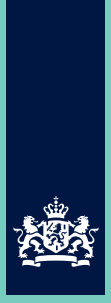

National Institute for Public Health
and the Environment
Ministry of Health, Welfare and Sport

A statistical assessment of the predictions of NO_2 concentrations by the Dutch standard air quality models

RIVM Letter Report 680705028/2013
J.A. Ferreira

National Institute for Public Health
and the Environment
Ministry of Health, Welfare and Sport

A statistical assessment of the predictions of NO_2 concentrations by the Dutch standard air quality models

RIVM Letter report 680705028/2013
J.A. Ferreira

Colophon

ISBN:

J.A. Ferreira, RIO/SMG

Contact:
Joost Wesseling
M&V/MIL/ILG
joost.wesseling@rivm.nl

This research has been carried out by order of the Ministry of Infrastructure and the Environment within the framework of the project Stedelijke Luchtkwaliteit.

Rapport in het kort

Een statistische beoordeling van de voorspelde NO_2 concentraties door de Nederlandse standaard luchtkwaliteit modellen

Schattingen van de gemiddelde jaarlijkse concentratie van stikstofdioxide (NO_2) worden regelmatig berekend voor de Nederlandse gemeentes met behulp van de zogenoemde standaard Nederlandse luchtkwaliteit modellen. Periodiek wordt de nauwkeurigheid van deze schattingen beoordeeld door ze te vergelijken met stikstofdioxidenconcentraties die op een aantal locaties in Nederland worden gemeten.

Volgens het RIVM voldoen de modelschattingen op basis van recente gegevens en meerdere statistische analyses (Wesseling *et al.*, 2013) aan de vereisten die hiervoor vanuit de Europese Commissie worden gesteld. In aanvulling daarop heeft het RIVM statistische analyses gemaakt van de percentuele afwijking tussen de berekeningen en metingen van de stikstofdioxidenconcentraties.

Boven de 35 microgrammen per kubieke meter blijkt de percentuele afwijking van een modelschatting gemiddeld rond -3.56 procent te zitten, ongeacht de bijbehorende stikstofdioxidemeting. Bij circa 95 procent van de modelschattingen is de percentuele afwijking tussen -25 en 18 procent, ongeacht de hoogte van de corresponderende stikstofdioxidemeting. Deze conclusies complementeren de analyse van Wesseling *et al*. ten aanzien van de kwaliteit van de modelberekeningen voor NO_2.

Abstract

A statistical assessment of the predictions of NO_2 concentrations by the Dutch standard air quality models

Estimates of the average annual concentration of nitrogen dioxide (NO_2) are produced regularly for the municipalities in the Netherlands by the so-called Dutch standard air quality models. The accuracy of these estimates is periodically assessed by comparing them with NO_2 measurements obtained at a number of locations.

The RIVM has recently provided an assessment of the model estimates based on recent data and various statistical analyses (Wesseling *et al*., 2013), concluding that the estimates comply with European criteria for air quality modeling. The present report complements the statistical analyses of Wesseling et al. by quantifying the percental error of the model estimates relative to corresponding measured NO_2 concentrations.

Although the characterization of the percental error over the whole range of NO_2 concentrations is an uncertain task, the error can be described approximately and rather simply for concentrations above 35 micrograms per cubic meter. It is concluded that, over this range of NO_2 concentrations, the percental error of a model estimate is around -3.56% on average, regardless of the value of the corresponding NO_2 measurement. About 95% of the model estimates have (irrespective of the corresponding NO_2 measurements) a percental error between -25% and 18%. These conclusions complement the assessment provided by Wesseling *et al*. about the quality of the model estimates of NO_2.

Contents

Summary

This report provides a statistical assessment of the quality of the predictions (estimates) of nitrogen dioxide (NO_2) produced by the Dutch standard air quality models during 2010 and 2011. More precisely, it attempts to characterize the statistical behaviour of the *relative* or *percental error* of the predictions, defined as the prediction of NO_2 minus the corresponding measured concentration divided by the latter times 100, conditionally on (or as a function of) the measured concentration. The focus on the error of a prediction relative to a measurement is justified by the fact that the measurements (despite eventual shortcomings) provide the best estimates available of actual NO_2 concentrations.

Our first conclusion is that the characterization of the relative error as a function of measured concentration is an uncertain task, due to statistical 'irregularities' in the relationship between measurements and predictions over the lower range of NO_2 concentration. Such irregularities could very well have an explanation, but the problem is that the available data do not afford any general explanation for them.

Fortunately, it appears that over the range of NO_2 concentrations above 35 $\mu g/m^3$ the relative error follows a rather regular pattern that can be described by a simple model. And from this model follow, in particular, the following two statements:

- On average, the relative error of a model prediction is about -3.56 %, irrespective of the corresponding measured concentration;
- About 95% of the relative errors lie between -25 and 18%.

Somewhat more elaborate but still simple and concrete statements are derived from the model. In particular, the two statements above can be qualified in terms of their uncertainty.

1 Introduction

The statistical comparison between *model predictions* (model estimates) and measurements of NO_2 concentrations can be carried out in a number of ways, depending on the quantities one wishes to look at and on the statistical models and methods one decides to adopt. Ideally one would like to set up a probabilistic model describing the joint behaviour of each pair of observations (prediction, measurement), possibly taking into account *contextual information* (e.g. location, presence or absence of vegetation) summarizing the conditions under which the observations are made. From such a model one would be able to derive all sorts of statements, at various levels of detail, about an arbitrary prediction and the corresponding measurement. From the data analyses summarized in this report it appears that a detailed model for the joint behaviour of predictions and measurements of NO_2 concentrations is difficult to construct and justify. Instead, it turns out to be more realistic, and fortunately also sufficient, to describe the *error of a model prediction relative to the corresponding measurement* as a function of the latter, provided only NO_2 measurements above 35 $\mu g/m^3$ are considered. In essence, if R (from 'reken') denotes a model prediction and M (from 'meet') the corresponding measurement, then, conditionally on the event that $M = m$ (i.e. that the measurement equals a given number m), the *relative* or *percental error*

$$100 \times \frac{R\text{-}M}{M}$$

is normally distributed with mean $-3.56\,\%$ and standard deviation 10.61%, provided m is ≥35 $\mu g/m^3$. This affords <u>general statements about the error of a model prediction conditionally on the value of a given measurement</u> over the range of greatest practical interest: For instance, if $M \geq 35\ \mu g/m^3$, then

- The mean, or expected, relative error of the model prediction R corresponding to the measurement M is about $-3.56\,\%$, regardless of the actual value of M;
- With about 95% probability, the relative error of the model prediction R corresponding to the measurement M lies between $-24.78 = -3.56 - 2 \times 10.61$ and $17.66 = -3.56 + 2 \times 10.61$, which is to say that with that same probability

$$-0.25 \times M \leq R\text{-}M \leq 0.18 \times M, \qquad \text{or} \qquad 0.75 \times M \leq R \leq 1.18 \times M.$$

These statements, which must be regarded as approximately correct (being based on estimates derived from the data and relying on certain assumptions), will be justified and qualified in the next section.
Focusing on the percental error of a prediction relative to a measurement is certainly meaningful if the measurements (despite possible errors) provide the best estimates available of actual NO_2 concentrations.

