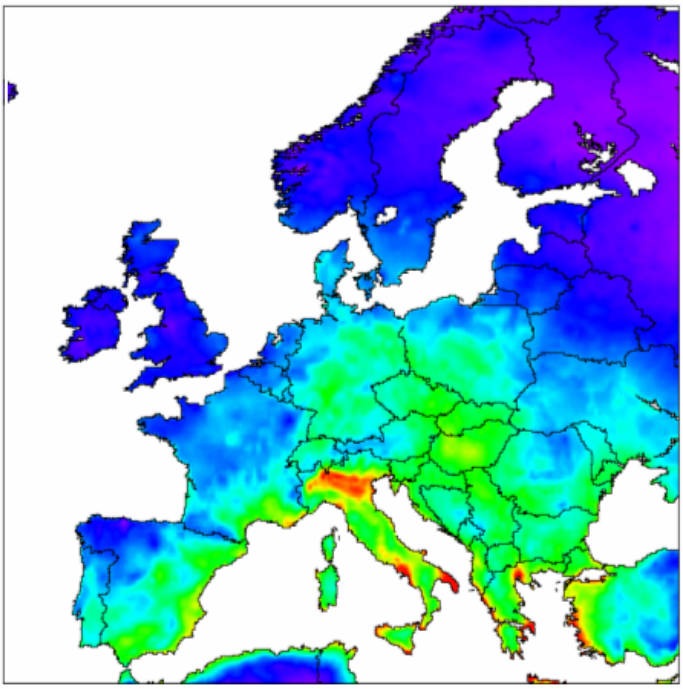
Formulation and quantification of uncertainties in air quality mapping



ETC/ACM Technical Paper 2011/9

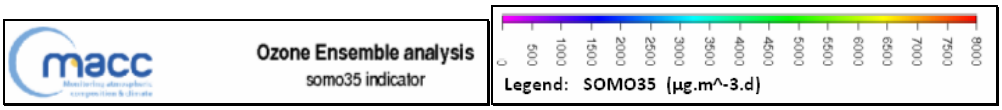
September 2011

Laure Malherbe, Anthony Ung,
Augustin Colette, Edouard Debry

European Topic Centre
on Air Pollution and
Climate Change Mitigation

The European Topic Centre on Air Pollution and Climate Change Mitigation (ETC/ACM)
is a consortium of European institutes under contract of the European Environment Agency
RIVM UBA-V ÖKO AEAT EMISIA CHMI NILU INERIS PBL CSIC

Front page picture:
SOMO35 ozone map for year 2008 (in µg.m^{-3}.d), obtained from an ensemble of analysed model simulations. This map has been produced as part of MACC project.



Author affiliation:
Laure Malherbe, Anthony Ung, Augustin Colette, Edouard Debry: L'Institut National de l'EnviRonnement Industriel et des Risques (INERIS), France

DISCLAIMER

This ETC/ACM Technical Paper has not been subjected to European Environment Agency (EEA) member country review. It does not represent the formal views of the EEA.

ETC/ACM Technical Paper 2011/9
European Topic Centre on Air Pollution and Climate Change Mitigation
PO Box 1
3720 BA Bilthoven
The Netherlands
Phone +31 30 2748562
Fax +31 30 2744433
Email etcacm@rivm.nl
Website http://acm.eionet.europa.eu/

CONTENTS

1. Introduction

Nowadays air quality mapping systems take different sources of data into account (monitoring, modelling). Mixing approaches are developed to combine chemistry-transport model outputs and monitoring data in a (geo)statistical framework.

In a previous ETC/ACC technical paper (Denby *et al.*, 2009), sources of uncertainty related to air quality mapping have been listed and described. Those are organized into a few topics: Monitoring, Modelling, Representativeness, Mapping methodology and Exposure and Heath.
In this new report, some of those items are discussed in the light of recent works conducted within European projects or linked with Air Quality monitoring and mapping systems (like Prev'Air, www.prevair.org). This report is neither a complete description of sources of uncertainty, nor a theoretical account of existing methodologies. It is more an analysis of how the methodologies and uncertainties are taken into account in recent papers/projects, leading to recommendations and propositions to come to more efficient and accurate mapping.
Section 2 is dedicated to in-situ measurements and the representativeness of a monitoring station according to the considered objectives/applications. Section 3 briefly addresses the question of measurement uncertainty and of its quantification. Section 4 deals with chemistry-transport models and recent studies devoted to the assessment of modelling uncertainty and variability. Section 5 is a review of mapping methodologies by kriging methods with focus on how uncertainties are involved in each step. Section 6 deals with analysis of sub-grid variability and exposure assessment and section 7 concludes with recommendations and propositions for further activities to come to improved mapping and assessment results**.**

2. Representativeness of air quality measurements

2.1 A polyvalent concept

Air quality measurement stations measure concentrations of pollutants and other chemical species in the atmosphere which may originate from local sources or be transported from long range sources. Networks of stations have been for decades a primary means for air quality monitoring and exposure assessment, and regulatory reporting on air pollution levels and exceedance is mainly based on measurement. Data from monitoring stations can t be coupled with information from complex air quality modelling systems to represent air pollution patterns. This can be helpful for understanding changes in air quality, determining the effectiveness of air quality regulations, and preparing adequate responses to new concerns. Finally, observation data are used for model verification and evaluation. Therefore, applications and uses of measurement data are various in the field of air quality and it is essential to know about the representativeness of the monitoring stations. It is actually the key concept which determines how to use the measurement data and how far we can rely on them.

One may accept that there are as many definitions of representativeness as uses. In section 2.2 four main uses of data from monitoring stations are distinguished and for each one the way of approaching representativeness is commented. In that part, representativeness is mainly considered from a spatial point of view. However, it is a complex concept which involves various aspects including temporal representativeness and measurement uncertainty. Those are addressed in sections 2.3 and 2.4 as well.

2.2 Definition of the representativeness according to data use

2.2.1 Reporting

For reporting purposes, measurement stations are used to assess compliance with air quality regulatory thresholds and identify areas in which air quality action plans are required according to the air quality Directives. As such, they should correctly account for pollutant concentrations in the surroundings or in geographical areas (or assessment zones and agglomerations) with similar characteristics as those where they are located. Depending on the station type, spatial extent for station representativeness, expressed in km^2 or km (case of traffic stations) is thus fixed by the Air Quality Directive 2008/50/EC (EU, 2008).
However, a major difficulty is that this concept is not explicitly defined in the European legislation which neither indicates for which time scale (averaging period) nor according to which criteria representativeness should be established. If the area of representativeness of a station is seen as a set of points (around the station or in similar locations) in which concentrations are “relatively close” to the station value, many questions arise: what does

"close" mean? For which pollutants? Over which period? How can the measurement uncertainty be addressed?

A concept for the representativeness of air quality monitoring stations has been developed in a study funded by the European Commission (UBA, 2007). In this concept, as summed up by a recent staff working paper from the European Commission (EC, 2011), the representativeness area of a station is defined in relation to a pollutant and an annual limit or target value according the following criteria:

- the concentration in this area stands within a certain range around the value measured at the station; propositions of ranges were made depending on the pollutants and the concentration thresholds;
- points included in this area belong to the same emission class as the station. A classification scheme was developed, based on the following main sources: local road traffic, domestic heating, industrial and commercial sources.
- dispersion conditions inside this area are similar. This criterion was not formulated in a quantitative way but a set of parameters related to climatic and topographic features and helpful for assessing similarity with regard to dispersion is provided in the report.

Besides, the area of representativeness is under average European conditions not much larger than an area with a radius of approximately 100 km around the measurement station. This constraint is based on average wind speeds and the residence time of air pollutants in the atmosphere related to the average transport distance of air masses over 12 hours; this time scale itself is related to the chemical conversion processes. More precisely, the maximal radius could be about 100 km in central Europe, higher in western and northern Europe and smaller in southern or south-eastern Europe (UBA, 2007).

Application of this methodology to an Austrian station is illustrated by Figure 1.

