$ and σ_z are calculated based on atmospheric background turbulence (*AirQUIS*, 2005; *Walker, S.E. and J. Berger*, 2007), with horizontal and vertical initial sizes of puffs or plumes, $\sigma_{x0} = \sigma_{y0}$ and σ_{z0}, calculated by the same semi-empirical equation for traffic-originated turbulence as used in the CAR-FMI model (*Härkönen, J. et al.*, 1996).

A comprehensive evaluation of the WORM model is given in *Berger, J. et al.* (2007) (this volume), where several open road models are compared using a Nordic data base of observed and model calculated values (the NORPAC study).

QUANTIFICATION OF UNCERTAINTIES

The WORM model is typically used to calculate hourly average concentrations in receptor points at most a few hundred meters away from the road. This makes the output concentration dependent on input data (emission, meteorology and background value) for the current hour, but virtually independent of data from previous hours. The model output concentration for each hour, can thus be viewed as a function of the input data from the current hour only.

If we let all input and internally calculated variables in the model for the current hour be denoted by the vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)$, the corresponding hourly average model output concentration C_r in a given receptor point $r = (x_r, y_r, z_r)$ can be written $C_r = C_r(\theta_1, \theta_2, \ldots, \theta_n)$. The typical approach in modelling is then, for each hour, to run the model using a set of nominal input and model calculated data $\boldsymbol{\theta}^0 = (\theta_1^0, \theta_2^0, ..., \theta_n^0)$ to produce model output concentrations $C_r^0 = C_r(\boldsymbol{\theta}^0) = C_r(\theta_1^0, \theta_2^0, ..., \theta_n^0)$ in one or more receptor points. The model results are usually given without any calculation, or estimation, of the uncertainties involved, i.e., errors or uncertainties associated with input and model calculated variables, and with the resulting model concentration values.

One of our aims in the development of the WORM model, however, as for other air pollution models in the future, is to produce concentration results with some form of quantification of the uncertainties involved. Uncertainties are generally best-treated using statistics and probability. Thus, describing uncertainties in results produced by a model generally means to assign probabilities to a range of possible model concentration values, instead of just

producing a single model result (with probability 1). The target is thus to provide model output concentrations, not as single numbers C_r, but rather as discrete or continuous probability distributions $\pi_r(c)$ of different values.

A model concentration result C_r is uncertain due to inevitable uncertainties associated with the model variables $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)$, and with the model formulation itself. These can generally be described using Bayesian statistics (*Box, G.E.P. and G.C. Tiao*, 1992). If $\boldsymbol{\theta}^t = (\theta_1^t, \theta_2^t, ..., \theta_n^t)$ denotes the correct (best input or true) values of the model variables for the current hour, and T_r denotes the correct or true concentration in receptor point r, we may write

$$T_r = C_r(\theta^t) + \varepsilon_r = C_r(\theta_1^t, \theta_2^t, ..., \theta_n^t) + \varepsilon_r \quad (2)$$

where ε_r denotes the model formulation error, i.e., the error induced by the model equations themselves, and not due to errors in the model data. If we are willing, and able, to put Bayesian subjective (prior) probabilities on all the model variables θ_i, for i = 1,…,n, and on model formulation errors ε_r, we obtain a Bayesian subjective (prior) probability distribution $\pi_r(T)$ associated with the true concentration T_r.

Generating the probability distribution $\pi_r(T)$ in the form of an explicit function, is, however, a difficult task. Instead an approximation is sought based on using a discrete set of points, or ensemble, $\{T^{(1)}, T^{(2)}, \ldots, T^{(N)}\}$, produced by randomly drawing values from the probability distribution $\pi_r(T)$, where N denotes the number of ensemble members (or ensemble size).

Based on the discrete ensemble of N concentration values $T^{(1)}, T^{(2)}, \ldots, T^{(N)}$, where each concentration value is associated with (discrete) probability mass 1/N, we may then calculate an estimate for the expected value E(T) by:

$$\hat{E} = \frac{1}{N}\sum_{i=1}^{N} T^{(i)} \quad (3)$$

Similarly, an estimate for the variance Var(T) can be calculated by:

$$\hat{V} = \frac{1}{N-1}\sum_{i=1}^{N} (T^{(i)} - \hat{E})^2 \quad (4)$$

Standard deviation is obtained by taking the square root of the variance. Using the ensemble of concentration values, other quantities associated with the probability distribution $\pi_r(T)$ such as different p-percentiles, e.g. the 90%-ile or the 95%-ile, can also be calculated.

