2010-06-03

Kalman Filter in the Real Time URBIS model

Date	June 2010
Author(s)	R. Kranenburg
Keywords	NOx, Real Time URBIS, Uncertainty, Rijnmond, Kalman filter, Screening process
Target	Master Thesis

Preface

In the context of my study applied mathematics. The last nine months I have done a graduate research for TNO. In this research, I have connected a model simulation for the air pollution in the Rijnmond area around Rotterdam, with a series of measurements to create a better view on the air pollution. This is described in the first part of this report. Further in the second part of this report, I have developed some strategies to create a better view on the air pollution.

It was a nice experience the past nine months. It was a challenge to connect mathematics with the different disciplines in this research. Therefore I want to thank a few people. First Michiel Roemer, Jan Duyzer and Arjo Segers from TNO, for the input of physical aspects in this research. Further Kees Vuik and Arnold Heemink for the mathemical input in this research.

Richard Kranenburg

Contents

Part I

Analysis of the uncertainty of the Real Time URBIS model with a Kalman filter

1 Introduction

In the first part of this report the use of a Kalman filter in the Real Time URBIS model will be discussed. The Real Time URBIS model is a model which calculates the concentration NO_x in a city or in an industrialized region. The concentration NO_x is assumed to be equal to the sum of the concentrations NO and NO_2. Nitrogen oxides are formed by the burning of fossil fuels in traffic and industry, they will arise if nitrogen from the air and from fuels reacts with oxygen. These nitrogen oxides reacts under influence of sunlight to air pollution, like smog and acid rain. Nitrogen oxides can also causes trouble for the eyes and lungs. Therefore the European commission has set out limit values for the concentrations of NO_2, thus it is important to have a good view on the concentrations NO_x and with that on the concentrations NO_2. The limit values for the concentration NO_2 are given in Appendix 2 of 'Wet Milieubeheer' [Cramer, 2007].

The Real Time URBIS model simulates the concentration NO_x by adding emissions from different sources like traffic, residents, shipping and industry. In this report a Kalman filter will be used to link the model simulations with a series of measurements made on 9 different monitoring stations. With this link a better simulation for the concentration NO_x can be given. Also a statistical uncertainty interval of the concentration NO_x can be given.

In Chapter 2, a more detailed explanation of the Real Time URBIS model is given. In Chapter 3, a statistical uncertainty analysis of the model is made, with this analysis some ideas for the Kalman filter are constructed. In Chapter 4, the general use of a Kalman filter is explained. In Chapter 5, the Kalman filter is applied on the background concentrations in the Real Time URBIS model. In Chapter 6, the Kalman filter is applied on all the different emission sources. In the last part of Chapter 6, the uncertainty intervals, calculated with the Kalman filter, will be connected with the population to give a functional application of this method. The conclusions and discussion are given in Chapter 7.

2 Model and measurements

2.1 Real Time URBIS model

Real Time URBIS is a model to determine the concentration NO_x in a city or in an industrialized region. The model calculates on each hour a concentration NO_x for the whole region, based on factors like wind, temperature and time. This study focuses on the Rijnmond area around Rotterdam; the domain of the study is shown in Figure 2.1.

The basis of the Real Time URBIS model is the URBIS model. The URBIS model calculates 88 annual mean concentrations NO_x. These 88 annual mean concentrations are concentrations caused by 11 different emission sources for 4 different wind directions and 2 different wind speeds. Further is this report, the 88 annual mean concentrations are called standard concentration fields. Plots of all standard concentration fields are included in Appendix B. Detailed information about the URBIS model can be found in [Wesseling and Zandveld, 2003].

With the Real Time URBIS model, the annual mean concentrations are used to calculate a hourly mean concentration. The state of the Real Time URBIS model consists the NO_x concentrations in a large number of grid points in the domain. Mathematically, the state is described by a vector:

$$\underline{c}_k \tag{2.1}$$

where k denotes the hour. In this study, the state vector is defined on about 94096 grid points covering the Rijnmond area, irregularly distributed over the grid. The state is computed as a linear combination of standard concentration fields. This is given in the state equation:

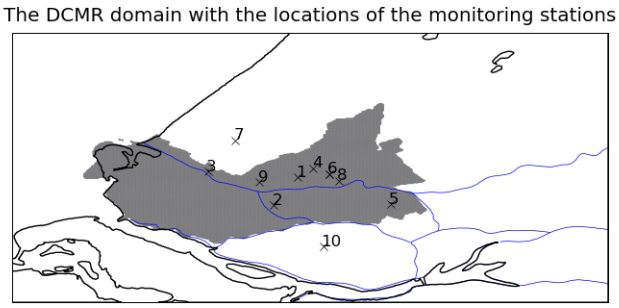
$$\underline{c}_k = M\underline{\mu}_k^{\mathrm{T}} \tag{2.2}$$

where each column of the matrix M is one of the standard concentration fields. The elements of vector μ_k represent the weights of each standard concentration field at hour k. The weights depends on the meteorological conditions (wind direction, wind speed, temperature) and the moment (month, day of the week, hour).

In Spaubek [2004] and [Kranenburg, 2009] a more detailed description of the Real Time URBIS model can be found. Note that the Real Time URBIS model described in those reports has an underlying URBIS model valid for the year 2000, while in this report the underlying URBIS model is valid for the year 2006. This URBIS model for 2006 has one large difference with the URBIS model for the year 2000. The model for 2000 consists of 10 different source categories instead of 11 in 2006, since an extra source in category traffic is added, namely the source 'Zone card'. Further is the source category 'Boundary' renamed into 'Rest'.

2.2 Measurements

In the Rijnmond area, there are also 11 monitoring stations, which monitores the concentrations NO and NO_2. The sum of these two concentrations is called NO_x. The locations of these 11 monitoring stations are also shown in Figure 2.1. Locations 1-6 are stations operated by DCMR[1], the locations 7-11 are stations operated by RIVM[2]. Monitoring stations 6 and 11 are located directly next to each other, thus in the Real Time URBIS model both locations have the same coordinates. Only 9 of these locations are in the domain covered in this study, locations 7 (Schipluiden) and 10 (Westmaas) are just outside of the domain and will only be used as background stations. These stations monitores the concentration which is blown into the area from the rest of the Netherlands. As will described in Chapter 3, the results of the measurements on the 9 locations inside the area can be used to estimate the uncertainty of the model, by comparing the model results with the results of the measurements. Further in this report the results of the measurements will be called observations.



DCMR - Locations	
1	Schiedam
2	Hoogvliet
3	Maassluis
4	Overschie
5	Ridderkerk
6	Bentinckplein

RIVM - Locations	
7	Schipluiden
8	Schiedamse Vest
9	Vlaardingen
10	Westmaas
11/6	Bentinckplein

Figure 2.1: *Domain of the working area for Real Time URBIS*

[1]DCMR: Dienst Centraal Milieubeheer Rijnmond. *www.dcmr.nl*
Environmental protection agency for the Rijnmond area around Rotterdam
[2]RIVM: RijksInstituut voor Volksgezondheid en Milieu. *www.rivm.nl*
Dutch institute for public health and environment

3 Statistical uncertainty of the Real Time URBIS model

3.1 Introduction

In [Kranenburg, 2009], a method is described to compute a bias correction for the simulations made by the Real Time URBIS model, by comparing the model simulations with the observations made on the 9 monitoring stations. After application of the Real Time URBIS model, the simulation is adjusted with a value dependent on the different meteorological conditions (wind direction, wind speed, temperature) and the moment (month, day of the week, hour). This correction is typically an example of post-processing; after applying the model, the model results are corrected with the aid of the observations. In addition, with the dependencies on the meteorological conditions and the moment, the origin of the uncertainties in the Real Time URBIS model can be found. In this chapter, the same method is applied on the Real Time URBIS model with the underlying URBIS model for 2006. For the year 2006 all the differences between the observations and the model simulations are calculated. All these differences are used to make a correction on the results of the Real Time URBIS model.

3.2 Log-normal distributions

For all 9 monitoring stations in the area, the observations are plotted in a histogram. This is shown in the left panel of Figure 3.1. In the right panel of Figure 3.1, the model simulations for the locations of this monitoring stations are plotted in a histogram. For each location, the model simulation is made by a weighted average of the model simulations on the grid points within a fixed distance from that monitoring station. The grid can be split into two parts: one grid with distance between two grid points equal to 100 meters and one special grid with a high resolution on busy local roads. Therefore the fixed distance to calculate the weighted average for each monitoring station will be equal to 150 meters. All 4 neighboring grid points from the first grid will then be involved in the weighted average. In Appendix A, the locations of all the monitoring stations are shown together with the surrounding grid points, which are involved in the weighted average.

It is important to notice that both the observations as well as the model simulations have a log-normal distribution. For this reason all corrections should be done in the log-domain. The main advantage of working in the log-domain is that, when a correction is added to the state, this correction is made on the logarithm of the concentration. After correction, the logarithm of the concentration could become negative but the corresponding concentration itself can not. In fact, an additive correction on the logarithm of the concentration is the same as a fractional correction on the absolute concentration:

$$\ln(c_k) \rightarrow \ln(c_k) + \lambda \tag{3.1}$$
$$c_k = e^{\ln(c_k)} \rightarrow e^{\ln(c_k)+\lambda} = e^{\lambda} c_k \tag{3.2}$$

where c_k is the concentration at time k and λ is the correction term. Since the correction factor e^{λ} is always positive, the concentrations will remain positive too. Detailed information about corrections in the log-domain is given in [Kranenburg, 2009].

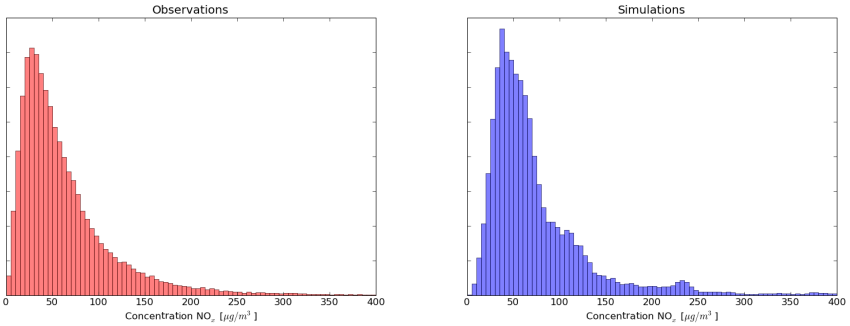


Figure 3.1: *Both the observations and the model simulations have a log-normal distribution.*

3.3 Uncertainty of the Real Time URBIS model

The method from [Kranenburg, 2009] to describe the uncertainty of the model is now applied on the Real Time URBIS model, with underlying URBIS model for 2006. For the year 2006 all logarithms of the observations made on the 9 monitoring locations are compared with logarithms of the model simulations. In total there are at most $9 \text{ stations} \times 8760 \text{ hours} = 78840$ of those differences. Due to some missing measurements or meteorological data, for 2006 there are only 67080 differences. With all these differences, the correction on the results of the Real Time URBIS model is made.

3.3.1 Structural bias

First the differences between the model results and the observations did not have mean zero, thus there is a systematical error in the model. This structural error causes a constant correction ($\lambda = \lambda_c$) of the logarithm of the model simulation, which corresponds with a constant fractional correction of the absolute model simulation.

3.3.2 Wind direction dependency

The differences between logarithms of the observations and the logarithms of the constant corrected simulations are plotted with respect to the wind direction. The wind directions are given by 10 degrees accurate, thus in total 36 wind directions are possible. For each wind direction, all differences which appears during that wind direction are taken. In Figure 3.2 all the means λ_i per wind direction are plotted as blue dots. The standard deviations σ_i of the differences per wind direction are represented by the length of the error bars. The values for λ_i and σ_i are given by:

$$\lambda_i = \frac{1}{n_i}\sum_{k=1}^{n_i}\left(\ln\left(y_{i,k}\right) - \ln\left(c_{i,k}^m\right)\right) \tag{3.3}$$

$$\sigma_i = \sqrt{\frac{1}{n_i}\sum_{k=1}^{n_i}\left(\left(\ln\left(y_{i,k}\right) - \ln\left(c_{i,k}^m\right)\right) - \lambda_i\right)^2} \tag{3.4}$$

where n_i represents the number of differences that appears during wind direction i, while y_i represents the observations and c_i^m the model simulations, during wind direction i.

The green line in Figure 3.2 forms the correction which is added to the logarithms of the model simulations. The correction on the model is now a function of the wind direction ϕ, thus $\lambda = \lambda_c + \lambda_{\text{wdir}}(\phi)$. This green line is a composed sinus function, that fits best on the differences between the model simulations and the observations. This best fitting is made with the blue dots and the wind rose in Figure 3.3, the weights for each blue dot are given in this wind rose. When a wind direction occurs a lot, the weight must be larger in the calculation of the best fitting sinus.

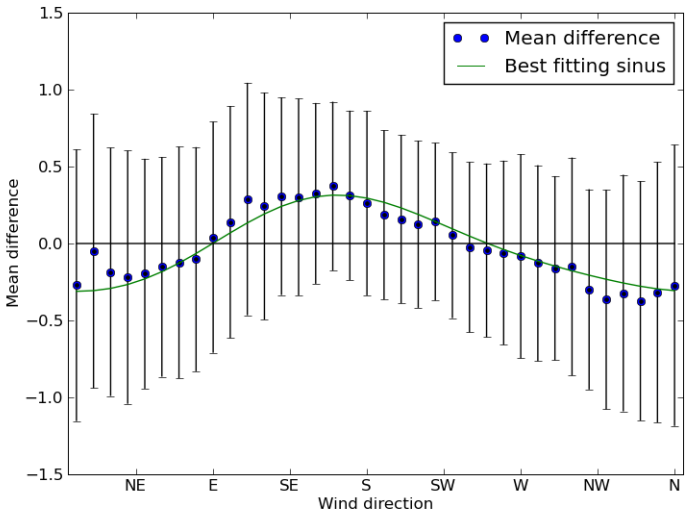


Figure 3.2: *Mean differences between the logarithms of the observations and the logarithms of the model simulations against the wind direction after constant correction.*

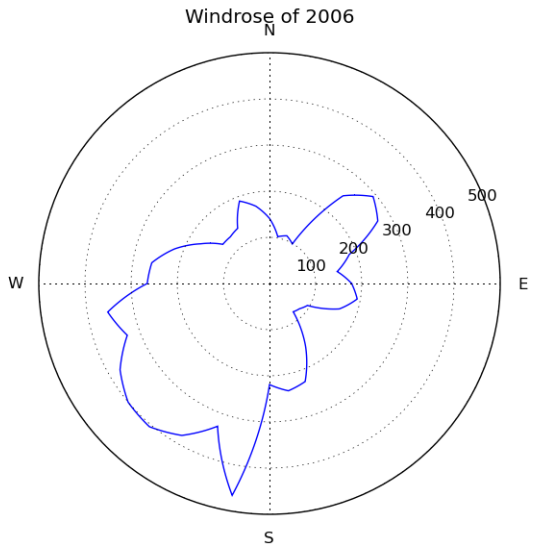


Figure 3.3: *Wind rose over 2006*

3.3.3 Hour dependency

After the correction on the wind direction, all the differences are plotted with respect to the hour of the day, this is shown in Figure 3.4. In this figure, the means of all differences per hour of the day are plotted with a blue dot and the standard deviations are represented by the widths of the error bars, computed similar to equations 3.3 and 3.4. The green line is again a composed sinus function, which fits best on the blue dots. Because of missing measurements or meteorological data, not every hour has the same contribution in calculating the best fitting sinus. This best fitting sinus forms the correction added to the logarithms of the model simulations as a function of the hour of the day. The total correction on the model is now built from three parts: a constant part, a function dependent on the wind direction (ϕ) and a function dependent on the hour of the day (h):

$$\lambda \quad = \quad \lambda_c + \lambda_{\text{wdir}}(\phi) + \lambda_{\text{hour}}(h) \tag{3.5}$$

After this correction, the differences were plotted against the other input parameters wind speed, temperature, month and day of the week. It was found that the differences are no longer dependent on one of these input parameters.

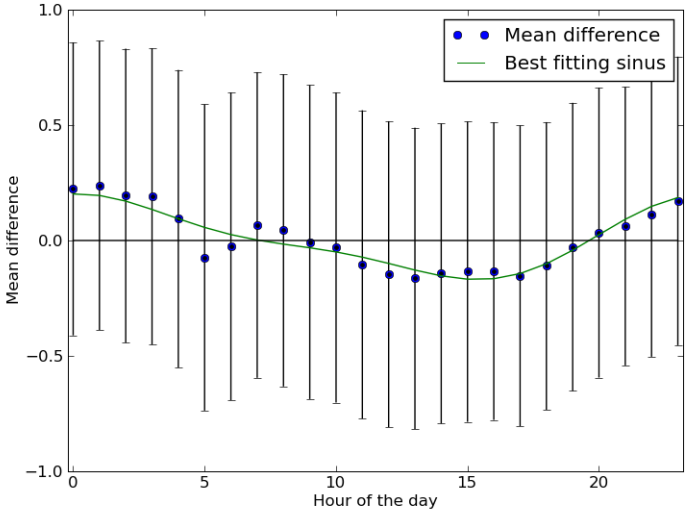


Figure 3.4: *Mean difference between the logarithms of the observations and the logarithms of the model simulations against the hour of the day, after correction on the wind direction*

3.3.4 Standard deviation of the differences

The standard deviation of the differences was found to be a function of the wind speed. This is shown in Figure 3.5. The blue dots represent the standard deviation of all differences as a function of wind speed. This is done with an equation similar to equation 3.4. The red line is the best fitting exponential function on the standard deviations per wind speed. In the calculation of this best fitting exponential the number of times that a wind speed occurs is also taken. When a wind speed occurs a lot, the weight must be larger in the calculation of the best fitting exponential.

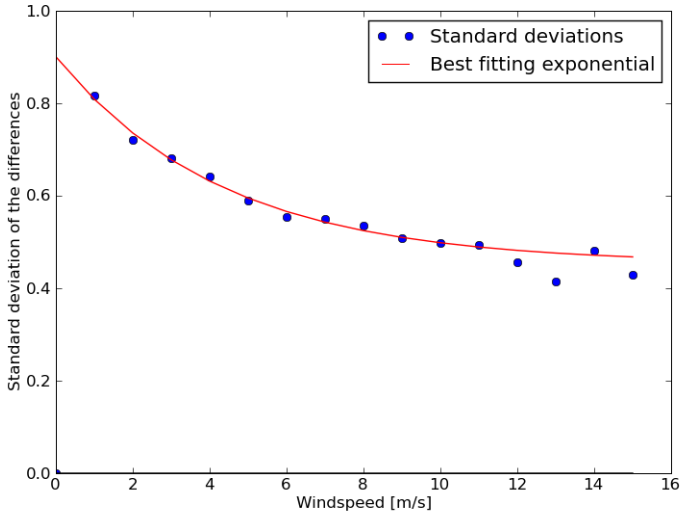


Figure 3.5: *The standard deviations of the differences between the logarithms of the observations and the logarithms of the model simulations against the wind speed, after correction on the hour of the day*

3.4 Discussion

With the method described above, an uncertainty interval for the concentration NO_x can be given. In Figure 3.6 the 1σ uncertainty interval for the first week of 2006 is given for the location of the monitoring station in Schiedam. The red dots represents the available observations made on the station in Schiedam. The described method gives a useful approximation of the uncertainty of the model, however the uncertainty of the model is very large at some times.

The next object is to decrease the uncertainty of the model by an improvement of the model. An indication for the largest inaccuracy of the model is given by the uncertainty analysis above. The differences between the observations and the model simulations are mainy dependent on the wind direction. For this reason it is assumed that the standard concentration fields for the source 'Background' are not accurate in the URBIS model. The source 'Background' corresponds with emission produced in the rest of the country which is blown into the Rijnmond area. Of course this source has a large dependency on the wind direction. In Chapter 5, a Kalman filter will be used to get better estimates of the background concentrations per wind direction. The advantage of using a Kalman filter is that also the uncertainty of the measurements is involved in the estimation. First in Chapter 4 the working of a Kalman filter is explained.

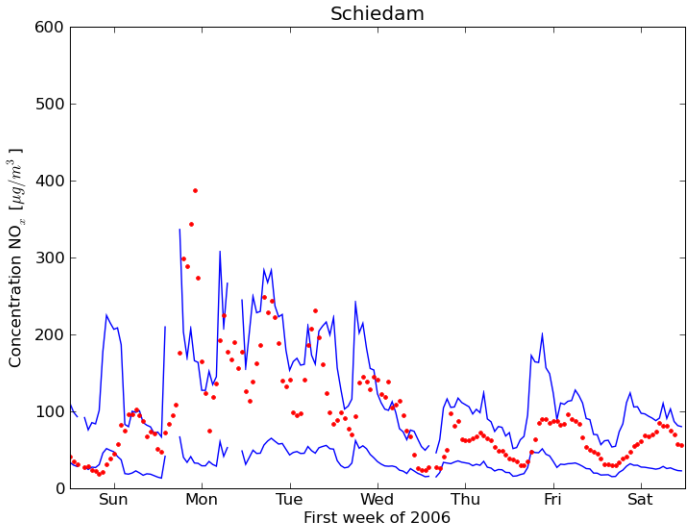


Figure 3.6: *Uncertainty interval for the first week on location Schiedam*

4 Kalman filter

4.1 Introduction

A Kalman filter is mostly used to smooth random errors in the model of a dynamical system. In Figure 4.1 a schematic representation of the working of a Kalman filter is given. The simulations made by the model for time step k are corrected with the aid of a measurement on time step k. In this correction, also the uncertainties of the model and the measurements are taken into account. This application is very useful in a real time application such as the Real Time URBIS model. Detailed information about a Kalman filter can be found in [Heemink, 1996] and [Welch and Bishop, 2006].

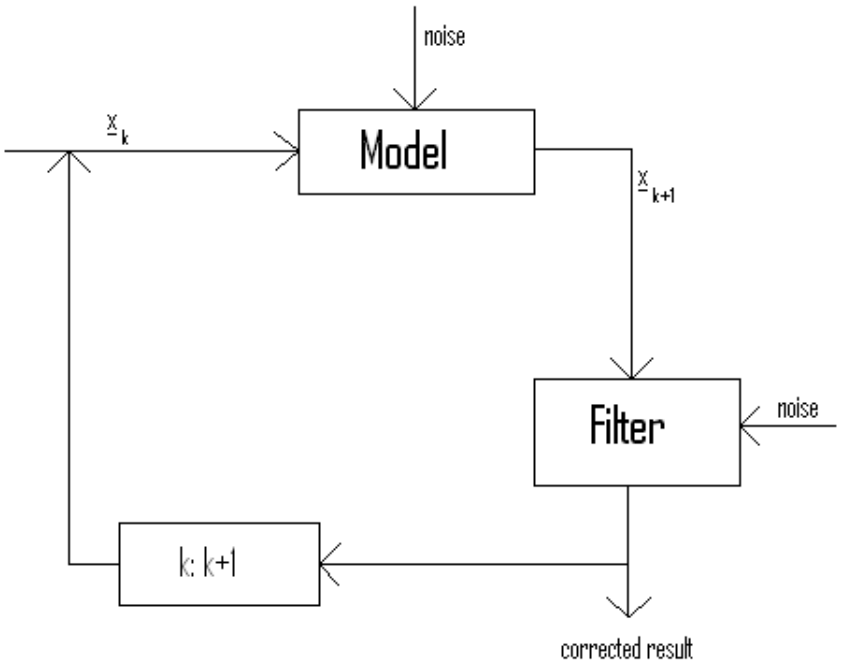


Figure 4.1: *Schematic representation of the Kalman filter*

4.2 Algorithm of Kalman filtering

In this section the working of a Kalman filter is explained with a simple one dimensional example. The simulations made with the Real Time URBIS model, for location Schiedam are compared with the observations on location Schiedam. This is done for the year 2006, in this year the Real Time URBIS model gives for 7906 of the 8760 hours a concentration NO_x. For the other hours, one of the meteo input data is missing, thus the model cannot give a result. For 8603 of the 8760 hours there is an observation, on the other hours, the measurement was incorrect or missing. When

the model does not give a result, the Kalman filter cannot give a result on that time step. When there is no measurement, the Kalman filter can give a result, which is computed from the previous time step. In the figures in this chapter there will be some 'holes', these are due to the missing model data.

4.2.1 Dynamical system

First of all it is important to have a well defined dynamical system. For the example in Schiedam, this dynamical system is given by:

$$\ln(c_k) = \ln(c_k^m) + \gamma_k \tag{4.1}$$

$$\gamma_{k+1} = \alpha\gamma_k + \beta_k\omega_k \qquad \omega_k \sim N(0,1) \tag{4.2}$$

In this equations the value c_k is the concentration of NO_x on time step k on location Schiedam. Every time step the Real Time URBIS model calculates a concentration NO_x on location Schiedam; these are called c_k^m. Because of the log-normal distribution of the concentration NO_x, the dynamical system deals with the logarithms of the concentration NO_x. More about this is discussed in Section 3.2 and in [Kranenburg, 2009]. The parameter γ_k is an estimate for the difference between the logarithm of the real concentration and the logarithm of the model result, also called the perturbation on the model. With a Kalman filter these perturbations γ_k will be estimated.

This estimation does not lead to a computation of the optimal value for γ_k. Instead the result after application of the Kalman filter is that the value of γ_k can be found in a Gaussian distribution with mean $\hat{\gamma}_k$ and a variance p_k^2. With this Gaussian distribution, the value for $\ln(c_k)$ can be found in a Gaussian distribution with mean $(\ln(c_k^m) + \hat{\gamma}_k)$ and variance p_k^2. This all leads to an uncertainty interval for the logarithm of the concentration NO_x at every time step and with that, an uncertainty interval for the absolute concentration NO_x .

For the perturbations, it is assumed that a perturbation at time k is correlated with the perturbation on time $k-1$, but that it also has a random component. A suitable mathematical description is an 'AR1' (auto-regressive 1) process, this is also called 'colored noise'. The temporal correlation is described by the parameter α, which also appears in the formula for the amplitude of the random contribution:

$$\beta_k = \sqrt{1-\alpha^2}\sigma_k \tag{4.3}$$

When $\alpha = 0$, the perturbation only has the random process, thus only white noise with standard deviation σ.

When α is close to one, the temporal correlation is strong and the fluctuations per time step are small. The value α is computed from:

$$\alpha = e^{-1/\tau} \tag{4.4}$$

where τ is a de-correlation scale. In this example the value of τ is chosen equal to 12, such that the perturbation is practically independent of the perturbation 12 time steps before.

4.2.2 Kalman filter form

For application of the Kalman filter, the dynamical system has to be written in the Kalman filter form:

$$\gamma_{k+1} = \alpha\gamma_k + \beta_k\omega_k \qquad \omega_k \sim N(0,1) \tag{4.5}$$

$$\ln(y_k) = H\left(\ln(c_k^m) + \gamma_k\right) + \nu_k \qquad \nu_k \sim N\left(0, r_k^2\right) \tag{4.6}$$

In here y_k is the observation on time step k, and H is the system operator, which projects the model state onto the observations. The observation error ν_k represents the error of the measurement, combined the instrumental error and the representation error, which is supposed to be Gaussian with zero mean and variance r_k^2.

For the example on location Schiedam, the system operator H is equal to 1, which means that the observation is just the model plus some perturbation. The value of r_k is assumed to be equal to 0.2. This means that the logarithms of the observations have an uncertainty of 20%. Also the value σ_k is set to 0.2 too, which means that the perturbation on the model also has an uncertainty of 20 %.

The Kalman filter process could be started with initial values $\gamma_0 = 0$, and $p_0^2 = 0$; this is equivalent to the assumption that the expected concentration at time 0 equals the model result and the uncertainty is zero at this time.

4.2.3 Forecast step

In the first step of the Kalman filter, a forecasted mean $\hat{\gamma}_k^f$ of the perturbation is calculated with the mean from the previous time step. This forecasted mean is the expectation of γ_k:

$$\begin{aligned}
\hat{\gamma}_{k+1}^f &= \mathrm{E}\left[\gamma_{k+1}\right] \\
&= \mathrm{E}\left[\alpha\gamma_k + \beta_k\omega_k\right] \\
&= \alpha\mathrm{E}\left[\gamma_k\right] + \beta_k\mathrm{E}\left[\omega_k\right] \\
&= \alpha\mathrm{E}\left[\gamma_k\right] \\
&= \alpha\hat{\gamma}_k
\end{aligned} \tag{4.7}$$

where is used that $\mathrm{E}\left[\omega_k\right] = 0$. For the example of Schiedam the temporal correlation is stated:

$$\alpha = e^{-1/12} \approx 0.92$$

Also a forecasted variance $\left(p_{k+1}^f\right)^2$ is calculated with the variance from the time

step before:

$$\begin{aligned}
\left(p_{k+1}^{f}\right)^{2} &= \mathrm{VAR}\left(\gamma_{k+1}\right) \\
&= \mathrm{E}\left[\left(\gamma_{k+1}-\mathrm{E}\left[\gamma_{k+1}\right]\right)^{2}\right] \\
&= \mathrm{E}\left[\left(\alpha\gamma_{k}+\beta_{k}\omega_{k}-\mathrm{E}\left[\alpha\gamma_{k}+\beta_{k}\omega_{k}\right]\right)^{2}\right] \\
&= \mathrm{E}\left[\left(\alpha\gamma_{k}+\beta_{k}\omega_{k}-\alpha\mathrm{E}\left[\gamma_{k}\right]\right)^{2}\right] \\
&= \mathrm{E}\left[\left(\alpha\left(\gamma_{k}-\mathrm{E}\left[\gamma_{k}\right]\right)+\beta_{k}\omega_{k}\right)^{2}\right] \\
&= \mathrm{E}\left[\alpha^{2}\left(\gamma_{k}-\mathrm{E}\left[\gamma_{k}\right]\right)^{2}+2\alpha\beta_{k}\left(\gamma_{k}-\mathrm{E}\left[\gamma_{k}\right]\right)\omega_{k}+\beta_{k}^{2}\omega_{k}^{2}\right] \\
&= \alpha^{2}\mathrm{E}\left[\left(\gamma_{k}-\mathrm{E}\left[\gamma_{k}\right]\right)^{2}\right]+2\alpha\beta_{k}\mathrm{E}\left[\gamma_{k}-\mathrm{E}\left[\gamma_{k}\right]\right]\mathrm{E}\left[\omega_{k}\right]+\beta_{k}^{2}\mathrm{E}\left[\omega_{k}^{2}\right] \\
&= \alpha^{2}\mathrm{E}\left[\left(\gamma_{k}-\mathrm{E}\left[\gamma_{k}\right]\right)^{2}\right]+\beta_{k}^{2} \\
&= \alpha^{2}\mathrm{VAR}\left(\gamma_{k}\right)+\beta_{k}^{2} \\
&= \alpha^{2}\left(p_{k}\right)^{2}+\left(1-\alpha^{2}\right)\sigma_{k}^{2} \qquad (4.8)
\end{aligned}$$

where the independency of γ_k and ω_k is used, as well as $\mathrm{E}\left[\omega_k\right]=0$ and $\mathrm{E}\left[\omega_k^2\right]=\mathrm{VAR}\left(\omega_k\right)=1$.

4.2.4 Analysis step

In the second step, the Kalman filter analyzes the results of the forecast step with an observation. A basic assumption in a Kalman filter is that the mean after the analyzing step $\hat{\gamma}_k^a$ is a linear combination of the forecasted mean and the difference between the logarithm of the observation and the logarithm of the model simulation. This results in an analyzed mean which is the forecasted mean plus a perturbation relative to the difference between the observation and its related simulation:

$$\hat{\gamma}_{k+1}^{a} = \hat{\gamma}_{k+1}^{f}+K_{k+1}\left(\ln\left(y_{k+1}\right)-H\left(\ln\left(c_{k+1}^{m}\right)+\hat{\gamma}_{k+1}^{f}\right)\right) \qquad (4.9)$$

The variance in this analyzing step $\left(p_{k+1}^{a}\right)^{2}$ is created by the variance from the forecast step and the variance from the representation error of the measurements:

$$
\begin{aligned}
\left(p_{k+1}^{a}\right)^{2} &= \mathrm{VAR}\left(\gamma_{k+1}\right) \\
&= \mathrm{E}\left[\left(\gamma_{k+1} - \mathrm{E}\left[\gamma_{k+1}\right]\right)^{2}\right] \\
&= \mathrm{E}\left[\left(\gamma_{k+1} - \hat{\gamma}_{k+1}^{a}\right)^{2}\right] \\
&= \mathrm{E}\left[\left(\gamma_{k+1} - \left(\hat{\gamma}_{k+1}^{f} + K_{k+1}\left(\ln\left(y_{k+1}\right) - H\left(\ln\left(c_{k+1}^{m}\right) + \hat{\gamma}_{k+1}^{f}\right)\right)\right)\right)^{2}\right] \\
&= \mathrm{E}\left[\left(\gamma_{k+1} - \left(\hat{\gamma}_{k+1}^{f} + K_{k+1}\left(H\left(\ln\left(c_{k+1}^{m}\right) + \gamma_{k+1}\right) + \nu_{k+1}\right)\right.\right.\right. \\
&\quad \left.\left.\left. - H\left(\ln\left(c_{k+1}^{m}\right) + \hat{\gamma}_{k+1}^{f}\right)\right)\right)^{2}\right] \qquad (4.10a) \\
&= \mathrm{E}\left[\left(\left(1 - K_{k+1}H\right)\left(\gamma_{k+1} - \hat{\gamma}_{k+1}^{f}\right) - K_{k+1}\nu_{k+1}\right)^{2}\right] \\
&= \mathrm{E}\left[\left(1 - K_{k+1}H\right)^{2}\left(\gamma_{k+1} - \hat{\gamma}_{k+1}^{f}\right)^{2}\right. \\
&\quad \left. - 2K_{k+1}\left(1 - K_{k+1}H\right)\left(\gamma_{k+1} - \hat{\gamma}_{k+1}^{f}\right)\nu_{k+1} + K_{k+1}^{2}\nu_{k+1}^{2}\right] \\
&= \left(1 - K_{k+1}H\right)^{2}\mathrm{E}\left[\left(\gamma_{k+1} - \mathrm{E}\left[\gamma_{k+1}\right]\right)^{2}\right] \\
&\quad + 2K_{k+1}\left(1 - K_{k+1}H\right)\mathrm{E}\left[\left(\gamma_{k+1} - \hat{\gamma}_{k+1}^{f}\right)\right]\mathrm{E}\left[\nu_{k+1}\right] + K_{k+1}^{2}\mathrm{E}\left[\nu_{k+1}^{2}\right] \\
&= \left(1 - K_{k+1}H\right)^{2}\left(p_{k+1}^{f}\right)^{2} + K_{k+1}^{2}r_{k+1}^{2} \qquad (4.10b)
\end{aligned}
$$

where the independency of γ_k and ν_k is used, as well as $\mathrm{E}\left[\nu_{k+1}\right] = 0$ and $\mathrm{E}\left[\nu_{k+1}^{2}\right] = 1$. In line 4.10a, the Formula 4.6 is used.