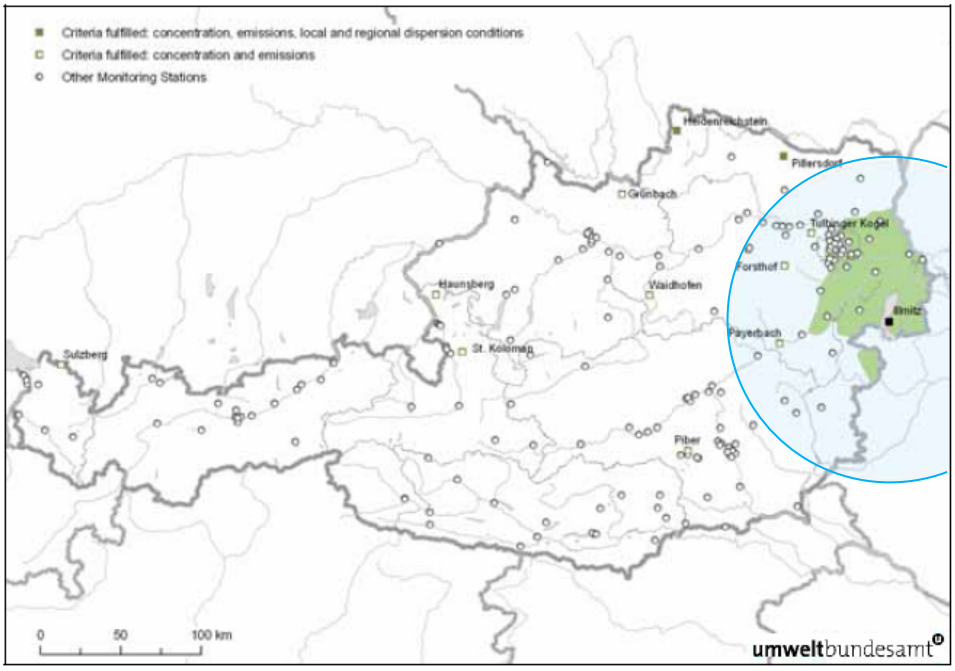


Figure 1 – Representativeness area (green) of the NO_x monitoring station located in Illmitz, Austria, depending on concentration, emission and dispersion criteria (UBA, 2007); (light blue circle represents the 100 km radial area).

Implementation of this methodology may be limited by the resolution of the available input data, in particular the resolution of concentration fields when those are provided by large-scale modelling. Results may also be sensitive to the definition of the emission classes used to characterize similarity in the emissions.

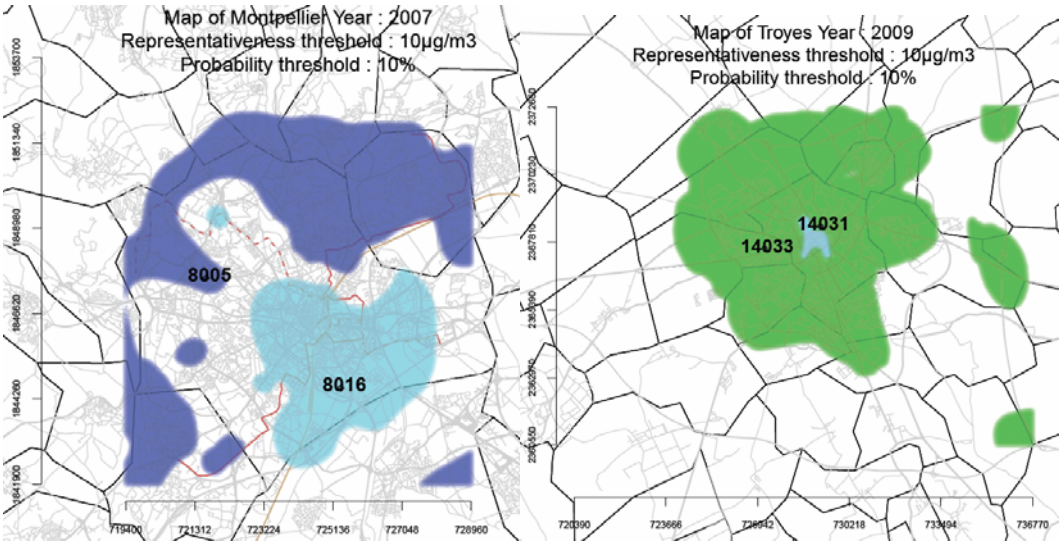
For pollutants like NO_2, measurement surveys by means of passive samplers may be an efficient way of collecting detailed information about concentrations around monitoring stations. Combined with other environmental variables they can be used as input to assess representativeness of each measurement site. A probabilistic methodology based on Bobbia *et al.*, (2008) was proposed for this purpose (Beauchamp *et al.*, 2011; Cárdenas and Malherbe, 2007). It consists of two main steps: 1) concentrations are estimated independently from the station measurements in a domain including the monitoring sites; multivariate geostatistics is applied to passive sampling data and high resolution ancillary variables; 2) concentration estimates are then compared to those measured by the stations taking the estimation uncertainty (kriging standard deviation) into account. Two parameters are set: δ, being a concentration criterion expressed as maximum authorized deviation from the station measurement; η, being the accepted risk that the true difference in concentration between a point and a station would be wrongly estimated at a value lower than δ. Making a hypothesis on the distribution of the estimation error (here assumed to be Gaussian), the condition for a point x to belong to the representativeness area of a station located at point x_0 is then:

$$|Z^*(x) - Z(x_0)| < \delta - \sigma_K(x) \cdot q_{1-\frac{\eta}{2}}$$

with $Z(x_0)$: measured concentration; $Z^*(x)$: estimated concentration; $\sigma_K(x)$: kriging standard deviation; q: quantile of the Gaussian law.

In the examples of Figure 2 (two French cities), application of the methodology resulted in two zones: one located in the city centre, in which NO_2 background concentrations are represented by the measurements of the central station; a peripheral one in which NO_2 background concentrations are represented by the measurements of a more outlying station.

Application of the methodology is dependent on some sampling characteristics. It requires that sampling data be numerous and well distributed in space. Otherwise locally higher estimation uncertainty may lead to unrealistic discontinuities in the representativeness areas. Besides, if the methodology is to be applied to larger domains than the urban areas considered here, criteria involving additional variables and distance may have to be introduced.



(source of measurement data: Air Languedoc-Roussillon and ATMO Champagne-Ardenne)

Figure 2 - Representativeness areas of NO_2 monitoring stations in the cities of Montpellier (purple area by station 8005; light blue area by 8016) and Troyes (green area by 14033; light blue area by 14031) for years 2007 and 2009 respectively. Assessment based on passive sampling surveys (Beauchamp et al., 2011). Representativeness threshold: maximum authorized difference in concentration with respect to the value measured at the station. Probability threshold: accepted risk that the concentration of a point included in the representativeness area of a station differs from the value measured at that station by more than the representativeness threshold.

In brief, representativeness for reporting is defined as a **geographical area** related to a pollutant and time scale.

Note that in both studies (UBA, 2007; Beauchamp *et al.*, 2011), sensitivity tests to the considered criteria were performed. The temporal stability of the representativeness areas was also examined. Those investigations should be pursued and the delimited zones be assessed against independent measurements wherever such data are available.

Representativeness for reporting thus appears as a geographical concept whereas representativeness for modelling and mapping applications is more a question of scale comparability between measurements and the studied phenomenon, as will be discussed hereafter. However, representativeness for reporting may be helpful in analysing representativeness for those last two uses.

2.2.2 Mapping

The definition of representativeness for mapping purposes differs from its definition for reporting purposes. Here the representativeness is more related to a selection of stations appropriate for mapping than to a precise geographical area.

For mapping purposes, measurement stations are used to build up the most reliable maps of pollutant concentrations by interpolating methods. The mapping process (based on kriging) is described in section 4. The first step is to define a set of measurement stations that will

provide input data for the kriging. This selection of stations will have an impact on the experimental variogram, the nugget, the interpolation and the resulting map. Therefore, it should be made in accordance with the geographical scale and spatio-temporal variability of the phenomenon to be mapped, with the spatial resolution of input variables (like results from chemistry-transport models) and the spatial resolution of the final representation (output grid or other output areas).

So far, representativeness for mapping has often been based on field expertise and usual station classifications like the AirBase one. AirBase (http://acm.eionet.europa.eu/databases/airbase) is the public air quality database system of the European Environment Agency. It contains air quality monitoring data and information submitted by the participating countries throughout Europe. This includes station metadata as the type of station and type of area.
A step forward is to consider results from more specific user-dedicated classification studies.

On regional scale (~10-50 km), "background" stations are often used in various European mapping exercises. In MACC project, two thirds of such stations were selected for mapping purpose (see red dots in the map in section 2.2.4), and one third for validation. The selection procedure was a random loop to put aside stations for validation. Only stations which were close (maximum of 100 km) to another one were removed. The remaining set of stations destined for mapping offers a good spatial distribution over the gridded map.
Further distinction can be made between rural and urban/suburban background stations to account for different spatial variability in rural and urban areas (Horálek *et al.*, 2010).

In Prev'Air system, "representative" stations for mapping background ozone on French domain were selected according to the recommendations of the local measurement operators. Most of them are rural and suburban background stations; some urban background and industrial stations were also included. All those measurement sites are designed to capture ozone concentrations prevailing over distances that are comparable to the chemistry-transport model (CTM) and output grid resolution (approx. 10 km).