2 A model for relative error as a function of measurement

The data set provided to us by Joost Wesseling (RIVM) consists of 436 pairs of measurements and model predictions, together with the corresponding data on a number of variables: Categorie, Code, Bomen, etc. For illustration, the following computer output shows the first 10 rows of the data file:

	Code	Bomen	Straattype	Background	MEET	REKEN	MEET_BIJ	REKEN_BIJ	Kwaliteit	Categorie
1	2	1	1	19.2	32	29.9	12.8	10.7	0	SRM2
2	2	1	1	19.1	24	22.1	4.9	3.0	0	Achtergrond
3	2	1	1	20.4	37	30.5	16.6	10.1	0	SRM2
4	2	1	1	20.4	23	21.9	2.6	1.5	0	Achtergrond
5	2	1	1	21.7	50	46.0	28.3	24.3	0	SRM2
6	2	1	1	22.2	33	26.9	10.8	4.7	0	SRM2
7	2	1	1	20.9	38	40.3	17.1	19.4	0	SRM2
8	2	1	1	23.5	29	28.8	5.5	5.3	0	SRM2
9	2	1	1	26.5	55	59.8	28.5	33.3	0	SRM2
10	2	1	1	26.4	32	32.8	5.6	6.4	0	SRM2

The variable MEET contains measured concentrations; the variable REKEN contains the predicted concentrations. The variable Code pertains to the geographical location of the observations, and Categorie to the type of model used to make the prediction. More information about the data and the meaning of the variables may be found in the report of Wesseling et al. (2013).

Figure 1 shows a *non-parametric estimate* of the *regression function* of relative error on measurement, the model prediction and the measurement being indicated by REKEN and MEET. The regression function is the mean (or expected value) of the relative error $100\times(R\text{-}M)/M$ of a model prediction R conditionally on $M = m$ (i.e. given that the value of the measurement M equals the number m); as m varies along the horizontal axis of the figure, the estimated regression function assumes different values, represented by the height (measured with reference to the vertical axis) of the full red line.

Above and below the estimate of the regression function are the upper and lower boundaries of the 95% *confidence and prediction bands*. The interpretation of a 95% confidence band is that if 100 data sets were generated from the same system that generated the present data set and each time a confidence band were constructed then the band would contain the true regression function about 95 times. The 95% prediction band pertains to a single observation of the relative error conditional on a given value of the measurement: if m is a given number and $M = m$, then the probability that $100\times(R\text{-}M)/M$ falls inside the prediction band *at the point* m is about 0.95. See **section 3** for more details about the estimates and the confidence and prediction bands for the regression curve.

Since the model predictions are intended to serve as substitutes for the measurements, it would be pleasing and convenient to find that the relative error is small in absolute value and varies regularly (e.g. is constant or follows a straight line with a small slope) with M. The estimated regression curve of

figure 1 indicates that this may be the case in the range of $M \geq 35\ \mu g/m^3$ or so, but over the lower range of measurements the relative error behaves in a *non-linear* way and reaches a mean value of about 11% around $28\ \mu g/m^3$. In spite of the non-linearity, the variability of the relative errors around their mean is rather constant, in the range of 8-12%, over the whole range of the measurements, as witnessed by the estimates of the standard deviation of relative error shown in **figure 2.**

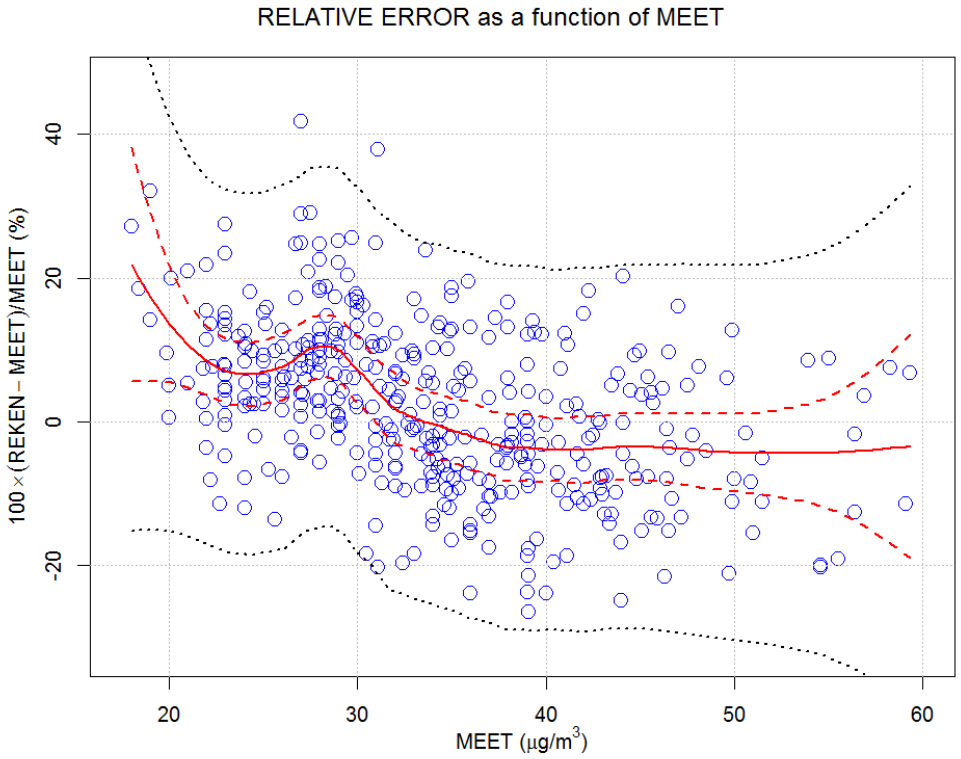

Figure 1: Non-parametric estimate (full red line), 95% confidence band (dashed red lines) and 95% prediction band (dotted black lines) for the regression function of the relative error on the measurement.

It is natural to try to explain the non-linearity of the relative error in the range $M < 35\ \mu g/m^3$ by looking at the error as a function of other variables besides M. Thus, it could be that by stratifying the data according to the values of Categorie, Code, Bomen, etc., or of combinations thereof, one would arrive at separate simple (e.g. linear) descriptions of the relative error as a function of the measurement. If this were so, then those descriptions would provide us with a characterization of the performance of the model *per stratum* (through statements such as the bulleted ones at the end of our introduction), which, though not so easy to summarize and communicate in a couple of statements, might be relatively accurate and informative. It appears, however, that the result of this procedure applied to the present data set is too fragmented and not very reliable. We elaborate a little on this now.