After this analyzing step, the values for $\hat{\gamma}_k^a$ and $\left(p_k^a\right)^2$ are the mean $\hat{\gamma}_k$ and the variance $\left(p_k\right)^2$ for the state of the system on time k.

A common choice for K_k is the minimum variance gain. For that gain, the value K_k is chosen such that the variance $\left(p_k^a\right)^2$ reaches a minimum. To obtain the minimum variance, the solution for K_{k+1} of

$$
\frac{\partial\left(p_{k+1}^{a}\right)^{2}}{\partial K_{k+1}} = 0 \qquad (4.11)
$$

has to be found. This is done in the next formula:

$$
\begin{aligned}
\frac{\partial\left(p_{k+1}^{a}\right)^{2}}{\partial K_{k+1}} &= 0 \\
2K_{k+1}r_{k+1}^{2} - 2H\left(1 - K_{k+1}H\right)\left(p_{k+1}^{f}\right)^{2} &= 0 \\
2K_{k+1}\left(r_{k+1}^{2} + H^{2}\left(p_{k+1}^{f}\right)^{2}\right) &= 2H\left(p_{k+1}^{f}\right)^{2} \\
K_{k+1} &= \frac{H\left(p_{k+1}^{f}\right)^{2}}{H^{2}\left(p_{k+1}^{f}\right)^{2} + r_{k+1}^{2}} \qquad (4.12)
\end{aligned}
$$

Because of the second derivative

$$\frac{\partial^2 \left(p_{k+1}\right)^2}{\partial K_{k+1}^2} = 2H^2 \left(p_{k+1}^f\right)^2 + 2r_{k+1}^2 > 0 \tag{4.13}$$

this extreme corresponds with a minimum.

If this minimal variance gain is used in the expression for the variance after the analysis step, this expression can be simplified to:

$$\left(p_{k+1}^a\right)^2 = \left(1 - K_{k+1}H\right)\left(p_{k+1}^f\right)^2 \tag{4.14}$$

4.2.5 *Simple example of Kalman filtering*

All steps of the Kalman filter are applied on location Schiedam for the first week of 2006. Figure 4.2 shows for the first week of 2006 all the model simulations and observations. At every hour the logarithm of the model result is shown together with the logarithm of the observation.

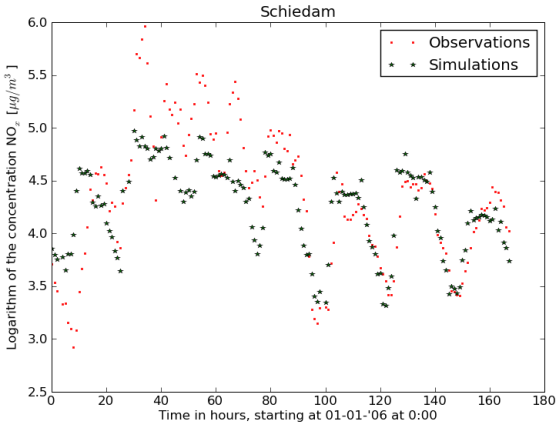


Figure 4.2: *Logarithms of the model simulations and the observations for the first week of 2006 on the monitoring station in Schiedam*

In Figure 4.3 all steps of the Kalman filter are applied on the model results and the observations for the first week of 2006. The logarithm of the real concentration can be found in a Gaussian distribution. The 1σ interval is given by the blue lines, this interval corresponds with:

$$[\ln(\text{model result}) + \hat{\gamma} - p, \ln(\text{model result}) + \hat{\gamma} + p] \tag{4.15}$$

where $\hat{\gamma}$ is the mean after the Kalman filter and p corresponds with the square root of the variance after the Kalman filter. It is clear that in this case the uncertainty interval is mostly between the model result and the observation. In Section 4.3, it will be shown how this interval depends on the several input parameters r^2, α and σ^2.

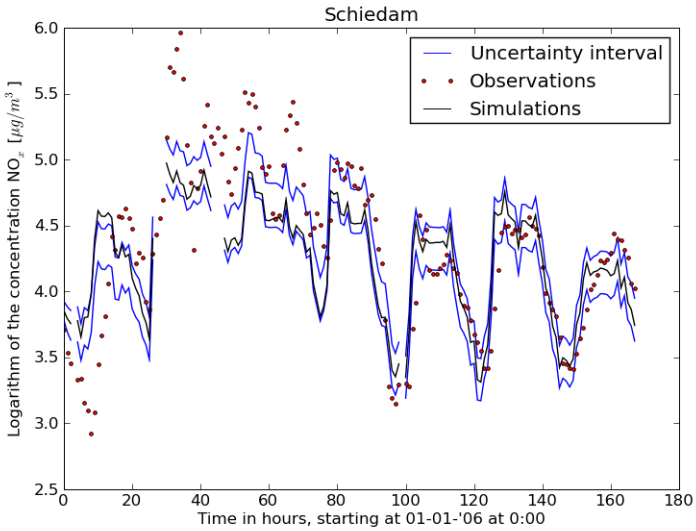


Figure 4.3: *Kalman filter applied on the first week of 2006 on location Schiedam*

In Figure 4.4, it is shown what happens when there is a certain period without observations. When there is no observation, only the forecast step of the Kalman filter is used. The mean value $\hat{\gamma}$ tends to the model results and the standard deviation p tends to the standard deviation of the model.

In the left panel of Figure 4.4 there are no observations analyzed, thus the uncertainty interval tends to be around the model and the width of the interval corresponds with the standard deviation of the model. In the right panel there are no observations analyzed between time step 15 and time step 90. Between those time steps the uncertainty interval after the Kalman filter tends to the model. After time step 90, the Kalman filter analyzes the observations again and the intervals are again between the model results and the observations.

In Figure 4.5 this phenomenon is better visible. In this figure, the differences between the model result and the measurement outcomes are shown, with black dots. The intervals in this figure are just the intervals $[\hat{\gamma} - p, \hat{\gamma} + p]$. When there are no measurements in the analysis step, the mean of the interval tends to zero and the width of the interval corresponds with the variance of the model. If the observations are analyzed again the mean of the interval lies between zero and the black dots

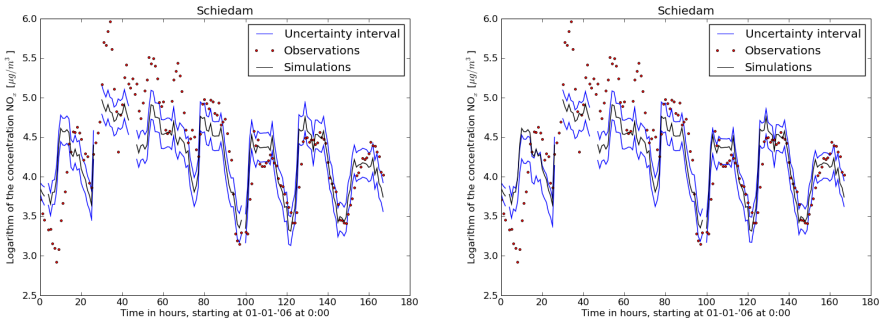


Figure 4.4: *Kalman filter applied on the first week of 2006 on location Schiedam, In the left panel there are no measurements in the analysis step. In the right panel there are no measurements in the analysis step between time steps 15 and 90*

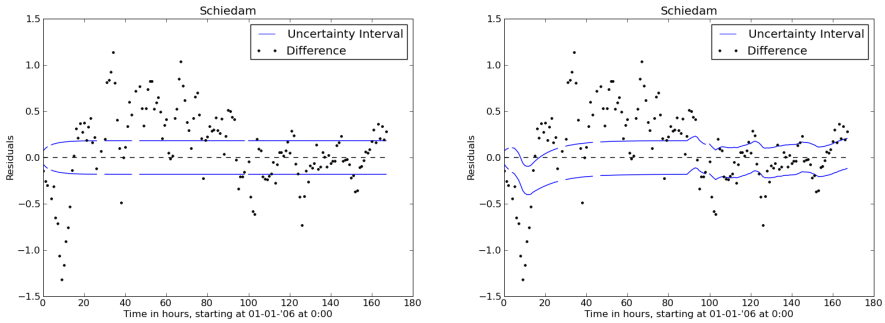


Figure 4.5: *Uncertainty interval of the perturbations, the black dots are the differences between the outcomes of the measurements and the model. In the left panel there are no measurements in the analysis step. In the right panel there are no measurements in the analysis step between time steps 15 and 90*

4.3 Sensitivity tests

In this example for Schiedam, some parameters can be changed to get a better view on their influence. When some parameters are changed the behavior of the Kalman filter is different.

4.3.1 Uncertainty of measurements (r^2)

The first parameter to change is the uncertainty of the measurements (r^2). In Figure 4.6, it is shown what happens when there is respectively a small and a large uncertainty. The left panel of Figure 4.6 shows that if the uncertainty is small, the interval after the Kalman filter is close around the measurements. The width of this interval is also small, due to the small uncertainty of the measurements. The right panel in Figure 4.6 shows that, when the measurements have large uncertainty, the interval after the Kalman filter is around the model results. The width of the interval is approximately as large as the uncertainty of the model.

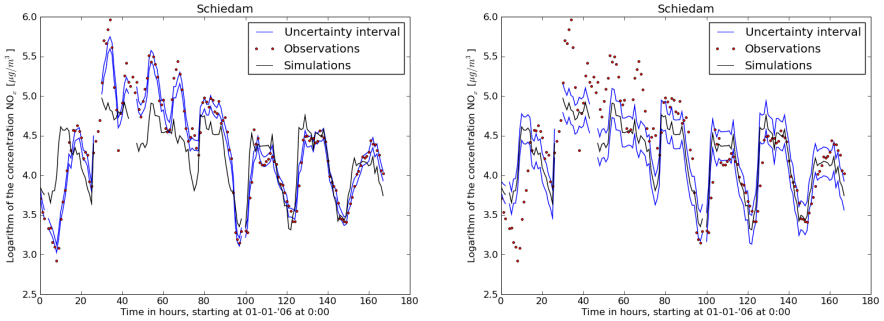


Figure 4.6: *Kalman filter applied on first week for location Schiedam with uncertainty of the measurements assumed small $r = 2\%$ (left panel) and large $r = 200\%$ (right panel)*

4.3.2 *Temporal correlation parameter (α)*

The second parameter to change is the temporal correlation parameter α. In Figure 4.7, the intervals of the perturbations are shown. In the left figure $\tau = 1$, thus $\alpha = e^{-1/\tau} \approx 0.37$, in the right figure $\tau = 250$, thus $\alpha = e^{-1/\tau} \approx 1.00$. The left panel of Figure 4.7 shows that, when α is small, there is hardly any temporal correlation. It is possible to get high fluctuations of the interval. If there are no observations analyzed from time step 15 till time step 90, the mean of the perturbation will tend rapidly to zero, the width of the interval will rapidly tend to the uncertainty of the model. The right panel of Figure 4.7 shows that if α is large, the temporal correlation is large and the interval does not make large fluctuations. For that reason the mean of the interval will tend slowly to 0 when there are no measurements analyzed from time step 15 till time step 90. Also the width of the interval will tend slowly to the width corresponding with the uncertainty of the model.

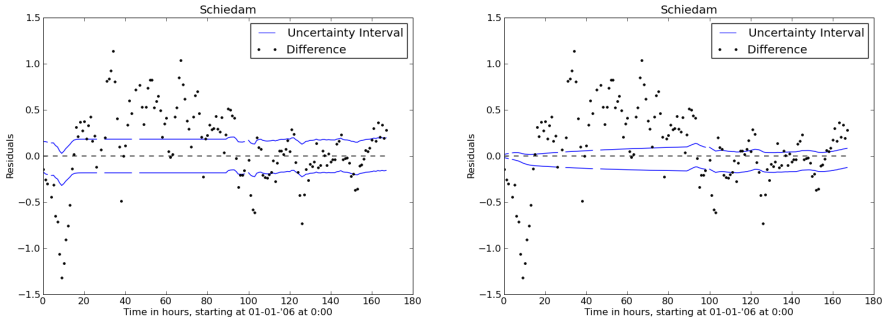


Figure 4.7: *Uncertainty interval for the perturbation for location Schiedam for the first week of 2006, with time correlation assumes small $\alpha \approx 0.37$ in the left panel and large $\alpha \approx 1.00$ in the right panel.*

4.3.3 *Uncertainty of the model (σ)*

The last parameter to change is the uncertainty of the model σ_k. In Figure 4.8, it is shown what happens when there is respectively a small and a large model uncertainty. In the left panel of Figure 4.8, the Kalman filter is applied with relatively small model uncertainty. The interval after the Kalman filter mostly follows the model and the width of the interval is also small because of the small uncertainty of the model. The right panel of Figure 4.8 shows the uncertainty interval when the model uncertainty is relatively high. The interval is close around the observations, while the width of the interval corresponds with the uncertainty of the measurements.

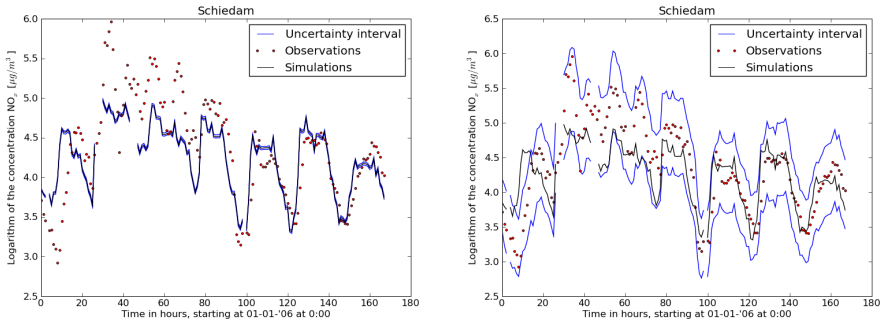


Figure 4.8: *Kalman filter applied on the first week of 2006 for location Schiedam, with model uncertainty assumed small $\sigma_k = 2\%$ in the left panel and large $\sigma_k = 200\%$ in the right panel*

4.4 Higher dimensional Kalman filtering

The algorithm described in Section 4.2 is an algorithm for a one dimensional problem. This algorithm can easily be extended to a higher dimensional problem. This is also explained with an example of the Real Time URBIS model. In the Rijnmond area there are 9 locations where the concentration NO_x is measured. On each of these 9 locations there is a real concentration NO_x called $\underline{c}_k$ (vector of length 9). Also the Real Time URBIS model gives for every hour a concentration NO_x on each location, called $\underline{c}_k^m$. Again it is necessary to work in the log-domain, thus $\underline{\gamma}_k$ is the perturbation on the model to estimate the real concentrations.

4.4.1 Dynamical system

The dynamical system of this problem:

$$\ln(\underline{c}_k) = \ln(\underline{c}_k^m) + \underline{\gamma}_k \tag{4.16}$$

$$\underline{\gamma}_{k+1} = A\underline{\gamma}_k + \underline{\omega}_k \qquad \underline{\omega}_k \sim N(0, Q_k) \tag{4.17}$$

where $\ln(\underline{c}_k)$ stands for a vector with logarithms of the concentrations.

The dynamical system has become a matrix-vector equation, where matrix A replaces α as time correlation parameter. In the example over all locations A is a diagonal matrix with time correlations α_i on the diagonal representing the temporal correlations for each entry of $\underline{\gamma}_k$. Q_k is a covariance matrix, built from the temporal correlation and the uncertainty of the model. The matrix Q is diagonal with elements $q_i^2 = \left(1 - \alpha_i^2\right)\sigma_i^2$.

4.4.2 Kalman filter form

The dynamical system has to be written in Kalman filter form:

$$\underline{\gamma}_{k+1} = A\underline{\gamma}_k + \underline{\omega}_k \qquad \underline{\omega}_k \sim N(0, Q_k) \tag{4.18}$$

$$\ln\left(\underline{y}_k\right) = H\left(\ln(\underline{c}_k) + \underline{\gamma}_k\right) + \underline{\nu}_k \qquad \underline{\nu}_k \sim N(0, R_k) \tag{4.19}$$

where R_k is a covariance matrix with uncertainty of the measurements. The matrix R_k is also a diagonal matrix with elements r_i^2, the uncertainty of each entry of the vector with observations $\ln(\underline{y}_k)$, H is now a higher dimensional system operator, which projects the model state onto the measurement outcomes.

The result after the Kalman filter is again that the vector $\underline{\gamma}_k$ can be found in a Gaussian distribution with mean $\tilde{\underline{\gamma}}_k$ and covariance matrix P_k. With this Gaussian distribution, the logarithm of the concentration NO_x can be found in a Gaussian distribution with mean $(\ln(\underline{c}_k^m) + \tilde{\underline{\gamma}}_k)$ and covariance matrix P_k. The value for the mean of $\underline{\gamma}_k$, called $\hat{\underline{\gamma}}_k$ is simply $\mathrm{E}[\underline{\gamma}_k]$. P_k is a covariance matrix with covariances between the entries of state vector $\underline{\gamma}_k$. On the main diagonal of a covariance matrix are variances. From this diagonal, the uncertainty interval for each entry of $\underline{\gamma}_k$ can be computed by taking the square root of these variance.

4.4.3 Forecast step

In the forecast step the mean $\hat{\underline{\gamma}}_{k+1}^f$ is computed with the mean $\hat{\underline{\gamma}}_k$ from the time step before:

$$\begin{aligned}
\hat{\underline{\gamma}}_{k+1}^f &= \mathrm{E}\left[\underline{\gamma}_{k+1}\right] \\
&= \mathrm{E}\left[A\underline{\gamma}_k + \underline{\omega}_k\right] \\
&= A\mathrm{E}\left[\underline{\gamma}_k\right] + \mathrm{E}\left[\underline{\omega}_k\right] \\
&= A\mathrm{E}\left[\underline{\gamma}_k\right] \\
&= A\hat{\underline{\gamma}}_k
\end{aligned} \tag{4.20}$$

where is used that $\mathrm{E}\left[\underline{\omega}_k\right] = 0$.

The covariance matrix P_{k+1}^f of $\underline{\gamma}_k$ is as in one dimension a function of the covariance matrix from the time step before:

$$\begin{aligned}
P_{k+1}^f &= \mathrm{COV}\left(\underline{\gamma}_{k+1}\right) \\
&= \mathrm{E}\left[\left(\underline{\gamma}_{k+1} - \mathrm{E}\left[\underline{\gamma}_{k+1}\right]\right)\left(\underline{\gamma}_{k+1} - \mathrm{E}\left[\underline{\gamma}_{k+1}\right]\right)^{\mathrm{T}}\right] \\
&= \mathrm{E}\left[\left(\left(A\underline{\gamma}_k + \underline{\omega}_k\right) - A\hat{\underline{\gamma}}_k\right)\left(\left(A\underline{\gamma}_k + \underline{\omega}_k\right) - A\hat{\underline{\gamma}}_k\right)^{\mathrm{T}}\right] \\
&= \mathrm{E}\left[A\left(\underline{\gamma}_k - \hat{\underline{\gamma}}_k\right)\left(\underline{\gamma}_k - \hat{\underline{\gamma}}_k\right)^{\mathrm{T}} A^{\mathrm{T}} + \underline{\omega}_k\left(\underline{\gamma}_k - \hat{\underline{\gamma}}_k\right) A^{\mathrm{T}}\right. \\
&\quad \left. + A\left(\underline{\gamma}_k - \hat{\underline{\gamma}}_k\right)\underline{\omega}_k^{\mathrm{T}} + \underline{\omega}\underline{\omega}_k^{\mathrm{T}}\right] \\
&= A\mathrm{COV}\left(\underline{\gamma}_k\right) A^{\mathrm{T}} + \mathrm{E}\left[\underline{\omega}_k\right]\mathrm{E}\left[\left(\underline{\gamma}_k - \hat{\underline{\gamma}}_k\right)\right] A^{\mathrm{T}} \\
&\quad + A\mathrm{E}\left[\left(\underline{\gamma}_k - \hat{\underline{\gamma}}_k\right)\right]\mathrm{E}\left[\underline{\omega}_k^{\mathrm{T}}\right] + \mathrm{COV}\left(\underline{\omega}_k\right) \\
&= AP_kA^{\mathrm{T}} + Q_k
\end{aligned} \tag{4.21}$$

where the independency of $\underline{\omega}_k$ and $\underline{\gamma}_k$ is used, as well as $\mathrm{E}\left[\underline{\omega}_k\right] = 0$ and $\mathrm{COV}\left(\underline{\omega}_k\right) = Q_k$.

4.4.4 Analysis step

In the analysis step the results of the forecast step are analyzed with a series of observations. The mean from the forecast step is analyzed with a linear Kalman gain K, such that the mean after the analysis step is similar with the one dimensional case:

$$\underline{\hat{\gamma}}^a_{k+1} = \underline{\hat{\gamma}}^f_{k+1} + K_{k+1}\left(\ln\left(\underline{y}_{k+1}\right) - H\left(\ln\left(\underline{c}^m_{k+1}\right) + \underline{\hat{\gamma}}^f_{k+1}\right)\right) \quad (4.22)$$

The covariance matrix after the analyzing step is as in the one dimensional case, a function of the covariance matrix from the forecast step:

$$\begin{aligned}
P^a_{k+1} &= \mathrm{COV}\left(\underline{\gamma}_{k+1}\right) \\
&= \mathrm{E}\left[\left(\underline{\gamma}_{k+1} - \mathrm{E}\left[\underline{\gamma}_{k+1}\right]\right)\left(\underline{\gamma}_{k+1} - \mathrm{E}\left[\underline{\gamma}_{k+1}\right]\right)^{\mathrm{T}}\right] \\
&= \mathrm{E}[\left(\underline{\gamma}_{k+1} - \underline{\hat{\gamma}}^a_{k+1}\right)\left(\underline{\gamma}_{k+1} - \underline{\hat{\gamma}}^a_{k+1}\right)^{\mathrm{T}}] \\
&= \mathrm{E}\left[\left(\underline{\gamma}_{k+1} - \left(\underline{\hat{\gamma}}^f_{k+1} + K_{k+1}\left(\ln\left(\underline{y}_{k+1}\right) - H\left(\ln\left(\underline{c}^m_{k+1}\right) + \underline{\hat{\gamma}}^f_{k+1}\right)\right)\right)\right)\right. \\
&\quad \left.\left(\underline{\gamma}_{k+1} - \left(\underline{\hat{\gamma}}^f_{k+1} + K_{k+1}\left(\ln\left(\underline{y}_{k+1}\right) - H\left(\ln\left(\underline{c}^m_{k+1}\right) + \underline{\hat{\gamma}}^f_{k+1}\right)\right)\right)\right)^{\mathrm{T}}\right] \\
&= \mathrm{E}\left[\left(\underline{\gamma}_{k+1} - \underline{\hat{\gamma}}^f_{k+1} - K_{k+1}\left(H\left(\ln\left(\underline{c}^m_{k+1}\right) + \underline{\gamma}_{k+1}\right) + \underline{\nu}_{k+1}\right.\right.\right. \\
&\quad \left.\left.\left. -H\left(\ln\left(\underline{c}^m_{k+1}\right) + \underline{\hat{\gamma}}^f_{k+1}\right)\right)\right)\right. \\
&\quad \left.\left(\underline{\gamma}_{k+1} - \underline{\hat{\gamma}}^f_{k+1} - K_{k+1}\left(H\left(\ln\left(\underline{c}^m_{k+1}\right) + \underline{\gamma}_{k+1}\right) + \underline{\nu}_{k+1}\right.\right.\right. \\
&\quad \left.\left.\left. -H\left(\ln\left(\underline{c}^m_{k+1}\right) + \underline{\hat{\gamma}}^f_{k+1}\right)\right)\right)^{\mathrm{T}}\right] \\
&= \mathrm{E}\left[\left(\left(I - K_{k+1}H\right)\left(\underline{\gamma}_{k+1} - \underline{\hat{\gamma}}^f_{k+1}\right) - K_{k+1}\underline{\nu}_{k+1}\right)\right. \\
&\quad \left.\left(\left(I - K_{k+1}H\right)\left(\underline{\gamma}_{k+1} - \underline{\hat{\gamma}}^f_{k+1}\right) - K_{k+1}\underline{\nu}_{k+1}\right)^{\mathrm{T}}\right] \\
&= \mathrm{E}\left[\left(I - K_{k+1}H\right)\left(\underline{\gamma}_{k+1} - \underline{\hat{\gamma}}^f_{k+1}\right)\left(\underline{\gamma}_{k+1} - \underline{\hat{\gamma}}^f_{k+1}\right)^{\mathrm{T}}\left(I - K_{k+1}H\right)^{\mathrm{T}}\right] \\
&\quad -\mathrm{E}\left[\left(I - K_{k+1}H\right)\left(\underline{\gamma}_{k+1} - \underline{\hat{\gamma}}^f_{k+1}\right)\underline{\nu}^{\mathrm{T}}_{k+1}K^{\mathrm{T}}_{k+1}\right] \\
&\quad -\mathrm{E}\left[K_{k+1}\underline{\nu}_{k+1}\left(\underline{\gamma}_{k+1} - \underline{\hat{\gamma}}^f_{k+1}\right)^{\mathrm{T}}\left(I - K_{k+1}H\right)^{\mathrm{T}}\right] \\
&\quad +\mathrm{E}\left[K_{k+1}\underline{\nu}_{k+1}\underline{\nu}^{\mathrm{T}}_{k+1}K^{\mathrm{T}}_{k+1}\right] \\
&= \left(I - K_{k+1}H\right)\mathrm{E}\left[\left(\underline{\gamma}_{k+1} - \mathrm{E}\left[\underline{\gamma}_{k+1}\right]\right)\left(\underline{\gamma}_{k+1} - \mathrm{E}\left[\underline{\gamma}_{k+1}\right]\right)^{\mathrm{T}}\right]\left(I - K_{k+1}H\right)^{\mathrm{T}} \\
&\quad +K_{k+1}\mathrm{E}\left[\underline{\nu}_{k+1}\underline{\nu}^{\mathrm{T}}_{k+1}\right]K^{\mathrm{T}}_{k+1} \\
&= \left(I - K_{k+1}H\right)\mathrm{COV}\left(\underline{\gamma}_{k+1}\right)\left(I - K_{k+1}H\right)^{\mathrm{T}} + K_{k+1}\mathrm{COV}\left(\underline{\nu}_{k+1}\right)K^{\mathrm{T}}_{k+1} \\
&= \left(I - K_{k+1}H\right)P^f_{k+1}\left(I - K_{k+1}H\right)^{\mathrm{T}} + K_{k+1}R_{k+1}K^{\mathrm{T}}_{k+1} \qquad (4.23)
\end{aligned}$$

where the independency of $\underline{\gamma}_k$ and $\underline{\nu}_k$, as well as $\mathrm{E}\left[\underline{\nu}_k\right] = 0$ and $\mathrm{COV}\left(\underline{\nu}_k\right) = R_k$.

Also in this higher dimensional problem it is common use to take for K_k the gain that minimizes the variance P_k^a in l_2 norm. This gain is expressed similar to the one dimensional case:

$$K_k = P_k^f H^{\mathrm{T}} \left(H P_k^f H^{\mathrm{T}} + R_k \right)^{-1} \tag{4.24}$$

As in the one-dimensional case, the expression for the covariance matrix can be simplified to:

$$P_{k+1}^a = \left(I - K_{k+1} H \right) P_{k+1}^f \tag{4.25}$$

More information about higher dimensional Kalman filtering can be found in [Segers, 2002]

5 Kalman filter on background concentrations

5.1 Introduction

As described in Section 2.1, the mathematical description of the Real time URBIS model is:

$$\underline{c}_k^m = M\underline{\mu}_k^{\mathrm{T}} \tag{5.1}$$

The value $\underline{c}_k^m$ is the concentration NO_x calculated by the model, while each column of M corresponds with a standard concentration field, computed by the URBIS model, shown in Appendix B. The vector $\underline{\mu}_k$ represents the weight for every standard concentration field on time k.

In Chapter 3, the comparison between the model simulations and the observations shows that the model simulations have some inaccuracies. In that chapter, it was shown that the differences between the model simulations and the observations depends on the wind direction and the wind speed. This analysis is done for all stations, thus it is assumed that the dependency on the wind direction is the same for the whole area. It is possible that the dependency on the wind direction is caused by an inaccurate local emission source, but it is not likely that an inaccurate local emission source influences all stations.

The figures in Appendix B shows that the standard concentration fields for the emission source 'Background' are the same for every wind direction and wind speed. This is contradicting with the ideas of Chapter 3. Therefore the emission from source 'Background' is marked as the inaccurate emission source. This source is typically a source that has to be dependent of the wind direction and the wind speed.

It is likely that the wind dependency found in Chapter 3, is caused by the lack of wind dependency in the source 'Background' in the URBIS model. In this chapter the standard concentration fields for this emission source will be corrected with a Kalman filter. The idea is that the correction is dependent on the wind direction and the wind speed.

Figure 3.2 gives an idea of how the standard concentration fields have to be corrected. When the wind is from direction north-west the model simulation is too high, thus the concentration fields from directions west and north have to be lower. When the wind is from direction south-east, the model simulation is too low, thus the standard concentration field from directions east and south has to be higher. Figure 3.5 gives the idea is that when the wind speed is high, the concentration is lower because of a larger dilution of the emission.

This application of the Kalman filter will also lead to an uncertainty interval of the total concentration NO_x for every time and location in the Rijnmond area.

5.2 Kalman filter

To make a correction on the standard concentration fields, every field gets a correction factor $e^{\gamma_{i,k}}$ for each hour k. These factors are larger then zero, thus there is no problem with negative concentrations. Adding these corrections to the model from Equation 5.1 leads to the following equation for the corrected model:

$$\underline{c}_k = \sum_{i=1}^{88} \mu_{i,k} \underline{m}_i e^{\gamma_{i,k}} \tag{5.2}$$

In this equation vectors $\underline{m}_i$ are the columns of M, representing the standard concentration fields. In the log-normal distribution the expected concentration $\mathrm{E}[\underline{c}_k]$ is given by:

$$\mathrm{E}[\underline{c}_k] = \sum_{i=1}^{88} \mu_{i,k} \underline{m}_i e^{\hat{\gamma}_{i,k} + 1/2 \hat{p}_{i,k}^2} \tag{5.3}$$

where $\hat{\gamma}_{i,k}$ is the median of $\gamma_{i,k}$ and $\hat{p}_{i,k}$ is the standard deviation of each entry of $\gamma_{i,k}$.

5.2.1 Dynamical system

In Chapter 3, the idea is that the background concentrations are not accurate in the model. The other fields were supposed to be good enough, thus the correction factor on those fields are stated equal to one ($\gamma_i = 0$) for $i = 9..88$. This leads to the following expression:

$$\sum_{i=1}^{8} \mu_{i,k} \underline{m}_i e^{\gamma_{i,k}} + \sum_{i=9}^{88} \mu_{i,k} \underline{m}_i e^{0} \tag{5.4}$$

The vectors $\underline{m}_i$ for $i = 1..8$ corresponds with the standard concentration fields for the source: 'Background', these fields have a correction $e^{\gamma_{i,k}}$. The second term of Equation 5.4 is not dependent of any γ_i, thus a constant called $\underline{c}_k^{m,d}$. This constant describes the model simulation for all sources, different from the source 'Background':

$$\underline{c}_k^{m,d} = \sum_{i=9}^{88} \mu_{i,k} \underline{m}_i \tag{5.5}$$

Because of the log-normal distribution of the model simulations, a transformation to the logarithms of the simulations is required:

$$\ln(\underline{c}_k^m) = \ln\left(\sum_{i=1}^{8} (\mu_{i,k} \underline{m}_i e^{\gamma_{i,k}}) + \underline{c}_k^{m,d} \right) \tag{5.6}$$

This is a non-linear equation for $\underline{\gamma}_k$. The Kalman filter requires a linear model, therefore a linearization of this equation is made around $\underline{\gamma}_k = 0$:

$$\ln\left(\sum_{i=1}^{8}\left(\mu_{i,k}\underline{m}_i e^{\gamma_{i,k}}\right)+\underline{c}_k^{m,d}\right) = \ln\left(\underline{c}_k^{m,b}+\underline{c}_k^{m,d}\right) + \left[\frac{\mu_{j,k}\underline{m}_j}{\underline{c}_k^{m,b}+\underline{c}_k^{m,d}}\right]_{j=1}^{j=8}\underline{\gamma}_k+\mathcal{O}\left(\underline{\gamma}_k\cdot\underline{\gamma}_k\right) \quad (5.7)$$

where $\underline{c}_k^{m,b}$ is the concentration calculated by the model for the source: 'Background':

$$\underline{c}_k^{m,b} = \sum_{i=1}^{8}\mu_{i,k}\underline{m}_i \quad (5.8)$$

In Equation 5.7, the quotient of two vectors is defined as the element wise quotient.