A few recent studies are also reported to help in refining the definition and selection of "representative" stations for mapping:

- In Henne *et al.* (2010) and GEOMON project report (Figure 3), the AirBase classification was reviewed and the authors proposed new classes to characterize the air quality measurement sites. The robustness of the site classification was tested by modifying different parameters used in the clustering procedure and the resulting groups did not change.
- In Peuch and Joly (2011), a classification of stations relevant for model validation and data assimilation was proposed, based on the statistical behaviour of temporal historical measurements.

Comparison and assessment studies about the impact of selecting different sets of "representative" stations according to new station classifications are still needed.

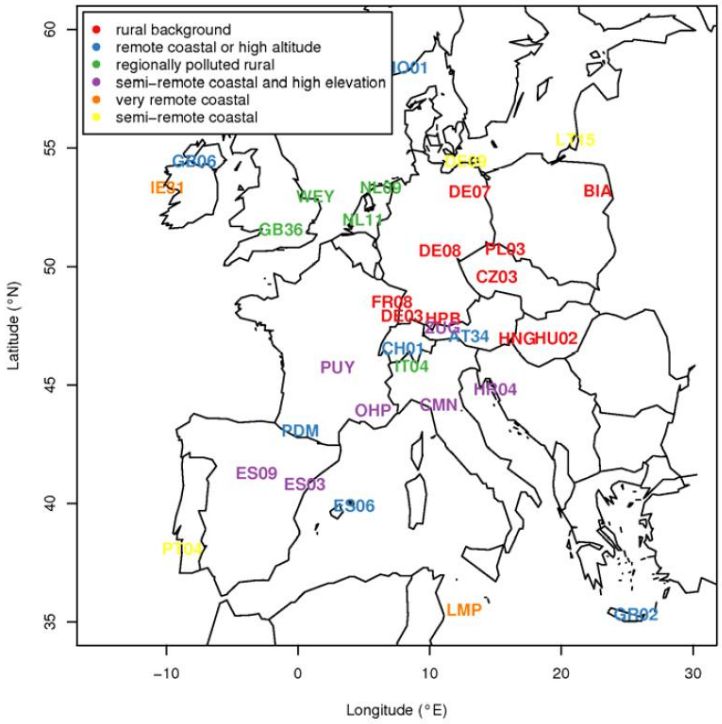


Figure 3 – Site classification in GEOMON project: Ward's clustering method (Ward, 1963) applied to selected sites for detailed assessment of representativeness.

To sum up, representativeness for mapping is the **ability of a station to capture a pollution process intended to be mapped with a desired spatial resolution.**

2.2.3 Data assimilation in models

Representativeness for data assimilation is different from the concepts described in 2.2.1 and 2.2.2. Those first two concepts mainly refer to similarity in absolute concentrations, leading to the delimitation of an area or a selection of stations. Here representativeness is related to the spatial properties of the CTM errors.

For data assimilation purposes, measurement stations data are assimilated into CTM to create the best analysis of a system state by combining incomplete or inaccurate measurements with imperfect simulations from a chemistry-transport model at the regional scale (GEMS project, 2010) or dispersion models at a lower scale. The data assimilation system performs a global time-space adjustment of the model to all selected observations (Tombette *et al.*, 2009; Wu *et al.*, 2008; Elbern and Schmidt, 2001). It minimizes a cost function penalizing the time-space misfits between the data and model values.
Particularly important is the construction of the background error covariance matrix, which contains information about the magnitude of the background errors (errors in the background field estimate) and about their spatial correlations (Singh *et al.*, 2011).
The correlation lengths in longitudinal and latitudinal directions indicate how far information spreads from an observation point to a finite domain surrounding it. A correct specification of those parameters is important for a good performance of the data assimilation system.

In that context, station representativeness describes the **knowledge of the correlation lengths** used to build the error covariance matrix.

2.2.4 Evaluation of model and mapping outputs

Two kinds of evaluation can be performed: an evaluation of the end products issued from the modelling process (raw simulation results) or an evaluation of the results issued from the mapping or analysis process through data assimilation. Once more, the set of observation data selected for model or mapping verification should be representative of the model resolution and be defined according the objective of modelling (Moussiopoulos and Isaken, 2007; AIR4EU reports). For instance, for evaluation of the modelled background air quality values, only background stations should be selected.

The question is also sensitive when evaluation of a mapping process is considered. Observation data that are included in the process should not directly be used when the quality of the results is verified. Generally, two sets of stations are defined (Figure 4). The first set is used as input in the mapping. The second set has to be located in the influence area of the first set. Figure 3 is taken from the MACC project (http://www.gmes-atmosphere.eu); validation stations were selected as close as possible to assimilated stations. In areas with low density of measurement sites, such as Scandinavia, stations were selected for mapping as a priority.

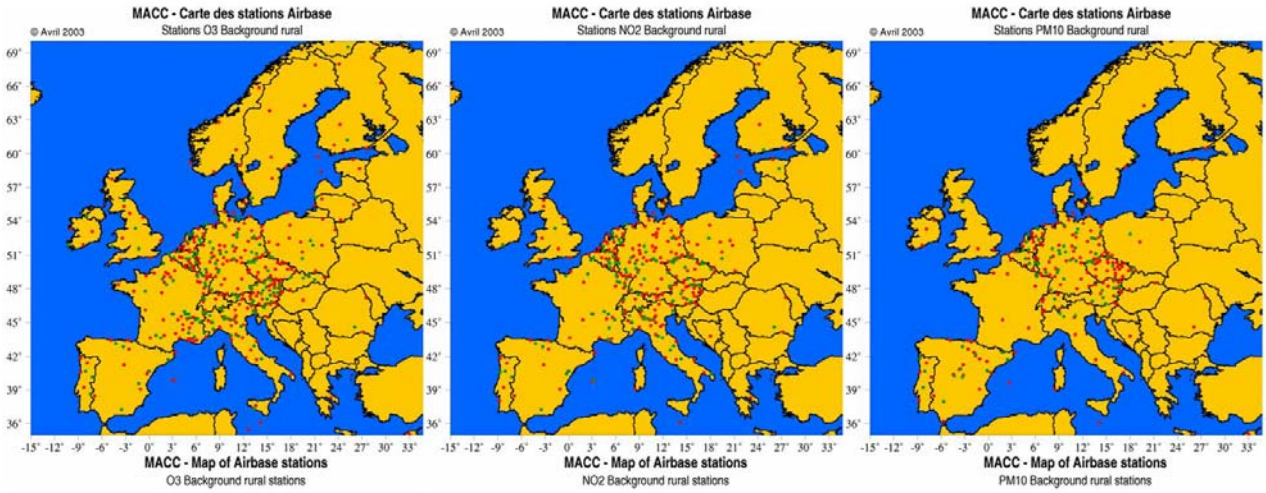


Figure 4 - AirBase background stations used in the MACC project for the provision of air quality analyses and re-analyses services: stations for assimilation (red dots) and stations for validation (green dots).

The definition of the representativeness here is not far from section (2) but this selection of station should be evaluated the assimilation process close to assimilated station.

As in section 2.2.2 representativeness for evaluation corresponds to the ability of a station to catch a pollution phenomenon modelled or mapped with a given spatial resolution. If the object being evaluated is the output of a mapping or data assimilation process (and not just

the raw CTM output), this definition has to be completed. It should include ability to assess the impact of that process around the assimilated stations.

2.3 Representativeness: a pollutant and time-dependent concept

The concept of representativeness for reporting as described in section 2.2.1 refers to a spatial knowledge of the station in relation with a pollutant and concentration statistics (and therefore with an averaging time). Dependency on pollutant and time is also true for the other uses. For example, a station flagged as "background" for NO_2, which is considered as a "local" pollutant due to its lifetime, might not be as relevant for mapping background concentrations of O_3 which is a long-lived species. Besides, given the diurnal or annual variations of pollutant concentrations and their changing spatial variability (as can be inferred from the variogram), the spatial representativeness of a station varies in time as well. For instance, during the night or the winter season, O_3 concentrations have a low spatial and temporal variability, which tends to increase spatial representativeness for that pollutant. Representativeness is therefore a space/time/pollutant dependent concept.

In section 3.1 mapping methodologies that take spatial representativeness into account to produce daily or hourly maps by kriging methods are discussed. De Fouquet *et al.* (2011) compared, for ozone, the temporal variability of the model outputs and the station data through the analysis of temporal variograms. Additional studies would be necessary to fully describe the link between spatial and temporal representativeness. Parallel to this paper, Gräler *et al.* (2011) explored options to describe such spatio-temporal relation for PM_{10} measurement data stored in AirBase. Interesting advantages with regard to mapping performance and knowledge of stations can be driven from a spatio-temporal approach but the conclusions of this work point out that the results are strongly related to the spatio-temporal dependence structure of the pollution phenomenon.