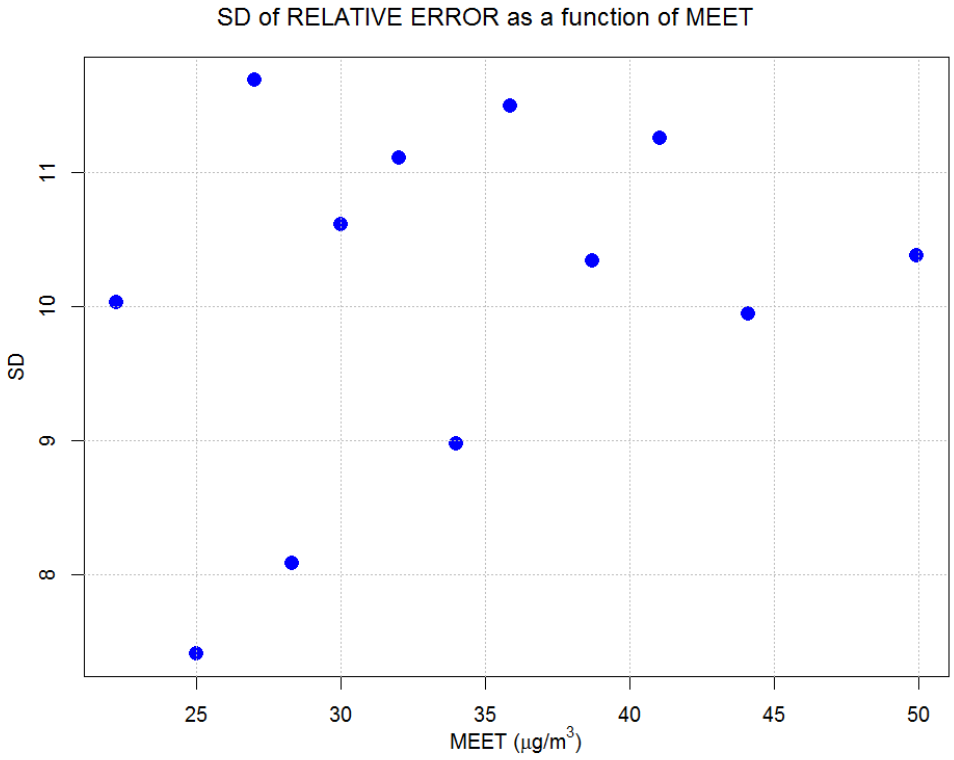

Figure 2: Estimates of the standard deviation (SD) of the relative error for varying values of the measurement (MEET).

First, it is clear that Categorie and Code are the only *bona fide* predictor variables, other than MEET (the measurement), containing a non-negligible amount of information on the relative error; this is the conclusion of certain *prediction analyses* which are explained briefly in **section 3**. This suggests that in order to get simpler (linear) descriptions of the relative error as a function of the measurement we should look separately at subsets of the data corresponding to the different levels of Categorie and/or Code. Stratifying the data in terms of Categorie alone does not help, as seen by the estimated regression function for data with Categorie=achtergrond shown in **figure 3** and by the analogous plots (not shown) of the relative error versus the measurement corresponding to Categorie=SMR1 and Categorie=SMR2—the 'non-linearities' around 25 and 28 $\mu g/m^3$ subsist despite the stratification.
Stratifying in terms of Code alone does not help either, as seen, for instance, by the estimate based on data with Code=3 shown in **figure 4**.

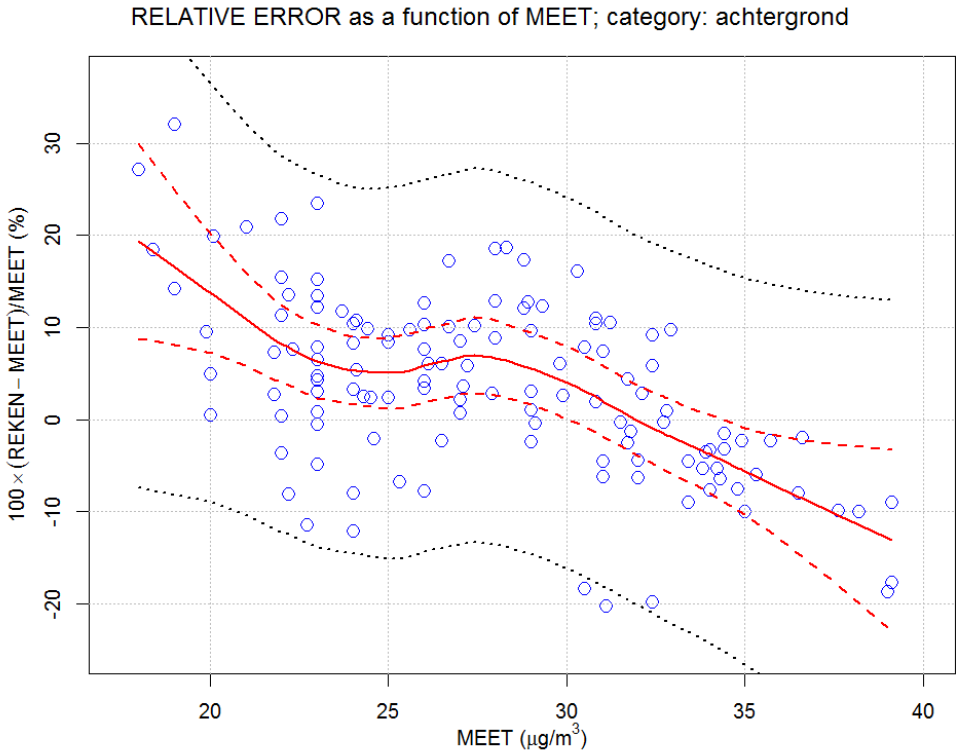

Figure 3: Non-parametric estimate and 95% confidence and prediction bands for the regression of relative error on measurement based on the data with Category=achtergrond.

Stratifying by Categorie and Code *jointly* explains the non-linearity to some extent, as seen by the straight lines fitted (by least squares) separately to the data sets corresponding to various levels of Code and Categorie=achtergrond shown in **figure 5**. Although each cluster of points appears to be somewhat homogeneous, it is clear that the fits (e.g. of Code=2 and Code=3) can be rather dubious, or at least uncertain (in the sense of their intercept and location parameter estimates having large variances). Similar observations can be made about the analogous fits done on the various subsets with Categorie=SMR1 and Categorie=SMR2. Thus, although the model predictions could be assessed separately within the various strata of Categorie and Code, the resulting statements (besides being too many) would not be very reliable, partly because of the small sample sizes and partly because of the relatively poor fits.