The perhaps simplest way to describe $\pi_r(T)$, and to make random draws from it, is to define $\pi_r(T)$ using a series of conditional probability distributions associated with the model variables θ_i and formulation errors ε. This can be done recursively, by viewing the model variables and calculations as a directed acyclic graph. This is depicted in Fig. 1, showing a conceived fragment of the model, where a model variable θ_k is calculated based on other model variables $\theta_1, \theta_2, \ldots, \theta_j$, which have been indexed here from 1 to j for simplicity of notation.

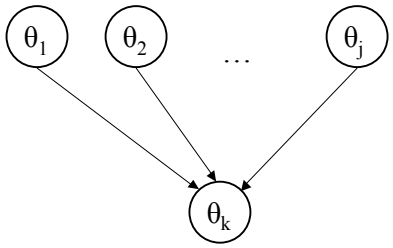

Fig. 1. Model calculation fragment (part of a directed acyclic graph).

The conditional probability distribution $\pi_k(\theta_k \mid \theta_1, \theta_2, \ldots, \theta_j)$ can then be defined recursively as follows. Assuming that the joint unconditional probability distribution $\pi_{1:j}(\theta_1, \theta_2, \ldots, \theta_j)$ is already defined in the acyclic directed graph for the variables $\theta_1, \theta_2, \ldots, \theta_j$, and that a probability distribution is locally defined for the local model formulation error ε_k, the conditional probability distribution for $\theta_k = \theta_k(\theta_1, \theta_2, \ldots, \theta_j) + \varepsilon_k$ is uniquely defined. The argument can be repeated until a (conditional) probability distribution has been defined for all model variables in the graph, and subsequently also for the last model variable calculated, the model output concentration.

This graph oriented recursive definition of the prior probability distribution $\pi_r(T)$ enables us also easily to simulate from this distribution, i.e., to draw samples from $\pi_r(T)$. This is done in the same manner as we defined the conditional probability distributions above. Consider again the directed acyclic graph model fragment given in Fig. 1. If we assume that we already have obtained a sample $(\theta_1, \theta_2, \ldots, \theta_j)$ from $\pi_{1:j}(\theta_1, \theta_2, \ldots, \theta_j)$, we may draw a sample of θ_k from $\pi_k(\theta_k \mid \theta_1, \theta_2, \ldots, \theta_j)$ by drawing a sample ε_k from the distribution of the local model formulation error, and adding this to the function value $\theta_k = \theta_k(\theta_1, \theta_2, \ldots, \theta_j)$. The resulting calculated value of θ_k will then be a sample from $\pi_k(\theta_k \mid \theta_1, \theta_2, \ldots, \theta_j)$. We may then continue this process of calculating (samples of) model variables until we obtain a sample of model output concentration $T^{(i)}$, from $\pi_r(T)$. By repeating the procedure N times we obtain our ensemble of N model concentrations, representing a set of N independent and identical (exact) samples from the prior probability distribution $\pi_r(T)$.

With recent and forthcoming advances in computer parallel processing capabilities, calculating a large number of such samples by randomly drawing model variables and running the model on a parallel basis, should become more practically feasible in the future. It is also important to use available air quality observations to calibrate and adjust the prior probability distribution $\pi_r(T)$. This could be done e.g., by making sure that calculated p% confidence intervals contains observed concentrations around p% of the time.

RESULTS

Fig. 2 shows the result of running the WORM model for NO_x, using data from an 850 m long 4-lane roadway at Nordbysletta, close to Oslo, Norway (*Walker, S.E. and J. Berger*, 2007).