3. Uncertainty of the measurements

Assessing the representativeness of a station on the basis of concentration levels only makes sense if the measurement data are reliable. The more uncertain they are, the more difficult it is to characterize station representativeness in space and time.

To evaluate the performance of the measuring devices, the uncertainty of the measurements can be approached following two main ways:

- an approach purely based on calculation, known as the GUM (ISO/IEC Guide 98-3:2008 to the expression of uncertainty measurement), which consists in combining uncertainties arising from different factors (calibration, drifts, interferences, etc).
- an approach based on direct experimentation through inter-comparison exercises (several measuring devices operating in parallel, see for example AQUILA activities, http://ies.jrc.ec.europa.eu/aquila-homepage.html).

Several norms describing uncertainty calculation for air quality measurements exist but careful reading reveals that they are not easy to implement and can be interpreted in different ways (Macé *et al.*, 2009a). Indeed there are some gaps and unsolved questions that need to be filled before implementation. For example, if the national calibration chain consists of several levels of comparison, from the national standard to the *in situ* analyser, the way of calculating uncertainty due to calibration requires additional details. The determination of some performance characteristics should be detailed as well if the user wants to include the performances of his own analyser and not those derived from standard approval or certification. Some other issues are: definition of relevant variation ranges for influent parameters and interfering species; how to include uncertainty due to the acquiring system and the rounding; how to take missing data into account when global uncertainty of a temporal mean has to be calculated.

With such a consideration, guidance reports have been published in France to facilitate the assessment of measurement uncertainties for all regulated pollutants, including those of the Directive 2004/107/EC (EU, 2005). They are based on existing documentation and methodologies and more particularly on the methods proposed by the European standardization groups CEN TC 264/WG12 and CEN TC 264/WG13. They also explain the way of assessing measurement uncertainties from the results of intercomparison exercises (Macé *et al.*, 2009b).
Those reports have been released as AFNOR documents (2007, 2008, 2011). They are currently undergoing revision.

Concrete results from interlaboratory comparison exercises carried out by JRC can be found in reports accessible or referenced on the following page: ftp://ipscftp.jrc.it/erlap/ERLAPDownload.htm. Quantitative information about measurement uncertainties is provided for different pollutants like SO_2, O_3, NO, NO_2, CO (*e.g.* Belis *et al.*, 2009a and/or 2009b); VOC (Pérez-Ballesta *et al.*, 2008), heavy metals (Gerboles *et al.*, 2011).

As a third approach, Gerboles and Reuter (2010) suggested using the spatial variability of the observation data to get information about measurement errors. Their argument is that the nugget value of the variogram provides an upper limit to the measurement error variance. Indeed it also includes a micro-scale variability that can be minimized by an accurate selection of stations. Tests were performed for several European countries using AirBase data. Daily variograms were computed, giving daily estimates of the measurement error variance. According to the authors, the obtained figures are consistent with expectations from laboratory and field estimation of uncertainty. The results suggest that measurement uncertainty is best estimated using all background stations of whatever area type. Further investigations could be carried out. In particular, the variogram nugget provides average information about the measurement error variance; it does not tell about the local variations of that term, especially if the error variance depends on the concentration level.
A potentially more stable procedure would be to use temporal variability, as in de Fouquet *et al.* (2011), possibly in combination with spatial variability (Gräler *et al.*, 2011).

4. Uncertainties and variability in chemistry-transport models

4.1 Monte Carlo methods to assess model uncertainties

Uncertainty of a dispersion model may arise from:

1. input data, among which mainly emissions, boundary conditions, and meteorological fields,
2. physical and chemical parameterizations, which generally constitute the core of the model,
3. numerical schemes for solving numerical equations.

Uncertainty studies are mostly focused on uncertainties arising from input data and model parameterizations.

The third point cannot be easily addressed together with the first two ones, since numerical schemes are usually hard-coded in the models and offer only few options for changes. So far, most studies dealing with numerical schemes have been limited to sensitivity studies and 0D test cases. Uncertainty due to numerical schemes in three-dimensional models is still an investigation field. Recent developments in that direction based on ensemble generation can be found in Garaud and Mallet (2010, 2011).

As regards point 1 and point 2 there are several methods to assess such uncertainties. Under given assumptions an adjoint model may be derived from the model using differentiating tools. It directly provides the sensitivity of modelled concentrations to given parameters. Monte-Carlo methods are based on the analysis of numerous model results obtained by randomly perturbating a given set of parameter values. They seem to be so far the most widely and commonly used techniques as they are relatively easy to implement and free from any limiting assumptions. Once the Monte Carlo process has reached a steady state, the model results enable to compute the standard deviation of concentrations with respect to parameter perturbations.

As part of the EC4MACS project (EU LIFE, www.ec4macs.eu), which aims at building a suite of modelling tools for a comprehensive integrated assessment of emission control strategies, a Monte-Carlo methodology has been implemented (Debry *et al.*, 2010). It was used to propagate the uncertainty of input parameters into the chemistry-transport CHIMERE model (French domain, resolution: 0.15° x 0.10°), providing the uncertainty of predicted concentrations. Considered pollutants were mainly ozone during summer and particles during winter. Main results delivered in this study were estimations for absolute and relative uncertainty of both considered pollutants (Figure 5 to Figure 6), and the list of input parameters to which they are most sensitive.

All parameters for which the uncertainty was reasonably well known were perturbed, to name but a few: emission, boundary and initial conditions, meteorological fields, chemical reaction constants, wet and dry deposition. The perturbation was applied as a given percentage relative to the value. In view of computational time resources, simulations were

performed for one summer month (August 2008) dealing with O_3 and for two winter weeks (the first two weeks of January 2009) as for PM_{10}. In the results the absolute and relative standard deviations usually stand for relative and absolute uncertainties respectively.

According to this study, influent parameters for both pollutants are temperature and deposition velocity. Influent parameters specific to O_3 concentrations are boundary conditions and chemical reaction rates, whereas those peculiar to PM_{10} concentrations are the fine fraction of carbon emissions and specific humidity.

Considering that not all parameters could be perturbed (e.g. numerical schemes) and that the uncertainty of some input parameters is still poorly known, the current estimates of uncertainty should be considered as bottom estimates, especially for PM_{10}.

For hourly O_3 concentrations averaged over the summer period, the average absolute and relative uncertainty provided by Monte-Carlo simulations is 16 µg/m^3 and 21% respectively. Relative uncertainty is mainly driven by the concentration level, decreasing when concentration increases, so that it is lowest at O_3 peak locations. It does not change significantly from one urban area to the other.
For hourly PM_{10} concentrations averaged over the winter period, the average absolute and relative uncertainty is 4 µg/m^3 and 17% respectively. Contrary to O_3, relative uncertainty depends not only on the concentration level, slightly decreasing in case of high pollution, but also on the topography and the vicinity of local sources. Uncertainty in chemistry-transport modelling for simulating PM_{10} concentrations is therefore higher in urban areas.
In the performed simulations (note that the simulation period was characterized by high PM_{10} pollution events), it was not able to explain the bias due to a lack in the emissions or to missing chemical processes for particles.

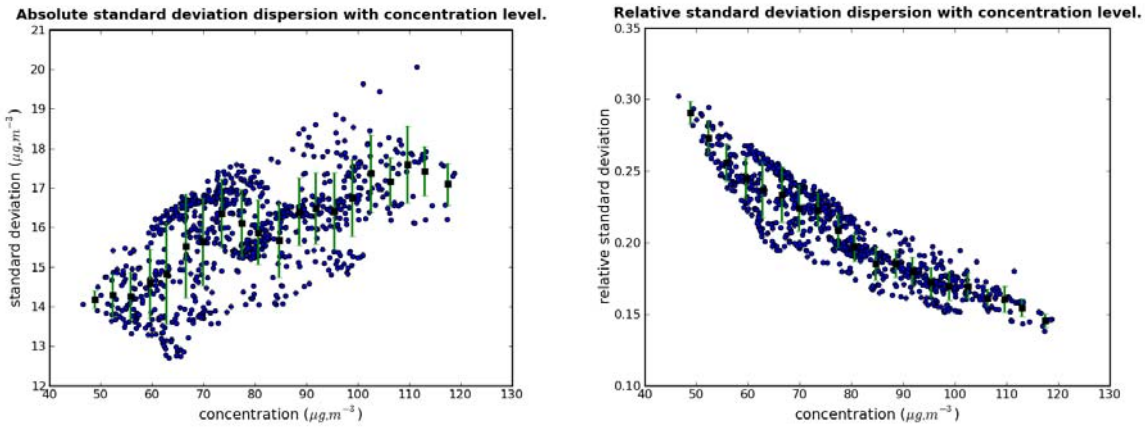


Figure 5 - Absolute and relative uncertainty as a function of concentration level. Simulations performed over the French domain for O_3 in summer.