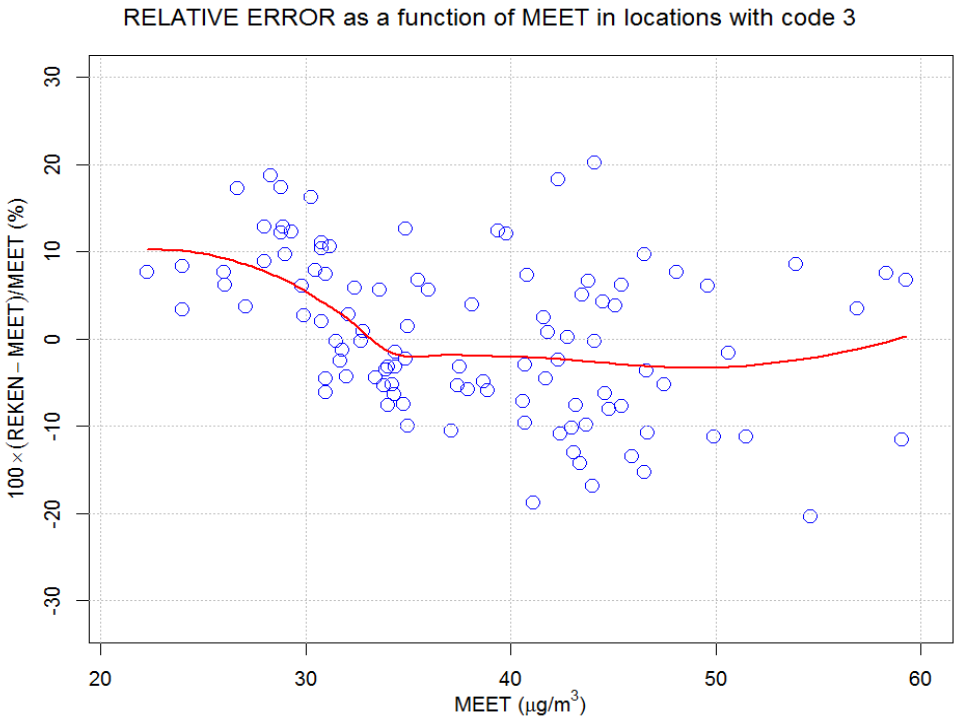

Figure 4: Non-parametric estimate of the regression function of relative error on measurement based on the data with Code=3.

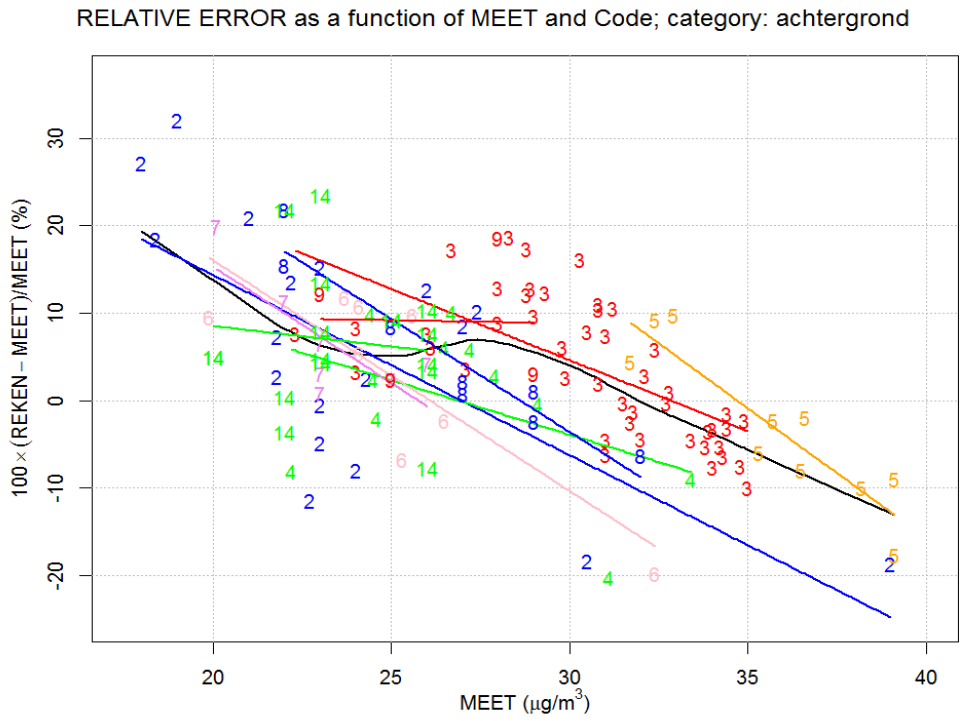

Figure 5: Relative errors with Categorie=achtergrond by Code with straight lines fitted by linear regression.

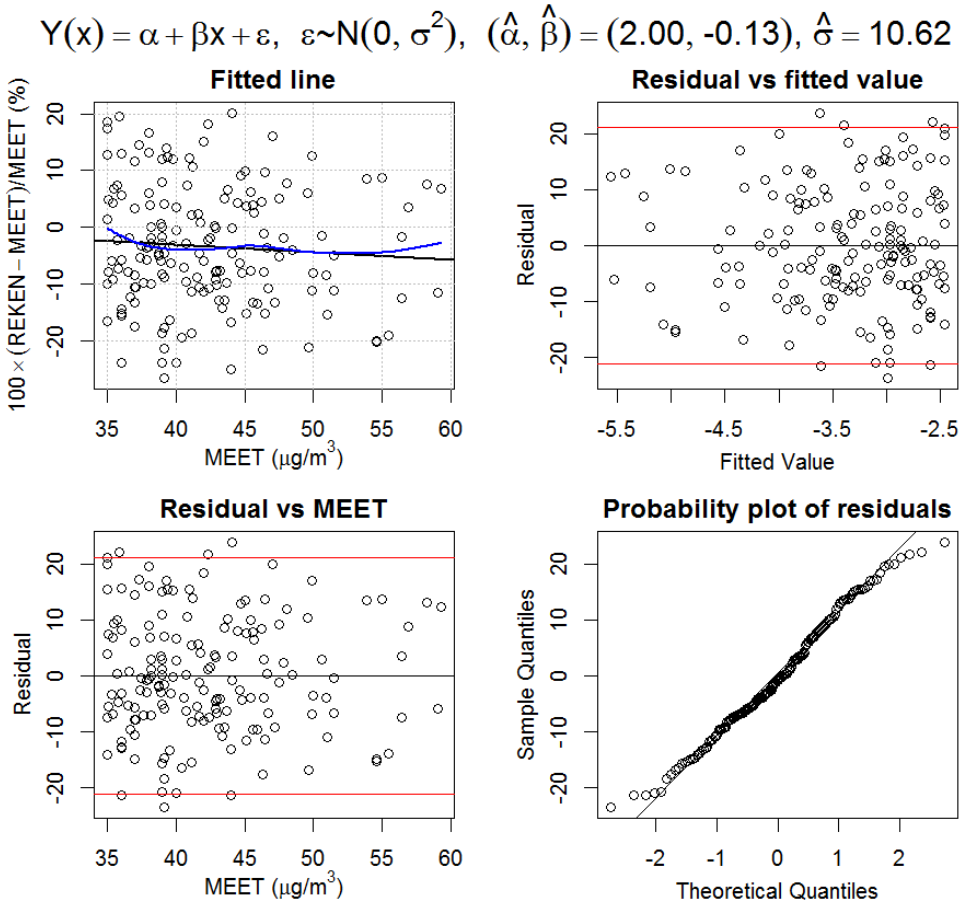

Figure 6: Results of fitting a linear regression model to relative error as a function of measurement in the range [35,60]. The top left panel shows the observations with the fitted line and a non-parametric estimate of the regression function (in blue); the other panels show how the residuals vary with the fitted values (values of the fitted line at the measurements) and with the measurements, and a probability plot of the residuals, designed to assess their approximate normality. The model and the parameter estimates appear at the top.