The graph contains hourly average observed concentrations (thick line), for a station situated 17 m from the roadway at a height of 3.5 m above the ground, together with model calculated concentrations for the same receptor point, in the form of an ensemble mean (broken line), together with a 90% confidence interval (lower and upper thin lines). The period covered is 3 February – 10 February 2002, but only hours with wind direction towards the station are included. The number of ensemble members used is N = 1000.

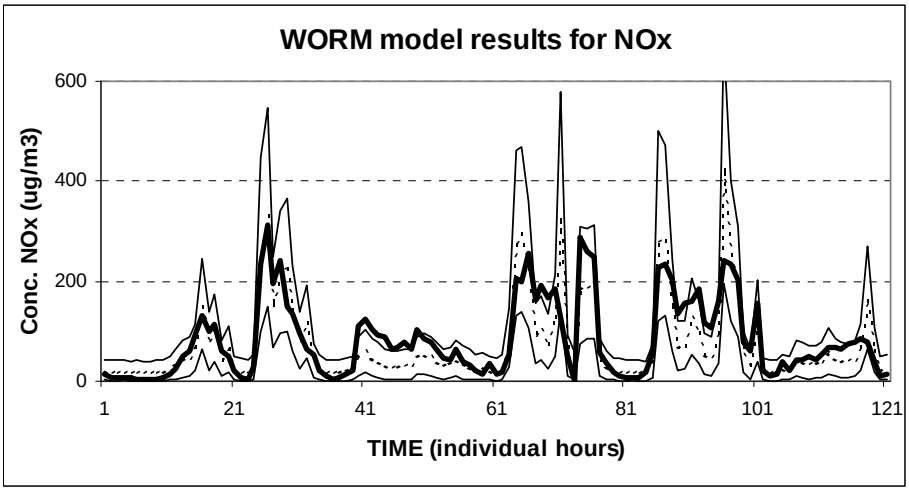

Fig. 2 WORM model results for NO_x.

The individual ensemble members are created by drawing the following model variables: U_{10m} (wind speed at 10 m), u_*, θ_*, L, $\theta_v = \tan^{-1}(\sigma_v/u))$, $\theta_w = \tan^{-1}(\sigma_w/u))$, and initial size of plume σ_{y0} ($\sigma_{z0} = \sigma_{y0}/2$) using Gaussian probability distributions around each nominal or model calculated (derived) value, with standard deviations set to 10% of respective mean values, except for U_{10m}, which is locally observed, and where the standard deviation has been set to 0.15 m/s. The standard deviation of model formulation error has been set to $\sqrt{20^2 + 0.3^2 \cdot C_r^2}$, where C_r is the corresponding model calculated concentration. Using this uncertainty model, 90% confidence intervals have been calculated using the ensembles (lower and upper thin lines in Fig. 2), and empirically here they contain the observed concentrations (thick line in Fig. 2) in about 85% of the hours.

SUMMARY AND CONCLUSION

A new integrated Gaussian line source model for open roads (WORM) is presented, which produces its output not as single concentration values, but rather as ensembles of model calculated values, based on quantification of uncertainties in the model variables and physical formulation. Based on such ensembles, ensemble mean values and p% (e.g., p = 90) confidence intervals for the true (or observed) concentrations can be calculated. Some preliminary, but encouraging, results using data from a 4-lane roadway at Nordbysletta, close to Oslo, Norway is presented.

REFERENCES

AirQUIS, 2005: Norwegian Institute for Air Research. http://www.airquis.com.

Berger, J., S.E. Walker, B. Denby, P. Løfstrøm, M. Ketzel, R. Berkowicz and J. Kukkonen, 2007: Evaluation and inter-comparison of open road line source models currently being used in the Nordic countries: A NORPAC project (this volume).

Box, G.E.P and G.C. Tiao, 1992: Bayesian Inference in Statistical Analysis, Wiley Classics Lib. Ed., New York.

Härkönen, J., E. Valkonen, J. Kukkonen, E. Rantakrans, K. Lahtinen, A. Karppinen and L. Jalkanen, 1996: A model for the dispersion of pollution from a road network. Finnish Meteorological Institute, Publications on Air Quality 23, Helsinki.

Walker, S.E. and J. Berger, 2007: Application of data assimilation in open road line source modelling, 6th International Conference on Urban Air Quality, Cyprus.