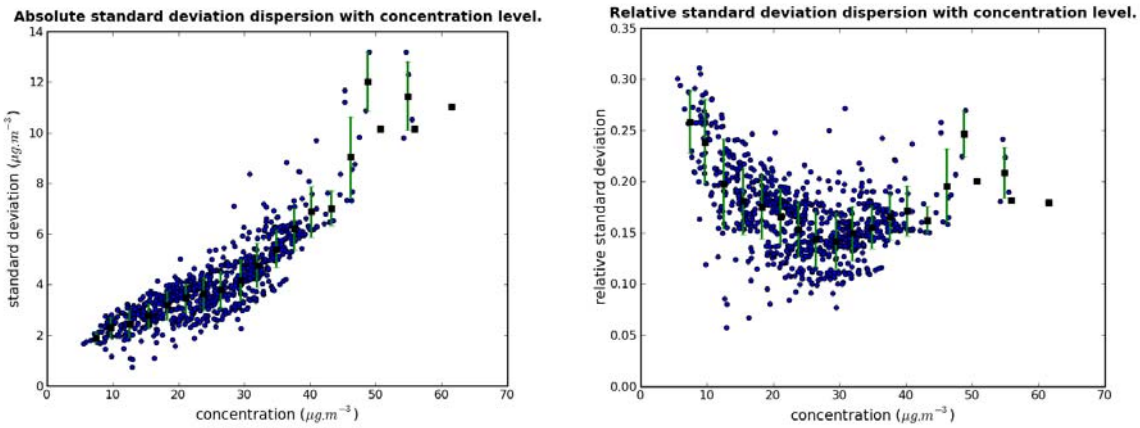


Figure 6 - Absolute and relative uncertainty as a function of concentration level. Simulations performed over the French domain for PM_{10} in winter.

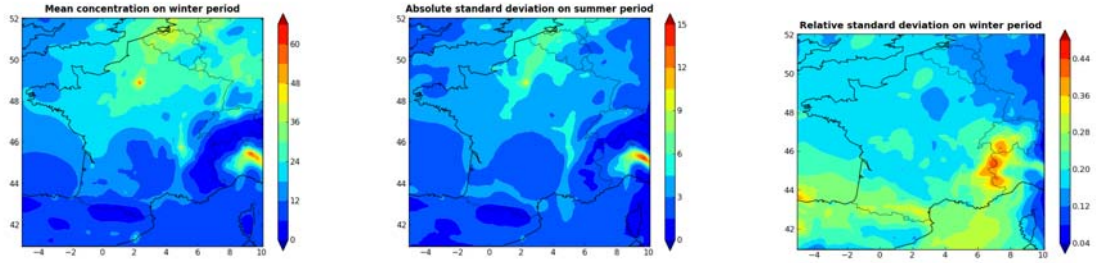


Figure 7 - Left the average PM_{10} concentrations ($\mu g/m^3$) calculated by CHIMERE over a winter period, and the absolute (centre) and relative (right) standard-deviation derived from 300 Monte-Carlo simulations.

4.2 Ensemble methods to assess model variability and parameterization

Within the framework of MACC project, a service provides a posteriori validated air quality assessments for Europe, based on re-analysed air pollutant concentration fields from chemistry-transport models (http://www.prevair.org/macc-reva). Model outputs are provided by the seven modelling teams involved in the regional operational production. The report "Evaluation Report of the Air quality assessments in Europe for 2007" (Rouïl et al., (2011a), http://www.gmes-atmosphere.eu/documents/deliverables/r-eva) contains a number of scores that allow the qualification of skills and performances of the MACC products. A systematic evaluation of the models and data assimilated results against a relevant set of dedicated observation data (not assimilated in the models) is done with AirBase database. This validation set is made of "background" stations located close to "background" stations used for analysis or assimilation. To overcome the variability of the models, an ensemble model is computed as the median of all the models. Two ensemble models are actually considered: ENS, which is the median of raw model results, and ENSa, which is the median of assimilated or analysed model results.

Figure 8 is an example of ozone time series and computed statistics for a rural station in Germany. It shows that the range of variability of the model results can spread out significantly for a given station.

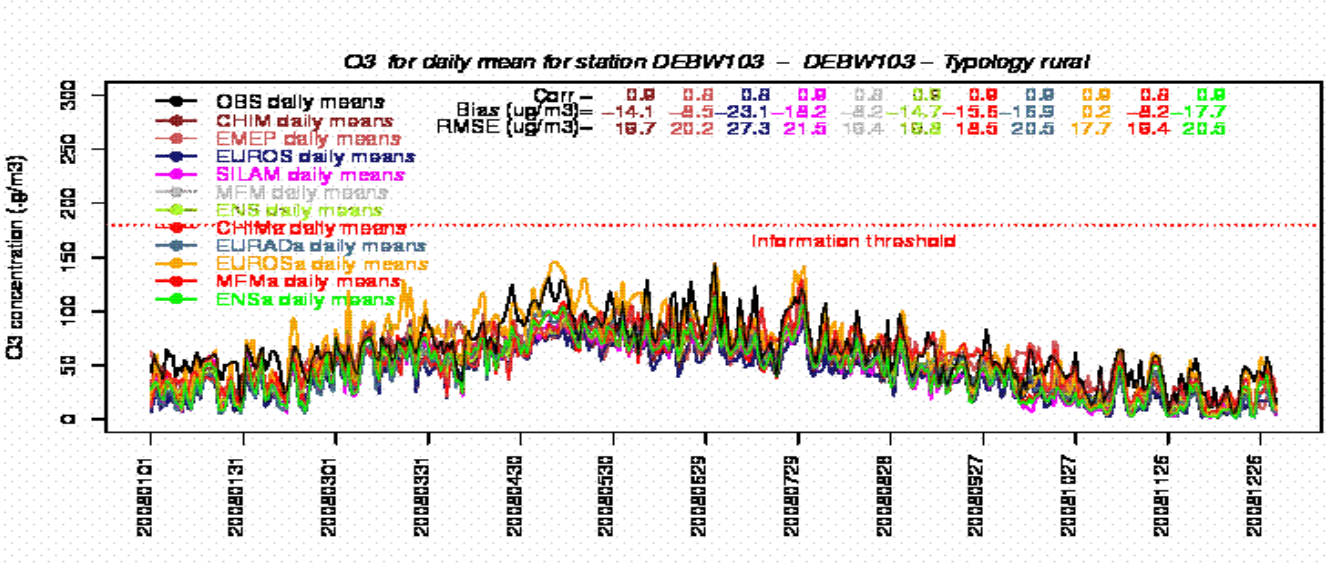


Figure 8 - Multi-model validation on station DEBW103 for year 2008. OBS : measurement, CHIM : Chimere model, EMEP : EMEP model, EUROS : EUROS-LOTOS model, SILAM model, MFM : MOCAGE without assimilation, ENS : ensemble mean of non assimilated models, CHIMa : Chimere with assimilation, EURADa : EURAD model with assimilation, EUROSa : EUROS-LOTOS with assimilation, MFMa : MOCAGE with assimilation, ENSa : ensemble mean of assimilated models.

Statistics can be plotted to have a quick view of the spatial variability of the skill scores (Figure 9) and help the modellers to identify the most uncertain regions in terms of modelling and mapping.

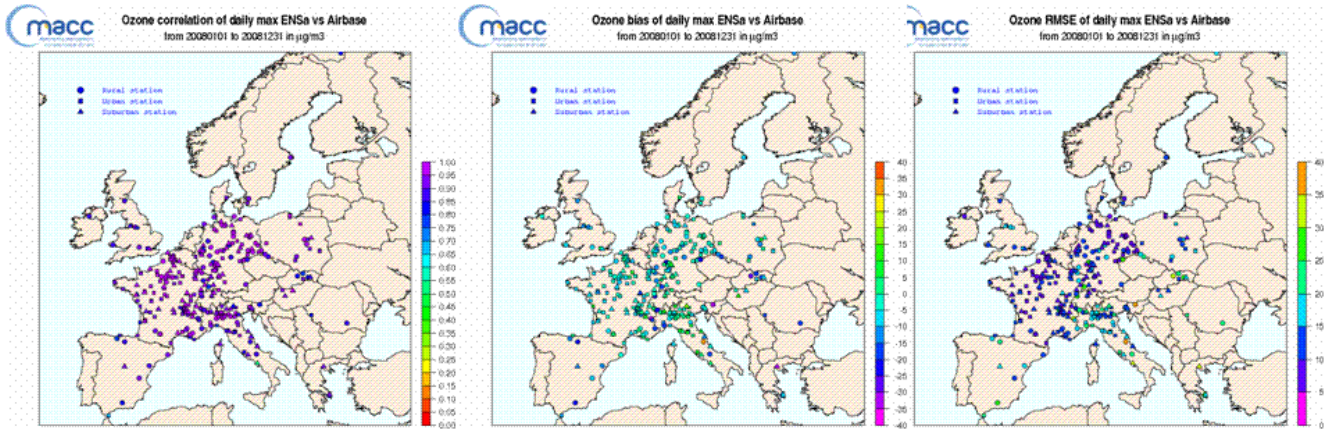


Figure 9 - Ensemble of assimilated models: Correlation (left), bias (centre), and RMSE (right) with validation stations for maximum daily ozone concentrations.