A more fruitful, if less ambitious, approach is to focus on describing the relationship between relative error and measurement in the range of $M \geq 35$ $\mu g/m^3$, over which about 180 observations are available. **Figure 1** suggests that this relationship is approximately linear in that range, and the results of fitting a straight line to the relative errors and measurements, shown in **figure 6**, give some support to that idea. The fitted line seems indeed to provide a simple and accurate description of relative error as a function of measurement, but we need to check for possible underlying patterns that could be explained by Categorie and Code. The prediction analyses already mentioned indicate that if the measurements are restricted to being above $35\ \mu g/m^3$ then Categorie, Code and MEET all have very little value for predicting relative error, which agrees with a model that is essentially equal to a constant plus a 'random error' (a term that is not susceptible of explanation), just like the model described in **figure 6**. In

fact, since the standard errors of the estimates of the intercept and slope parameters are 6.20 and 0.15, respectively, the relative errors appear to be compatible with such a model. (Note that with $\alpha = 2$, $\beta = -0.13$, $A = 35$ and $B = 60$ the 'mean value' of the straight line $y \equiv y(x) = \alpha + \beta x$ over the interval $[A, B]$ is $(B-A)^{-1}\int_A^B y(x)\,dx = \alpha + \beta(A+B)/2 = -5.8$, so the straight line represented in **figure 6** is close to a horizontal line at height -5.8.)

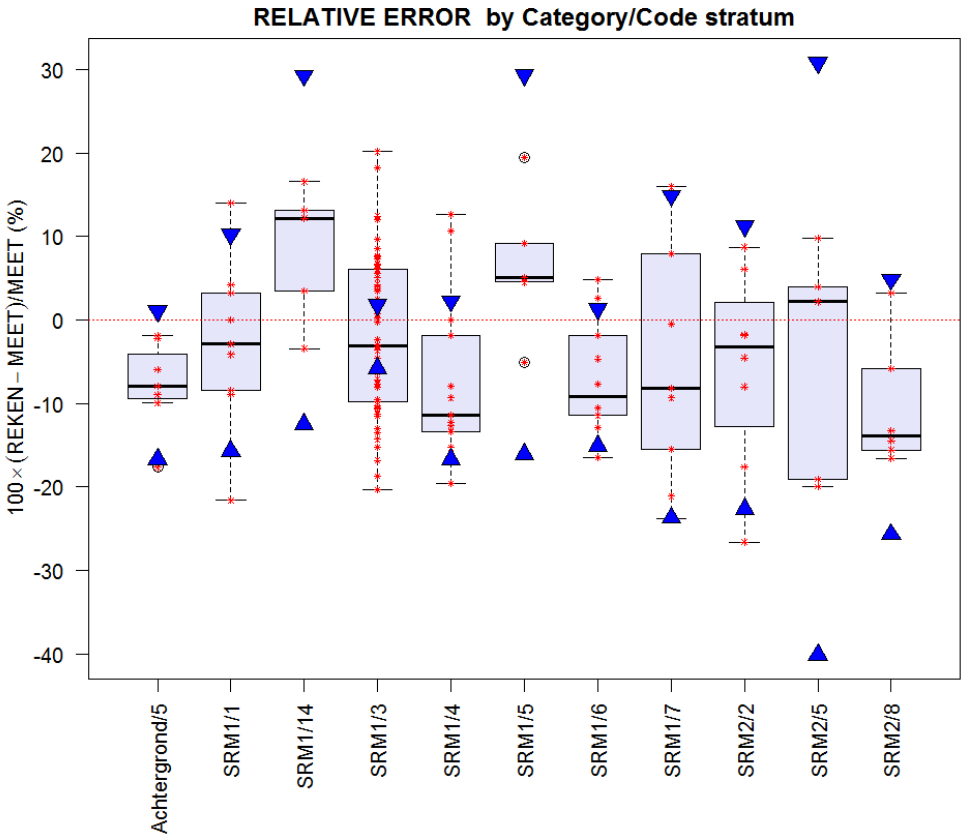

Figure 7: Box plots of relative error by Categorie/Code stratum; 95% confidence intervals for the population means are indicated by pairs of blue triangles. Only strata with at least five observations are considered here.

Another way of checking for patterns that may be explainable by Categorie and Code is to look at the distribution of relative error by Categorie/Code stratum. **Figure 7** indicates that the relative errors have roughly the same distribution, with a negative median near zero, in all strata, with the exception of those in stratum SRM1/14 (Categorie=SMR1 and Code=14), which seem to be concentrated around 10%. **Figure 8** provides an overview of the data labelled according to their levels of Categorie and Code; apart from the cluster of observations with Code=14 in the top left corner of the plot and the cluster of three observations with Code=11 in the bottom left corner, both of which are implausibly "far out", we see no other obvious signs of systematic deviations from the straight line. While an ANOVA test gives some evidence that the relative errors with Code=14 are out of tune with the errors from the other strata, when those data are

discarded the same test gives no evidence of a difference between the different populations of relative errors.

It thus seems that most of the data follow a straight line model that is almost a horizontal line, and that the data with Code=14, and probably those with Code=11, deviate systematically from it. Of course, it is important to try to find reasons—*other than statistical ones*—for the apparent 'lack of conformity' of these data. Assuming that there are indeed reasons to regard them as 'faulty' in some sense, it may be appropriate to fit the linear regression model only to the rest of the data. Instead of doing that, however, we propose to use a *simpler model* to describe the relative errors corresponding to measured concentrations $\geq 35\ \mu g/m^3$, namely a constant plus a normal random error (which corresponds to a linear regression model with slope equal to zero).