The dynamical system for the background concentration will then become the following:

$$\ln\left(\underline{c}_k\right) = \ln\left(\underline{c}_k^{m,b}+\underline{c}_k^{m,d}\right)+\left[\frac{\mu_{j,k}\underline{m}_j}{\underline{c}_k^{m,b}+\underline{c}_k^{m,d}}\right]_{j=1}^{j=8}\underline{\gamma}_k \quad (5.9)$$

$$\underline{\gamma}_{k+1} = A\underline{\gamma}_k+\underline{\omega}_k \qquad \underline{\omega}_k\sim N\left(0,Q\right) \quad (5.10)$$

The first equation is the linearization of the equation for the logarithm of the concentration, the second equation is the auto-correlation process for the series of perturbations $\left\{\underline{\gamma}_k\right\}_{k=1}^{k=n}$, with $n=8760$, the number of hours in a year. In the Kalman filter, an estimate of the uncertainty interval of the vector $\underline{\gamma}_k$ will be found. This uncertainty interval of $\underline{\gamma}_k$ will then be used to get a better uncertainty interval for the total concentration at time k.

The interpretation of the dynamical system is now that the logarithm of the real concentration is the logarithm of the model simulation plus a correction on the background. The correction on the background is a temporal correlated process, the temporal correlation is calculated in Section 5.4. The covariance matrix Q is assumed to be independent from time and this matrix is built from the temporal correlation and the model uncertainty. Matrix Q is a diagonal matrix with on the main diagonal elements q_i^2. This is a colored noised process driven by a white noise process, assuming that both the temporal correlation and the uncertainty of the model are independent of time:

$$q_i = \sqrt{1-\alpha_i^2}\sigma_i \quad (5.11)$$

where σ_i corresponds with the overall uncertainty of the perturbations.

5.2.2 *Kalman filter form*

The dynamical system in Equation 5.9 and 5.10 has to be written in a Kalman filter form. There are 9 series of observations $\underline{y}$, which are made on the 9 monitoring

stations in the domain. This series of observations have to be compared with the model results.

This leads to the following system of equations in Kalman filter form:

$$\begin{aligned} \underline{\gamma}_{k+1} &= A\underline{\gamma}_k + \underline{\omega}_k & (5.12)\\ \ln\left(\underline{y}_k\right) &= H\left(\ln\left(\underline{c}_k^{m,b} + \underline{c}_k^{m,d}\right)\right. & \\ &+ \left.\left[\frac{\mu_{j,k}\underline{m}_j}{\underline{c}_k^{m,b} + \underline{c}_k^{m,d}}\right]_{j=1}^{j=8} \underline{\gamma}_k\right) + \underline{\nu}_k \qquad \underline{\nu}_k \sim N\left(0, R_k\right) & (5.13) \end{aligned}$$

Matrix H is the system operator which projects the model state onto the observations. The covariance matrix R represents the uncertainty of the logarithms of the observations, combined the instrumental error and the representation error. This matrix R is a diagonal matrix with diagonal elements r_i^2, the values for r_i will be estimated in Section 5.3. To simplify notations, the system is rewritten to:

$$\begin{aligned} \underline{\gamma}_{k+1} &= A\underline{\gamma}_k + \underline{\omega}_k & (5.14)\\ \underline{\tilde{y}}_k &= \tilde{H}_k\underline{\gamma}_k + \underline{\nu}_k \qquad \underline{\nu}_k \sim N\left(0, R_k\right) & (5.15) \end{aligned}$$

where vector $\underline{\tilde{y}}_k$ and matrix $\tilde{H}_k$ are defined by:

$$\begin{aligned} \underline{\tilde{y}}_k &= \ln\left(\underline{y}_k\right) - H\ln\left(\underline{c}_k^{m,b} + \underline{c}_k^{m,d}\right) & (5.16)\\ \tilde{H}_k &= H\left[\frac{\mu_{j,k}\underline{m}_j}{\underline{c}_k^{m,b} + \underline{c}_k^{m,d}}\right]_{j=1}^{j=8} & (5.17) \end{aligned}$$

5.2.3 Forecast of background correction

On this Kalman filter form the algorithm for the Kalman filter can be applied. The forecast step gives then the following formulas for the expected median $\underline{\hat{\gamma}}_k^f$ and the variance P_k^f of $\underline{\gamma}_k$:

$$\begin{aligned} \underline{\hat{\gamma}}_{k+1}^f &= A\underline{\hat{\gamma}}_k & (5.18)\\ P_{k+1}^f &= APA^{\mathrm{T}} + Q & (5.19) \end{aligned}$$

5.2.4 Analysis of background correction

In the analyzing step, the filter makes a comparison with a series of observations, in this case 9 observations per time step for the 9 monitoring stations in the domain. This leads to the following formulas for the expected median $\underline{\hat{\gamma}}_k^a$and variance P_k^a of

$\underline{\gamma}_k$:

$$\begin{aligned} \underline{\hat{\gamma}}^a_{k+1} &= \underline{\hat{\gamma}}^f_{k+1} + K_{k+1}\left(\underline{\tilde{y}}_{k+1} - \tilde{H}_{k+1}\underline{\hat{\gamma}}^f_{k+1}\right) && (5.20) \\ P^a_{k+1} &= \left(I - K_{k+1}\tilde{H}_{k+1}\right)\left(P^f_{k+1}\right)\left(I - K_{k+1}\tilde{H}_{k+1}\right)^{\mathrm{T}} \\ &+ K_{k+1}R_{k+1}K^{\mathrm{T}}_{k+1} && (5.21) \end{aligned}$$

where K_{k+1} is the Kalman gain that minimizes the variance P^a_{k+1}. This Kalman gain is given by:

$$K_k = P^f_k \tilde{H}^{\mathrm{T}}_k \left(\tilde{H}_k P^f_k \tilde{H}^{\mathrm{T}}_k + R_k\right)^{-1} \quad (5.22)$$

The values $\underline{\hat{\gamma}}^a$ and P^a_k, are stated as the mean and the covariance matrix for $\underline{\gamma}$ on time step k, and will be used as input for the next time step.

5.3 Uncertainty of the observations

The observation error (R in Equation 5.13) is an important parameter in the Kalman filter. Section 4.3 shows the influence on the solution when parameter r^2 is changed. Because of R is built from all r^2_i, the observation errors of each entry of the observation, the influence of covariance matrix R is also large.

The uncertainty of the measurements is assumed to be the square of a percentage (r_{frac}) of the outcome of the measurement:

$$R_{ii,k} = r^2_{\mathrm{frac}} y^2_{i,k}$$

where r_{frac} will contain both the instrumental error and the representation error.

At location Bentinckplein in Rotterdam, two monitoring stations are located directly next to each other, one station from DCMR and one from RIVM. With the two series of observations made on these two stations, an indication of the instrumental error can be found. In Figure 5.1, the logarithms of the observations made on these two stations are shown in a scatter plot. An assumption for the logarithm of the real concentration at this location is the mean of the logarithms of the two observations:

$$\ln\left(y^r_k\right) = \frac{\ln\left(y_k\right) + \ln\left(z_k\right)}{2} \quad (5.23)$$

where y^r_k is the real concentration at time k and y_k, z_k are respectively the observations on the DCMR and the RIVM station.

In Figure 5.2, a histogram with differences between the logarithms of the observations at the DCMR station and the assumed logarithms of the real concentrations is shown. The red line is the probability density function of the normal distribution with mean 0 and standard deviation 0.08, this standard deviation is the same as the standard deviation of the differences plotted in the histogram.

The peak of the histogram is not located on zero, which means that the annual mean concentration is not the same on both stations. The annual mean on the RIVM station

is larger than the annual mean on the DCMR station. Although the normal distribution did not fit very well with the histogram, the assumption that the differences are normal distributed with standard deviation 0.08 is at least a good approximation. This corresponds with an uncertainty of the logarithm of the measurements of 8 %. This will be used as an estimate for the instrumental error in r_{frac}, a random noise on the observation. The histogram for the RIVM station is the same as for the DCMR station, but then the negative version so that the histogram is mirrored in the y-axis.

The contribution of the representation error is not easy to calculate, this will be done by a method of trial and error. The Kalman filter will be applied with different values for $r_{\text{frac}} > 0.08$ to obtain the optimal value for r_{frac}.

The last assumption is that the observation error is the same for all stations, and not correlated between the stations. The matrix R is then a diagonal matrix with on the main diagonal elements $r_{\text{frac}} y_j^2$.

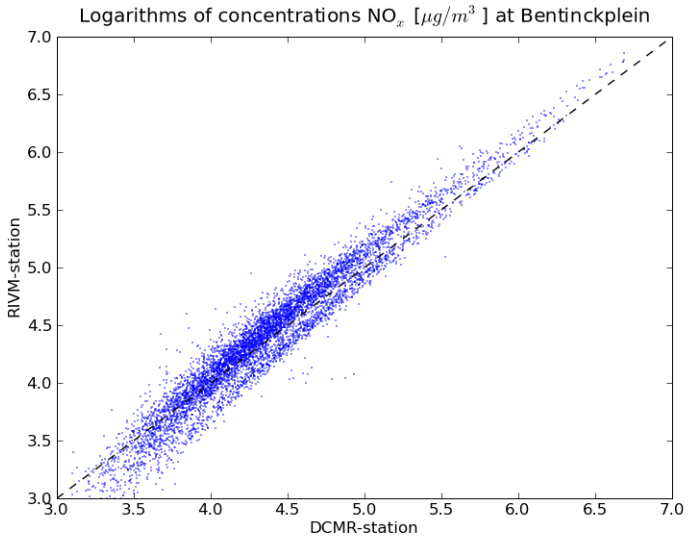


Figure 5.1: *The logarithms of the observations of the two monitoring stations at location Bentinckplein*

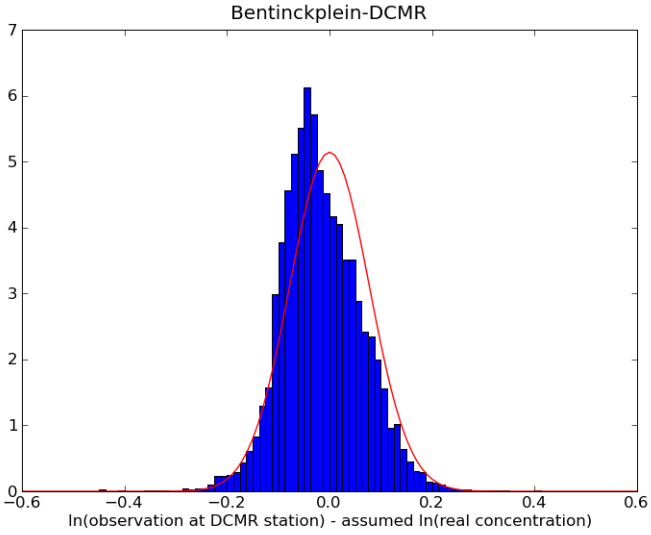


Figure 5.2: *Histogram of the differences between the logarithms of the observations and the assumed logarithm of the real concentration at location Bentinckplein. The red line is the probability density function of the normal distribution with mean zero and standard deviation 0.08.*

5.4 Temporal correlation parameter

Another important parameter is the temporal correlation. In the dynamical system given by Equation 5.9 and 5.10, the matrix A contains the temporal correlation parameters $\alpha_{i,j}$ for the perturbation on the logarithm of the several background concentrations. The monitoring stations in Schipluiden and Westmaas, numbers 7 and 10 in Figure 2.1, are two stations which are located far away from industry sources or main roads. These locations are chosen to obtain estimates of the background concentrations. With the observations made on this stations, it is possible to get an estimate for the temporal correlation parameters.

In general, the correlation (ρ) between two series of measurements $\{y_i\}_{i=1}^n$and $\{z_i\}_{i=1}^n$ could be computed with the following formula:

$$\rho \quad = \quad \frac{1}{n}\sum_{i=1}^{n}\frac{(y_i-\bar{y})}{\sigma_y}\frac{(z_i-\bar{z})}{\sigma_z} \tag{5.24}$$

where $\bar{y}$, $\bar{z}$ are the mean of the series $\{y_i\}_{i=1}^n$and $\{z_i\}_{i=1}^n$, and σ_y, σ_z are the standard deviations of the series $\{y_i\}_{i=1}^n$and $\{z_i\}_{i=1}^n$.

Assumed is that there is no correlation between the perturbations from different wind directions and wind speeds. The matrix A will then be a diagonal matrix with on the main diagonal elements α_i. An estimate for α_i is made with Equation 5.24 from two series of measurements $\{z_k\}_{k=1}^{n-m}$ and $\{z_k\}_{k=m}^n$, where z_k is the difference between the logarithm of the observation and the logarithm of the model simulation at time step k on location Schipluiden:

$$z_k \quad = \quad \ln(y_k) - \ln(c_k^m) \tag{5.25}$$

Another estimate for α_i is made with the differences on location Westmaas. For both locations this is done for $m \in [0, 60]$.

Both locations Westmaas and Schipluiden are outside the model domain, thus there is no model simulation. Because of both stations are assumed to be background stations, the concentration is caused mainly by the background. This background is assumed constant, therefore the calculation of the correlation could be done with $z_k = \ln(y_k)$.

For every period, the correlation is calculated for the perturbation on time k compared with the perturbation on time $k+m$. In Figure 5.3 these correlations are plotted with respect to the period m. This figure shows some peaks at period 24 hours, and period 48 hours. This means that the correlation has a daily pattern. This is a reasonable idea, because it is expected that the emission in Schipluiden and Westmaas is mostly produced by people living in Schipluiden and Westmaas.

A reasonable assumption is that the concentration on time step k does not depend on the concentration on time $k - m$ when m is large. Therefore the correlation between the perturbation on time k and the perturbation on time $k+m$ should go to zero when $m \to \infty$. Mathematically there is a correlation between the concentration on time step k and time step $k+m$, this can be understood by the time patterns in the emission and diurnal cycles in meteorological parameters. For example the concentration on Monday at 08:00 in the morning is roughly the same as the concentration at Tuesday 08:00 in the morning. Mathematically this gives a high temporal correlation for $\Delta t = 24$, but physically this concentrations are not related. For that reason it is only important to look at the temporal correlation for a few hours.

In Figure 5.3, a fitting exponential function is drawn for the first few periods. In this case the formula for this function is $\alpha(\Delta t) = e^{-\Delta t/12}$. The de-correlation parameter $\tau = 12$ gives the idea that the concentration on time $k + 12$ is not dependent on the concentration at time k.

Finally this de-correlation parameter $\tau = 12$ must be seen as an estimate. This estimate is obtained with varying wind speeds and wind directions. When the wind is with constant speed from the same direction, the correlation is perhaps different. In the application an optimal value for each α_i will be found by testing the Kalman filter with different values for each α_i.

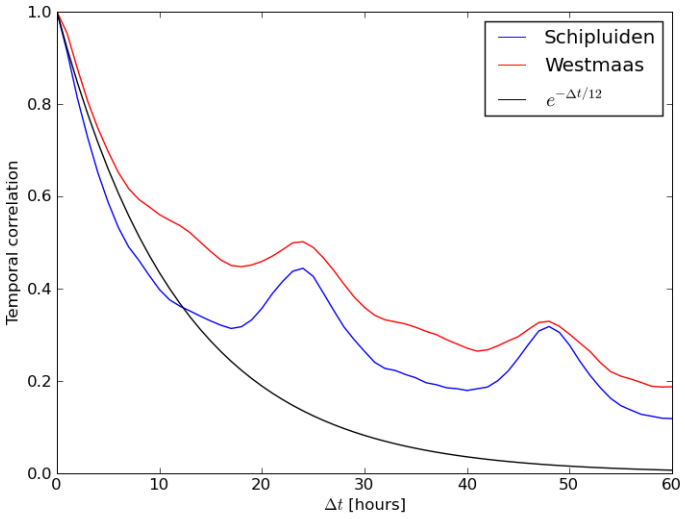


Figure 5.3: *The temporal correlation for background stations Schipluiden and Westmaas. The black line corresponds with correlation* $\alpha(\Delta t) = e^{-\Delta t/12}$

5.5 Kalman filter runs

The application of the Kalman filter, assuming the model uncertainty σ equal to 0.4, leads to a calculation of $\underline{\hat{\gamma}}_k$, the expected median of vector $\underline{\gamma}_k$ and P_k the covariance matrix of vector $\underline{\gamma}_k$ at every time step k. In this application the vector $\underline{\gamma}$ represents the perturbations on the logarithms of the background concentration for four different wind directions and two different wind speeds.

For the first week of 2006, the 1σ intervals of these eight perturbations are given in Figure 5.4. On every time step the weight $\mu_{i,k}$ for a standard concentration field is different, the values for $\mu_{i,k}$ are also given in Figure 5.4. When the contribution of a standard concentration field is high, the change in the correction factor γ_i is also high. If for a longer period a standard concentration field has no contribution, the mean of the correction factor γ_i tends to zero.

An interesting aspect of this result is that the values for γ_i are relatively high for some time steps, this means that the background concentration receives a relatively high correction factor for that time step. This is due to the fact that in this application it is assumed that the difference between the observation and the model simulation is completely depending on the background concentration. A better assumption is that when the difference between the observation and the model simulation is large, that there are some other errors in the model.

Another aspect is that the linearization of the dynamical system around $\underline{\gamma} = 0$ has accuracy $\mathcal{O}\left(\underline{\gamma} \cdot \underline{\gamma}\right)$, when $\underline{\gamma}$ become large the accuracy of the linearization decreases quadratically. For those reasons a screening process is implemented in the Kalman filter. When the difference between the observation and the model simulation is too large, the analysis step will not (or partly) be executed. The result is that the values for γ_i are limited. This screening process is explained in Section 5.6.

In Figure 5.5, the problems with large values for γ_i are shown. In this figure, at every time step the concentrations are calculated with the values for γ_i and with Equation

5.2 and 5.4. In each figure the yellow line represents the largest contribution on the correction of the background $\max\left(\mu_{i,k}e^{\gamma_{i,k}}\right)$ for every times step k. In these figures it is clear that the concentrations after applying the Kalman filter are not accurate in the regions where the values for γ_i become large.

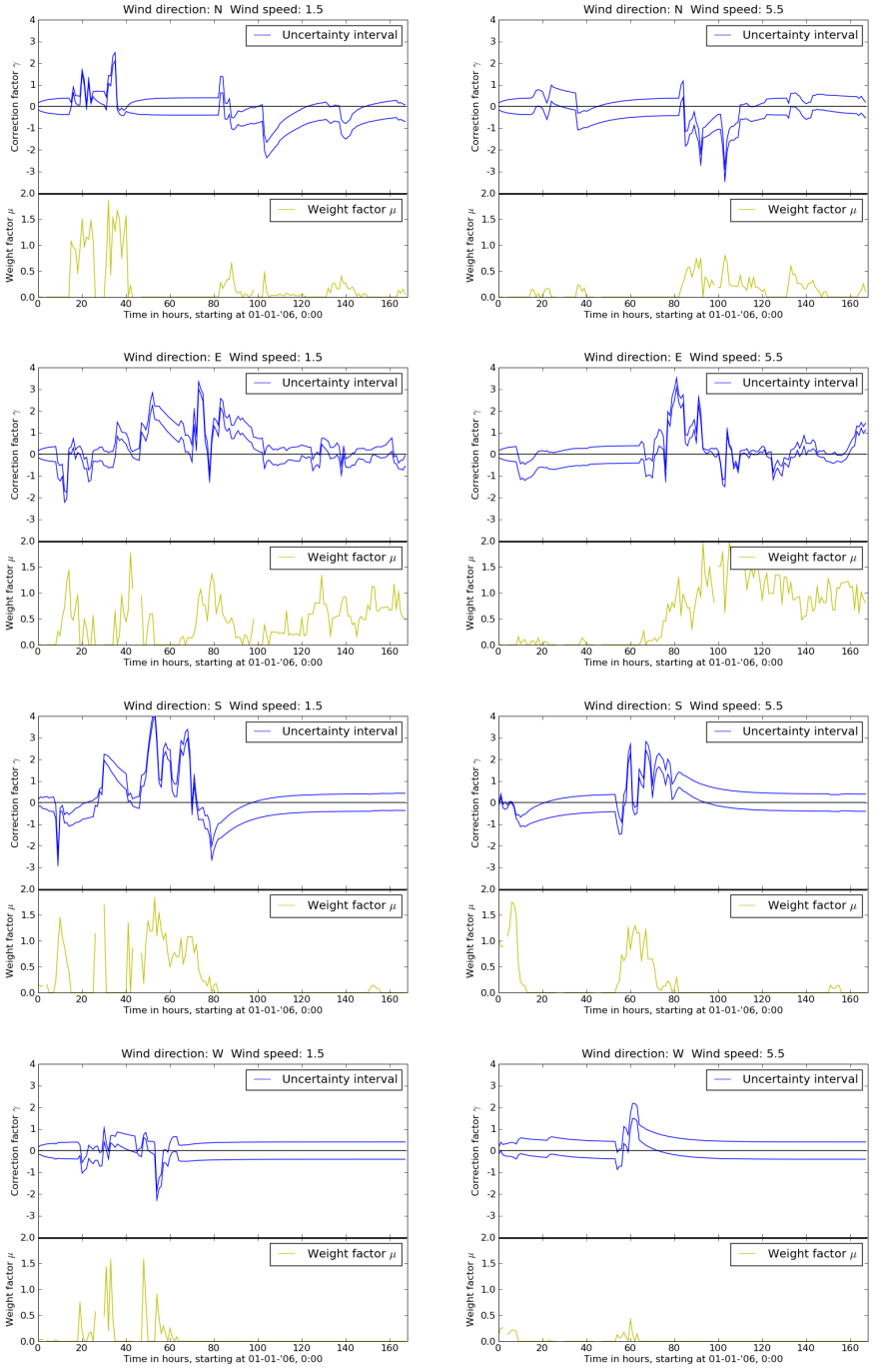


Figure 5.4: *Uncertainty intervals for the correction factors γ_i, together with the weights for each standard concentration field, for the first week of 2006*

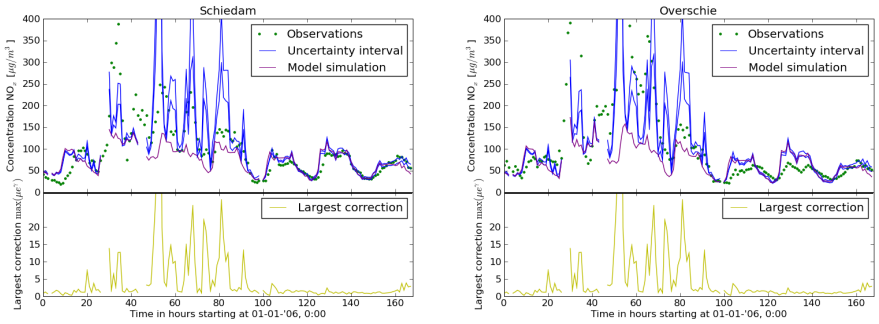


Figure 5.5: *Concentrations for the first week of 2006 after application of the Kalman filter on the background concentrations, for locations Schiedam and Overschie.*

5.6 Screening process

As mentioned in Section 5.5, for some time steps the difference between the observation and the model simulation can not only be explained by inaccuracies of the background concentrations. For that reason, a screening process is implemented in the Kalman filter. If the difference between the observation and the forecasted concentration is too high, the difference between the observation and the model simulation is not only caused by the inaccurate background, but also by some other sources or incidental occasions. If this situation occurs, the analysis step will only be executed on the observations which are close to the forecasted mean, such that the values for γ_i will stay small. The result is that the background concentrations will not get large correction factors and the linearization still have good accuracy. It is also important to have a view on which observations are screened, this could give an idea of other inaccuracies in the model. For example, if many observations are screened in the weekend, the model have large uncertainty in the weekend. More information about a screening process in a Kalman filter can be found in [de Haan et al., 1999].

To implement a screening process, a criterion has to be made, whether a difference between an observation and a model simulation is too high. For the Kalman filter on the background concentration the assumption is the following:

$$\underline{y}_k - \left(\sum_{i=1}^{8} \underline{m}_i \mu_{i,k} e^{\gamma^f_{i,k} + 1/2 (p^f_{i,k})^2} + c^{m,d}_k \right) \sim N\left(0, P_{abs,k} + R_{abs,k}\right) \tag{5.26}$$

In here, $P_{abs,k}$ and $R_{abs,k}$ represents the variance for respectively the model simulations and the observations. The variance for the model is not known explicitely, due to the log-normal distribution. Therefore this variance is assumed to be equal to the square of the difference between the upper band and the median of the 1σ interval of the concentration. The variance of the observations corresponds with the uncertainty of the measurements, $r_{frac} > 0.08$. Thus $P_{abs,k}$ and $R_{abs,k}$ are calculated as follows:

$$P_{abs,k} = \left(\sum_{i=1}^{8} \underline{m}_i \mu_i e^{\gamma^f_{i,k}} + \underline{c}^{m,d}_k - \left(\sum_{i=1}^{8} \underline{m}_i \mu_{i,k} e^{\gamma^f_{i,k} + p^f_{i.k}} + \underline{c}^{m,d}_k \right) \right)^2 \tag{5.27}$$

$$R_{abs,k} = \left(r_{\text{frac}} y_k\right)^2 \tag{5.28}$$

where $p^f_{i,k}$ is the standard deviation of $\gamma^f_{i,k}$. With the assumption from Equation 5.26 a criterion is chosen whether an observation is 'good' enough:

$$\left(\underline{y}_k - \left(\sum_{i=1}^{8} \underline{m}_i \mu_{i,k} e^{\gamma^f_{i,k} + 1/2 (p^f_{i,k})^2} + \underline{c}^{m,d}_k\right)\right)^2 \quad < \quad \beta^2 \left(P_{abs,k} + R_{abs,k}\right) \tag{5.29}$$

The parameter β defines the screening criterion. If the square of a difference is more than β^2 times the variance of model simulation plus the variance of the observations, the observation does not fit on the assumption that the difference is only caused by the inaccuracy of the background. In that case the observation is not involved in the analysis step. This is a vector inequality, which means that if for an entry of both vectors this inequality holds, the observation corresponding with that entry is not involved in the analysis step.

This screening process is implemented in the Kalman filter with parameter $\beta = 2$, this means that the square of a difference may not be larger than 4 times the sum of variations. The value $\beta = 2$ is chosen because in the normal distribution approximately 95% of the data lies in the 2σ interval.

The application of the Kalman filter with this screening process results in concentrations for Schiedam and Overschie for the first week of 2006, as shown in Figure 5.6. The largest correction $\max\left(\mu_{i,k} e^{\gamma_{i,k}}\right)$ become much smaller, and the concentrations have less extremes. In these figures, it is also shown that a lot of observations are not taken into account during the analysis step. About 68% of the observations are not taken into the analysis step.

A possibility is that the temporal correlation is too large, a large temporal correlation in the Kalman filter leads to a result without large fluctuations in the concentration. When the observations have large fluctuations, it is possible that many observations will be screened. If the temporal correlation is set with de-correlation parameter $\tau = 1$, there are still 66% of the observations, which are not taken into account during the analysis step. So it is not expected that the large number of screened observations is caused by a large temporal correlation.

Another idea is that the differences are not completely caused by the inaccuracy of the background concentrations. In Chapter 6, the Kalman filter is applied on all the different emission sources, to get a better estimate of all the different concentration fields of the URBIS model. The idea is that the large differences between the observations and the simulations are caused by inaccuracies of one of the standard concentration fields.

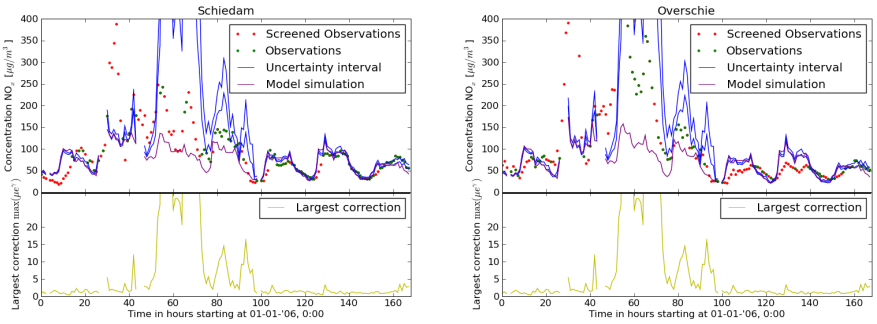


Figure 5.6: *Concentrations for the first week of 2006 after application of the Kalman filter on the background concentrations, with a screening process, at locations Schiedam and Overschie*

5.7 Discussion

The ideas from Chapter 3 were that the differences between the model simulations and the observations are caused by inaccuracies of the source 'Background'. In this chapter, it is shown that this assumption does not hold for most of the differences. A correction on this source is not sufficient to eliminate the most of the differences between the model simulations and the observations made on the 9 monitoring stations.

With the corrections made on the background, it is possible to create better standard concentration fields. Because of the large number of measurements which are not involved in the Kalman filter process, it is not expected that the new standard concentration fields for the background will be very accurate. For that reason, the Kalman filter will be applied to all different emission sources to get a 'better' standard concentration field for every source. This application will be explained in Chapter 6, together with the different runs to obtain the optimal values for each α_i, σ and R.

6 Kalman filter on all emission sources

6.1 Introduction

In Chapter 5, it is shown that a correction on the background concentrations leads to a better estimate of the real concentrations for only a small number of time steps. For the other time steps, the Kalman filter do not give a correction on the background because the difference between the observation and the model simulation is not only caused by the inaccurate background.

Therefore an additional analysis on the differences between the observations and the model simulations is made. In Figure 6.1, it is shown in which cases the differences between the observations and the model simulations on location Schiedam are relatively large. The red bar give the percentage of the situations where the observation is more than two times the model simulation. The blue bar give the total number of differences that occurs for every input parameter (wind direction, wind speed, temperature, hour of the day, day of the week and month of the year).

For the wind direction, a high percentage of the differences is relatively large when the wind is from the south-east, but the total number of wind directions from the south-east is not very high. Thus it is assumed that the contribution to the total inaccuracy is not very large. For the wind speed, a high percentage of the differences is relatively large when the wind speed is below 2 $^m/_s$. Also the total number of times that the wind speed is below 2 $^m/_s$, is relatively large. This suggests that the inaccuracies in the model when the wind speed is low, have a large contribution to the total inaccuracy.

Another notable parameter is the hour of the day, in the morning and the end of the evening there are relatively many large differences. This is an indication that there are some inaccuracies in the sources which are time dependent (traffic and residents). The last interesting parameter is the month of the year. In the autumn and the winter are relatively many large differences. This is also an indication that the time dependent sources have inaccuracies. In Section 2.2 of [Kranenburg, 2009], it was already mentioned that the sources industry and shipping do not have a time dependency in the Real Time URBIS model and that this could be a shortcoming of the model. Thus also the sources shipping and industry may have inaccuracies.

The idea in this chapter is that the uncertainty of the model is caused by several different emission sources, therefore the Kalman filter will be applied on all the different sources. With this application, all the standard concentration fields for all emission sources will be estimated. These estimates are again calculated by multiplying each field with a correction factor, which leads to a corrected state equation:

$$\underline{c}_k = \sum_{i=1}^{88} \mu_{i,k} \underline{m}_i e^{\gamma_{i,k}} \tag{6.1}$$

In this chapter it is no longer assumed that some of the entries of $\underline{\gamma}$ are equal to zero.

The Kalman filter process will estimate all values of $\underline{\gamma}$ by a comparison of the model with the observations.

An advantage of this application is that the uncertainty intervals for the total concentration is a combination of uncertainty intervals for the different emission sources. This leads to an uncertainty interval which is different for every location. For example, on locations where the concentration is mostly caused by emission from traffic, the uncertainty interval is approximately equal to the uncertainty interval of the concentration from traffic sources. If the uncertainty for the source traffic can be reduced, the uncertainty on all locations with high traffic emission will be reduced.

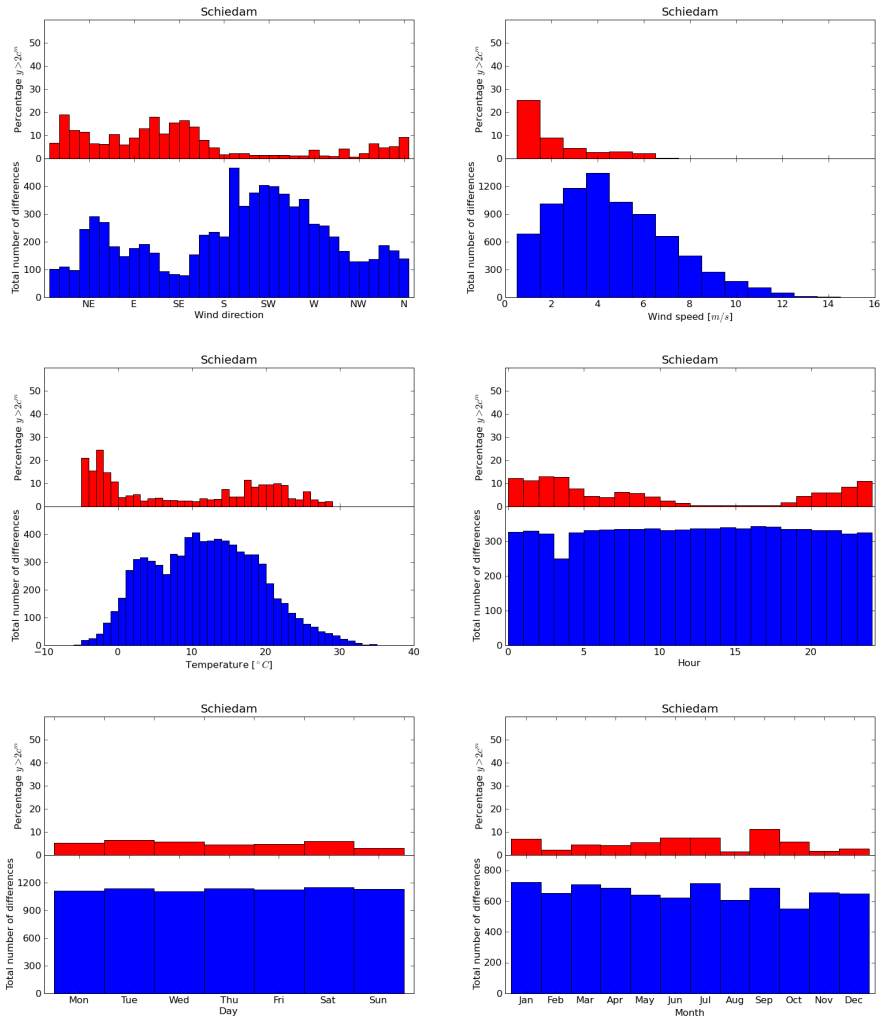


Figure 6.1: *Bar plots of relatively large differences between observations and model simulations at location Schiedam. In the lower graphs is the total number of differences plotted for each input parameter, in the upper graphs is the percentage given when the observation is more than two times the model simulation.*

6.2 Kalman filter

The application of the Kalman filter is nearly the same as in the application for the background concentrations. Every standard concentration fields gets a correction factor. So for each of the entries of $\underline{\gamma}$ a temporal correlation parameter has to be

found. For some sources this will be difficult because there is no good series of measurements to calculate the correlation. In Section 6.4 the different values for α will be calculated.