The Europe domain is split into 5 sub-domains: EUW: Western Europe, EUC: central Europe, EUS: Southern Europe, EUN: Northern Europe, EUE: Eastern Europe. This multi-model highlights the most sensitive geographical areas where the ozone peak re-analyses are the most uncertain (Figure 10). The Root-Mean Square error (RMSE) indicator can differ largely from a sub-region to another. It is lowest in Northern, Western and central Europe,

ranging from 10 to less than 20 µg/m^3 for the best models. As expected, all models show poorer performances in Southern Europe: complexity of the meteorological patterns and topography, uncertainties on some sources (for instance biogenic sources) and uncertainties related to some chemical mechanisms could explain this “well-known” limitation of the

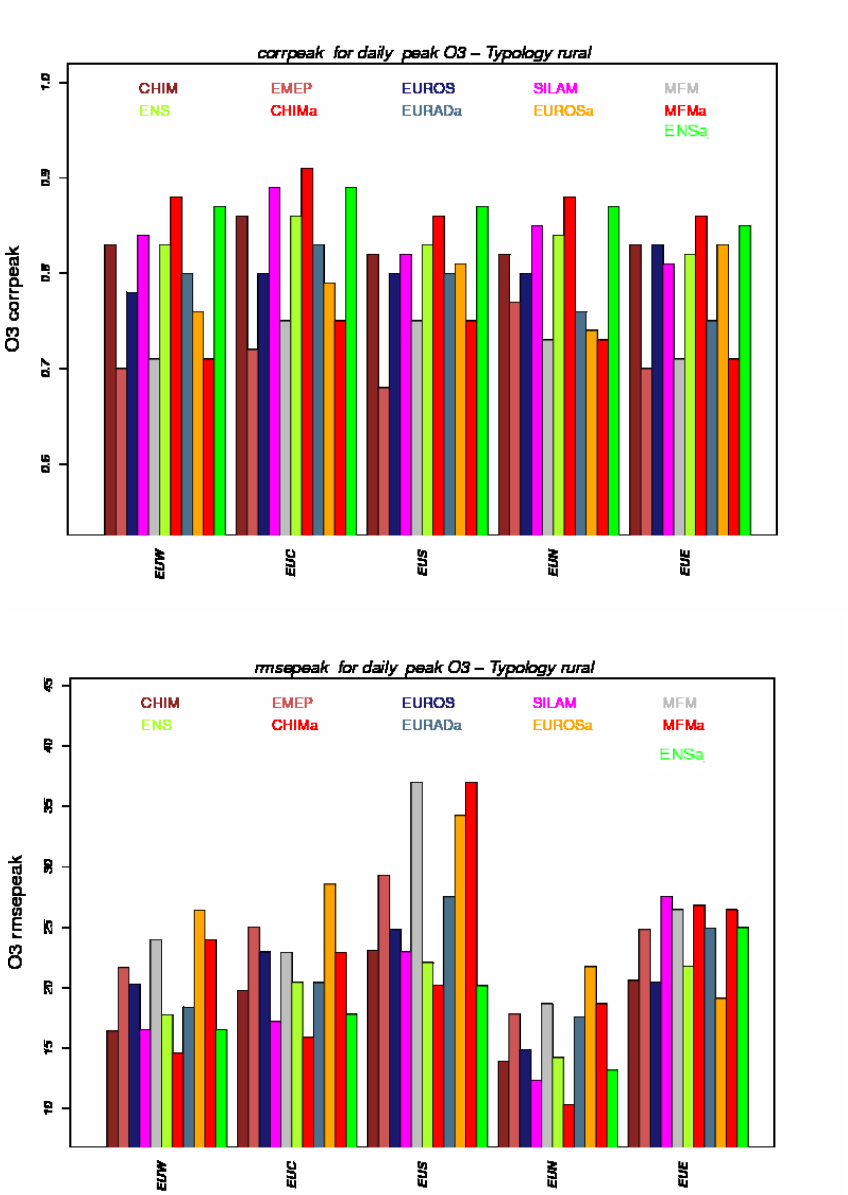


Figure 10 - Correlation (top) and RMSE (bottom) coefficients for the daily ozone peak for all ensemble members for 5 sub-regions EUW, EUC, EUS, EUN and EUE (left to right on the x-axis).

current modelling systems. Results show that the best model (better correlation with lower root mean square error RMSE) is obtained by analysis of CHIMERE model (CHIMa in red; central red bar on the plots). The ensemble model computed as the median of assimilated or analysed model results (ENSa, in dark green color – right green bar on the plots) performs almost as well as the best model (CHIMa). Compared to CHIMa it exhibits slightly lower correlation and higher RMSE except for Southern (EUS) and Eastern Europe (EUE) where it shows lower RMSE. In those regions where the individual models perform less efficiently

general patterns are correctly represented by the ensemble model: the episode situations are generally predicted but their intensity is underestimated. In conclusion, the ensemble model (median of the models) is a good candidate to take all the models into account and keep good scores.

This assessment exercise is to be reconducted every year in order to check the performance of each individual model and of the ensemble forecast.

4.3 Trend analysis

In the CITYZEN project (www.cityzen-project.eu), the capability of chemistry-transport models (CTMs) to capture the observed trends of major atmospheric pollutants has been assessed (Colette *et al.,* 2011). The purpose of the study was to investigate past modelled trends in air pollution hotspots in order to demonstrate the potential and limitations of existing models for assessing the impact of future air pollution control strategies.
A numerical experiment was set up involving six chemistry-transport models that reflect a variety of approaches: regional or global coverage, online or offline chemistry and transport coupling. Only anthropogenic emissions were prescribed uniformly for all models while the choice of remaining forcing data (meteorology, biogenic emissions, boundary conditions, etc.) was left open. The idea was to have an ensemble of models that best represents CTM ability to capture air quality trends. As such, the experiment was not a model inter-comparison initiative but an attempt to assess the uncertainties in air pollution trend modelling. Moreover, it was an opportunity to perform a multi-annual model evaluation.

The study was focused on the European scale, background stations and on aggregated metrics such as daily and monthly means over a period of ten years, from 1998 to 2007. Observation data were extracted from AirBase.

One striking result is the consistency of model performances between regional and global CTMs induced by the scope of the study and the use of a common emission inventory. Further analysis indicates that the ensemble of models covers a wide envelope of behaviours. The ensemble shows a relatively large spread, with a wide range of biases compared to the measurements, and encompasses the observed values. Those results give confidence in the representativeness of this set of models with regard to current scientific knowledge and existing tools.

The CTMs proved to be quite successful in capturing the decreasing concentration trends of primary pollutants, especially in the emission hotspot areas around the Benelux region.

For NO_2, the main feature is a decrease from 1998 to 2007 over most of Western Europe (Figure 11), more specifically United Kingdom, Germany, Benelux and Italy except France and Spain, reflecting the trend of primary emission reductions reported in the inventory. By contrast NO_2 tends to increase over the main ship tracks. Those dominating patterns are consistently accounted for by all models. In particular, compared to AirBase data, downwards trends of NO_2 are successfully captured at 68% to 80% of the stations depending on the model.

For O_3, the strongest pattern is an increase of daily O_3 in the Southern UK, Benelux and Germany (Figure 12). This behaviour is reproduced by all models and is also detected in the observations, hence demonstrating the robustness of this statement. Over France, the observed trend is very noisy for suburban and rural background stations. Over Northern Italy, the modelled geographical patterns are highly variable, as well as trends in observations. The very different behaviour in the Mediterranean region highlights the much larger model uncertainty in this area. However it is not possible to conclude on the most reliable trends due to the lack of measurements.

For PM_{10}, the model results in Figure 13 are quite successful in capturing the trend of PM_{10}, apart from the Czech Republic and Spain, with a fraction of significant trends with correct sign of 65, 62, 68, and 71% for Bolchem, Chimere, Emep and Eurad respectively.