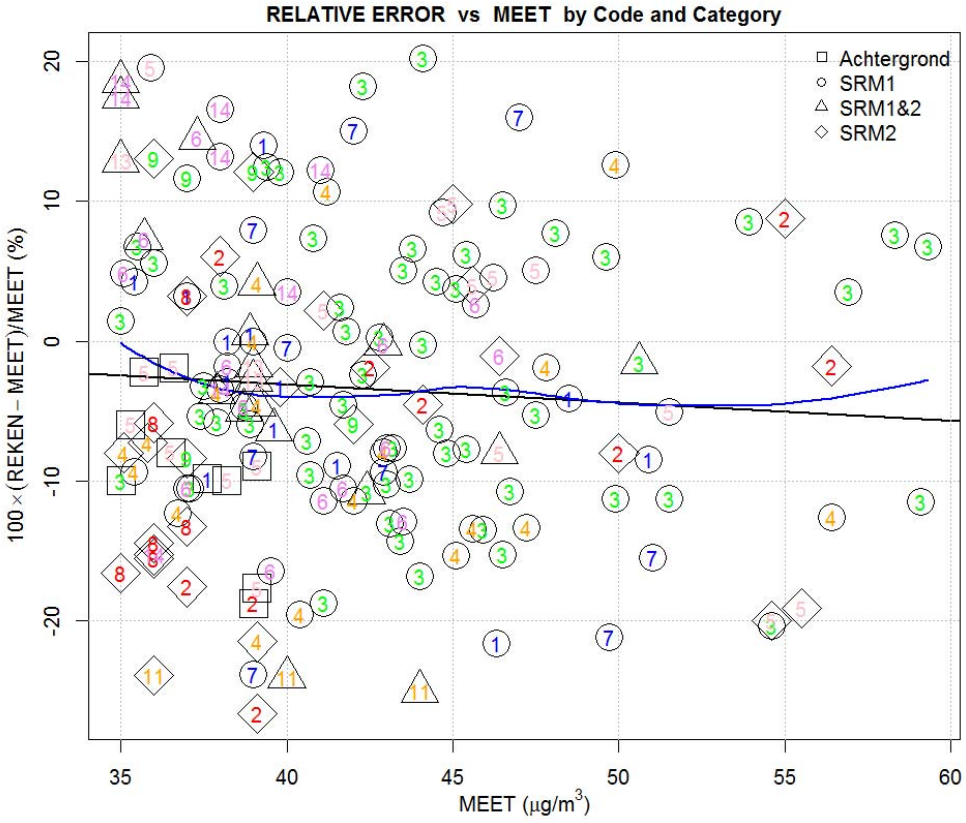

Figure 8: Scatter plot of relative errors versus measurements with fitted straight line and non-parametrically estimated regression function, indicating the levels of Categorie (by different symbols) and Code (by code numbers and colours).

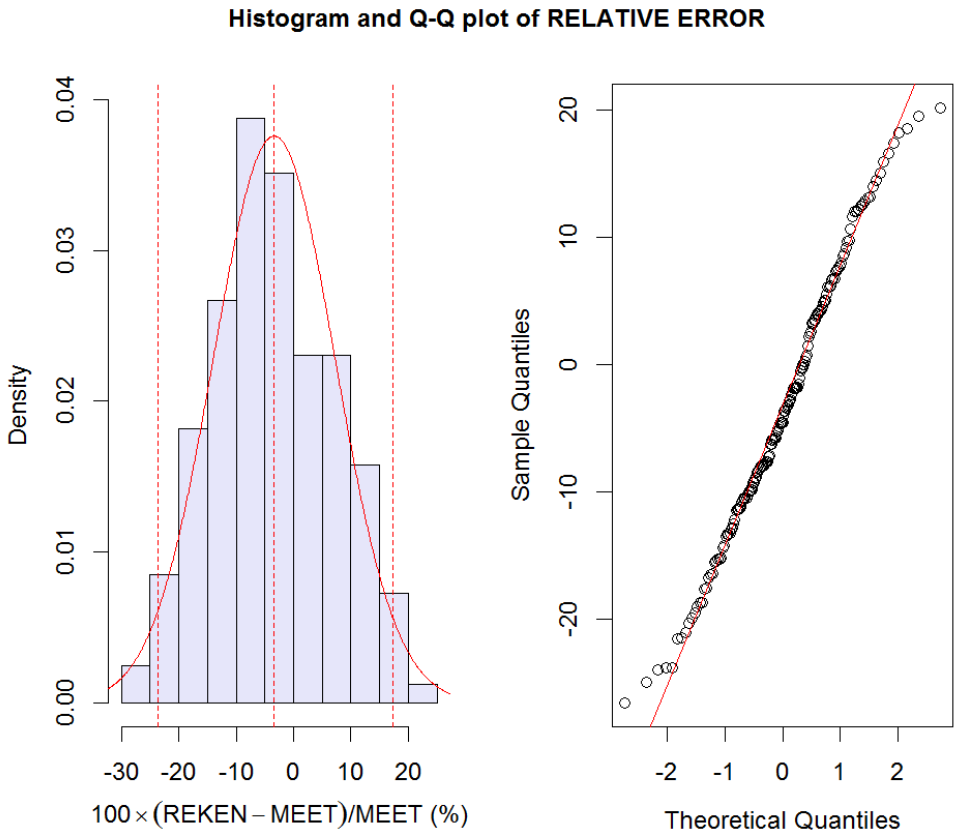

Figure 9: Histogram and normal Q-Q plot of the relative errors. Superimposed on the histogram are the fitted normal density and vertical dashed lines indicating the estimates of the mean and of the 0.025 and 0.975 quantiles.

To estimate this 'constant model' we simply take the sample mean of the relative errors as an estimate of the constant and the sample standard deviation as an estimate of the standard deviation of the errors; these are -3.56 and 10.61 $\mu g/m^3$ and, together with the normality assumption, yield the statements made in the introduction. **Figure 9** indicates that the model is approximately correct (Shapiro's test provides no evidence at all against normality); the vertical dashed lines superimposed on the histogram indicate the mean estimate, -3.56, and the 95% prediction interval for an arbitrary relative error, which is used in the second bulleted statement of the introduction.

For the sake of simplicity, the two bulleted statements in the introduction ignore the uncertainty about the constant. A 95% confidence interval for it is $[-5.14, -1.97]$; according to this, the mean underestimation of the predictions relative to the measurements may be as small as -1.97 $\mu g/m^3$ and as large as -5.14 $\mu g/m^3$ in the range of $M \geq 35$ $\mu g/m^3$.

Despite the difference in estimates, this model is rather close (in the range 35-60 $\mu g/m^3$) to the linear regression model presented above. Furthermore, if

the suspicious data with Code=14 are discarded then the regression line is hardly distinguishable from a horizontal line. Thus, it is of little consequence for purposes of making approximate statements about the relative error whether those statements are based on the regression model or on the constant model. We prefer the latter because of the simplicity of the statements it affords and of the greater robustness of its estimate as compared to the estimates of the regression model, which are somewhat sensitive to the inclusion or exclusion of the observations with Code=14. For completeness, however, the regression model is represented once more in **figure 10** with a 95% confidence band for the straight line and a 95% pointwise prediction band for the relative error associated with an arbitrary measurement (see **section 3** below for more details on these), from which approximate statements about the relative error can be made.