Not all the values for α can be computed exactly and also the uncertainty of the model σ and the uncertainty of the observations (R), estimated in Section 5.3, are not known exactly. Therefore some sensitivity runs are done to find the optimal values for α, σ and R, this will be explained in Section 6.5

6.2.1 *Dynamical system*

To create a dynamical system for $\underline{\gamma}$, it is again necessary to make a switch to the logarithms of the concentrations. This is done in the next formula:

$$\ln\left(\underline{c}_k\right) = \ln\left(\sum_{i=1}^{88} \mu_{i,k}\underline{m}_i e^{\gamma_{i,k}}\right) \tag{6.2}$$

This equation is non-linear for $\underline{\gamma}_k$, therefore a linearization is made around $\underline{\gamma}_k = 0$:

$$\ln\left(\sum_{i=1}^{88} \mu_{i,k}\underline{m}_i e^{\gamma_{i,k}}\right) = \ln\left(\underline{c}_k^m\right) + \left[\frac{\mu_{j,k}\underline{m}_j}{\underline{c}_k^m}\right]_{j=1}^{j=88} \underline{\gamma}_k + \mathcal{O}\left(\underline{\gamma}_k \cdot \underline{\gamma}_k\right) \tag{6.3}$$

where $\underline{c}_k^m$ is the total concentration, calculated by the model. The dynamical system will then become:

$$\ln\left(\underline{c}_k\right) = \ln\left(\underline{c}_k^m\right) + \left[\frac{\mu_{j,k}\underline{m}_j}{\underline{c}_k^m}\right]_{j=1}^{j=88} \underline{\gamma}_k \tag{6.4}$$

$$\underline{\gamma}_{k+1} = A\underline{\gamma}_k + \underline{\omega}_k \qquad \underline{\omega}_k \sim N\left(0, Q\right) \tag{6.5}$$

Matrix A contains the temporal correlation parameters, these are calculated in Section 6.4. Covariance matrix Q is again a diagonal matrix, with diagonal elements q_i^2. This is a colored noise process, driven by a white noise process, like in the application for the background. Furthermore it is assumed that both the temporal correlation and the model uncertainty are independent of time:

$$q_i = \sqrt{1-\alpha_i^2}\sigma_i^2 \tag{6.6}$$

where σ_i corresponds with the model uncertainty for each entry of $\underline{\gamma}$.

6.2.2 *Kalman filter form*

The dynamical system has to be written in Kalman filter form, for the implementation of the Kalman filter. There are still 9 series of measurements available, these series will be compared with the model simulations to get a better estimate of the NO_x concentration. The dynamical system in Kalman filter form is defined as follows:

$$\underline{\gamma}_{k+1} = A\underline{\gamma}_k + \underline{\omega}_k \qquad \underline{\omega}_k \sim N(0, Q) \tag{6.7}$$

$$\ln\left(\underline{y}_k\right) = H\left(\ln\left(\underline{c}_k^m\right) + \left[\frac{\mu_{j,k}\underline{m}_j}{\underline{c}_k^m}\right]_{j=1}^{j=88}\underline{\gamma}_k\right) + \underline{\nu}_k \quad \underline{\nu}_k \sim N(0, R_k) \tag{6.8}$$

In these equations matrix H is the system operator which projects the model state onto the observations. Covariance matrix R corresponds to the uncertainty of the logarithms of the observations, the instrumental error combined with the representation error. Matrix R will be a diagonal matrix, the elements on the diagonal are estimated in Section 5.3.

To simplify the notations from Equation 6.7 and 6.8, the Kalman filter equations are written as follows:

$$\underline{\gamma}_{k+1} = A\underline{\gamma}_k + \underline{\omega}_k \qquad \underline{\omega}_k \sim N(0, Q) \tag{6.9}$$

$$\underline{\tilde{y}}_k = \tilde{H}\underline{\gamma}_k + \underline{\nu}_k \qquad \underline{\nu}_k \sim N(0, R_k) \tag{6.10}$$

where vector $\underline{\tilde{y}}$ and matrix $\tilde{H}$ are defined as follows:

$$\underline{\tilde{y}}_k = \ln\left(\underline{y}_k\right) - H\ln\left(\underline{c}_k^m\right) \tag{6.11}$$

$$\tilde{H} = H\left[\frac{\mu_{j,k}\underline{m}_j}{\underline{c}_k^m}\right]_{j=1}^{j=88} \tag{6.12}$$

6.2.3 Forecast step

In the forecast step, a prediction is made for the values of $\underline{\gamma}_{k+1}$ with information from the time step before. The expected median and variance of $\underline{\gamma}_{k+1}$ are given by:

$$\underline{\hat{\gamma}}_{k+1}^f = A\underline{\hat{\gamma}}_k \tag{6.13}$$

$$P_{k+1}^f = AP_kA^{\mathrm{T}} + Q \tag{6.14}$$

6.2.4 Analysis step

In the analysis step the forecasted concentrations are compared with the observations. Like in the application for the background concentrations, it is not expected that the values for $\underline{\gamma}$ will become very large. Also the linearization around $\underline{\gamma} = 0$ is of order $\mathcal{O}(\underline{\gamma} \cdot \underline{\gamma})$, thus large values for γ_i will cause stability problems. For those reasons a screening process as described in Section 5.6 is implemented. In Section 6.3 the screening process for this application will be explained. The analysis step is the same as in the application for the background:

$$\underline{\gamma}_{k+1}^a = \underline{\gamma}_{k+1}^f + K_{k+1}\left(\underline{\tilde{y}}_{k+1} - \tilde{H}_{k+1}\underline{\hat{\gamma}}_{k+1}^f\right) \tag{6.15}$$

$$P_{k+1}^a = \left(I - K_{k+1}\tilde{H}_{k+1}\right)P_{k+1}^f\left(I - K_{k+1}\tilde{H}_{k+1}\right)^{\mathrm{T}} + K_{k+1}R_{k+1}K_{k+1}^{\mathrm{T}} \tag{6.16}$$

where K_{k+1} is again the minimal variance gain, the gain that minimizes P^a_{k+1}defined as follows:

$$K_{k+1} = P^f_{k+1}\tilde{H}^{\mathrm{T}}_{k+1}\left(\tilde{H}_{k+1}P^f_{k+1}\tilde{H}^{\mathrm{T}}_{k+1} + R_{k+1}\right)^{-1} \tag{6.17}$$

6.3 Screening process

If the difference between the observation y and the model simulation $\underline{c}^m_k$ is large, the Kalman filter will produce a large correction factor for one or more standard concentration fields. This is not wanted, because it is likely that a large difference is caused by another inaccuracy in the model or by an incidental occasion. For example when a road is blocked, the traffic pattern is different and thus the emissions are different from the expectations calculated by the model. Like in the application for the background in Chapter 5, a criterion has to be made whether a measurement is good enough.

After the forecast step, it is possible to make an uncertainty interval of the forecasted concentration. Further an uncertainty interval for the observation can be calculated with the uncertainty of the measurements (R). When both intervals have an empty intersection, the difference between the simulation and the observation is too large. If for both intervals the 2σ uncertainty interval is taken, the screening criterion corresponds with the screening criterion in Section 5.6:

$$\left[\sum_{i=0}^{88} m_i\mu_{i,k}e^{\gamma^f_{i,k}-\beta p^f_{i,k}}, \sum_{i=0}^{88} m_i\mu_{i,k}e^{\gamma^f_{i,k}+\beta p^f_{i,k}}\right] \cap [y_k - \beta r_{\mathrm{frac}}y_k, y_k + \beta r_{\mathrm{frac}}y_k] \neq \emptyset \tag{6.18}$$

The value for r_{frac} is optimized in Section 6.5 and equal to 0.34. The application of the Kalman filter with this screening process with $\beta = 2$, results in a concentration for the first week of 2006 for locations Schiedam and Overschie as shown in Figure 6.2. Contradicting to Figure 5.6, only 12% of the observations are not executed in the analysis step of the Kalman filter. This means that the inaccuracies in the model could be well described by inaccuracies of the several standard concentration fields.

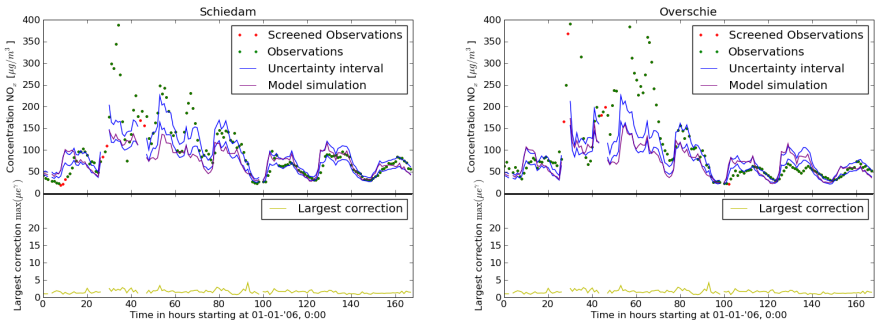


Figure 6.2: *Concentrations for the first week of 2006 after application of the Kalman filter on all the sources, with a screening process, at locations Schiedam and Overschie*

6.4 Correlation parameters α

In Section 5.4, an estimate value for the parameters α_i corresponding with the source background is calculated. In this section the same procedure will be done to obtain estimated values for the other parameters α_i. The temporal correlation for the source 'Rest' is not possible to calculate with a series of measurements, therefore some runs of the Kalman filter has to be applied to get the optimal values for α_i corresponding with the source 'Rest', this will be done in Section 6.5.

6.4.1 Traffic sources

The temporal correlation for the traffic sources will be obtained by looking at the observations from the monitoring stations in Overschie and Ridderkerk. The station in Overschie is located close to main road A20. In Ridderkerk the station is located close to main roads A15 and A16, therefore both stations will give a good approximation of the emission from traffic sources.

In Figure 6.3, the temporal correlation is given for both stations, this is done with the same method as for the background sources, described in Section 5.4. The best fitting exponential function has de-correlation parameter $\tau = 10$, thus an estimated value for each α_i corresponding with a traffic source is equal to $e^{-1/10}$.

The high peaks at 24 and 48 can be explained by the fixed traffic pattern. Each day the amount of traffic is roughly the same, thus there is a high mathematical temporal correlation for periods of one day.

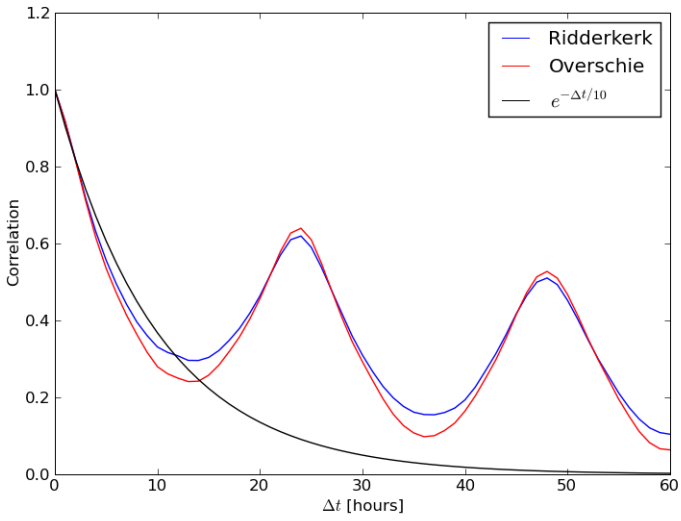


Figure 6.3: *The temporal correlation for traffic stations Overschie and Ridderkerk. The black line corresponds with the de-correlation parameter* $\tau_{tr} = 10$

6.4.2 Industry source

The temporal correlation for the source 'Industry' is not easy to determine. There is no monitoring station, which is placed on a location with a dominating industry emission. The monitoring station in Vlaardingen is the best station to calculate the

correlation. This only results in an estimated correlation which is not very accurate. In Figure 6.4, the temporal correlation on location Vlaardingen is given. The best fitting exponential has de-correlation parameter $\tau = 10$, thus an estimated value for α_i corresponding with the source 'Industry' is equal to $e^{-1/10}$.

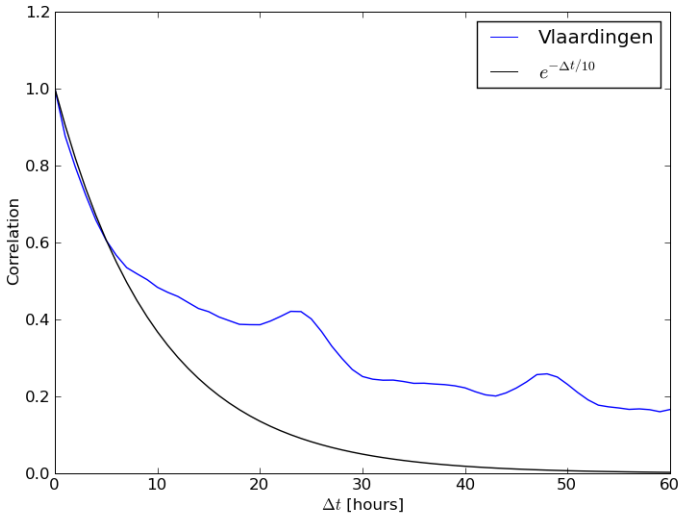


Figure 6.4: *The temporal correlation for station Vlaardingen, the station which matches best with the source industry. The black line has de-correlation parameter $\tau_{in} = 10$.*

6.4.3 Shipping sources

Also for the temporal correlation of the shipping sources, it is not easy to determine an estimate value for α_i. The monitoring station Maassluis is the best to calculate the correlation, the temporal correlation at Maassluis is given in Figure 6.5. In here the same holds as for the industry, the de-correlation parameter $\tau = 8$ is only an inaccurate estimate. This leads to an estimated value for α_i corresponding with the shipping sources which is equal to $e^{-1/8}$.

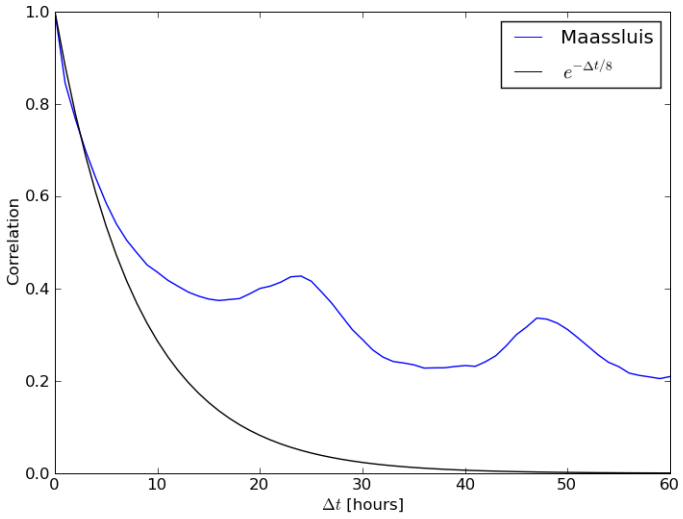


Figure 6.5: *The temporal correlation for station Maassluis, the station which corresponds best with the emission from shipping. The black line corresponds with de-correlation parameter* $\tau_{sh} = 8$.

6.4.4 Rest source

It is clear that the temporal correlation for the source 'Rest' is not easy to declare. The only idea of this temporal correlation is that the de-correlation parameter τ_{re} will be around the values for the de-correlation parameters $\tau_{bg}, \tau_{tr}, \tau_{in}$ and τ_{sh}. The de-correlation parameter τ_{re} is estimated equal to 10, thus α_i corresponding with source 'Rest' is estimated equal to $e^{-1/10}$.

6.5 Sensitivity runs

In Section 5.4 and 6.4, all the parameters $\alpha_{i,j}$ for the matrix A are not computed exactly. Also the parameter σ for the uncertainty of the model and the uncertainty of the measurements r_{frac} are not known exactly. Therefore the Kalman filter is applied for different values of $\tau_{bg}, \tau_{tr}, \tau_{sh}, \tau_{in}, \tau_{re}$ and different values of σ and r_{frac}.

In Section 5.4, it has been shown that an estimated value for τ_{bg} is equal to 12. In Section 6.4 the estimated values for $\tau_{tr}, \tau_{sh}, \tau_{in}$ and τ_{re} were found. The uncertainty of the model is assumed between 10% and 30%. The instrumental error in the observations calculated in Section 5.3 is equal to 8%, but the representation error may be larger, due to a grid with a low resolution. A trial and error process leads to the conclusion that the Kalman filter gives an optimal result with total uncertainty of the measurements between 20% and 40%. This will lead to applications of the Kalman filter with the following values:

$$\begin{aligned} \tau_{bg} &\approx 12 \\ \tau_{tr} &\approx 10 \\ \tau_{sh} &\approx 8 \\ \tau_{in} &\approx 10 \\ \tau_{re} &\approx 10 \\ \sigma &\in [0.10, 0.20] \\ r_{frac} &\in [0.20, 0.40] \end{aligned} \tag{6.19}$$

To obtain which combination of values for $\tau_{bg}, \tau_{tr}, \tau_{sh}, \tau_{in}, \sigma$ and r_{frac} is the best, there are three criteria which have to be optimized: The Root Mean Squared Error ($RMSE$), the mean of the differences between the Kalman filter results and the observations ($Mean$) and finally the standard deviation of these differences (Std).

In the analysis to find the optimal values for the input parameters, the monitoring stations in Overschie and Ridderkerk are not involved. At those two locations the differences between the observations and the simulations are very large. This could lead to inaccuracies in the calculation of the optimal values for the Kalman filter parameters.

6.5.1 Root Mean Squared Error (RMSE)

The first criterion is to minimize the value for $RMSE$:

$$RMSE = \frac{1}{\sqrt{7n}} \sqrt{\sum_{k=1}^{n} \sum_{i=1}^{7} \left(y_{i,k} - c_{i,k}^{KF} \right)^2} \tag{6.20}$$

where $c_{i,k}^{KF}$ is the concentration after the application of the Kalman filter. The number of time steps corresponds with the value of n, which is equal to 8760 for a whole year. The summation over i is up to 7, the number of monitoring stations. The value for the RMSE is a measure for the absolute difference between the results after application of the Kalman filter and the observations. When this is minimized the Kalman filter results have the smallest distance the observations.

6.5.2 Mean

Another criterion is the mean of the differences between the Kalman filter results and the observations:

$$Mean = \frac{1}{7n} \sum_{k=1}^{n} \sum_{i=1}^{7} \left(y_{i,k} - c_{i,k}^{KF} \right) \tag{6.21}$$

The value for this mean is in the optimal situation equal to 0. In that case the Kalman filter results have the same mean as the observations.

6.5.3 Standard deviation

The final criterion is the standard deviation of the differences between the Kalman filter results and the observations:

$$Std = \frac{1}{\sqrt{7n}} \sqrt{\sum_{k=1}^{n} \sum_{i=1}^{7} \left(\left(y_{i,k} - c_{i,k}^{KF} \right) - Mean \right)^2} \tag{6.22}$$

In the optimal situation the value for the standard deviation is equal to 1.

6.5.4 Optimal results for RMSE, Mean and Standard deviation

If the Kalman filter is applied with different values for each parameter, the numbers $RMSE$, Std and $Mean$ determine the optimal value for each parameter. In the left upper panel of Figure 6.6 the influence of parameter τ_{bg} is shown. It can be seen that the $RMSE$ is nearly independent of this parameter, the value for $Mean$ is close to zero for $\tau_{bg} = 2$ and the value for Std is close to 1 for $\tau_{bg} = 10$. Therefore the optimal value for the parameter τ_{bg}, according to this sensitivity run will be equal to 6, which is the average between 2 and 10.

In the right upper panel of Figure 6.6, the influence of parameter τ_{tr} is shown. This figure shows that an optimal value for τ_{tr} will be equal to 4. The lower left panel and the lower right panel shows the influence of parameters σ and r_{frac}. The same analysis as for the temporal correlation parameters leads to optimal values $\sigma = 0.19$ and $r_{\text{frac}} = 0.34$.

The parameters, τ_{sh}, τ_{in} and τ_{re} does not have a large impact on the three criteria. Therefore they will be chosen equal to the estimated values from Section 6.4.

In Table 6.1, all the information about the input parameters is given: the second column contains the estimated values for these parameters and the third column contains the optimal values according to some sensitivity runs. Finally, the last column shows the input parameters which are used in the rest of this report. These values are determined by the estimate and by the optimal values from the sensitivity runs.

Table 6.1: *Input parameters for the Kalman filter. In the final column are the values which are involved in the Kalman filter.*

Parameter	Estimate	Calculation by sensitivity runs	Value implemented in the Kalman filter
r_{frac}	0.20-0.40	0.34	0.34
σ	0.10-0.30	0.19	0.19
τ_{bg}	12	6	10
τ_{tr}	10	4	8
τ_{sh}	8	—	8
τ_{in}	10	—	10
τ_{re}	10	—	10

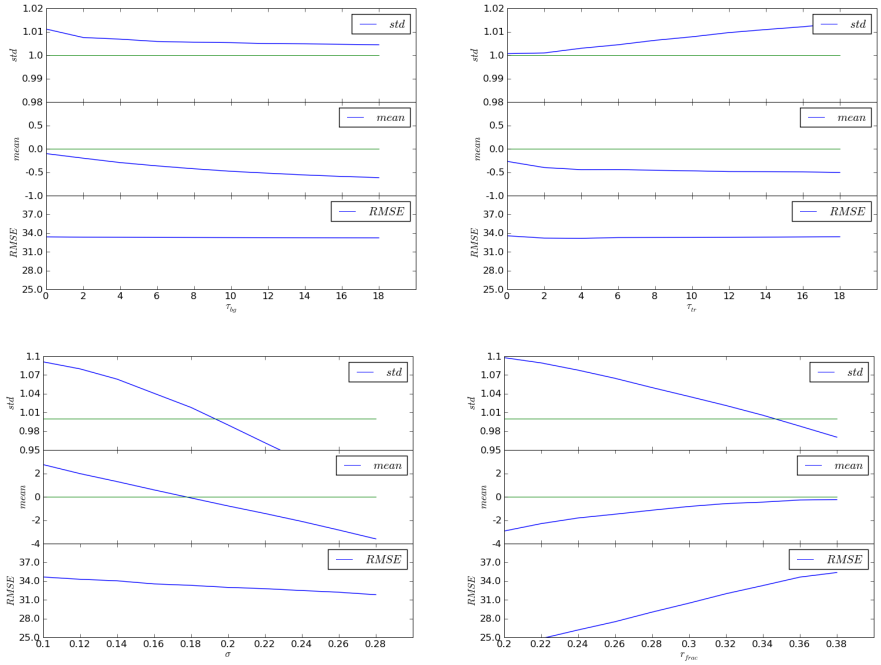


Figure 6.6: *Sensitivity of the three criteria against the Kalman filter parameters. Upper left:* τ_{bg}*, upper right:* τ_{tr}*, lower left:* σ*, lower right:* r_{frac}

6.6 Connection with population

After application of the Kalman filter, it is possible to calculate an uncertainty interval for the concentration NO_x for each grid point in the area of interest. The next objective is to connect the interval on a certain grid point with the number of people living nearby that grid point.

6.6.1 Population density

A map of the population of the area is given in Figure 6.7. This figure represents the density of postal zip codes per grid point instead of the number of people per grid point. The total number of zip codes in this area is equal to 595.396. According data from CBS [1], the total number of residents in this region is equal to 1.186.306 on the first of January of 2006. Thus the average number of people per zip code is equal to 1.99. Further in this report it is assumed that the number of people per zip code is equal, thus the number of residents per grid point is 1.99×the number of zip codes per grid point.

$$\text{pop}_j = 1.99 \times \#\text{of zip codes} \tag{6.23}$$

[1] CBS: Centraal Bureau voor de Statistiek. *www.cbs.nl*
Dutch organization for statistics

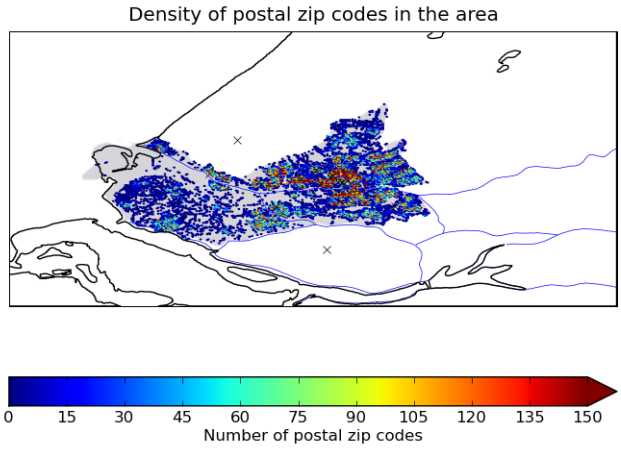


Figure 6.7: *Density of postal zip codes in the DCMR area*

6.6.2 Absolute uncertainty connected with population density

For every grid point, on every hour an uncertainty interval is calculated by the Kalman filter application. The width of these intervals is a measure for the uncertainty of the concentration NO_x, if the width of the interval is small, the estimate of the concentration NO_x is accurate, thus little uncertainty. The idea is now to have small intervals on locations where the population density is high, in that case there is a good estimate of the exposure of the population to the concentration NO_x. The width of an uncertainty interval in grid point j on time k is the upper bound of the 1σ interval minus the lower bound of the 1σ interval:

$$u_{j,k}^{abs} = \sum_{i=1}^{88} \mu_{i,k} m_{i,j} e^{\gamma_{i,k}+p_{i,k}} - \sum_{i=1}^{88} \mu_{i,k} m_{i,j} e^{\gamma_{i,k}-p_{i,k}} \tag{6.24}$$

where $m_{i,j}$ is the standard concentration of emission source i in grid point j. Further the width of an uncertainty interval is called the absolute uncertainty.

In Figure 6.8, the annual mean $\bar{u}_j^{abs}$ of $\left\{u_{j,k}^{abs}\right\}_{k=1}^{8760}$ is plotted for each grid point:

$$\bar{u}_j^{abs} = \frac{1}{8760} \sum_{k=1}^{8760} u_{j,k}^{abs} \tag{6.25}$$

In this figure, it can be seen that relatively many grid points have an annual mean of absolute uncertainty above 40. This large uncertainty mostly occurs on main roads and industrial regions. So there are not that many people that lives nearby grid points with a large uncertainty. This is shown in Figure 6.9, where the annual mean $\bar{u}_j^{abs}$ is compared with the population. On the x-axis are the values of $\bar{u}^{abs}$, on the y-axis are the number of people living nearby a grid point with that annual mean. For $\bar{u}_j^{abs} \in \left[u_i^{abs}, u_{i+1}^{abs}\right]$, the number of people for that width range of $\bar{u}^{abs}$ is equal to:

$$\sum_{j=1}^{n_{gc}} \text{pop}_j \mathcal{I}_{\{\bar{u}_j^{abs} \in [u_i^{abs}, u_{i+1}^{abs}]\}} \tag{6.26}$$

where n_{gc} is the number of grid points and $\mathcal{I}$ is the indicator function. In this figure, it can be seen that unless a lot of grid points have a large uncertainty, there are not many people living nearby those grid cells. The histogram is centered around $\bar{u}_j^{abs} = 14$, which means that most of the people lives nearby a grid point with annual mean of the absolute uncertainty around 14.

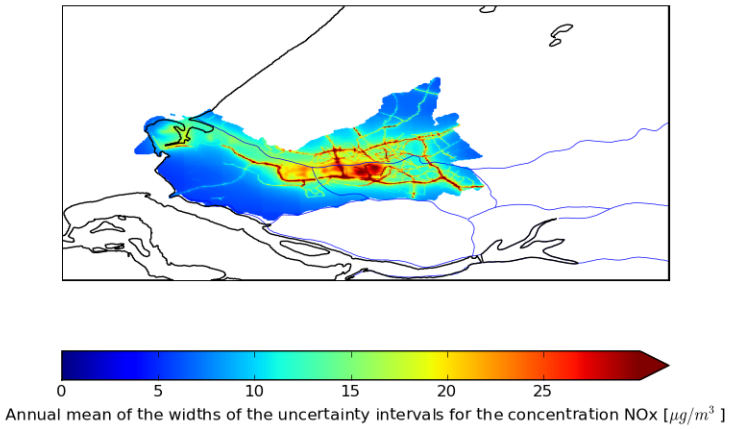


Figure 6.8: *The values for $\bar{u}_j^{abs}$ over the whole area of interest.*

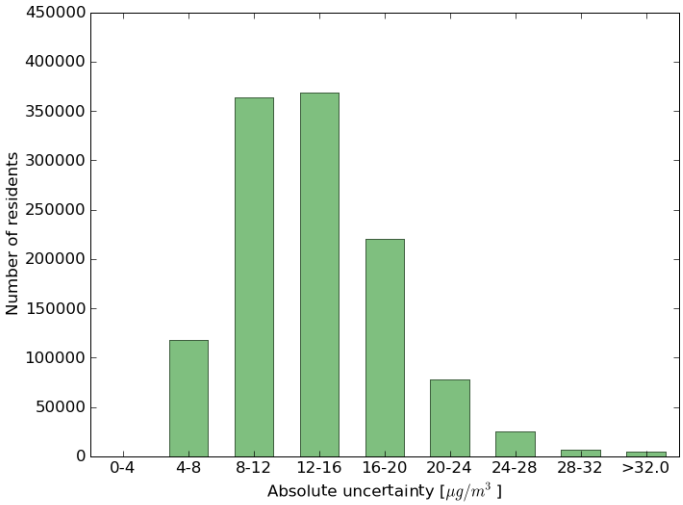


Figure 6.9: *Histogram of the number of people against the annual means of the absolute uncertainties $\bar{u}^{abs}$. On the x-axis are the ranges of $\bar{u}^{abs}$, on the y-axis are the number of people living nearby a grid point with $\bar{u}_j^{abs}$ in that range.*

6.6.3 *Relative uncertainty connected with the population density*

Furthermore it is interesting to look at the relative uncertainty in the whole region. On each grid point the absolute uncertainty could be divided by the expected concentration. This is a measure for the relative uncertainty:

$$u_{j,k}^{rel} = \frac{u_{j,k}^{abs}}{c_{j,k}^{KF}} \tag{6.27}$$

where $c_{j,k}^{KF}$ is the expected concentration on grid point j at time k after the application of the Kalman filter:

$$c_{j,k}^{KF} = \sum_{i=1}^{88} \mu_{i,k} m_{i,j} e^{\gamma_{i,k} + 1/2 p_{i,k}^2} \tag{6.28}$$

The expected concentration is calculated with the expectation of the log-normal distribution, therefore the term $1/2 p_{i,k}^2$ is taken into the exponential.

In Figure 6.10 the annual mean $\bar{u}_j^{rel}$ of $\left\{u_{j,k}^{rel}\right\}_{k=1}^{8760}$ is plotted for each grid point in the domain. It is clear that the relative uncertainty has the smallest values on the main roads and around the 'Nieuwe Waterweg' the harbor entry of Rotterdam. At the 9 monitoring stations, the contribution from the shipping and traffic sources is large. Therefore the Kalman filter produces a smaller relative uncertainty of these sources.

In Figure 6.11, the relative uncertainty is connected with the population. On the $x-$axis are the annual means of the relative uncertainty, on the $y-$axis the number of people living nearby a grid cell with that relative uncertainty. Most of the population have almost the same relative uncertainty. This is because, the main roads have the smallest relative uncertainty, but this does not a have large impact on the population.

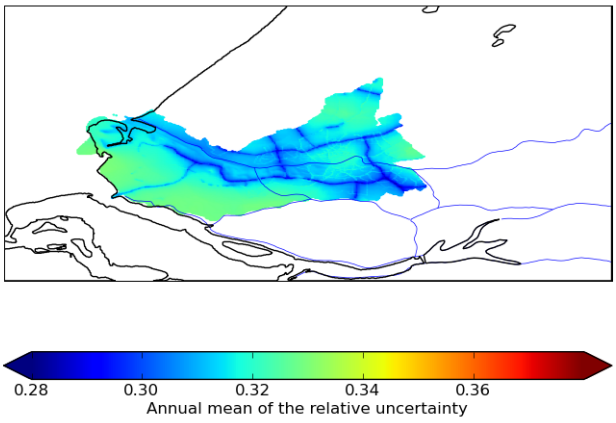


Figure 6.10: *The values for $\bar{u}_j^{rel}$ over the whole area of interest*

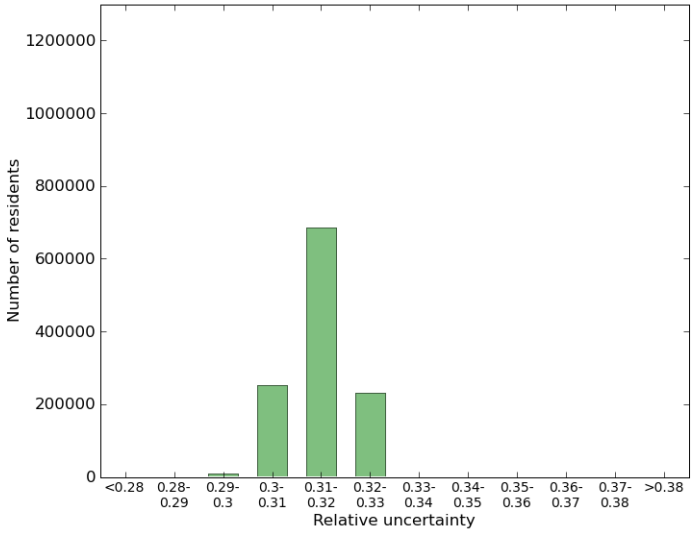


Figure 6.11: *Histogram of the number of people against the annual mean of the relative uncertainties $\bar{u}_j^{rel}$. On the x-axis are the ranges of $\bar{u}_j^{rel}$, on the y-axis are the number of people living nearby a grid cell with $\bar{u}_j^{rel}$ in that range.*

7 Conclusions and discussion

The present application of the Real Time URBIS model causes some problems like the possible occurrence of negative concentrations. These problems are caused by some inaccuracies of the model. In Chapter 3, it is shown that the inaccuracies of the model are depends on the wind direction, the wind speed and the hour of the day. For that reason a Kalman filter is applied on the standard concentration fields from the URBIS model to eliminate these inaccuracies. With the Kalman filter the model simulations are connected with a series of measurements. This connection leads to a corrected simulation of the concentration NO_x, together with an uncertainty interval for this concentration.