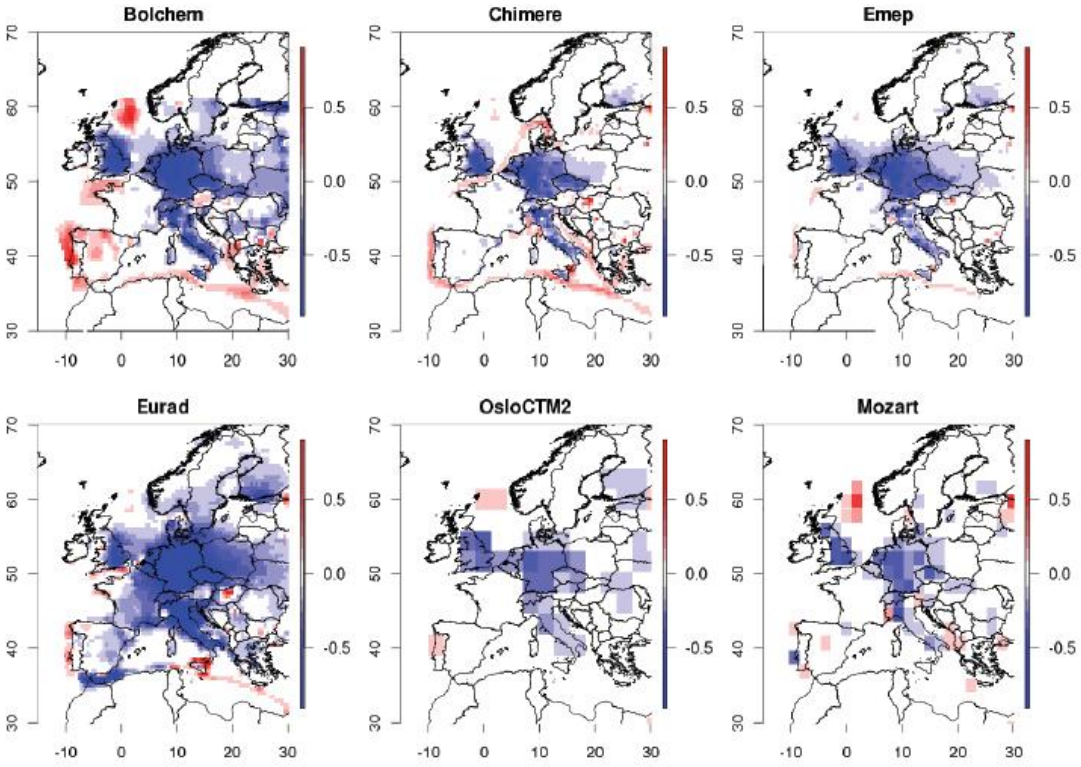


Figure 11 - Modelled NO_2 trend (µg/m^3/yr) for each CTM and at each grid point computed on the basis of monthly means of daily means over the 1998–2007 period with a linear least square fit of de-seasonalised values. (Blue stands for concentration reductions and red for concentration increases. The higher the colour intensity the steeper the trend.)

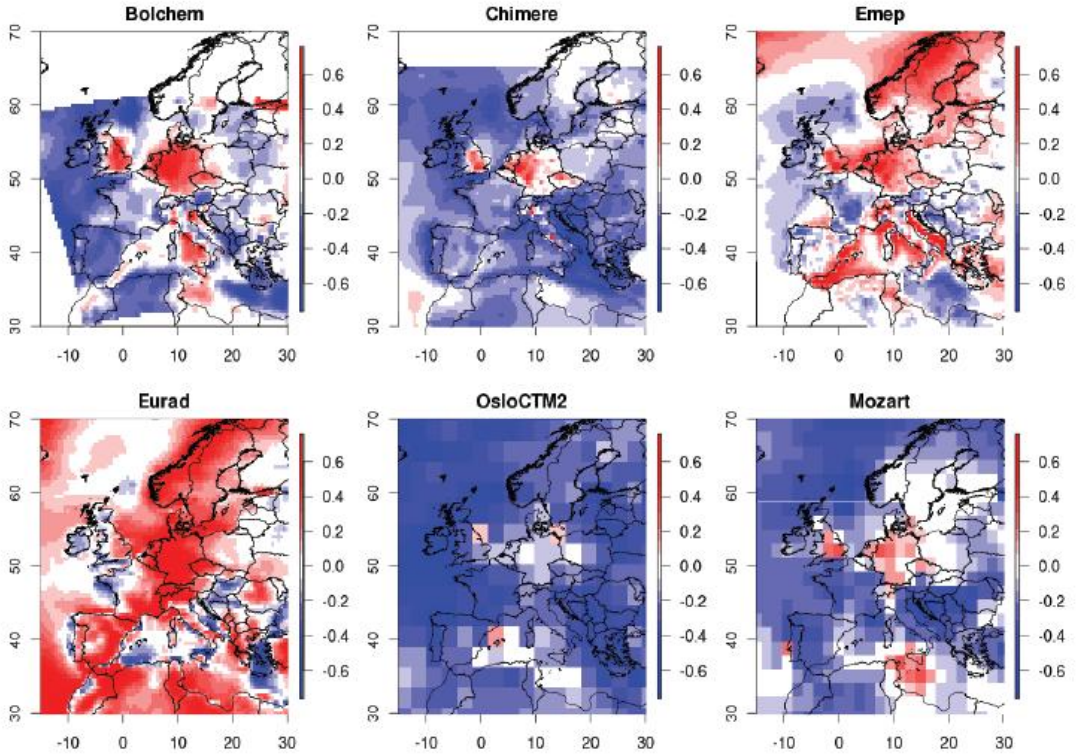


Figure 12 - Modelled O_3 trend (µg/m^3/yr) for each CTM and at each grid point computed on the basis of monthly means of daily means over the 1998–2007 period with a linear least square fit of de-seasonalised values.

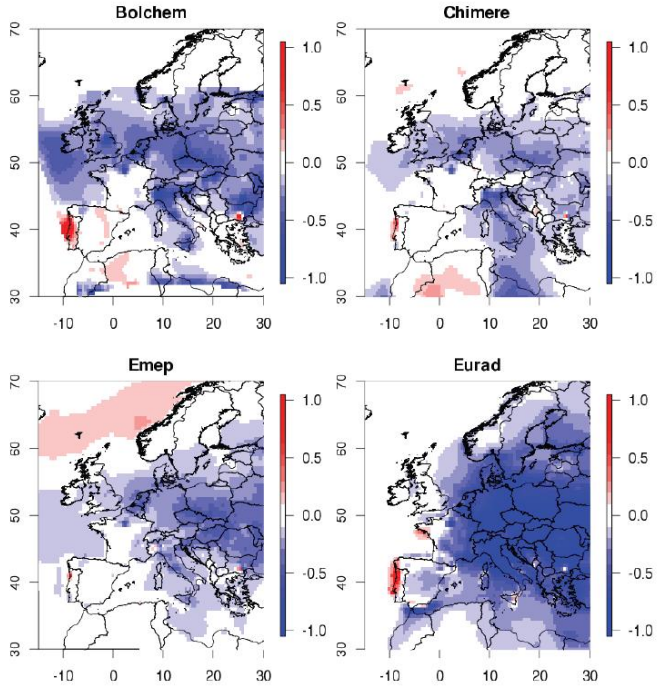


Figure 13 - Modelled PM_{10} trend (µg/m^3/yr) for each CTM and at each grid point computed on the basis of monthly means of daily means over the 1998–2007 period with a linear least square fit of de-seasonalised values.

5. Mapping methodology and related uncertainties

This chapter is mainly focusing on geostatistical kriging methods used for mapping air pollutant concentration fields. Those methods have been used for many years to combine outputs from chemistry-transport model (CTM) outputs and monitoring data, both for (near) real-time air quality mapping in operational systems (e.g. PREV'AIR) and annual production of assessment maps (de Smet *et al.*, 2010 and previous technical papers; Denby *et al.*, 2008a). In evaluation studies against independent measurements, kriging outputs are usually in better agreement with observation data than CTM raw simulated fields, leading to an overall improvement of the skill scores (Rouïl, 2011a, 2011b). Other technical and practical advantages contributing to the wide use of kriging methods are: easy implementation, possibility of introducing additional information by means of ancillary variables, low computational time.

However, beyond the conclusions drawn from statistical scores, one should be aware that:

- the quality of kriging results is linked to the quality of the inputs, in particular to the number, the spatial distribution and the precision of the measurement data;
- assumptions are made about the isotropy and stationarity of the phenomenon;
- assumptions are made about the relationships between the measurements and the CTM, between the measurements and the other ancillary variables.