Finally, we note that *it is* of consequence that the data with Code=14 probably do not follow any of the two models, while the rest of the data probably do.

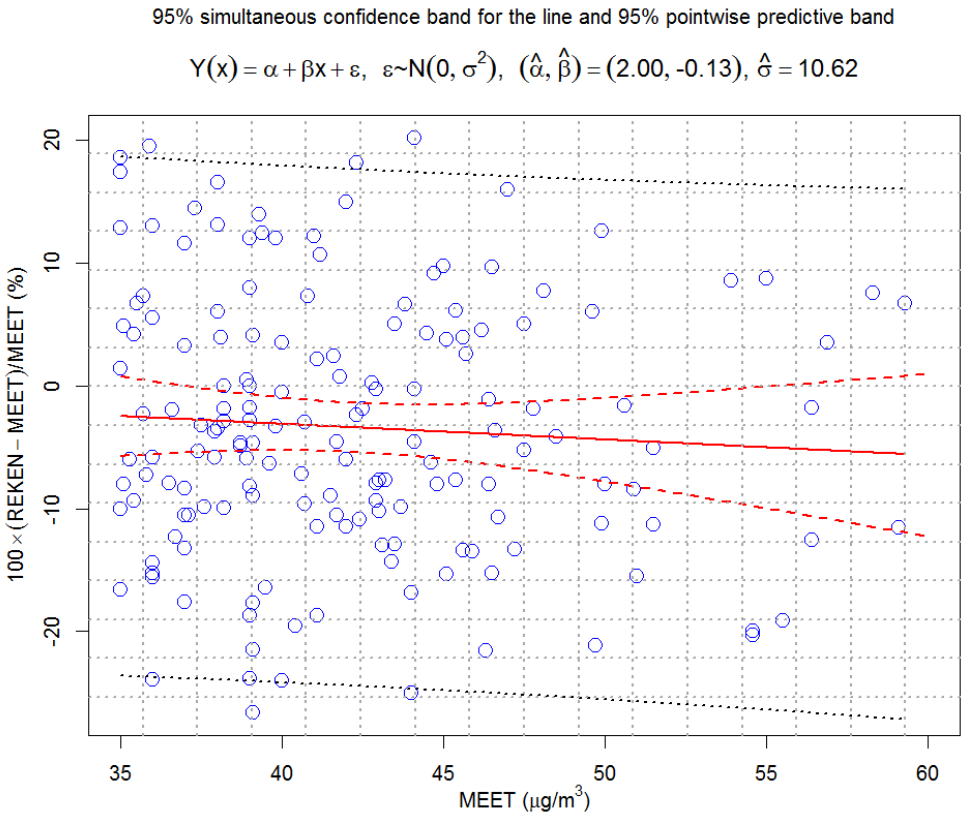

Figure 10: Fitted regression line, 95% confidence band for the true line (red dashed lines) and 95% prediction band for the relative error (black dotted lines), based on the same data used in the fit of **figure 6**.

3 Some details

The results reported here were obtained in R (R Core Team, 2012). In this section we describe briefly the computation of confidence and prediction bands and the prediction analyses mentioned in **section 2.**

The regression function of relative error on measurement is defined by

$$\varphi(m) = E\left[100 \times \frac{R\text{-}M}{M} \;\middle|\; M = m\right];$$

this is the expectation of the random variable $100 \times (R\text{-}M)/M$ conditionally on the event that the random variable M assumes the numerical value m. The non-parametric estimates of φ shown in **figures 1**, **3** and **4** were computed with the R function loess.

The calculation of non-parametric confidence bands for a continuous regression function is far from trivial. The difficulty is due to the requirement that *the whole curve* $m \to \varphi(m)$ be contained within the plane region delimited by *lower* and *upper* boundary curves $m \to \hat{\varphi}_L(m)$ and $m \to \hat{\varphi}_U(m)$ (the problem of finding a confidence *interval* for the *number* $\varphi(m)$, corresponding to a *single, fixed* m, is straightforward). The use of the bootstrap is probably the only general and feasible solution (cf. pp. 139-157 of Härdle (1990)). I have implemented a version of the bootstrap method to compute confidence bands of the form

$$[\hat{\varphi}(m) - c \cdot sd(\hat{\varphi}(m)), \hat{\varphi}(m) + c \cdot sd(\hat{\varphi}(m))],$$

where $\hat{\varphi}$ is the estimate of φ, $sd(\hat{\varphi}(m))$ an estimate of the standard deviation of $\hat{\varphi}$ at m, and c a positive constant determined by a bootstrap algorithm (c will typically be larger than the factor 2 that applies to a confidence interval for $\varphi(m)$ at a fixed m). Although this method is only approximate, some simulation experiments based on sample sizes of 400 indicate that the approximation it provides is rather good.

The prediction bands are computed by assuming that conditionally on $M = m$ the random variable $100 \times (R\text{-}M)/M$ has a normal distribution with mean $\varphi(m)$ and standard deviation σ (independent of m), estimating σ by a certain $\hat{\sigma}$ (obtained by averaging the estimates shown in **figure 2**), and then adding/subtracting $2\hat{\sigma}$ to/from the upper/lower boundaries of the confidence bands. The assumption of normality seems to be more or less realistic (e.g. **figure 9**), and so is the constancy of the standard deviation (e.g. **figure 2**). If anything, the prediction bands are somewhat conservative because they assume the "worst case scenarios" of the true φ being equal to the upper and lower boundaries of the confidence band (and, indeed, in **figure 1** only a couple of observations fall beyond the boundaries of the predictive band).

The confidence band for the regression line over the interval $[35,60]$ shown in **figure 10** was computed by a standard method explained on pp. 143-4 of Seber and Lee (2003).

Finally, let us say something about the prediction analyses mentioned in **section 3.** In our context, by a *prediction analysis* we mean (*i*) the construction of a *predictor* (a prediction algorithm) that predicts the relative error using the knowledge of the measurement and other predictor variables, such as Code and Categorie, (*ii*) the assessment of that predictor (namely in terms of how accurate the predictions are), and (*iii*) the ranking of the predictor variables in terms of their usefulness in predicting the relative error. Our analyses were based on *random forest* predictors, which are implemented in the R package randomForest. In essence, a random forest is a nearly unbiased and very flexible non-parametric regression model that describes a response (in this case the relative error) in terms of a set of predictor variables, and which therefore can be used to compute predictions for the former on the basis of the latter.