In Chapter 5, the Kalman filter was only applied on the correction factors for the background concentrations, this was not sufficient to eliminate all inaccuracies. Therefore in Chapter 6, the Kalman filter is applied on all emission sources. The corrected model simulations fits better on the observations, thus the Kalman filter is a good instrument to make a real time correction of the Real Time URBIS model.

The application of the Kalman filter results in an uncertainty interval for each correction factor, with these uncertainty intervals it is possible to calculate an uncertainty interval for the concentration NO_x on the whole domain covered by DCMR. The widths of the uncertainty intervals depends on the contribution of each emission source to the total emission.

The uncertainty interval has a large width on the main roads and in the industry region around Pernis. On this locations the concentration is relatively large, thus also the absolute uncertainty will be large. The application of the Kalman filter reduces the relative uncertainty, this mainly happened on the main roads and around the 'Nieuwe Waterweg'. This is because the concentrations on the 9 monitoring stations have large contributions from the traffic and the shipping sources, thus the relative uncertainty of these sources is decreased.

The connection between the uncertainty intervals and the population density leads to some extensions of the Kalman filter. The extensions will be discussed in the next part of this report.

Part II

Extended applications of the Kalman filter to reduce the uncertainty

8 Introduction

In the first part of this report, the theory and the results of the use of a Kalman filter in the Real Time URBIS model is given. The basic idea of a Kalman filter is to produce a Gaussian distribution for a certain unknown variable. In the Real time URBIS model the unknown variable is the vector with correction factors for each standard concentration field from the URBIS model. With the Gaussian distribution for the correction factor, an uncertainty interval for the concentration NO_x is found.

In this part some methods are described to reduce the uncertainty, the main idea is that the uncertainty should be as small as possible on locations with high population density. In Chapter 9, some extra monitoring stations are added to the present monitoring system. If these stations are placed on well chosen locations, the total uncertainty connected with the population can be minimized. In this chapter also a method is described to create an optimal setting of monitoring stations. In Chapter 10, the Kalman filter is applied on some different time scales. Using this, it is possible to add monitoring stations, which measures the concentration on different time scales. This will lead to a description of an optimal placement of extra monitoring stations with different time scales. In Chapter 11, another extension of the Kalman filter will be described. In that chapter the correction factors, calculated as in the first part, will be analyzed. This analysis leads to some other ideas of inaccuracies in the Real Time URBIS model. With this ideas the model could be improved, such that the uncertainty will decrease. Finally in Chapter 12, the conclusions of the extensions of the Kalman filter are given.

9 Extra monitoring stations

9.1 Introduction

The aim is to reduce the width of the uncertainty intervals. The idea is that a reduction of the uncertainty can be done by a reduction of the uncertainty of one of the emission sources. To reduce the uncertainty of an emission source, it is possible to add a monitoring station. If this station is located nearby a grid point with a dominating emission from one of the sources, the idea is that the uncertainty of that source will be reduced. This will then lead to a reduction of the total uncertainty.

To have an optimal reduction of the total uncertainty, it seems obvious to reduce the uncertainty of the important emission sources. In Section 9.2, the exposure on the population is shown for each of the 11 emission sources. This leads to an insight into the importance of each emission source. The sources with the largest contribution to the exposure causes the largest uncertainty, thus a reduction in the uncertainty of those sources will cause an effective reduction in the total uncertainty in relation with the population.

Further it is important to look at the influence of a monitoring station on the uncertainty, therefore in Section 9.3, a simulation is made without any measurements involved in the Kalman filter. The uncertainty of the model (the uncertainty p of the correction factors γ were stated equal to 19 %) will be used to get the uncertainty intervals for each grid point for each hour. After that in Section 9.4, the present stations will be added to the Kalman filter to see the influence of a monitoring station on the uncertainty.

If the influence of the different stations on the uncertainty and the importance of each emission source are known, some virtual monitoring stations will be added to the system in Section 9.6.1. The locations of these virtual monitoring stations will be determined by the analysis of the influences of the other stations and by the analysis of the importance of each emission source.

9.2 Exposure per emission source

In this section the exposure to the concentration caused by each emission source is determined. The exposure per emission source can be used to determine the importance of each source. When the exposure of the population on the concentration NO_x caused by a specific emission source is small, it is not useful to decrease the uncertainty of that specific emission source. A reduction of the emission from that source will not lead to a large reduction of the total uncertainty connected with the population.

To determine the emission per source, a simulation is made for the whole year for each emission source separately. A measure for the exposure to each emission source can be given by the number E_s:

$$E_s = \sum_{i=1}^{n_{gc}} \bar{c}_{s,j} \times \text{pop}_j \tag{9.1}$$

where $\bar{c}_{s,j}$ is the annual mean of the concentration caused by source s on grid point j. The number of people living nearby grid point j corresponds with pop_j. In Table 9.1, the numbers E_s are given for each emission source, as well as the contribution to the total exposure. This table shows that the sources 'Zone card', 'Background' and 'Ships sea' have the largest contribution to the exposure. The source 'Zone card' is an emission sources which covers the emission of highway traffic, the source 'Background' covers the emission which is blown into the area from the rest of the Netherlands and the source 'Ships sea' covers the emission from sea ships in the harbor of Rotterdam. The figures in Appendix B shows the standard concentration fields of each of the emission sources.

Table 9.1: *Exposure caused by each of the different emission sources*

Source	Exposure E $\times 10^6$	Percentage on total contribution
Abroad	0	0
Background	5.3	21.0 %
Zone card	5.0	20.0 %
CAR	1.5	6.0 %
Roads nearby	2.4	9.6 %
Roads far	1.5	6.0 %
Industry	0.92	3.7 %
Domestic	1.3	5.2 %
Ships inland	0.016	0.1 %
Ships sea	4.4	17.9 %
Rest	2.6	10.5 %
Total	24.9	100.0 %

9.3 Annual mean of the uncertainty without a Kalman filter

For the year 2006, the concentration is simulated without the Kalman filter. The uncertainty of the model forms the basis of the annual mean of the uncertainty on each grid point. The annual means of the absolute uncertainties are shown for each grid point in the left panel of Figure 9.1. The absolute uncertainty in grid point j at time k, is simply the width of the uncertainty interval of the total concentration NO_x, as in Equation 6.24:

$$u_{j,k}^{abs} = \sum_{i=1}^{88} \mu_{i,k} m_{i,j} e^{\gamma_{i,k}+p_{i,k}} - \sum_{i=1}^{88} \mu_{i,k} m_{i,j} e^{\gamma_{i,k}-p_{i,k}} \tag{9.2}$$

There is also a relative uncertainty as in Equation 6.27, this is the absolute uncertainty divided by the expected concentration NO_x:

$$u_{j,k}^{rel} = \frac{u_{j,k}^{abs}}{c_{j,k}^{KF}} \tag{9.3}$$

Here $c_{j,k}^{KF}$ represents the expected concentration NO_x in grid point j at time k after the application of the Kalman filter, this expected concentration is again calculated from the expectation of the log-normal distribution:

$$c_k^{KF} = \sum_{i=1}^{88} \mu_{i,k} m_{i,j} e^{\gamma_{i,k} + 1/2 p_{i,k}^2} \tag{9.4}$$

The annual mean of the relative uncertainty for only the model simulation is given in the right panel of Figure 9.1. Because there are no measurements involved in the Kalman filter, the the parameters γ and p are known and equal to $\gamma = 0$ and $p = 0.19$.

Both the absolute and the relative uncertainty can be connected with the population in the area to get an idea of the influence of this uncertainty on the population. In Figure 9.2 these connections are shown. In the left panel the connection with the absolute uncertainty is shown, for each annual mean of the absolute uncertainty the number of people that lives nearby a grid point with that annual mean is given in the histogram.

The total absolute uncertainty on the population could also be expressed as a single number. This number U^{abs} is the sum over all grid points of the annual mean of the absolute uncertainty per grid point multiplied with the number of people living nearby that grid point:

$$U^{abs} = \sum_{j=1}^{n_{gp}} \bar{u}_j^{abs} \times \text{pop}_j \tag{9.5}$$

Here n_{gp} is the number of grid points and $\bar{u}_j^{abs}$ is the annual mean of the absolute uncertainty on grid point j. The variable pop_j is the number of people living nearby grid point j. The idea is to minimize this number U^{abs}. When the uncertainty is large in a sparsely populated grid point, this number will not get large. For the simulation without any measurements taken into the Kalman filter application this number is equal to 19.0×10^6.

The connection between the relative uncertainty and the population is shown in the right panel of Figure 9.2. For each annual mean of the relative uncertainty, the number of people that lives nearby a grid point with that relative uncertainty is given in the histogram. Because the relative uncertainty is only determined by the model uncertainty($\gamma = 0, p = 0.19$), the relative uncertainty is constant on the whole domain and equal tot 0.38. Therefore the histogram has only one peak, all the people lives on a location with relative uncertainty equal to 0.38.

The total relative uncertainty on the population could also be expressed as a single number, this single number U^{rel} is given by:

$$U^{rel} = \sum_{j=1}^{n_{gp}} \bar{u}_j^{rel} \times \text{pop}_j \tag{9.6}$$

where $\bar{u}_j^{rel}$ is the annual mean of the relative uncertainty in grid point j. For the model simulation without any measurements involved in the Kalman filter, this number U^{rel} is equal to 445×10^3.

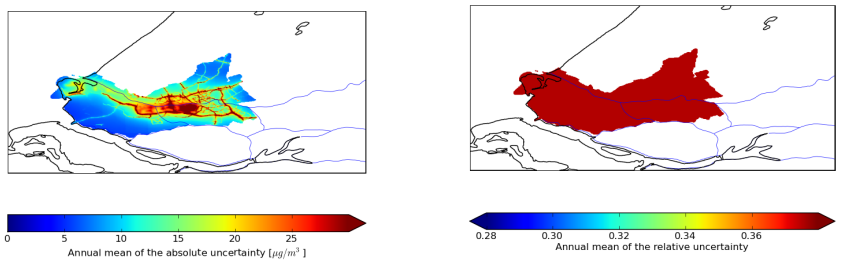


Figure 9.1: *Annual mean of the uncertainties for only the model simulation without any measurements involved in the Kalman filter. Left panel: absolute uncertainty. Right panel: relative uncertainty.*

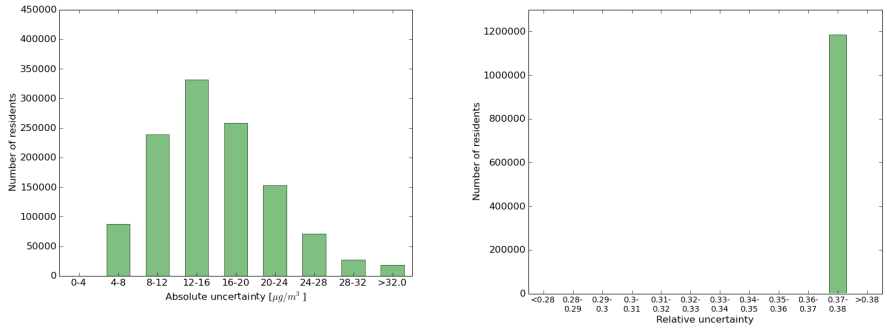


Figure 9.2: *Histograms of the number of people living nearby a grid point per range of uncertainty. Left panel: absolute uncertainty. Right panel: relative uncertainty.*

9.4 Influence of original stations on the absolute and relative uncertainty

In this section the influence on the uncertainty of the measurements made on the current monitoring stations is determined. In the current situation there are 4 locations with a dominating emission from the traffic sources. These locations are Ridderkerk, Overschie and the two stations at Bentinckplein. The monitoring station in Maassluis have a dominating shipping source. The other stations have more than 1 significant contribution of the several emission sources, these stations are so called combined stations.

In Section 9.4.1, the influence of the traffic sources will be determined. The idea is that the uncertainty of the traffic sources will decrease, such that the uncertainty on each point in the domain with dominating traffic emission will decrease. In Section 9.4.2, the influence of the shipping station in Maassluis will be determined. The influence of the combined stations will be determined in Section 9.4.3. Finally in Section 9.4.4, some combinations of stations will be analyzed.

9.4.1 Traffic stations

The stations in the domain with a dominating traffic source are located in Overschie, Ridderkerk and Bentinckplein. In Overschie and Ridderkerk the emission from highway traffic dominates the total emission. The stations on Bentinckplein have dominating emission from local traffic. Now the Kalman filter is applied on the model with the series of measurements made on one of these stations. Per station the influence on the uncertainty is measured by the number U^{abs} defined in Equation 9.5, and by the number U^{rel} defined in Equation 9.6. According to the theory about a Kalman filter, the relative uncertainty will always decrease in l^2-norm, after the application of the Kalman filter, while the absolute uncertainty could both increase and decrease. More about this is explained in Section 9.5.

Overschie

At location Overschie the traffic sources have a total contribution about 51 %, the largest contribution is from the highway traffic which has a contribution about 43 % on the total emission. The station in Overschie is located next to main road A13 not far from the junction of the main roads A13 and A20 (Kleinpolderplein). The problem with this monitoring station is that the observations on this location are much higher than the model simulations. The annual mean of all the observations is equal to 88.4 $\mu g/m^3$, while the annual mean of the model simulations is equal to 54.8 $\mu g/m^3$. The observations are 62 % higher than the simulations. Unless this large difference, only 4 % of the available observations is thrown out the analysis step of the Kalman filter by the screening criterion.

For the absolute uncertainty the number U^{abs} becomes equal to 19.5×10^6 which is an addition of 3.2 %. In Section 9.5, an explanation of this increase of uncertainty will be given.

For the relative uncertainty, the number U becomes 425×10^3, this is a reduction of 4.4 %. The idea is that this reduction is mainly caused by a reduction of the uncertainty of the traffic sources. This idea is confirmed by the fact that the relative uncertainty is most decreased on the main roads. This is shown in Figure 9.3, the left panel shows a reduction of the relative uncertainty for the main roads. Further the uncertainty in the rest of the area also decreased a little, this is caused by the other sources which have small contribution on the total emission in Overschie. The connection with the population is shown in the right panel of Figure 9.3, the first peak corresponds with the people that lives not far from the main roads, the other peak corresponds with the rest of the people.

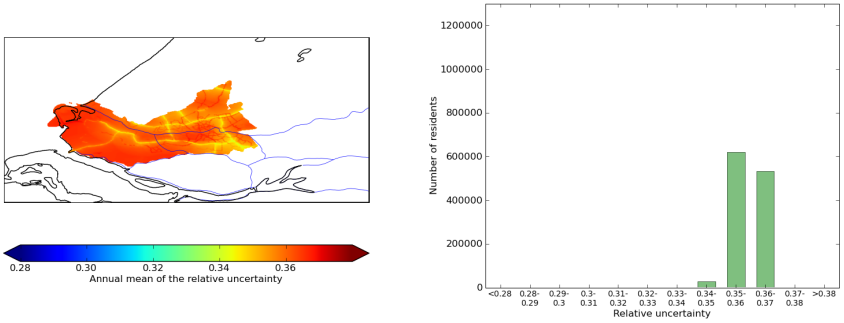


Figure 9.3: *Left panel: The annual mean of the relative uncertainty, with the measurements from Overschie in the Kalman filter, in the right panel these annual means are connected with the population.*

Ridderkerk

At location Ridderkerk the emission from highway traffic also dominates the total emission, about 85 % of the total emission is caused by highway traffic. The total contribution of all the traffic sources is about 91 %. The location Ridderkerk is located next to the main roads A15 and A16, not far from the junction Ridderster.

At this location, a similar problem exists as in Overschie. There is a large difference between the model simulations and the observations, at this locations the observations are much lower than the model simulations. The annual mean of the model simulations is equal to 203.9 $\mu g/m^3$, while the annual mean of the observations is equal to 94.6 $\mu g/m^3$. The simulations are 116 % higher than the observations.

In this situation the screening criterion throws 38 % of the observations out of the analysis step of the Kalman filter. The absolute uncertainty gets an extra reduction, because the expected concentration is lower than the model simulation. An explanation for this will be given in Section 9.5. The number U^{abs} becomes equal to 18.0×10^6, which is a reduction of 5.0 %.

For the relative uncertainty, the number U^{rel} has become equal to 434×10^3 which is a reduction of 2.4 % with respect to the situation without any observations involved in the Kalman filter. Also for this station the reduction is mainly caused by a reduction of the uncertainty of the traffic sources. Figure 9.4 shows the decrease of the relative uncertainty and the influence on the population of this reduction. Because of the large contribution from the traffic sources, the uncertainty on locations with small traffic emission is not decreased. Therefore, the uncertainty is little decreased for a lot of people, also the small number of observations which are used by the Kalman filter causes this small reduction.

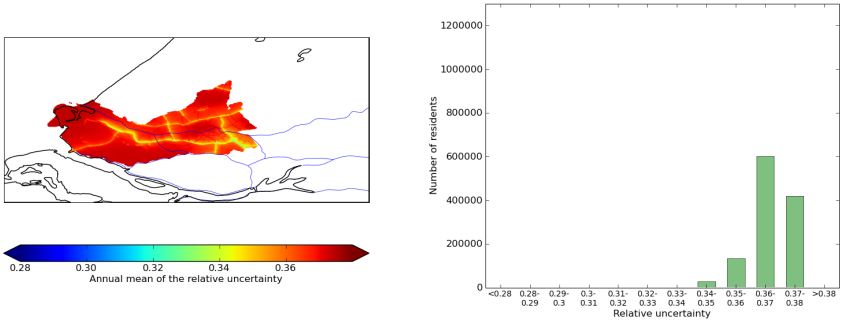


Figure 9.4: *Left panel: The annual mean of the relative uncertainty, with the measurements from Ridderkerk in the Kalman filter, in the right panel these annual means are connected with the population.*

Bentinckplein

The monitoring stations at Bentinckplein are located directly next to a busy local road in the center of Rotterdam. Therefore the emission from local traffic dominates the total emission, in the model the emission from local traffic is determined by two different sources 'CAR' and 'Roads nearby'. The total contribution of these two sources is 59 % of the total emission.

The annual mean of the model simulations is equal to 92.8 $\mu g/m^3$ for the DCMR station and 104.9 $\mu g/m^3$ for the RIVM station. Those two annual means should be the same, but only the model values with a corresponding observation are taken into these means. Because of some missing measurements the annual means of the concentrations for both stations are not taken for the same series of model simulations.

The annual mean of the observations made by DCMR is equal to 80.3 $\mu g/m^3$, which means that the simulations are 16 % higher than the observations. The annual mean of the observations made by RIVM is equal to 96.9 $\mu g/m^3$, which means that the observations are 8 % higher than the simulations. The screening criterion throws out 5 % of the observations made on the DCMR station and also 5 % of the observations made on the RIVM station.

For the relative uncertainty, the number U^{rel} becomes equal to 433×10^3 for the DCMR station and 436×10^3 for the RIVM station, this are reductions of 2.7 % and 2.0 %. Both these reductions are theoretically nearly independent of the observations, thus they must be nearly the same. This is not the case, the difference between those reductions could be caused by the difference in the number of observations taken into the analyzing step of the Kalman filter (7431 against 5818). This is explained in mor detail in Section 9.5.

The reductions are mainly caused by the reduction in the local traffic sources. Because these sources do not have large contributions on the total emission on the whole domain (see Section 9.2), the decrease in the relative uncertainty will not be very large. Further the decrease will be nearly the same in the whole area, as shown in Figure 9.5. The reason for this is that the contribution of the local traffic sources is nearly the constant on the whole area. Therefore the connection with the population shows that all people lives nearby grid points which have nearly the same (little reduced) relative uncertainty.

For the absolute uncertainty the number U^{abs} becomes equal to 17.9×10^6 for the DCMR station, this is a reduction of 5.3 % with respect to the situation without any

measurements. For the RIVM station, the number U^{abs} is equal to 18.2×10^6, a reduction of about 4.0 %.

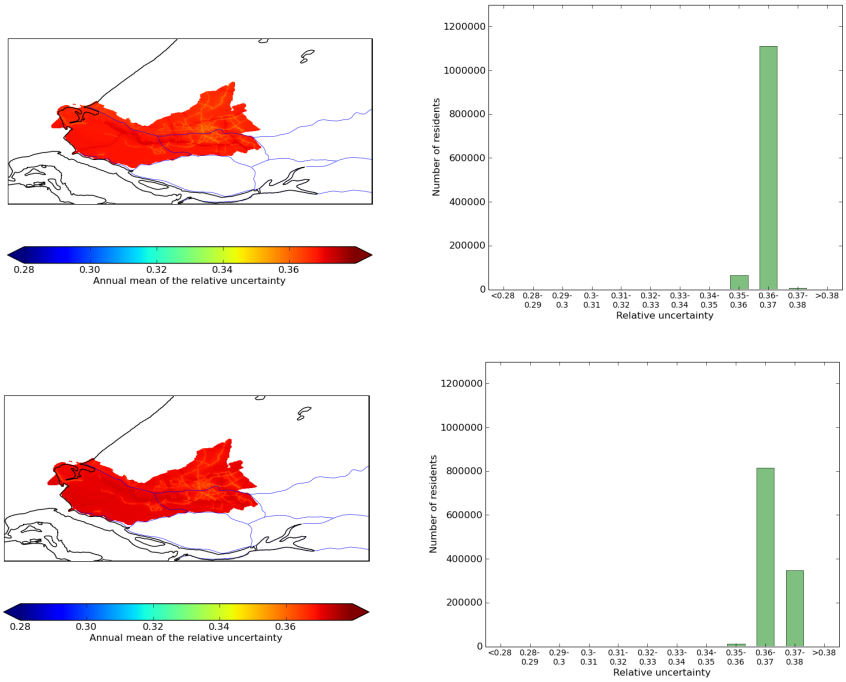


Figure 9.5: *Left panel: The annual means of the relative uncertainty, with the measurements from Bentinckplein (DCMR upper and RIVM lower) in the Kalman filter, in the right panel these annual means are connected with the population.*

9.4.2 Shipping stations

Only the station in Maassluis has a large contribution from the emission sources in category shipping. At location Maassluis, the source 'Ships sea' has a contribution about 44 % of the total emission. This source represents the emission from sea ships. The monitoring station is located in a quiet residential area in Maassluis not far from the 'Nieuwe Waterweg', the harbor entry of Rotterdam. Therefore the emission from maritime ships is large with respect to the other sources.

The annual mean concentration calculated by the model is equal to 42.2 $\mu g/m^3$, while the annual mean of the observations equals 51.6 $\mu g/m^3$. Therefore the Kalman filter causes a little increase in the calculated concentration.

The number U^{abs} has become equal to 18.4×10^6, which is a reduction of 3.0 % with respect to the situation without any measurements. The number U^{rel} is now equal to 421×10^3, a reduction of 5.4 %. These large reductions are caused by the large contribution from the sea ships to the total emission and by the large number of observations which are involved in the analysis step of the Kalman filter. The relative uncertainty is mostly decreased in the region with a large emission from the shipping sources, this is shown in Figure 9.6. This figure shows also the connection with the population, because the emission from shipping sources has a large influence on the exposure, the uncertainty is decreased for a lot of people.

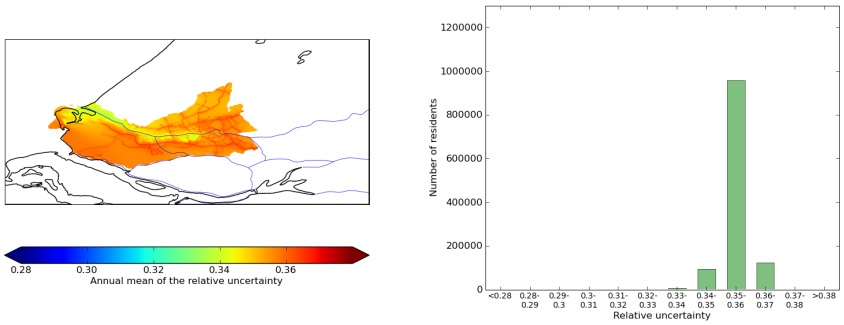


Figure 9.6: *Left panel: The annual mean of the relative uncertainty, with the measurements from Maassluis in the Kalman filter, in the right panel these annual means are connected with the population.*

9.4.3 Combined stations

There are several stations which have not one dominating emission source, these stations are Schiedam, Hoogvliet, Schiedamsevest and Vlaardingen. When one of these stations is added to the Kalman filter, the uncertainty of some different sources will be reduced.

Vlaardingen

One of the stations with no dominating source is Vlaardingen. The contribution from the maritime ships is 24 % but also the sources local traffic (20 %), background (16 %), highway traffic (14 %) and 'Rest' (14%) have significant contributions on the total emission. If the series of measurements from the monitoring station in Vlaardingen is added to the Kalman filter, the relative uncertainty of all these sources will decrease a little.

The annual mean of the observations at this location is equal to 57.2 $\mu g/m^3$, while the annual mean of the model simulation is equal to 54.1 $\mu g/m^3$. The observations are a little higher than the simulations. The number U^{abs} has become equal to 18.1×10^6, which is a reduction of 4.6 %, while the number U^{rel} becomes equal to 429×10^3, a reduction of 3.5 % with respect to the situation without measurements involved in the Kalman filter. Because there is no dominating source, the relative uncertainty decreases on almost all locations with the same rate. Only the uncertainty in the region around the 'Nieuwe Waterweg' is a little more reduced. Together with the connection with the population this is shown in Figure 9.7.

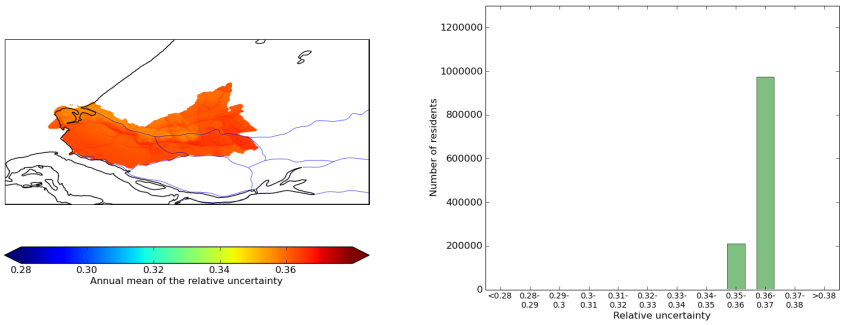


Figure 9.7: *Left panel: The annual mean of the relative uncertainty, with the measurements from Vlaardingen in the Kalman filter, in the right panel these annual means are connected with the population.*

9.4.4 Other stations and combinations of stations

For the other stations in the area, the same sources, highway traffic, maritime shipping and local traffic dominates. In Table 9.2, the reductions of the numbers U^{abs} and U^{rel} are shown for each station in the Kalman filter. Also some combinations on the nine monitoring stations are shown in this table.

The reduction of the relative uncertainty when the station Overschie is added is equal to 4.4 %, and for the station in Ridderkerk this reduction is 2.4 %. Combining these two reductions should lead to $0.956 \times 0.976 = 0.933$, a reduction of 6.7 %. The actual reduction when both stations Overschie and Ridderkerk are added to the Kalman filter is 6.0 %. This shows that the efficiency for each extra station becomes smaller.

In Figure 9.8, it is shown what happens if one series of measurements is added several times. For the station in Hoogvliet the reduction of number U^{rel} is equal to 4.4 %, if this series of measurements is added twice (two series of measurements with the same values on the same location), the reduction of the number U^{rel} is 5.3 %. If this series of measurements is added more times, the extra reduction of number U^{rel} becomes smaller. This gives the idea that an addition of stations with the same contributions from each source is ineffective.

Therefore an idea is that an extra monitoring station should be on a location with a domination source, which does not dominate on other monitoring stations.

If all the stations are involved in the Kalman filter the maximal reduction of the relative uncertainty should be $0.960 \times 0.956 \times 0.946 \times 0.956 \times 0.976 \times 0.973 \times 0.971 \times 0.965 \times 0.980 = 0.724$, a reduction of 27.6 %. The actual reduction is equal to 16.1 %.

Table 9.2: *Reductions of the number U^{abs} and the number U^{rel} by application of the Kalman filter on different monitoring stations*

Monitoring stations involved in the Kalman Filter	Dominating source(s) at the monitoring station	Number $U^{abs} \times 10^6$	Reduction with respect to the situation without Kalman filter	Number U^{rel} $\times 10^3$	Reduction with respect to the situation without Kalman filter
No Stations	—	19.0	—	445	—
Schiedam	Zone card (22 %) Ships sea (19 %) Background (15 %)	17.8	6.2 %	427	4.0 %
Hoogvliet	Zone card (21 %) Ships sea (21 %) Background (16 %)	18.2	4.1 %	426	4.4 %
Maassluis	Ships sea (44 %)	18.4	3.0 %	421	5.4 %
Overschie	Zone card (43 %)	*19.5*	*-3.2 %*	425	4.4 %
Ridderkerk	Zone card (85 %)	18.0	5.0 %	434	2.4 %
Bentinckplein (DCMR)	Roads nearby (35 %) CAR (24 %)	17.9	5.3 %	433	2.7 %
Schiedamsevest	Roads nearby (25 %) Background (16 %) Ships sea (14 %)	18.0	5.1 %	432	2.9 %
Vlaardingen	Ships sea (24 %) Background (16 %) Zone card (14 %) Rest (14 %) CAR (14 %)	18.1	4.6 %	429	3.5 %
Bentinckplein (RIVM)	Road nearby (35 %) CAR (24 %)	18.2	4.0 %	436	2.0 %
All stations	—	16.2	14.5 %	373	16.1 %
Overschie + Ridderkerk	—	18.8	0.5 %	418	6.0 %
Bentinckplein (2×)	—	17.8	6.1 %	431	3.2 %

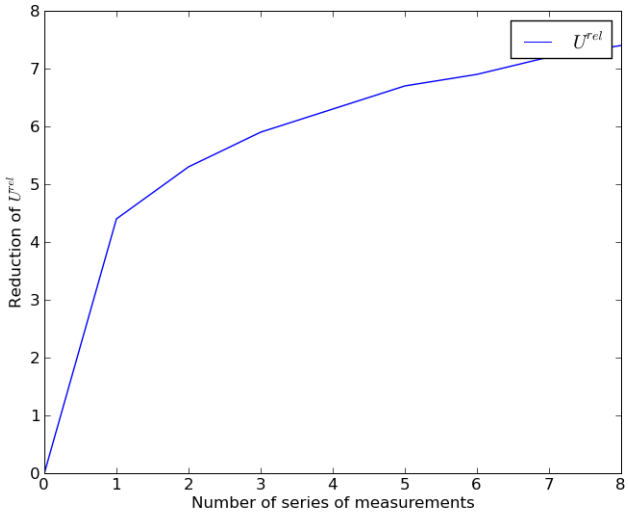


Figure 9.8: *Reduction of the number U^{rel} for several times the monitoring station in Hoogvliet involved in the Kalman filter.*

9.5 Influence of measurements on uncertainties

The results in Table 9.2 show that the relative uncertainty decreases if a series of measurements is added to the Kalman filter application. The absolute uncertainty could both decrease and increase when a series of measurements is added to the Kalman filter. This are results of Kalman filtering theory, which will be explained in this section.

9.5.1 Absolute uncertainty

The absolute uncertainty on grid point j at time k is defined as in Equation 9.2:

$$u_{j,k}^{abs} = \left(\sum_{i=1}^{88} \mu_{i,k} m_{i,j} e^{\gamma_{i,k}+p_{i,k}} - \sum_{i=1}^{88} \mu_{i,k} m_{i,j} e^{\gamma_{i,k}-p_{i,k}} \right) \tag{9.7}$$

Because of the minimal variance gain in the Kalman filter, the covariance matrix P^a will always be smaller in l_2-norm than the initial covariance matrix P^f from the model simulation:

$$\begin{aligned} \|P^a\|_2 &= \left\| (I - KH) P^f \right\|_2 \\ &= \left\| P^f - KHP^f \right\|_2 \\ &\leq \left\| P^f \right\|_2 + \left\| KHP^f \right\|_2 \\ &\leq \left\| P^f \right\|_2 \end{aligned} \tag{9.8}$$

On the main diagonal of P_k^a are the values for $p_{i,k}^2$.

When the observations at time step k are larger than the model simulations, the correction factors $\gamma_{i,k}$ will become positive. Therefore it is possible that the absolute uncertainty increases. For the monitoring station in Overschie the observations are much higher than the simulations, but only a small number of this large observations is thrown out by the screening criterion. The values of $p_{i,k}$ have decreased by the Kalman filter, but the correction factor $\gamma_{i,k}$ became large enough to increase the absolute uncertainty.

At location Ridderkerk the observations are much lower than the model simulations, therefore the correction factor $\gamma_{i,k}$ became negative, and the reduction of the absolute uncertainty will be strengthened.

When the difference between the observations and the simulations is very large, the total reduction of the absolute uncertainty could be inaccurate. In that case it is useful to improve the model or the representation of the measurements, such that the model fits better on the observations.

9.5.2 Relative uncertainty

The relative uncertainty in grid point j at time k is defined as in Equation 9.3:

$$u_{j,k}^{rel} = \frac{u_{j,k}^{abs}}{c_{j,k}^{KF}} \tag{9.9}$$

where $c_{j,k}^{KF}$ is the expected concentration on grid point j at time k from equation 9.4. Thus for the whole domain:

$$\begin{aligned} \underline{u}_k^{rel} &= \frac{\mu_{1,k}\underline{m}_1 e^{\gamma_{1,k}}\left(e^{p_{1,k}} - e^{-p_{1,k}}\right) + \cdots + \mu_{88,k}\underline{m}_{88} e^{\gamma_{88,k}}\left(e^{p_{88,k}} - e^{-p_{88,k}}\right)}{\mu_{1,k}\underline{m}_1 e^{\gamma_{1,k}} e^{0.5p_{1,k}^2} + \cdots + \mu_{88,k}\underline{m}_{88} e^{\gamma_{88,k}} e^{0.5p_{88,k}^2}} \\ \underline{u}_k^{rel} &= \frac{\sum_{i=1}^{88} \mu_{i,k}\underline{m}_i e^{\gamma_{i,k}}\left(e^{p_{i,k}} - e^{-p_{i,k}}\right)}{\sum_{i=1}^{88} \mu_{i,k}\underline{m}_i e^{\gamma_{i,k}} e^{0.5p_{i,k}^2}} \end{aligned} \tag{9.10}$$

dividing this by $\mu_{i,k}\underline{m}_i e^{\gamma_{i,k}}$ leads to:

$$\underline{u}_k^{rel} = \sum_{i=1}^{88} \frac{\left(e^{p_{i,k}} - e^{-p_{i,k}}\right)}{e^{0.5p_{i,k}^2} + \sum_{j=1, j\neq i}^{88} T_{i,j}} \tag{9.11}$$

where the tail $T_{i,j}$ is equal to:

$$T_{i,j} = \frac{\mu_j \underline{m}_j e^{\gamma_{j,k}} e^{0.5p_{i,k}^2}}{\mu_i \underline{m}_i e^{\gamma_{i,k}}} \tag{9.12}$$

The values for $T_{i,j}$ are all positive thus equation 9.11 becomes:

$$\underline{u}_k^{rel} \leq \sum_{i=1}^{88} \frac{e^{p_{i,k}} - e^{-p_{i,k}}}{e^{0.5p_{i,k}^2}} \tag{9.13}$$

In Figure 9.9, the formula $\frac{e^x - e^{-x}}{e^{0.5x^2}}$ is plotted with respect to x, this figure shows that the relative uncertainty will decrease if $p_{i,k}$ decreases and $p_{i,k} \leq 1.2$.

The Kalman filter is built with the minimal variance gain such that P^a decreases in l_2-norm, this is shown in Equation 9.8. The values for $p_{i,k}^2$ are on the main diagonal of P^a, so the values for $p_{i,k}^2$ will also decrease. Further the assumed value for $p_{i,k}$ from the model simulation is 0.19, as shown in Section 6.5. Thus the relative uncertainty will decrease if a series of measurements is added to the system.

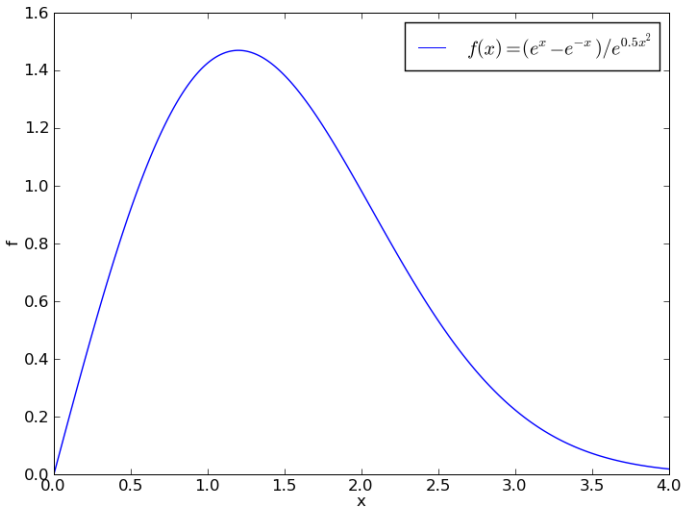


Figure 9.9: *The relative uncertainty will decrease if p decreases and p is below 1.2*

9.5.3 Influence of the measurements on the uncertainties

Independent of the values of the observations, the relative uncertainty will decrease if a series of observations is added to the system. The rate of this decrease is still unknown and possibly dependent of the values of the observations.

First it is shown in Table 9.2, that the reduction of the relative uncertainty is small for the station in Ridderkerk. At this station 38 % of the observations is screened, thus the Kalman filter did not have many possibilities to decrease the uncertainty. Therefore the first conclusion is that, if the difference between the model simulations and the observations is large, the screening criterion throws out a lot of observations and the relative uncertainty will not have a large reduction.

For the situation that the observations and the simulations does not differ a lot, the rate of decrease of the relative uncertainty is possibly dependent of the observations. As shown in Section 9.5.2, the relative uncertainty is dependent of the values for p. These values are calculated in the analysis step of the Kalman filter, the values for p are equal to the variances on the main diagonal of covariance matrix P. The matrix P is calculated as follows:

$$P = (I - KH)\, P^f \tag{9.14}$$

where K is the minimal variance gain:

$$K = P^f H^{\mathrm{T}} \left(H P^f H^{\mathrm{T}} + R \right)^{-1} \tag{9.15}$$

Because of the matrix R is dependent of the observations, the relative uncertainty depends also of the observations. The influence of this dependency is shown by a special application of the two stations at Bentinckplein. In this application only the observations are involved on the time steps that both of the stations have a valid observation. This leads to 6062 observations for both stations. The annual mean of the model simulations on this 6062 time steps is equal to 104.8, the annual mean of the observations made on the DCMR station is 92.3 and for the RIVM station the annual mean of the observations is 97.0 At the DCMR station the screening process throws out 289 observations, this is 4.8 %. At the RIVM station 312 observations are thrown out by the screening process, which is 5.1 %.

The numbers U^{rel} becomes equal to 436×10^3 for both the DCMR station and the RIVM station, the difference between these two numbers is a measure for the influence of the values of the observations. The difference between these numbers is significant equal to 0, which means that the relative uncertainty is nearly independent of the measurements. This conclusion only holds if the difference between the observations and the model simulations is small.

This application is also done without the screening process, which also leads to numbers $U^{rel} = 436 \times 10^3$ for both the DCMR station and the RIVM station. The difference between these two numbers is also significant equal to zero, thus the relative uncertainty is again nearly independent of the measurements.

This leads to the following conclusion: If the model simulations do not differ a lot from the observations, the rate of decrease of the relative uncertainty is nearly independent of the values of the observations. If the difference between the simulations and observations is large, the screening criterion throws out a lot of observations and

the rate of decrease is smaller. This is also an indication that the model must be improved to get a more accurate simulation.

9.6 Setting an optimal placement of monitoring stations

9.6.1 Reduce uncertainty of important sources

In this section the theory about adding virtual monitoring stations will be explained. The idea is to add a virtual monitoring station on a well-chosen location. This location is chosen such that the uncertainty of one of the important sources will be decreased. This will lead to a large decrease of the total uncertainty. In Section 9.2 and 9.4, some ideas are found for the placement of such virtual monitoring stations.

In Section 9.2, it is shown that the emission from highway traffic, sea ships and background have the largest contribution on the exposure of the population to the concentration NO_x. Therefore the virtual monitoring stations have to be located, such that the uncertainty of these three sources will decrease.

Some virtual monitoring stations are added to the system on locations with dominating emission from respectively highway traffic, background and maritime shipping. With this virtual monitoring stations, the influence on the uncertainty can be calculated.

If a virtual monitoring station will be added to the Kalman filter, it is important to have a deliberate choice for the simulated series of measurements. Section 9.5 shows that the absolute uncertainty depends on the measurements, also the relative uncertainty depends a little on the measurements. Therefore the series of measurements is simulated around the model simulation. This means that the difference between the observations and the simulations is small, thus the reduction in absolute uncertainty will be accurate. The relative uncertainty is nearly independent of the measurement, thus the reduction in this uncertainty will also be accurate.

Traffic station

The emission from highway traffic has the largest contribution to the exposure as shown in Section 9.2. A reduction in the uncertainty of this source leads to an efficient reduction of the total uncertainty. In the present system of monitoring stations, there are already two stations which cover the emission from highway traffic, these stations are located in Overschie and Ridderkerk. As shown in Section 9.4.1, the screening process throws out a lot of observations, especially for the station in Ridderkerk. This is an indication that the model is not accurate at those two stations, therefore the model could be improved to solve this problem and to reduce the uncertainty.

Another method to reduce the uncertainty of the emission from highway traffic is to add another station which cover the emission from this source. At the Harmsen bridge, on the junction of the main road A15 and local road N57, according to the model, the contribution from this source is 93 %. This is the largest contribution in the whole area. If a monitoring station is placed near this bridge the uncertainty of the emission from highway traffic will have a large reduction. Now a virtual monitoring system will be added to the system to have a look at the reduction of the uncertainty.

In the right panel of Figure 9.10, the relative uncertainty is shown for the whole area. This relative uncertainty is calculated as an average of 5 different series of virtual measurements for the virtual station near the Harmsen bridge. The figure shows the relative uncertainty for one of these five simulations, the other 4 have nearly the same pattern. In the left panel, the relative uncertainty is shown for the present situation with only the 9 present monitoring stations. It is obvious that the uncertainty is mainly decreased around the main roads.

The numbers U^{abs} and U^{rel}, calculated as the average of the 5 different simulations, become equal to respectively 15.7×10^3 and 359×10^3. This are reductions of 3.9 % with respect to the present situation with 9 monitoring stations in the area (row 11, all stations, in Table 9.2). The reductions with respect to the situation without any measurements in the Kalman filter (row 1, no stations, in Table 9.2) are 17.8 % for the absolute uncertainty and 19.4 % for the relative uncertainty.

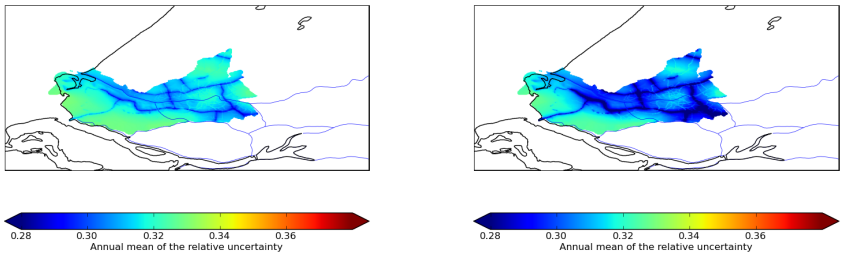


Figure 9.10: *Relative uncertainty for the whole domain. Left panel: The Kalman filter applied on the present stations. Right panel: The Kalman filter applied on the present stations plus an extra monitoring station near the Harmsen bridge.*

Background station

The emission from source background have a large influence on the exposure of the population to the concentration NO_x. Table 9.1, shows that the contribution of this source to the total exposure is equal to 21 %. Because none of the present monitoring stations is dominated by this source, a good idea will be to add an extra station on a location where the background dominates the total emission. At the zeedijk in Bernisse, south west of Rotterdam, the contribution from this source is about 55 % of the total emission, this is the largest contribution in the whole area. The rest of the emission at this location is mainly caused by the shipping sources (19 %) and by source rest (13 %).

In the right panel of Figure 9.11, the relative uncertainty is shown when an extra station is added in Bernisse. The left panel shows the relative uncertainty if only the present monitoring stations are involved in the Kalman filter. The region with the largest contribution of source 'Background' has the largest reduction of the uncertainty. These regions are mainly located south west of Rotterdam.

The numbers U^{abs} and U^{rel} are again calculated as an average of five different series of simulated measurements for the Zeedijk in Bernisse. The average of this numbers are $U^{abs} = 14.9 \times 10^6$ and $U^{rel} = 346 \times 10^3$, this are reductions of respectively 8.0 % and 7.2 %, with respect to the situation with only the present monitoring stations. The reductions with respect to the situation without measurements are respectively 21.3 % and 22.2 %.

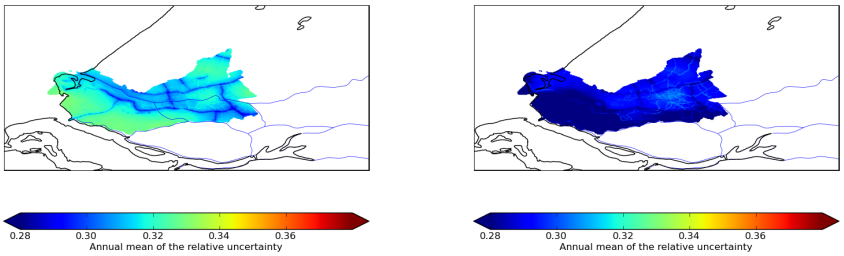


Figure 9.11: *Relative uncertainty for the whole domain. Left panel: The Kalman filter applied on the present stations. Right panel: The Kalman filter applied on the present stations plus an extra virtual monitoring station in Bernisse.*

Shipping station

Another source with a large contribution to the exposure is the source 'Ships sea'. This source covers the emission from maritime ships. In the present situation, only the station in Maassluis has a significant contribution from this source. Therefore an extra station on a location with a large contribution from the shipping sources would be a good choice to reduce the uncertainty. The location with the largest contribution is the Missouriweg nearby Hoek van Holland. On this location 57 % of the emission is from maritime ships, the rest of the emission is mainly caused by source background (14 %) and source rest (24 %).

In the right panel of Figure 9.12, the relative uncertainty is shown for the situation with an extra station at the Missouriweg nearby Hoek van Holland. In the left panel the relative uncertainty is shown for the situation without extra monitoring stations. The figure shows that the relative uncertainty is mostly decreased in the region around the 'Nieuwe Waterweg', the harbor entry of Rotterdam.

Also for this virtual monitoring station, the numbers U^{abs} and U^{rel} are calculated from the average of 5 runs of the Kalman filter, with each a different series of virtual measurements. These number became equal to $U^{abs} = 15.8 \times 10^6$ and $U^{rel} = 363 \times 10^3$, reductions of 3.0 % and 2.5 % with respect to the present situation and reductions of 15.8 % and 17.7 % with respect to the model simulation without any measurement in the Kalman filter.

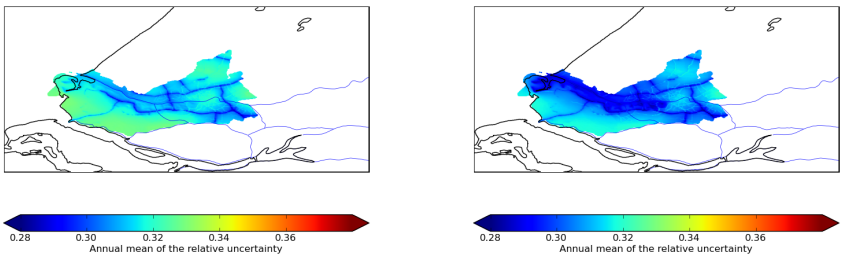


Figure 9.12: *Relative uncertainty for the whole domain. Left panel: The Kalman filter applied on the present stations: Right panel: The Kalman filter applied on the present stations plus an extra virtual monitoring station at the Missouriweg nearby Hoek van Holland*

9.6.2 Create an optimal setting of monitoring stations

The results from Section 9.6.1 shows that the total uncertainty can be reduced by the addition of monitoring stations which covers the important sources. Now, a method will be described to create an optimal setting of monitoring stations in a (new) area.

Suppose a city wants a monitoring system for the concentration NO_x, which can be connected with the Real Time URBIS model. The two main questions are: how many stations are necessary? and where should those stations be located to have an uncertainty which is smaller than the required uncertainty? The use of virtual monitoring stations can give an idea of the placement of the stations.

The process start with one monitoring station at a random place. If this (virtual) station is replaced to another place in the area, the total uncertainty will change. With an optimization algorithm it is possible to find the location for this station such that the total uncertainty is minimal. A possible method for this optimization is the gradient method, with this method the stations is moved in the direction, such that the uncertainty had the largest decrease. After the translation in that direction, the station is again translated in the direction with the largest reduction of the uncertainty.

If the station is located on a (local) optimal place the process, there criterion for acceptable uncertainty must be checked. If this criterion is not fullfilled, the optimization process can be restarted with two randomly placed monitoring stations. The gradient method can still be applied to find the (local) optimal combination of these two monitoring stations. At the end of each optimization process, it must be checked if the uncertainty is smaller than the required uncertainty. When the target is reached, the setting of the stations will then be the optimal placement of the monitoring stations.

This method also causes some troubles. The gradient method finds a local optimum, which is not necessary the same as the global optimum. This can be avoided by several runs of the process. If the same optimum is found several times, it is reasonable that this optimum is the global optimum. Another way to avoid the problems with local optima is to use a global optimization algorithm. More about global optimization algorithms is described in Weise [1988].

9.7 Conclusion

The relative uncertainty decreases if an extra monitoring station is added to the system. In Section 9.4 and 9.6, it is shown that the uncertainty decreases most if the uncertainty of an important source is decreased. An important source is defined as a source with a large contribution to the exposure of the population.

In Section 9.4, it is also shown that the efficiency of the reduction decreases if more stations with the same dominating sources are added to the system. This leads to the conclusion that an optimal setting of the monitoring stations is such that each of the important sources is covered by at least one station. If more stations cover the same source, the reduction will be less efficient. In Section 9.4.4, it is shown that more stations for the same source will result in a diminishing return of extra reduction of the uncertainty.

For the Rijnmond area covered in this study, the important sources are highway traffic, maritime shipping and background. The present stations covers mainly the

sources highway traffic and maritime shipping. The other important source background is not covered by one of the monitoring stations. In Section 9.6.1, it is shown that an extra station which covers the source background will lead to a significant reduction of uncertainty. Also an extra location which cover the emission from maritime ships will lead to a useful reduction.

Another station which cover the emission from highway traffic will also have an efficient reduction of the uncertainty, this is because many observations are screened in the two present traffic stations Overschie and Ridderkerk. More stations for the highway traffic are not expected to be very efficient, also stations which cover the less important sources will not have a very efficient reduction of the relative uncertainty.

The absolute uncertainty will have an accurate reduction if the model simulations does not differ too much from the observations. If the model simulations are lower than the observations, the absolute uncertainty will be overestimated. Otherwise the absolute uncertainty will be underestimated if the model simulation are larger than the observations. If this situation occurs it will be more efficient to change the model such that the model fits better on the observations. Therefore a critical view on the differences between the observations and the simulations is necessary to say something about the reduction in the absolute uncertainty.

10 Other time resolutions

The exposure of the population is mostly determined with annual mean concentrations. In the previous chapters the annual mean of the uncertainty is determined by the average of the hourly mean uncertainties. It could be useful to look at some different time scales. It is trivial that the uncertainty is smaller if the time scale is larger, the hourly mean concentrations fluctuates a lot and will have some extreme values. If the time scale is larger, the extremes will be averaged thus the uncertainty is smaller. The idea is that the annual mean of the uncertainty is smaller by this calculations, therefore the limit values for the annual mean concentrations stated in Appendix 2 of 'Wet Milieubeheer' [Cramer, 2007] can be checked more accurate. The limit values for the hourly mean concentrations can not be checked more accurate with measurements of daily, weekly or monthly mean concentrations.

In Section 10.1, the annual mean of the uncertainty will be determined by the average of daily mean uncertainties, in Section 10.2 with weekly mean uncertainties and in Section 10.3 with monthly mean uncertainties.

If the time scale is larger, the number of available observations in a year will be smaller. With 9 monitoring stations in the area, there are a maximum of 9×8760 = 78840 hourly mean concentrations available. For the daily mean concentrations, a maximum of 9×365 = 3285 observations are available. For the weekly mean, a maximum of 9×52 = 468 observations are available. For the monthly mean concentrations, the number of observations became equal to 9 × 12 = 108. In the previous chapter, it is already shown that less number of observations results in smaller reduction of the uncertainty.

10.1 Daily mean concentrations

If the Kalman filter is applied for the daily mean concentrations, the maximum available number of observations is still 3285. Therefore the application of the Kalman filter will still be useful to calculate the annual mean of the uncertainty. The annual mean of the uncertainty is now determined by the average of daily mean uncertainties. In Section 10.1.2, the annual mean of the uncertainties are determined. First in Section 10.1.1, the Kalman filter equations for this application are constructed.

10.1.1 Dynamical system and Kalman filter form

The Kalman filter equations have to be changed such that those equations fits the daily mean concentrations. The dynamical system for the logarithm of the concentration NO_x at day k becomes the following:

$$\ln\left(\underline{c}_k\right) = \ln\left(\sum_{i=1}^{88} \bar{\mu}_{i,k}\underline{m}_i e^{\gamma_{i,k}}\right) \tag{10.1}$$

The correction factor $\gamma_{i,k}$ is the correction factor on the standard concentration field i at day k, $\underline{m}_i$ is the standard concentration field i. The total number of standard concentration fields is still equal to 88. The value $\bar{\mu}_{i,k}$ is the average of all weight factors μ_{i,j_k} at day k:

$$\bar{\mu}_{i,k} = \frac{1}{24} \sum_{j=1}^{24} \mu_{i,j_k} \tag{10.2}$$

in here the parameter μ_{i,j_k} is calculated for each hour j of day k. This parameter is calculated with the wind speed, the wind direction, the temperature, the hour of the day, the day of the week and the month of the year.

The dynamical system 10.1 is again non-linear, therefore a linearization is made:

$$\ln\left(\sum_{i=1}^{88} \bar{\mu}_{i,k} \underline{m}_i e^{\gamma_{i,k}}\right) = \ln\left(\underline{c}_k^m\right) + \left[\frac{\bar{\mu}_{j,k}\underline{m}_j}{\underline{c}_k^m}\right]_{j=1}^{j=88} \underline{\gamma}_k + \mathcal{O}\left(\underline{\gamma}_k \cdot \underline{\gamma}_k\right) \tag{10.3}$$

the model concentration (also daily mean) at day k is denoted by $\underline{c}_k^m$. The dynamical system then becomes the following:

$$\ln\left(\underline{c}_k\right) = \ln\left(\underline{c}_k^m\right) + \left[\frac{\bar{\mu}_{j,k}\underline{m}_j}{\underline{c}_k^m}\right]_{j=1}^{j=88} \underline{\gamma}_k \tag{10.4}$$

$$\underline{\gamma}_{k+1} = A\underline{\gamma}_k + \underline{\omega}_k \qquad \underline{\omega}_k \sim N\left(0, Q\right) \tag{10.5}$$

where matrix A contains the temporal correlation parameters. The covariance matrix Q is a diagonal matrix, with the model uncertainties, both A and Q are assumed to be independent of time.

To implement this dynamical system in the Kalman filter, it has to be written in Kalman filter form:

$$\underline{\gamma}_{k+1} = A\underline{\gamma}_k + \underline{\omega}_k \qquad \underline{\omega}_k \sim N\left(0, Q\right) \tag{10.6}$$

$$\underline{\tilde{y}}_k = \tilde{H}\underline{\gamma}_k + \underline{\nu}_k \qquad \underline{\nu}_k \sim N\left(0, R_k\right) \tag{10.7}$$

where R_k represents the uncertainty of the measurements. Furthermore $\underline{\tilde{y}}_k$ and $\tilde{H}$ are defined as follows:

$$\underline{\tilde{y}}_k = \ln\left(\underline{y}_k\right) - H\ln\left(\underline{c}_k^m\right) \tag{10.8}$$

$$\tilde{H} = H\left[\frac{\bar{\mu}_{j,k}\underline{m}_j}{\underline{c}_k^m}\right]_{j=1}^{j=88} \tag{10.9}$$

where $\underline{y}_k$ are the daily mean observations on day k.

10.1.2 Uncertainties after Kalman filtering

When the Kalman filter is applied on the daily mean concentrations with the Kalman filter equations from Section 10.1.1, it appears that the uncertainties are smaller. The extreme hourly mean values are averaged out and the daily mean concentrations have less uncertainty.

Input parameters for the Kalman filter

To obtain the temporal correlation parameters for the traffic sources and for the shipping sources, the temporal correlations of the daily mean concentrations are calculated for the monitoring stations Overschie and Ridderkerk (traffic) and for Maassluis (shipping). This is done similar to the method in Section 6.4. In Figure 10.1, the temporal correlations are shown, in the left panel for monitoring stations Overschie and Ridderkerk and in the right panel for Maassluis. This figures shows that the daily mean concentrations are nearly uncorrelated, the parameters τ_{tr} and τ_{sh} seems to be close to 2. Calculating the correlation on the other monitoring stations leads to parameters τ_{bg}, τ_{in} and τ_{re} all equal to 2.

The input parameters r_{frac} and σ are still unknown, therefore the Kalman filter will be applied for different values of these parameters and for the temporal correlation parameters close to 2. Similar to Section 6.5, the values for the $Mean$, Std and $RMSE$ have to be optimized to find the optimal values for all input parameters.

Also for the daily mean concentrations, the differences between the observations and the simulations are very large at the monitoring stations Overschie and Ridderkerk. Therefore the optimal values for $r_{\text{frac}}, \sigma, \tau_{bg}, \tau_{tr}, \tau_{in}, \tau_{sh}$ and τ_{re} are calculated with only the observations from the other 7 monitoring stations. This leads to the following series of parameters: $r_{\text{frac}} = 0.26$, $\sigma = 0.13$ and $\tau_{bg} = \tau_{tr} = \tau_{sh} = \tau_{in} = \tau_{re} = 2$.

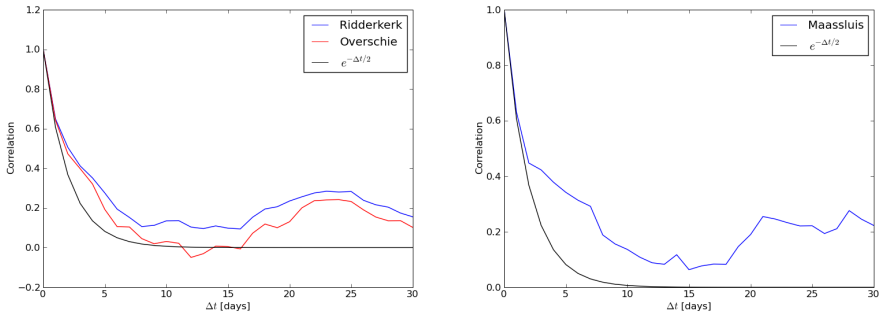


Figure 10.1: *Temporal correlation for the daily mean concentrations at monitoring stations Overschie and Ridderkerk in the left panel and for Maassluis in the right panel*

Uncertainty of the model

For the model uncertainty the parameter σ is important. The optimal value for this parameter was determined to be equal to 0.15. With this parameter the absolute and relative uncertainty of the model simulation could be calculated for each day on each point in the domain. The annual mean of all those daily mean uncertainties are shown in Figure 10.2. In the left panel the absolute uncertainty is shown, in the right panel the relative uncertainty is shown.

As a result of the smaller uncertainty of the daily mean concentrations, both the absolute and the relative uncertainty are smaller than for the applications with hourly means as in Figure 9.1. The absolute uncertainty is still very large on the main roads and in the industrial region around Pernis, the relative uncertainty is (as assumed) constant on the whole area.

Also for this application it is possible to calculate the number U^{abs} and U^{rel}, for the model simulation. These values are equal to $U^{abs} = 13.1 \times 10^6$ and $U^{rel} = 306 \times 10^3$.

These numbers are about 31 % smaller than these numbers for the model simulations with hourly mean concentrations from Section 9.3.

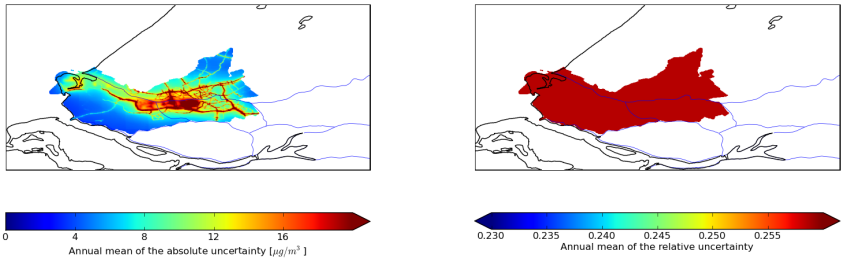


Figure 10.2: *Annual mean of the absolute and the relative model uncertainty calculated as an average of daily mean uncertainties.*

Uncertainty after Kalman filter application

In Figure 10.3, the absolute and relative uncertainty are shown for the whole area with all 9 monitoring stations involved in the Kalman filter. The absolute uncertainty is somewhat decreased on the main roads and the industrial region around Pernis.

The relative uncertainty is mostly decreased on the main roads and in the region around the 'Nieuwe Waterweg'. This is due to the monitoring stations which covers mostly the emission from traffic and shipping sources. The number U^{abs} become equal to 12.0×10^6 and the number U^{rel} became equal to 280×10^3, these are reductions of 8.3 % and 8.6 % with respect to the model simulations.

These reductions are useful, but smaller than the same reductions for the application with hourly mean concentrations. The row 'All stations' from Table 9.2, shows that these reductions were equal to 14.5 % and 16.1 %. The application with daily mean concentrations has less observations, thus less possibilities to reduce the uncertainties.

In Table 10.1, the reductions of the numbers U^{abs} and U^{rel} are given for all the different monitoring stations involved in the Kalman filter application for the daily mean concentrations. Comparing this table with Table 9.2 shows that the uncertainties are smaller due to the smaller uncertainty for the daily mean concentrations. The reductions with respect to the model simulations are also smaller, this is due to the smaller number of observations which are involved in the Kalman filter.

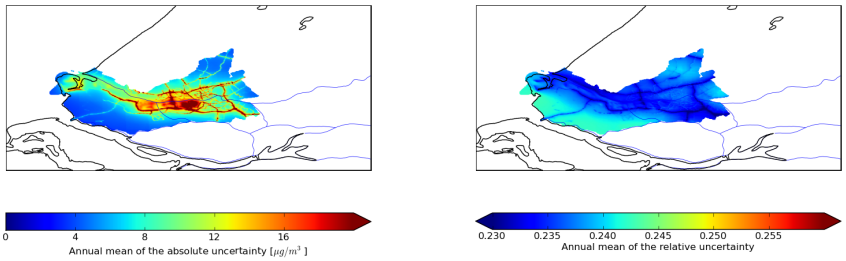


Figure 10.3: *Annual mean of the absolute and the relative uncertainty calculated as an average of daily mean uncertainties. All nine monitoring stations are involved in the Kalman filter application.*

Table 10.1: *Reductions of the numbers U^{abs} and U^{rel} for the daily mean concentrations in the Kalman filter application.*

Monitoring stations involved in the Kalman Filter	Dominating source(s) at the monitoring station	Number U^{abs} $\times 10^6$	Reduction with respect to the situation without Kalman filter	Number U^{rel} $\times 10^3$	Reduction with respect to the situation without Kalman filter
No Stations	—	13.1	—	306	—
Schiedam	Zone card (22 %) Ships sea (19 %) Background (15 %)	12.7	3.1 %	301	1.8 %
Hoogvliet	Zone card (21 %) Ships sea (21 %) Background (16 %)	12.9	1.5 %	300	2.0 %
Maassluis	Ships sea (44 %)	12.9	1.6 %	298	2.8 %
Overschie	Zone card (43 %)	13.3	-1.4 %	301	1.6 %
Ridderkerk	Zone card (85 %)	12.8	2.4 %	304	0.8 %
Bentinckplein (DCMR)	Roads nearby (35 %) CAR (24 %)	12.7	2.6 %	303	0.9 %
Schiedamsevest	Roads nearby (25 %) Background (16 %) Ships sea (14 %)	12.8	2.4 %	302	1.3 %
Vlaardingen	Ships sea (24 %) Background (16 %) Zone card (14 %) Rest (14 %) CAR (14 %)	12.8	1.8 %	301	1.8 %
Bentinckplein (RIVM)	Road nearby (35 %) CAR (24 %)	12.8	2.2 %	303	1.1 %
All stations	—	12.0	8.3 %	280	8.6 %
Overschie + Ridderkerk	—	12.9	1.0 %	299	2.3 %
Bentinckplein (2×)	—	12.6	3.3 %	302	1.5 %

10.2 Weekly mean concentrations

If the measurements and the model simulations covers the weekly mean concentrations, it is trivial that the uncertainty is smaller than the daily mean concentrations.

The Kalman filter equations are the same as the equations in Section 10.1.1. Of course the averages are now taken over a whole week instead of a day. Analysis of the values for $Mean$, Std and $RMSE$ as in Section 10.1.2 leads to input parameters $r_{\text{frac}} = 0.19$, $\sigma = 0.105$ and $\tau_{bg} = \tau_{tr} = \tau_{sh} = \tau_{in} = \tau_{re} = 2$.

If the Kalman filter is applied with this input parameter, the number U^{abs} becomes equal to 10.5×10^6 for the situation without any observations involved. The number U^{rel} becomes equal to 248×10^3, this are reductions about 44 % with respect to the application with hourly mean concentrations. The absolute and the relative uncertainty are shown for the whole domain in Figure 10.4. The uncertainties are smaller but they have the same patterns as the uncertainties calculated with the daily mean or the hourly mean concentrations.

A disadvantage of this application is that the number of observations which is available for the Kalman filter is only 9×52 = 468. Therefore the Kalman filter cannot reduce this uncertainty very much. In Figure 10.5, the relative and the absolute uncertainty are shown for all the measurements made on the 9 monitoring stations involved in the Kalman filter. The numbers U^{abs} and U^{rel} become equal to 0.98×10^6 and 233×10^3. This are reductions of 5.4 % and 6.1 %, these reductions are smaller than the reductions in the applications with hourly mean or daily mean concentrations, but they still have nearly the same pattern. Finally in Table 10.2 all reductions are shown for each different monitoring station.