Figure 11 illustrates the results of two prediction analyses, one based on all the data and all the potential predictor variables, the other based only on data with MEET≥35 $\mu g/m^3$ and only on 'bona fide' predictor variables. We call *bona fide* predictor variables to all the predictor variables except REKEN_BIJ, MEET_BIJ and Background. These three variables represent numerical quantities which are involved directly in the computation of NO_2 predictions and hence are trivially related to the relative error, for which reason it would not make sense to use them for the creation of the strata mentioned in **section 3.** The left panels of the figure illustrate the agreement between the *actual* relative error and the *predicted* relative error and show estimates of the *mean square error* (MSE), *mean absolute error* (MAE), *bias*, and *proportion of explained variance*. In order to define these, let $\hat{Y}$ denote the *prediction* of the relative error, Y the *actual* relative error, and Var(Y) the variance of Y; then the MSE is the expected value of the random variable $(Y-\hat{Y})^2$, the MAE is the expected value of $|Y-\hat{Y}|$, the *bias* is the expected value of $Y-\hat{Y}$, and the *proportion of explained variance* is a standardized version of the MSE defined by 1-MSE/Var$Y$ (the closer this is to 1 the more accurate the predictions are). The right panels show graphs of 'variable importance', which provide a *relative* ranking of the predictor variables in terms of their usefulness in predicting the relative error.

As expected, the predictions based on all the predictor variables are quite accurate (0.82 explained variance), and REKEN_BIJ and MEET_ BIJ are by far the strongest predictors; if we remove these, then the quality of the predictions decreases very substantially. In particular, if only data with MEET≥35 $\mu g/m^3$ and bona fide predictor variables are considered then the proportion of explained variance gets close to zero; and if only data with Code≠14 are used then the proportion of explained variance is practically zero (0.07) and the relative ranking of the variables becomes completely irrelevant. This last observation supports the conclusion that in the range of measurements ≥35 $\mu g/m^3$ the relative error is essentially randomly distributed around a constant (≈ -3.56).

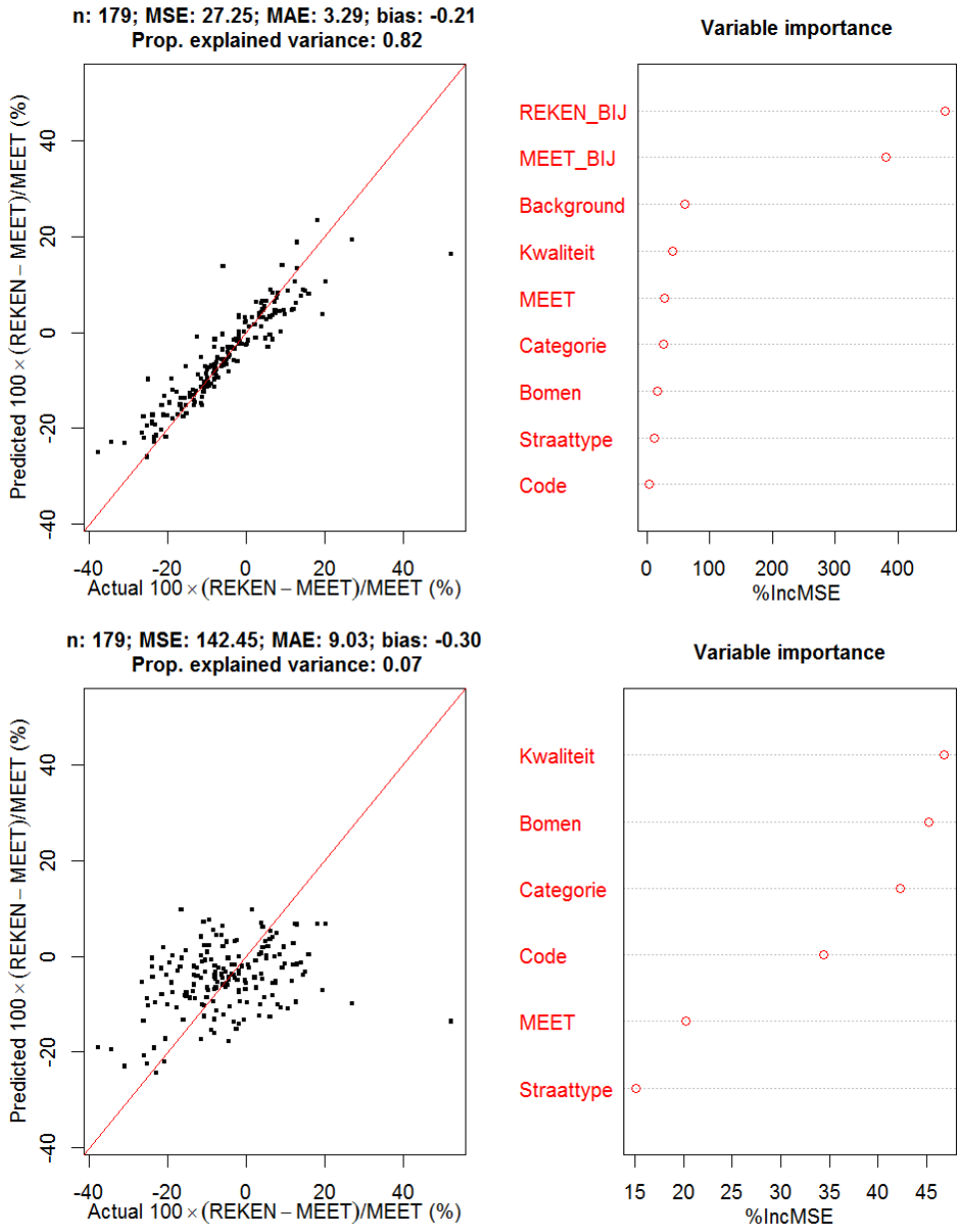

Figure 11: Results of some prediction analyses. Top: prediction of relative error using all the data and all the predictor variables. Bottom: prediction using only data with measurements $\geq 35\ \mu g/m^3$ and only the 'bona fide' predictor variables. The plots on the left compare the predicted with the actual relative errors; n gives the sample size; the other quantities are defined in the text. The plots on the right provide a ranking of the variables regarding their contribution in predicting relative error; the greater the importance of a predictor variable, the greater the percental increase in mean square error (%IncMSE) that results from 'confounding' that variable in the original data set.

4 References

Härdle, W. (1990). *Applied non-parametric regression*, Cambridge University Press.

Liaw A. and Wiener, M. (2002). Classification and Regression by randomForest, R *News*, Vol. 2, No. 3. (Available at http://CRAN.R-project.org/doc/Rnews/)

R Core Team (2012). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, url http://www.R-project.org

Seber, G.A.F. and Lee, A.J. (2003). *Linear regression analysis*, 2nd Ed., Wiley.

Wesseling, J., van Velze, K., Hoogerbrugge, R., Nguyen, L., Beijk, R. and Ferreira, J. (2013). Gemeten en berekende NO2 concentraties in 2010 en 2011, RIVM report.

National Institute for Public Health and the Environment
P.O. Box 1 | 3720 BA Bilthoven
www.rivm.com