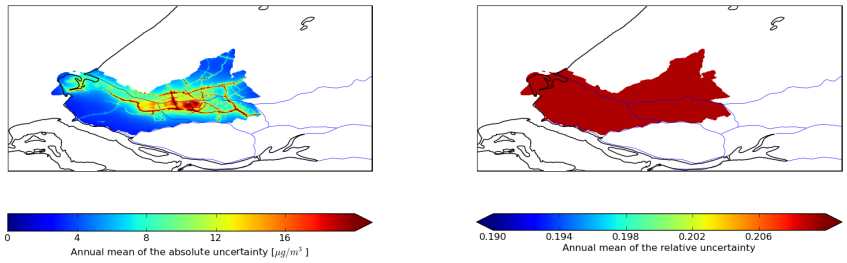


Figure 10.4: *Annual mean of the absolute and the relative model uncertainty calculated as an average of weekly mean uncertainties.*

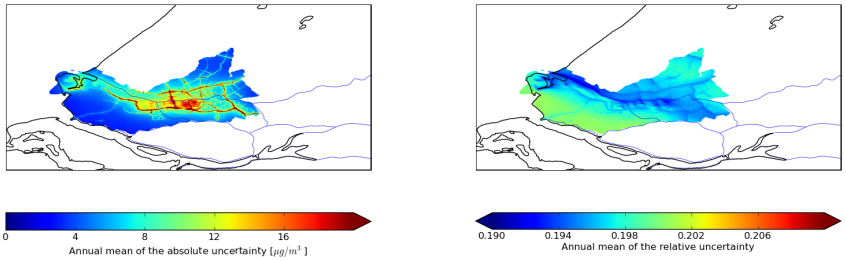


Figure 10.5: *Annual mean of the absolute and the relative uncertainty calculated as an average of weekly mean uncertainties. All nine monitoring stations are involved in the Kalman filter application.*

Table 10.2: *Reductions of the numbers U^{abs} and U^{rel} for the weekly mean concentrations in the Kalman filter application.*

Monitoring stations involved in the Kalman Filter	Dominating source(s) at the monitoring station	Number U^{abs} $\times 10^6$	Reduction with respect to the situation without Kalman filter	Number U^{rel} $\times 10^3$	Reduction with respect to the situation without Kalman filter
No Stations	—	10.5	—	248	—
Schiedam	Zone card (22 %) Ships sea (19 %) Background (15 %)	10.1	2.7 %	245	1.4 %
Hoogvliet	Zone card (21 %) Ships sea (21 %) Background (16 %)	10.4	0.9 %	245	1.1 %
Maassluis	Ships sea (44 %)	10.4	0.3 %	243	2.2 %
Overschie	Zone card (43 %)	10.7	-2.1 %	245	1.0 %
Ridderkerk	Zone card (85 %)	10.3	1.4 %	247	0.3 %
Bentinckplein (DCMR)	Roads nearby (35 %) CAR (24 %)	10.2	2.3 %	246	0.6 %
Schiedamsevest	Roads nearby (25 %) Background (16 %) Ships sea (14 %)	10.1	3.1 %	245	1.2 %
Vlaardingen	Ships sea (24 %) Background (16 %) Zone card (14 %) Rest (14 %) CAR (14 %)	10.4	0.8 %	244	1.5 %
Bentinckplein (RIVM)	Road nearby (35 %) CAR (24 %)	10.3	1.2 %	246	0.9 %
All stations	—	0.98	5.4 %	233	6.1 %
Overschie + Ridderkerk	—	10.5	-0.5 %	245	1.3 %
Bentinckplein (2×)	—	10.3	1.8 %	245	1.1 %

10.3 Monthly mean concentrations

When the time resolution is changed into monthly mean concentrations, the uncertainty becomes again smaller than with weekly or daily mean concentrations. A disadvantage of monthly mean concentrations is that the maximum number of measurements in a year is equal to $9 \times 12 = 108$. Therefore the Kalman filter do not have much possibilities to reduce the uncertainty. Analysis of the values for $RMSE$, $Mean$ and Std, leads to input parameters $\tau_{bg} = \tau_{tr} = \tau_{in} = \tau_{sh} = \tau_{re} = 2$, $\sigma = 0.07$ and $r_{\text{frac}} = 0.165$.

If the Kalman filter is applied with this input parameters on the monthly mean concentrations, then the number U^{abs} becomes equal to 6.9×10^6 and U^{rel} becomes equal to 166×10^3, this holds when no observations are involved in the Kalman filter.

If all the observations from the 9 monitoring stations are involved the following uncertainties are found: $U^{abs} = 6.7 \times 10^3$ and $U^{rel} = 160 \times 10^3$. This equals with reductions of respectively 2.9 % and 3.1 %, again smaller reductions than in the application with hourly, daily or weekly mean concentrations. In Figures 10.6 and 10.7, the absolute and relative uncertainties are shown for both the situation without any

measurements and the situation with all measurements involved. In Table 10.3, all the reductions of the absolute and relative uncertainty are shown.

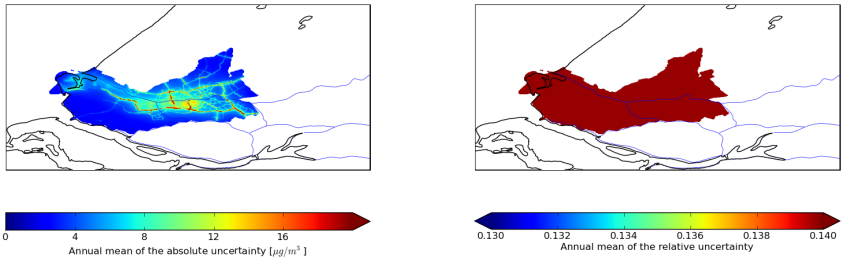


Figure 10.6: *Annual mean of the absolute and the relative model uncertainty calculated as an average of monthly mean uncertainties.*

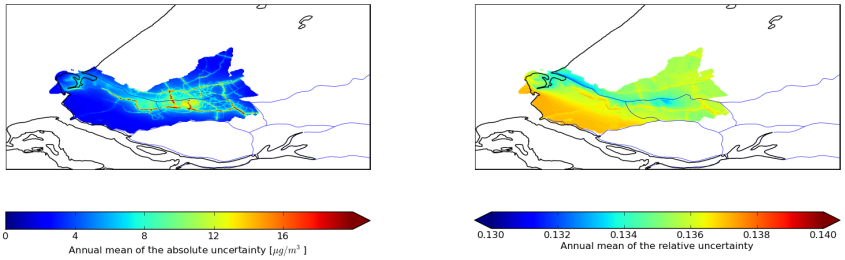


Figure 10.7: *Annual mean of the absolute and the relative uncertainty calculated as an average of monthly mean uncertainties. All nine monitoring stations are involved in the Kalman filter application.*

Table 10.3: *Reductions of the numbers U^{abs} and U^{rel} for the monthly mean concentrations in the Kalman filter application.*

Monitoring stations involved in the Kalman Filter	Dominating source(s) at the monitoring station	Number U^{abs} $\times 10^6$	Reduction with respect to the situation without Kalman filter	Number U^{rel} $\times 10^3$	Reduction with respect to the situation without Kalman filter
No Stations	—	6.9	—	166	—
Schiedam	Zone card (22 %) Ships sea (19 %) Background (15 %)	6.8	1.9 %	164	0.8 %
Hoogvliet	Zone card (21 %) Ships sea (21 %) Background (16 %)	6.9	1.2 %	165	0.5 %
Maassluis	Ships sea (44 %)	7.0	-0.9 %	163	1.3 %
Overschie	Zone card (43 %)	7.0	-1.6 %	165	0.2 %
Ridderkerk	Zone card (85 %)	6.9	0.2 %	166	0.0 %
Bentinckplein (DCMR)	Roads nearby (35 %) CAR (24 %)	6.8	2.2 %	165	0.4 %
Schiedamsevest	Roads nearby (25 %) Background (16 %) Ships sea (14 %)	6.8	2.2 %	165	0.6 %
Vlaardingen	Ships sea (24 %) Background (16 %) Zone card (14 %) Rest (14 %) CAR (14 %)	6.9	0.3 %	164	0.9 %
Bentinckplein (RIVM)	Road nearby (35 %) CAR (24 %)	6.9	0.5 %	165	0.5 %
All stations	—	6.7	2.9 %	161	3.1 %
Overschie + Ridderkerk	—	7.0	-1.6 %	165	0.2 %
Bentinckplein (2×)	—	6.8	2.1 %	165	0.7 %

10.4 Combining various time resolutions

In the previous sections, it is shown that the efficiency of the Kalman filter application reduces if the time resolution becomes larger. On the other hand, the automatic monitoring system to create hourly mean concentrations is relatively expensive. Therefore a cheap alternative is to extend the present system with 9 monitoring stations with hourly mean concentrations with a system of monitoring stations with monthly mean concentrations. To imply this in the Kalman filter application, the state equation and also the Kalman filter equations must be changed.

10.4.1 Kalman filter equations for combined time scaled

In the situation with hourly mean and monthly mean observations, the state equation must contain all the hourly mean concentrations from the past month. At the end of the month all concentrations from the past month will be corrected with the information from the monthly mean observations. The state equation becomes the following:

$$\underline{\tilde{c}}_k = \begin{bmatrix} \underline{c}_k \\ \underline{c}_{k-1} \\ \vdots \\ \underline{c}_{k-719} \end{bmatrix} \tag{10.10}$$

The vectors $\underline{c}_k$ contains two different parts, $\underline{c}_{h,k}$ and $\underline{c}_{m,k}$. The vector $\underline{c}_{h,k}$ corresponds with the hourly mean concentrations on time k on locations with hourly mean measurements. The vector $\underline{c}_{m,k}$ corresponds with the hourly mean concentrations on locations with monthly mean measurements. In the present situation the vector with hourly mean concentrations is of length 9. The length of vector $\underline{c}_{m,k}$ depends on the number of (new) stations which covers monthly mean concentrations.

$$\underline{c}_k = \begin{bmatrix} \underline{c}_{h,k} = \sum_{i=1}^{88} \mu_{i,k} \underline{m}_{h,i} e^{\gamma_{i,k}} \\ \underline{c}_{m,k} = \sum_{i=1}^{88} \mu_{i,k} \underline{m}_{m,i} e^{\gamma_{i,k}} \end{bmatrix} \tag{10.11}$$

where $\underline{m}_{h,i}$ and $\underline{m}_{m,i}$ are the standard concentrations of source i on the locations with respectively hourly and monthly mean measurements.

For the hourly mean concentration on time k, the same state equation as in Equation 5.2 still holds:

$$\underline{c}_k = \sum_{i=1}^{88} \mu_{i,k} \underline{m}_i e^{\gamma_{i,k}} \tag{10.12}$$

This leads to a total state $\underline{\tilde{c}}_k$ with state equation:

$$\underline{\tilde{c}}_k = \begin{bmatrix} \sum_{i=1}^{88} \mu_{i,k} \underline{m}_i e^{\gamma_{i,k}} \\ \vdots \\ \sum_{i=1}^{88} \mu_{i,k-719} \underline{m}_i e^{\gamma_{i,k-719}} \end{bmatrix} \tag{10.13}$$

The total vector with unknowns is:

$$\underline{\tilde{\gamma}} = \begin{bmatrix} \underline{\gamma}_k \\ \vdots \\ \underline{\gamma}_{k-719} \end{bmatrix} \tag{10.14}$$

Further there have to be a vector with observations:

$$\underline{y}_k = \begin{bmatrix} \underline{y}_{h,k} \\ \underline{y}_{m,k} \end{bmatrix} \tag{10.15}$$

in here $\underline{y}_{h,k}$is the hourly mean concentration on time k. The vector $\underline{y}_{m,k}$ compares with the monthly mean concentration from time step $k-719$ to time step k (the mean concentration over the past 30 days).

These observations have to be connected with the concentrations $\underline{c}_k$. Because the observations have a log-normal distribution, the following equation holds:

$$\ln\left(\begin{bmatrix} \underline{y}_{h,k} \\ \underline{y}_{m,k} \end{bmatrix}\right) = \ln\left(\begin{bmatrix} \mathrm{I}_h & \emptyset & \cdots & \emptyset & \emptyset \\ \emptyset & {}^1\!/_{720}\mathrm{I}_m & \cdots & \emptyset & {}^1\!/_{720}\mathrm{I}_m \end{bmatrix} \begin{bmatrix} \underline{c}_{h,k} \\ \underline{c}_{m,k} \\ \vdots \\ \underline{c}_{h,k-719} \\ \underline{c}_{m,k-719} \end{bmatrix}\right) + \begin{bmatrix} \underline{\nu}_{h,k} \\ \underline{\nu}_{m,k} \end{bmatrix} \tag{10.16}$$

in here I_h is the identity matrix with size as large as the number of stations with hourly mean measurements, I_m is the identity matrix with size as large as the number of stations with monthly mean measurements.

These equations are again not linear in the variable $\underline{\tilde{\gamma}}$, therefore a linearization is made around $\underline{\tilde{\gamma}} = 0$. This leads to the following system of equations:

$$\begin{bmatrix} \underline{\tilde{y}}_{h,k} \\ \underline{\tilde{y}}_{m,k} \end{bmatrix} = \tilde{H}\underline{\tilde{\gamma}} + \begin{bmatrix} \underline{\nu}_{h,k} \\ \underline{\nu}_{m,k} \end{bmatrix} \tag{10.17}$$

with:

$$\underline{\tilde{y}}_{h,k} = \ln\left(\underline{y}_{h,k}\right) - \ln\left(\underline{c}^m_{h,k}\right) \tag{10.18}$$

$$\underline{\tilde{y}}_{m,k} = \ln\left(\underline{y}_{m,k}\right) - \ln\left(\frac{1}{720}\left(\underline{c}^m_{m,k} + \cdots + \underline{c}^m_{m,k-719}\right)\right) \tag{10.19}$$

$$\tilde{H} = \begin{bmatrix} \left[\frac{\mu_{i,k}\underline{m}_{h,i}}{\underline{c}^m_{h,k}}\right]_{i=1}^{i=88} & \emptyset \cdots & \emptyset \\ \left[\frac{\mu_{i,k}\underline{m}_{m,i}}{\frac{1}{720}\left(\underline{c}^m_{m,k}+\cdots+\underline{c}^m_{m,k-719}\right)}\right]_{i=1}^{i=88} & \cdots & \left[\frac{\mu_{i,k-719}\underline{m}_{m,i}}{\frac{1}{720}\left(\underline{c}^m_{m,k}+\cdots+\underline{c}^m_{m,k-719}\right)}\right]_{i=1}^{i=88} \end{bmatrix}$$

Finally the temporal correlation matrix A between the vectors $\underline{\tilde{\gamma}}_{k+1}$ and $\underline{\tilde{\gamma}}_k$ and the matrix Q, corresponding with the uncertainty of the model, must be determined. The matrix A will have the following form:

$$A = \begin{bmatrix} A_1 & & & \\ \mathrm{I} & \emptyset & & \\ & \ddots & \ddots & \\ & & \mathrm{I} & \emptyset \end{bmatrix} \tag{10.20}$$

where the matrix A_1, size 88×88, corresponds with the temporal correlation matrix between vectors $\underline{\gamma}_{k+1}$ and $\underline{\gamma}_k$, this must be the same matrix as in Section 6.2. The identity matrices have also size 88×88, these matrices shifts the vectors $\underline{\gamma}_{k-1}$ through the large vector $\underline{\tilde{\gamma}}$.

The matrix Q corresponds with the uncertainty of the state vector at time k thus this matrix is:

$$Q = \begin{bmatrix} Q_1 & & & \\ & \emptyset & & \\ & & \ddots & \\ & & & \emptyset \end{bmatrix} \tag{10.21}$$

with Q_1, size 88×88, as in Section 6.2.

10.4.2 Results of the Kalman filter with different time scales

If the Kalman filter is applied with the Kalman filter equations as described in the Section 10.4.1, two important results occurs. At the end of each month the uncertainty of all correction factors will decrease, because of the presence of the monthly mean observations. Further, at each time step the uncertainty of the correction factors from the 720 time steps before are decreased. The reason for this reduction is the structure of the covariance matrix P. The covariance between the correction factors on time k and the correction factors on time $k-1, k-2 \ldots k-719$ are not equal to 0. Therefore the minimum variance gain minimizes the variances of the correction factors of the previous 719 time steps.

The rate of reduction of the uncertainty from the previous time steps is therefore dependent on the temporal correlation. If the temporal correlation is large, the Kalman filter will cause a large reduction of the uncertainty from the previous time steps. The idea behind this reductions is that a small uncertainty of a correction factor on time k must lead to a small uncertainty of the same correction factor on the next time step.

The results of this application is very difficult to determine because the state vector is a vector of length $88 \times 720 = 63360$. Therefore the matrices A,Q and P are of size 63360×63360, the present computer systems can not (yet) handle this large matrices.

Therefore the application is not done with hourly mean concentrations but with the average concentration over 3 hours. In that case the state vector has length $88 \times 3 = 264$, which can be handled by the present computers. The patterns of the reductions are again the same as for the situation with only hourly mean concentrations. Thus the reduction of the uncertainty by adding stations with monthly mean concentrations is a reasonable idea.

From now on it is again possible to create an optimization algorithm such that the total uncertainty connected with the population will be minimized. This algorithm must build in the same way as in Section 9.6.2.

11 Structural inaccuracies of the Real Time URBIS model

11.1 Correction factors per standard concentration field

In the first part of this report, the Kalman filter is applied to the Real Time URBIS model. The result of this application is a correction factor for each hour for each standard concentration field. It is now possible to calculate an annual mean of these correction factors. The annual mean of the correction factors for standard concentration field i is given by the following expression:

$$\frac{1}{8760}\sum_{k=1}^{8760} e^{\gamma_{i,k}} \tag{11.1}$$

If the annual mean of a certain correction factor is significantly different from 1, it is possible that the corresponding standard concentration field have some structural inaccuracies.

In Table 11.1, the annual means of all correction factors are given for each standard concentration field. The results in this table shows that the correction factors for the source 'Background' are little larger than 1, while the correction factors for the traffic sources are little smaller than 1. This could give an indication that the emissions from traffic and the background concentration should be better estimated by the model.

Another interesting fact is that the correction factors for the wind directions east and south are mostly larger than the correction factors for the wind direction north and west. This can also be caused by some inaccuracies in the model. If the wind is from the east or the south, there is mostly not much turbulence in the air (stable weather). This little turbulence causes less dispersion of the air pollution, thus larger concentrations NO_x. In the model, there is no distinction between stable and unstable weather. The larger correction factors for the wind directions east and south indicates that the model should make a difference between stable and unstable weather.

Table 11.1: *Mean of the correction factors for each standard concentration field.*

Wind speed	1.5 m/s				5.5 m/s			
Source\Wind Direction	N	E	S	W	N	E	S	W
Abroad (*)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Background	1.01	1.03	1.04	1.01	0.99	1.02	1.07	1.01
Zone card	0.97	0.99	1.00	0.98	0.97	0.99	0.98	0.96
CAR	0.98	1.00	1.00	0.99	0.95	0.98	0.96	0.94
Roads nearby	0.95	1.02	0.98	0.96	0.94	0.99	0.94	0.96
Roads far	0.99	1.02	1.01	1.00	1.00	1.00	1.00	1.00
Industry	1.00	1.01	1.01	1.00	1.00	1.01	1.01	1.00
Domestic	1.00	1.02	1.01	1.00	1.00	1.00	1.00	1.00
Ships inland (*)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Ships sea	0.98	1.04	1.04	0.98	0.99	1.01	1.01	0.99
Rest	0.99	1.02	1.01	1.00	0.99	1.01	1.00	0.99

(*) The sources Abroad and Ships sea only have a very small contribution to the total concentration, therefore the correction factor for these sources is equal to 1.

11.2 Correction factors per emission source

Another application of the Kalman filter in the Real Time URBIS model is that the original model only gets 11 correction factors, each of the sources gets a correction factor instead of 8 per source. The corrected concentration ($\underline{c}_k^{\text{corr}}$) is given by the following formula:

$$\underline{c}_k^{\text{corr}} = \sum_{j=1}^{11} \left(\sum_{i=1}^{8} \mu_{ji,k} \underline{m}_{ji} \right) e^{\gamma_{j,k}} \tag{11.2}$$

where $\mu_{ji,k}$ is the weight for the standard concentration field on time k of source j and wind combination i. The vector $\underline{m}_{ji}$is the standard concentration field of source j and wind combination i. The correction factor on time k for source j is given by $e^{\gamma_{j,k}}$.

The result of this application is that each emission source gets a correction factor on each hour. The annual means for 2006 of this correction factors are given in Table 11.2. In this table also the annual means of the correction factors are calculated as an average of correction factors with other time scales. The correction factors for the other time scales are calculated with Equation 11.2 with k as time step day, week or month.

In this table, the correction factors for the source 'Background' are little larger than 1, while the correction factors for the traffic sources are little smaller than 1. This leads to the same ideas as in Section 11.1, the emissions from the traffic sources and the background concentrations have possible inaccuracies.

Table 11.2: *Mean of the correction factors per emission source*

Source	Hourly	Daily	Weekly	Monthly
Abroad (*)	1.00	1.00	1.00	1.00
Background	1.05	1.05	1.03	1.02
Zone card	0.96	0.98	1.01	1.03
CAR	0.95	0.95	0.98	0.99
Roads nearby	0.94	0.94	0.93	0.88
Roads far	1.00	1.00	0.98	0.96
Industry	1.01	1.01	1.00	0.99
Domestic	1.01	1.00	0.99	0.97
Ships inland (*)	1.00	1.00	1.00	1.00
Ships sea	1.00	1.01	1.02	1.10
Rest	1.00	0.99	0.98	0.97

(*) The sources Abroad and Ships sea only have a very small contribution to the total concentration, therefore the correction factor for these sources is equal to 1.

12 Conclusions and discussion

In this part, three extensions of the Kalman filter in the Real Time URBIS model are described. With these three extensions it is possible to reduce the uncertainty of the estimated concentration NO_x.

In the first extension, a method is described to add some extra monitoring stations. The main conclusion is that the total uncertainty connected with the population will be minimized if the extra monitoring station are located such that the emission from the important sources is covered. In the Rijnmond area are the sources 'Background', 'Ships sea' and 'Zone card' stated as important sources. Therefore some possible locations for extra monitoring stations are: The Zeedijk in Bernisse, the Harmsen Bridge on the junction of the A15 and the N57 and the Missouriweg in Hoek van Holland.

The second extension on the Kalman filter is the application of the Kalman filter with different time scales. The uncertainty of the model simulation is smaller if the model covers daily, weekly or monthly concentrations. This is because the extremes which can occur in hourly mean calculations are averaged out. On the other hand, the Kalman filter has less information from measurements to reduce the uncertainty. The patterns of the uncertainty are nearly the same for each time scale, therefore it is possible to add some monitoring stations to the system with different time scales. In a Kalman filter, it is possible to combine different time scales. With this combination it is again possible to find an optimal setting of extra monitoring stations with another time scale.

Finally the correction factors, calculated in the first part of this report are analyzed. If the annual mean of a correction factor is significantly different from 1, this is possibly caused by an inaccuracy of the comparing standard concentration field. One of the possibilities is that the emission from a certain source is not accurate in the model, but this is not necessary. It is also possible that other assumptions in the model causes an inaccuracy, or the representativity of some measurements is not sufficient. Therefore the correction factors only leads to some ideas of the origin of the inaccuracies.

In total the Kalman filter is a good instrument to reduce the uncertainty of the model simulation. One method is: extra monitoring stations which corrects the model simulation. The other method is: analyze the information about the inaccuracies of the model and use this information to improve the model.

Bibliography

Cramer. Wetsbesluit LMV 2007.109578, Ministerie van VROM, 2007.

A.F.J. de Haan, J. van der Velde, W.B. Geven, C. Festen, and A.L.M. Verbeek. A Multistate Kalman Filter for Neonatal Clotting Time Prediction and Early Detection of Coagulation Disturbances. Technical Report 1999; 38, University of Nijmegen, 1999.

A.W. Heemink. Filtertheorie. PhD thesis A42, TU Delft, 1996.

R. Kranenburg. Statistische onzekerheidsanalyse van Real Time URBIS voor het Rijnmondgebied. TNO-Report 2009-01347, TNO, May 2009.

A.J. Segers. Data assimilation in atmospheric chemistry model using Kalman filtering. Technical report, TU Delft, 2002.

Vera Spaubek. Opzet en test van een Real Time URBIS in de Rijnmond. TNO-rapport 2004/229, TNO, may 2004.

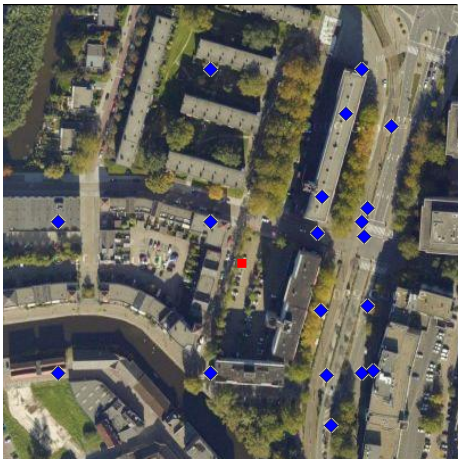
T. Weise. Global Optimization Algorithms - Theory and Application - . e-book, University of Kassel, 1988.

Greg Welch and Gary Bishop. An Introduction to the Kalman Filter. Technical Report TR 95-041, University of North Carolina, 2006.

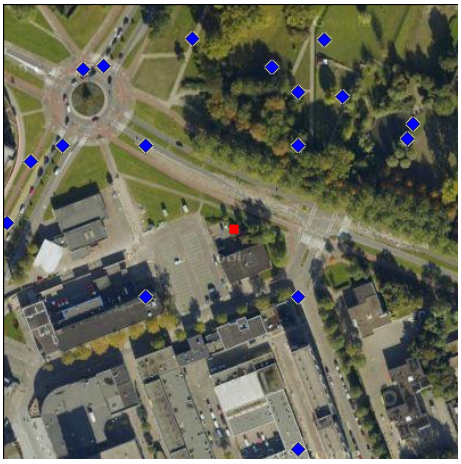
P. Wesseling and P.Y.J. Zandveld. URBIS Rotterdam Rijnmond; A pilot study. TNO-rapport 2003/245, TNO, 2003.

A Locations of the monitoring stations

The locations of the 9 monitoring stations, in the domain covered by DCMR. The two stations at Bentinckplein are located in the same building, given in Figure A.2b. The blue diamonds represents the locations of the grid points, while the red squares are the monitoring stations.



(a) Schiedam



(b) Hoogvliet

(c) Maassluis

(d) Overschie

Figure A.1

(a) Ridderkerk

(b) Bentinckplein

(c) Schiedamsevest

(d) Vlaardingen

Figure A.2

B Standard concentration fields

Standard concentration fields for the 11 different sources in the URBIS model, each source has 8 standard concentration fields valid for 4 different wind directions (N, E, S, W) and 2 different wind speeds (1.5 $^{m}/_{s}$ and 5.5 $^{m}/_{s}$).

Figure B.1: Emission source: Abroad

Figure B.2: Emission source: Background

Figure B.3: Emission source: Zone card

Figure B.4: Emission source: CAR

Figure B.5: Emission source: Roads nearby

Figure B.6: Emission source: Roads far

Figure B.7: Emission source: Industry

Figure B.8: Emission source: Domestic

Figure B.9: Emission source: Ships inland

Figure B.10: Emission source: Ships sea

Figure B.11: Emission source: Rest

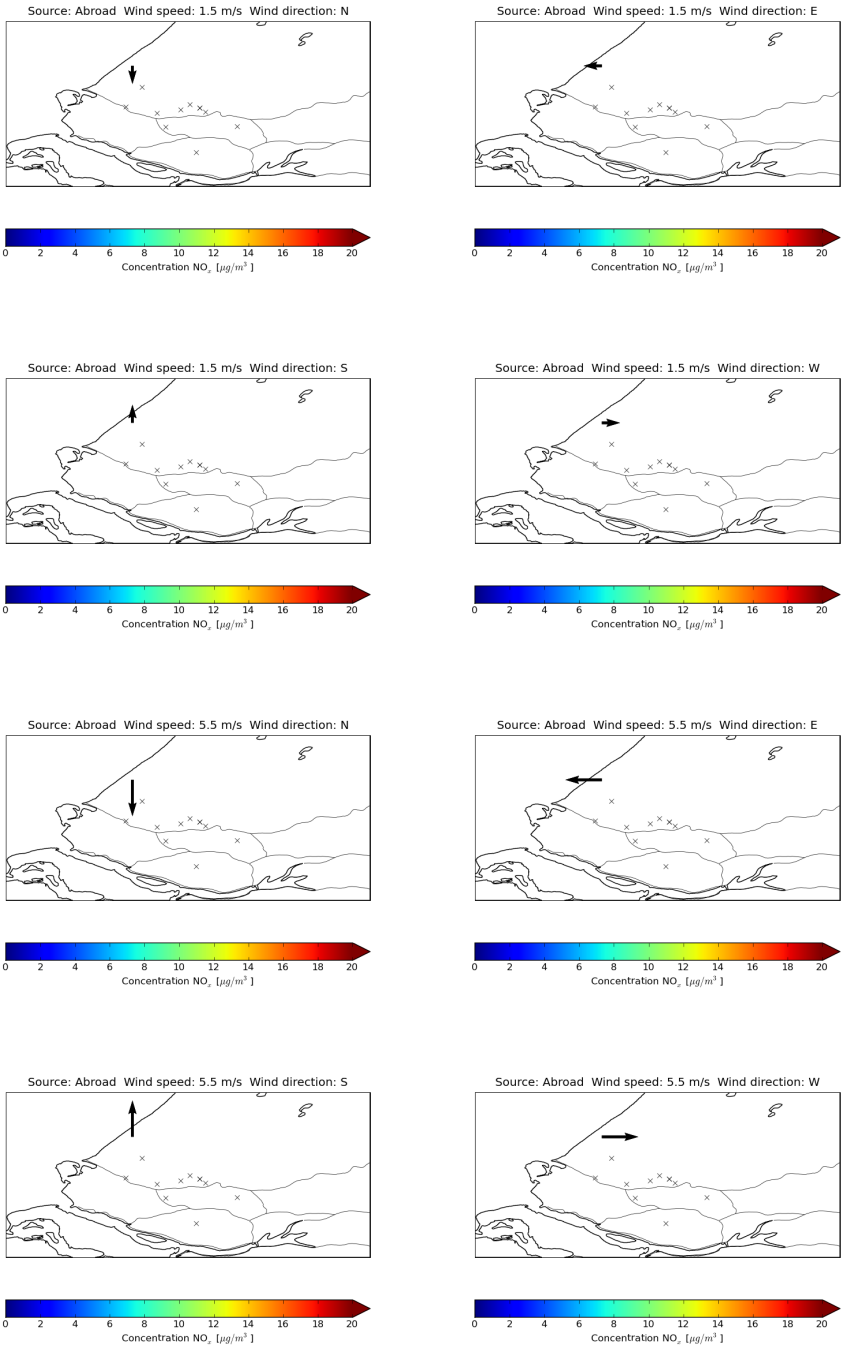


Figure B.1: *Emission Source: Abroad*

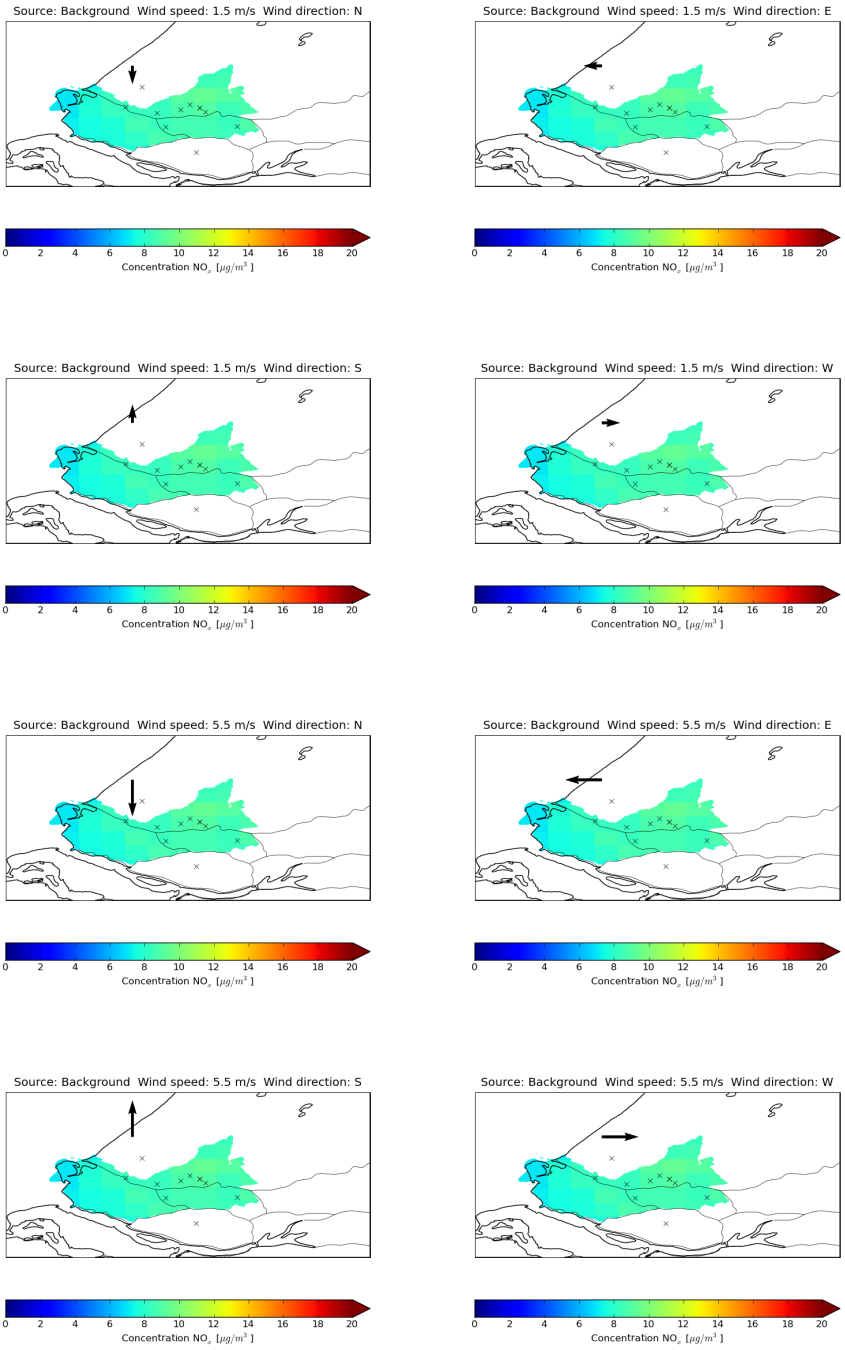


Figure B.2: *Emission Source: Background*

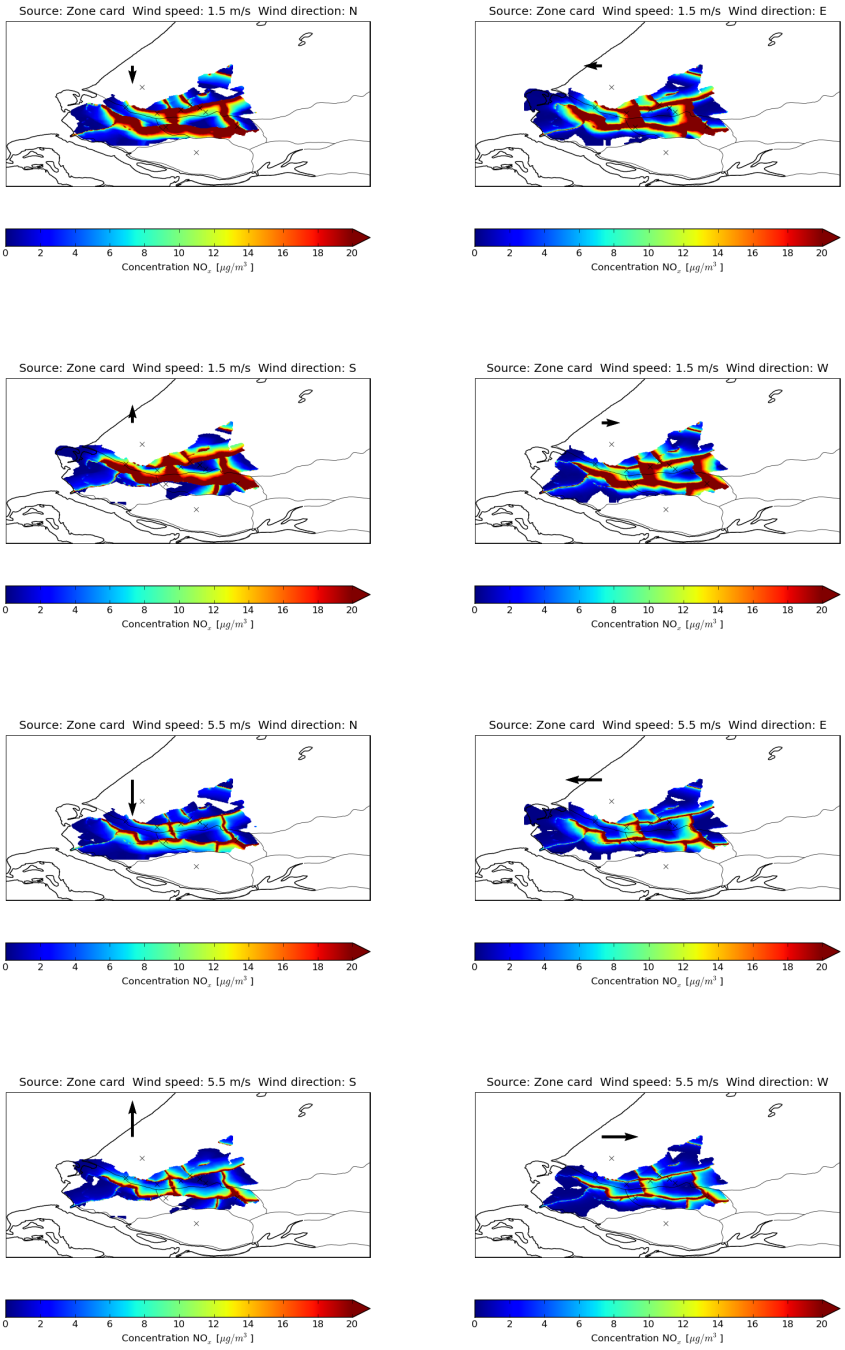


Figure B.3: *Emission Source: Zone Cards*

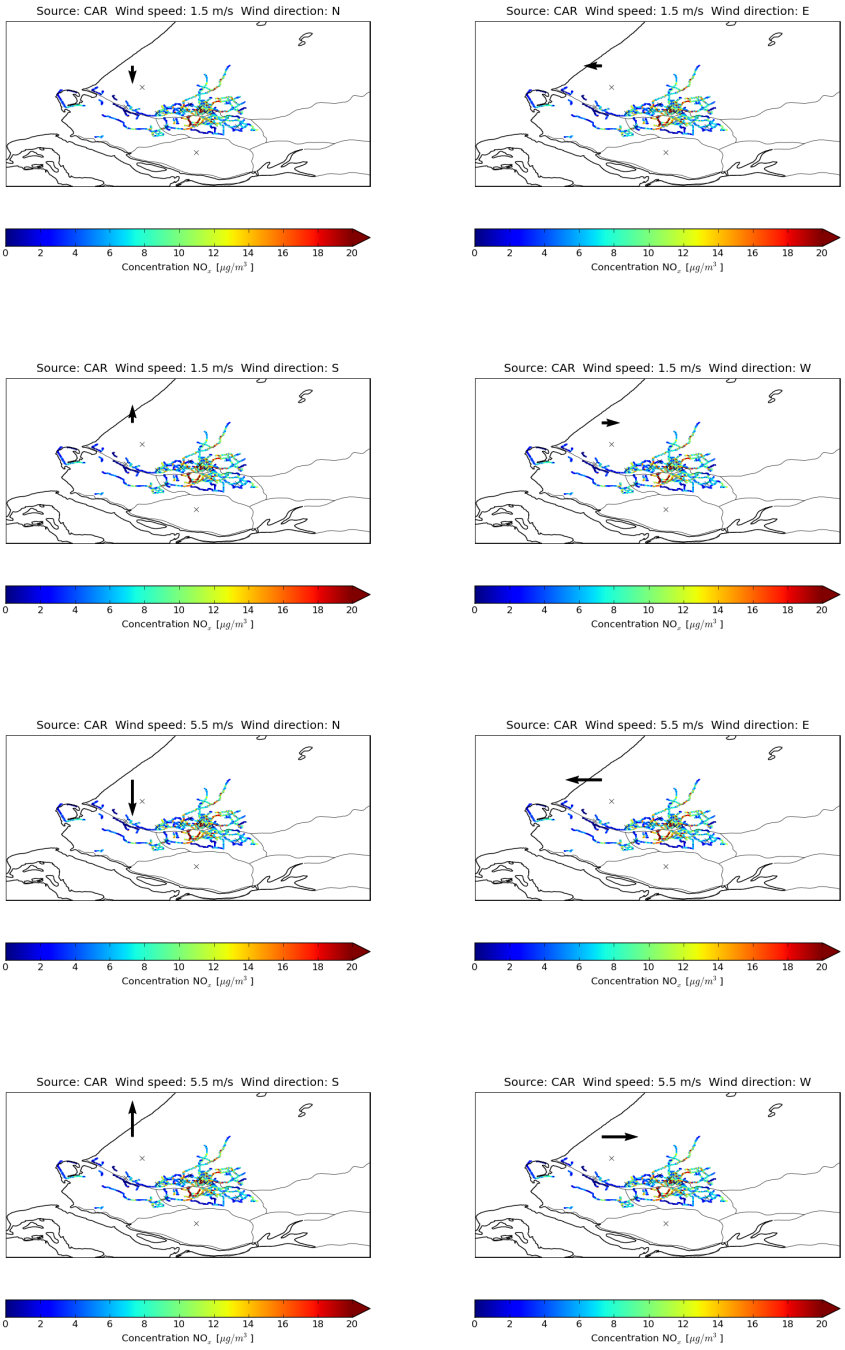


Figure B.4: *Emission Source: CAR*

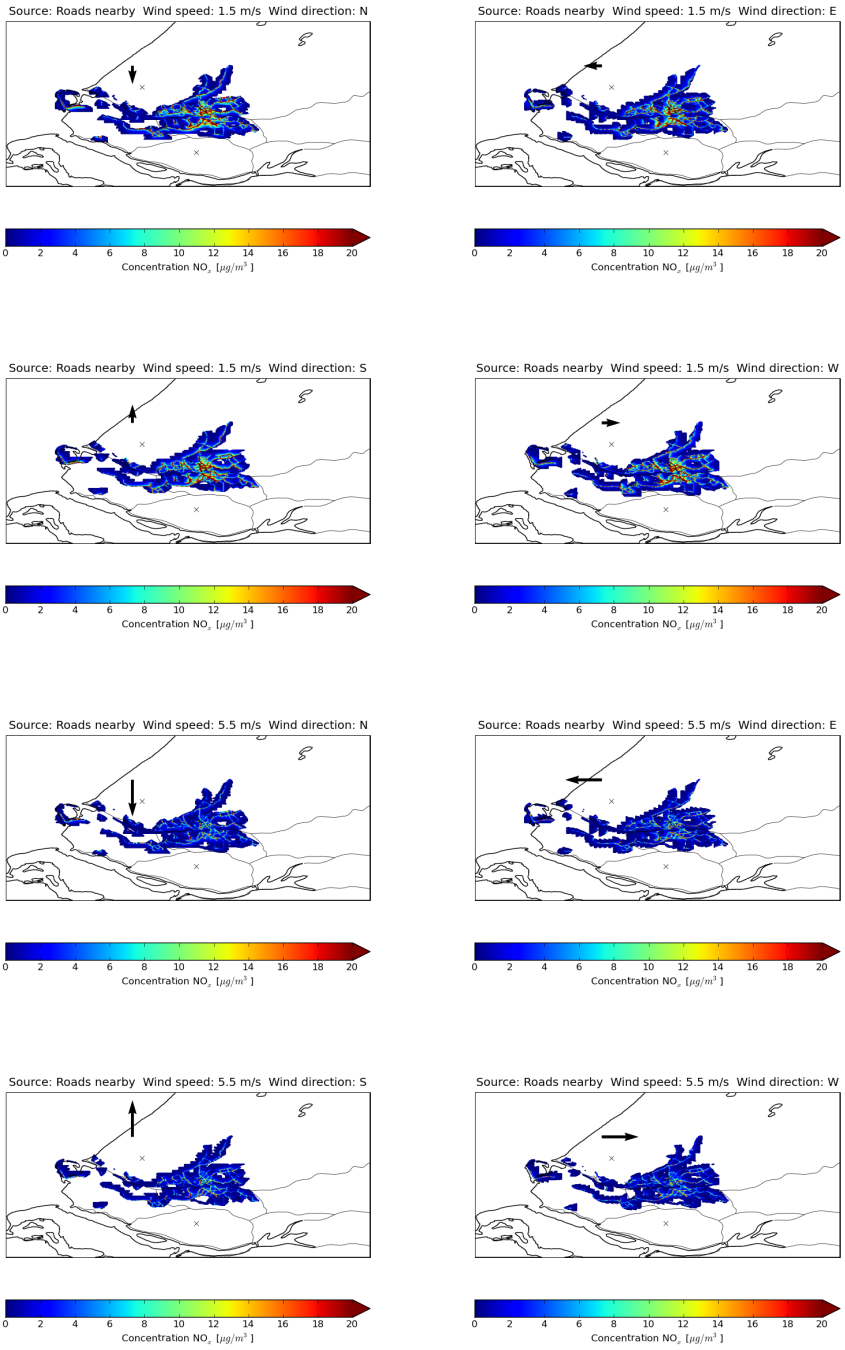


Figure B.5: *Emission Source: Road nearby*

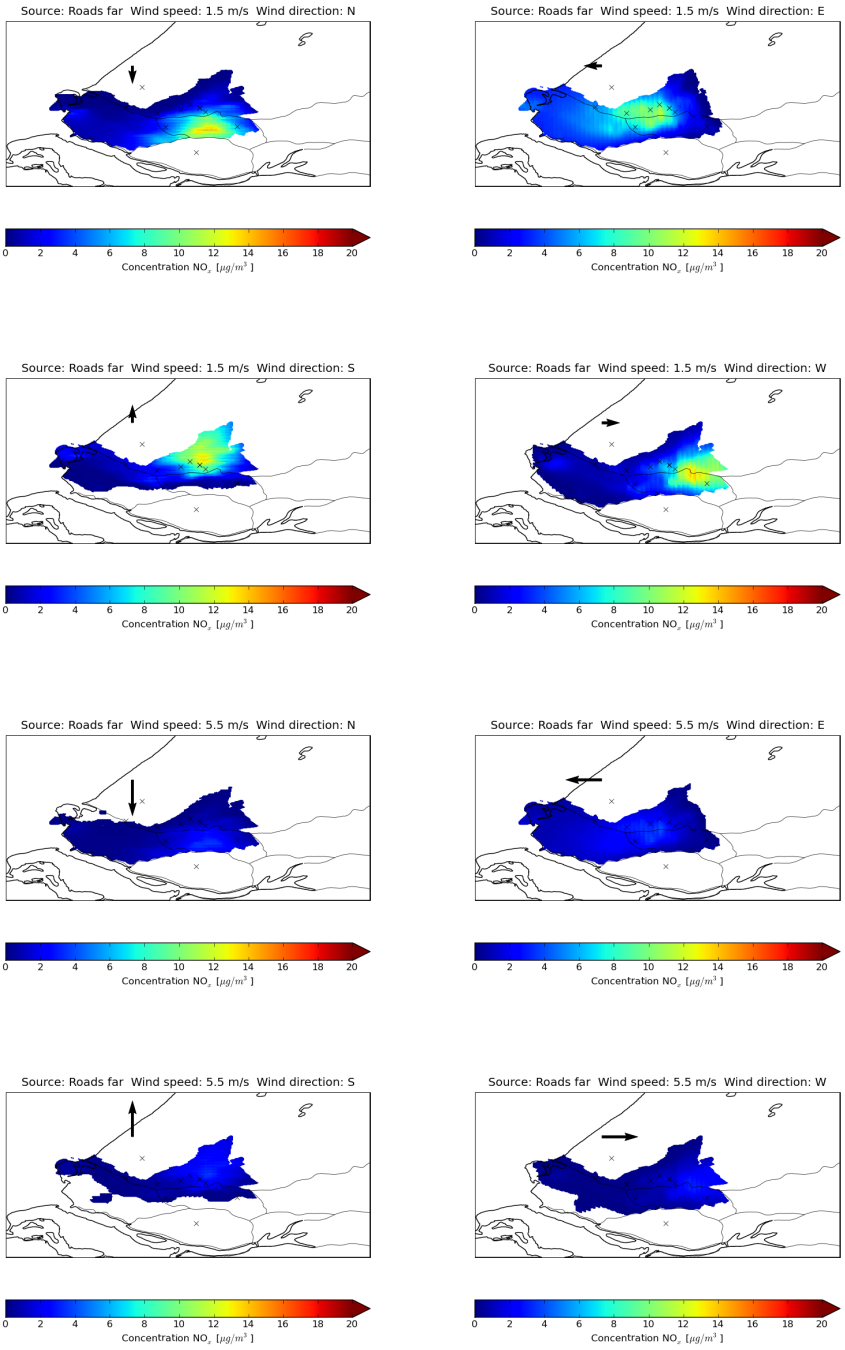


Figure B.6: *Emission Source: Road far*

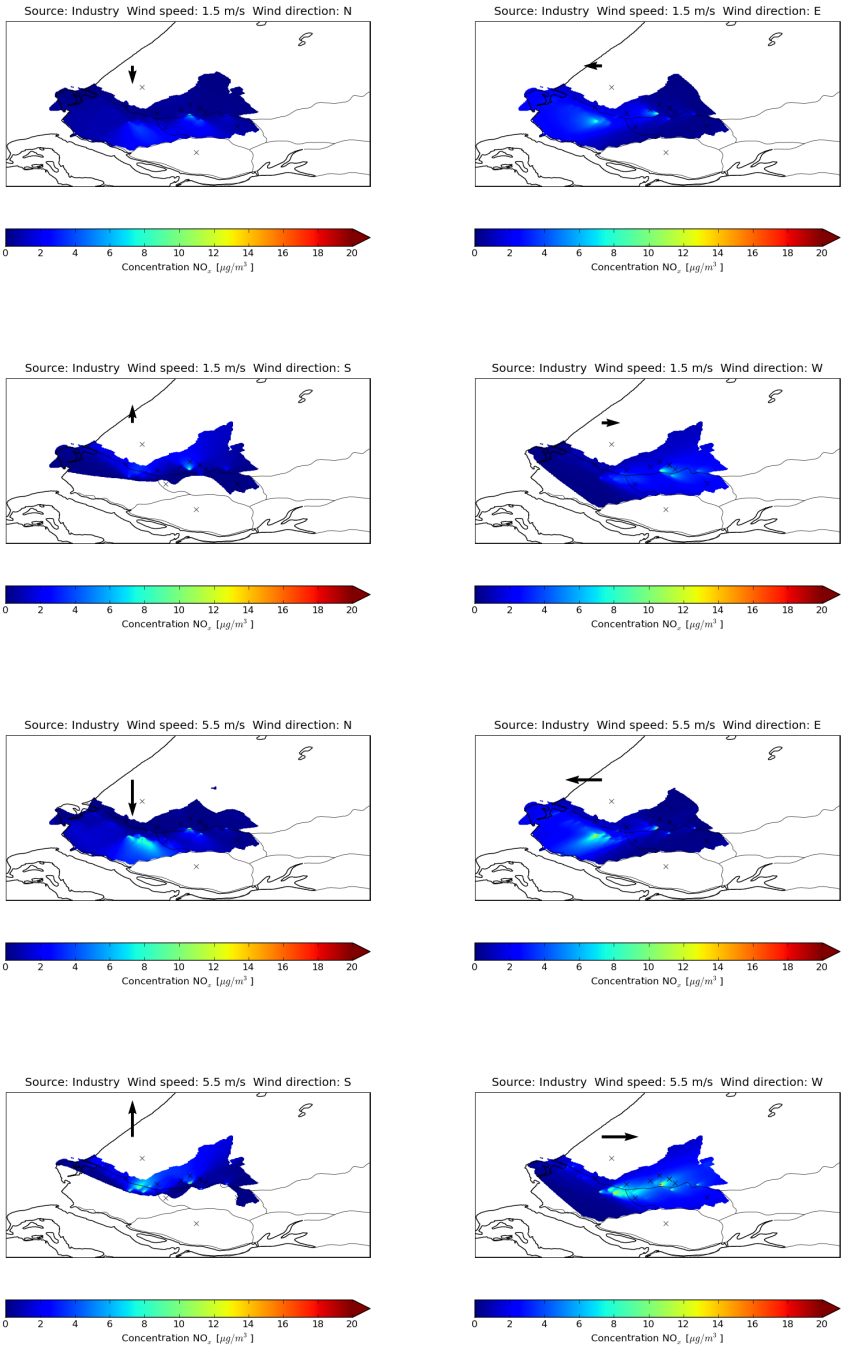


Figure B.7: *Emission Source: Industry*

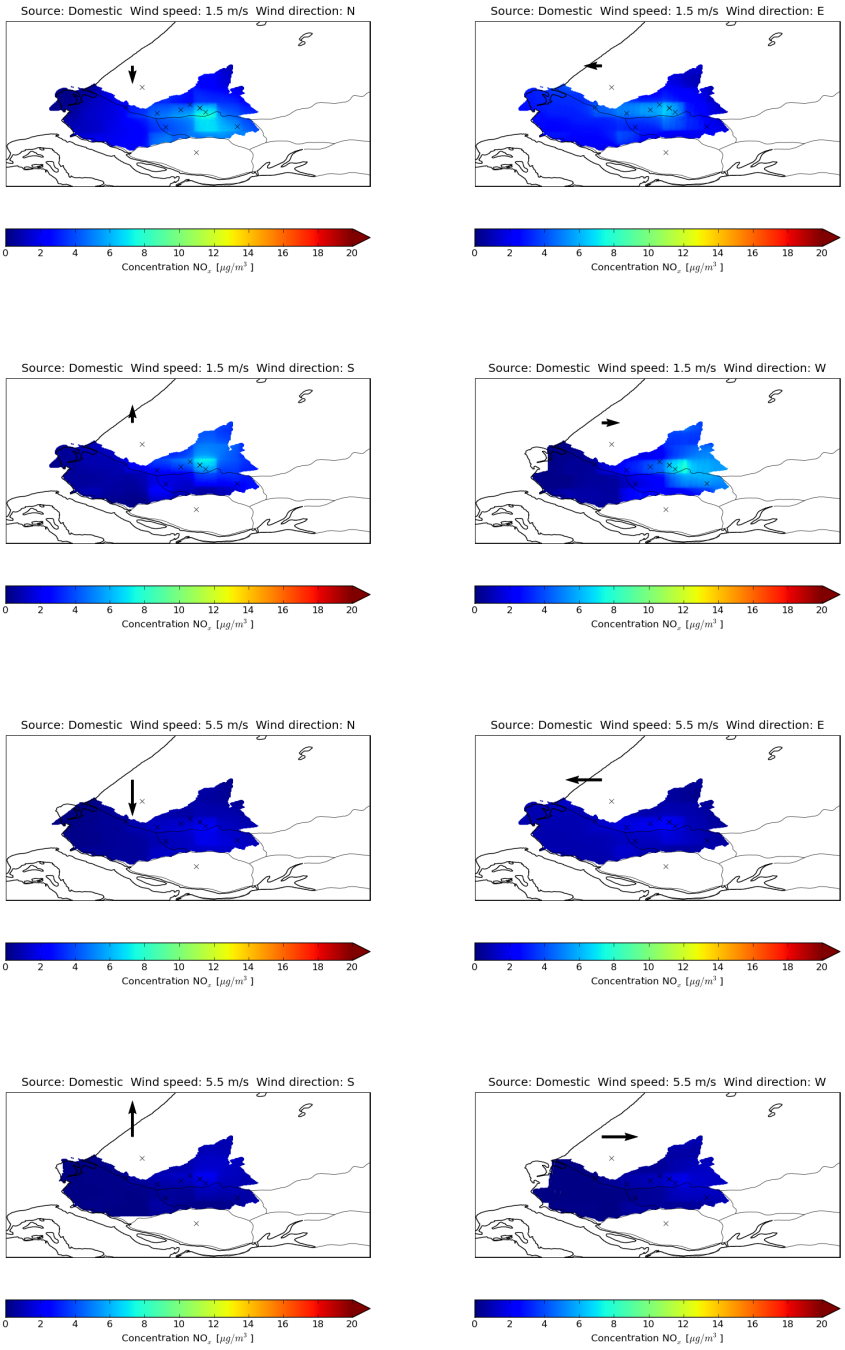


Figure B.8: *Emission Source: Domestic Rijnmond*

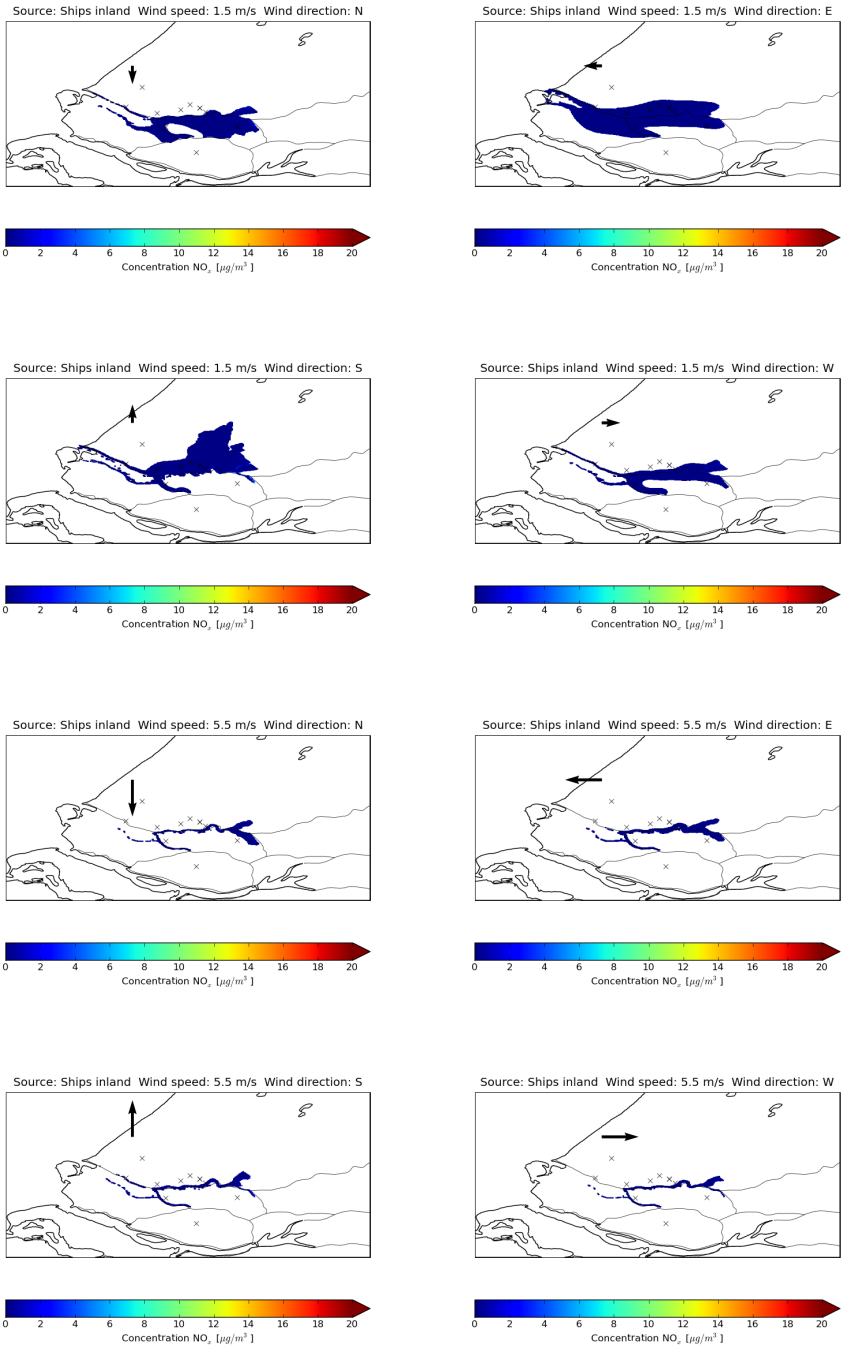


Figure B.9: *Emission Source: Ships inland*

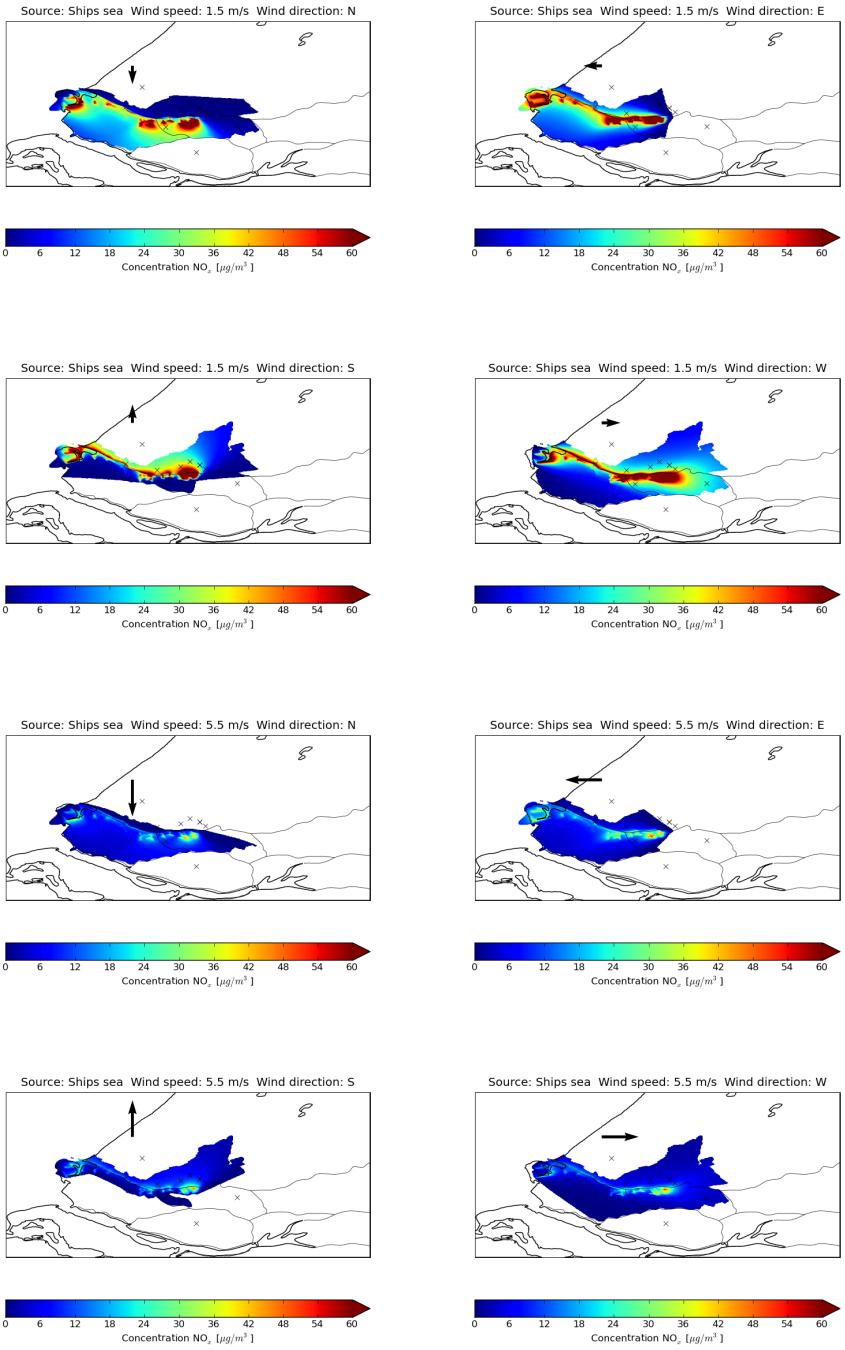


Figure B.10: *Emission Source: Ships sea*

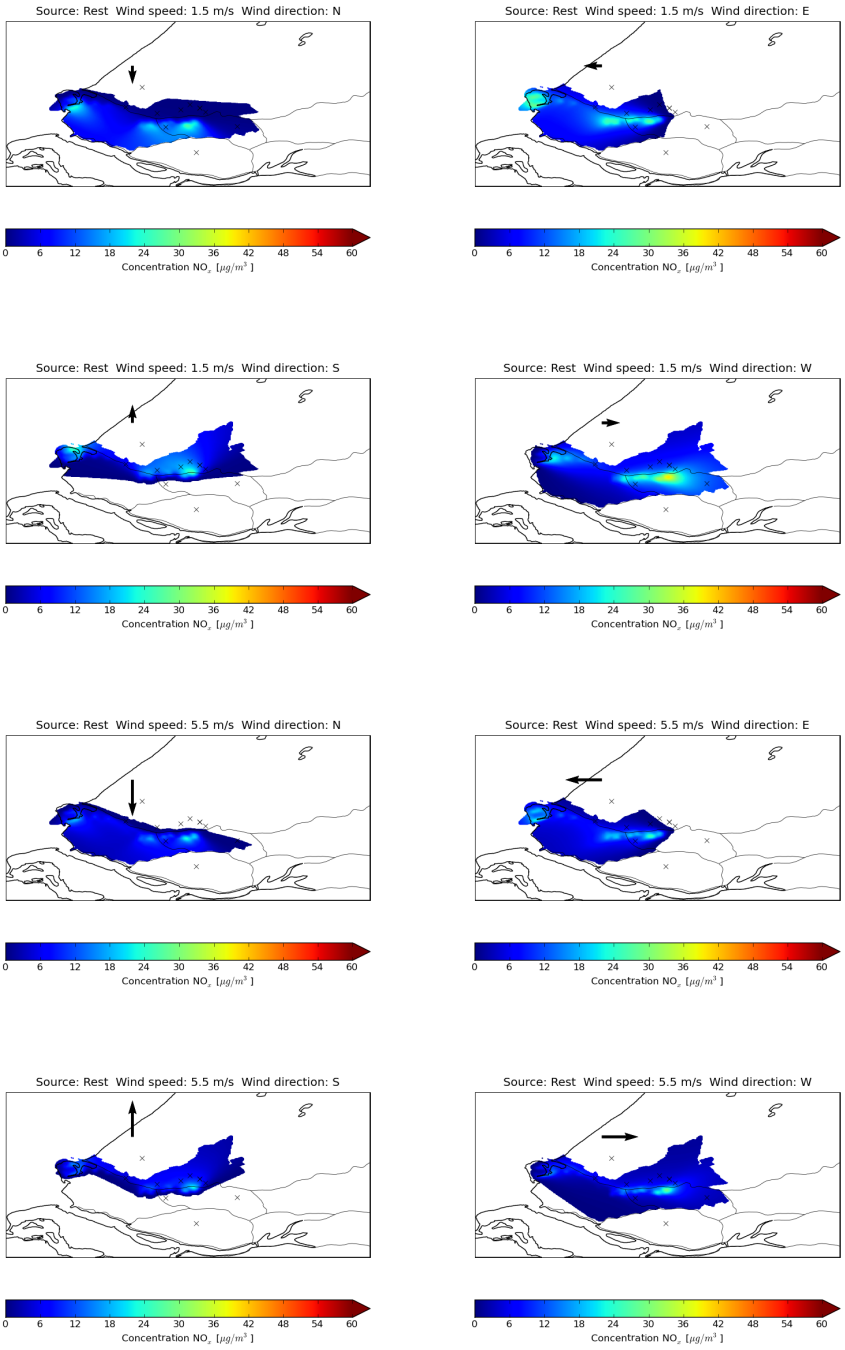


Figure B.11: *Emission Source: Rest*