Consequently, a comprehensive analysis of mapping uncertainties implies a review of the different hypotheses underlying the application of kriging techniques.

5.1 Methods and assumptions

5.1.1 Missing information: how to deal with it

Concentrations of pollutants are measured at a number of monitoring sites:

- which describe pollution on a small, medium, or large scale according to their location and type of pollutant;
- which are distributed over the domain more or less regularly, according to the region and the country and depending on their purpose (research data needs, reporting obligation or optimum representation by spatial interpolation).

The purpose of mapping is to produce a continuous picture of pollution fields over a given area and on a given scale (e.g. background pollution over Europe) from this punctual information.

Between monitoring sites, missing data can be completed by:

- taking advantage of CTM outputs;
- taking advantage of additional variables likely to explain concentration levels: land cover, population, emissions, altitude, meteorological fields, etc.;
- building a geostatistical model from the available data;
- applying this model to interpolate concentrations (kriging).

It is important to note that such mapping:

- is based on the information contained in the input data. In other words, it cannot reflect pollution phenomena that are not accounted for in the measured data or chemistry-transport model outputs. For example, even though auxiliary variables can be helpful in building more detailed maps, it is impossible to represent small-scale pollution from background measurements and large-scale deterministic simulations.
- applies within a spatial domain delimited by the data points. Outside the convex hull of the observation sites, kriging results are highly uncertain.

5.1.2 Different types of kriging

Kriging is a generic term which actually encompasses different techniques. The choice of a method can be related to the experience of the expert team; and/or driven by the characteristics of the data. Inter-comparison studies between kriging variants usually do not show major differences in the results, at least in the convex domain delimited by the observation points (example Figure 14, Malherbe and Ung, 2009).

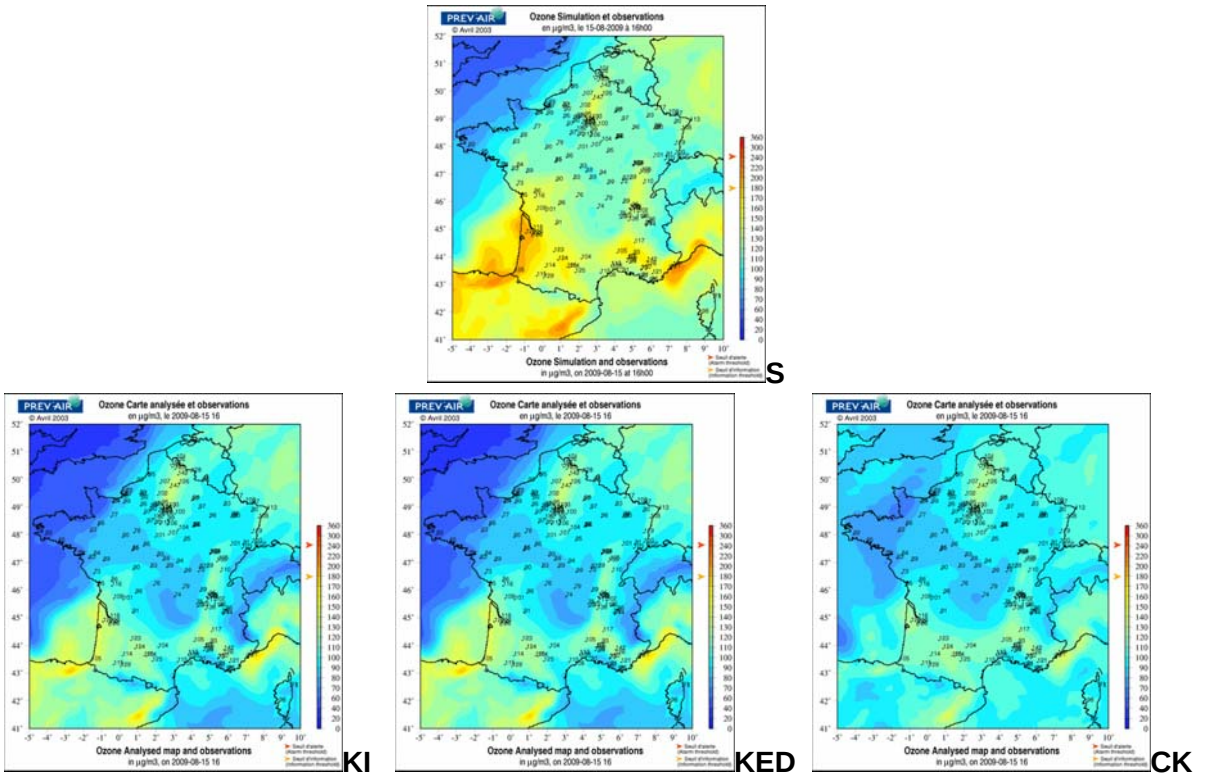


Figure 14 – Ozone concentrations simulated by CHIMERE (S); after kriging of the innovations (KI); after external drift kriging (KED); after cokriging (CK). 15 August 2009, 16h.

However, to select the most suitable variant wherever possible and have a better insight in the results, it is important to recall the geostatistical model and hypotheses on which each type of kriging is based, as briefly summarized hereafter:

- kriging of residuals

- kriging with external drift
- cokriging with unknown mean.

This summary is limited to kriging methods most commonly used for combining measurement data and chemistry-transport model (CTM) outputs to provide near real-time analyses of air pollution fields in a fully operational process. Some considerations about the variance of the kriging error (shortly called *kriging variance*) are also provided.

In the following formulas, $Z(x)$ denotes the pollutant concentration at a point x; $f_k(x)$, k=1...p, denote the ancillary variables at that point (including the CTM).

❒ Kriging of the residual

Model and hypotheses:

$$Z(x) = \sum_{k=1}^{p} a_k f_k(x) + b + R(x)$$

- The relationship between Z and f_k is linear and identical for the whole domain. Coefficients a_k and b denote the weights and constant of this relationship.
- There is no spatial correlation between the f_k and the residual R.

Estimation:

At an estimation point x_0, x_i being the measurement sites in a given search region V_0 around x_0 (kriging neighbourhood):

$$Z^*(x_0) = \sum_{k=1}^{p} \hat{a}_k f_k(x_0) + \hat{b} + \sum_{x_i \in V_0} \lambda_i R(x_i)$$

$R(x_i)$ are the regression residuals at data points x_i.

Coefficients a_k and b are usually estimated by least squares regression.

$\sum_{x_i \in V_0} \lambda_i R(x_i)$ is the ordinary kriging estimate of the residual, with λ_i: kriging weights.

Kriging variance:

- It represents the estimation error variance of the residual R.
- It does not include the uncertainty on the variogram model. In particular the least-squares estimation of coefficients a_k is not optimal as it ignores the spatial correlation of R. The variogram of R is therefore biased (Chilès and Delfiner, 1999).
- It does not include the uncertainty on the drift estimation.

Examples: Application examples on the European scale can be found in Horálek *et al.* (2010, 2008), de Smet *et al.* (2010, 2009).

Kriging of the innovations which was formerly used in PREV'AIR (Honoré *et al.*, 2008) is actually a particular case of kriging of the residual. The model is:

$$Z(x) = f_1(x) + R(x)$$

where the trend is entirely provided by the CTM: $f_1(x)=CTM(x)$ and there is no spatial correlation between the CTM and R, which is a rather strong assumption.

❒ Kriging with external drift

Model and hypotheses:

$$Z(x) = \sum_{k=1}^{p} a_k(x) f_k(x) + b(x) + R(x)$$

- The relationship between Z and f_k is locally linear (within a given neighbourhood).
- The coefficients a_k and b are unknown and the residual R is inaccessible.

Estimation:

At an estimation point x_0, x_i being the measurement sites in a given search region V_0 around x_0 (kriging neighbourhood):

$$Z^*(x_0) = \sum_{x_i \in V_0} \lambda_i Z(x_i)$$

with the additional condition on the kriging weights λ_i:

$$f_k(x_0) = \sum_{x_i \in V_0} \lambda_i f_k(x_i) \quad \forall k$$

- The relationship between Z and the f_k is locally adjusted within the kriging neighbourhood V_0, though it is only implicitly estimated (the unknown coefficients a_k and b are filtered by the kriging process).

Kriging variance:

- It represents the estimation error variance of concentration Z.
- It does not include the uncertainty on the variogram model. In particular, the true residual is inaccessible, its experimental variogram can only be approached prior to model fitting or the modelling be performed indirectly (cross-validation trials).
- It integrates the uncertainty on the drift estimation.

Examples: External drift kriging provides a more general estimation framework than kriging of the innovations. It has been implemented for MACC and PREV'AIR (Figure 15).

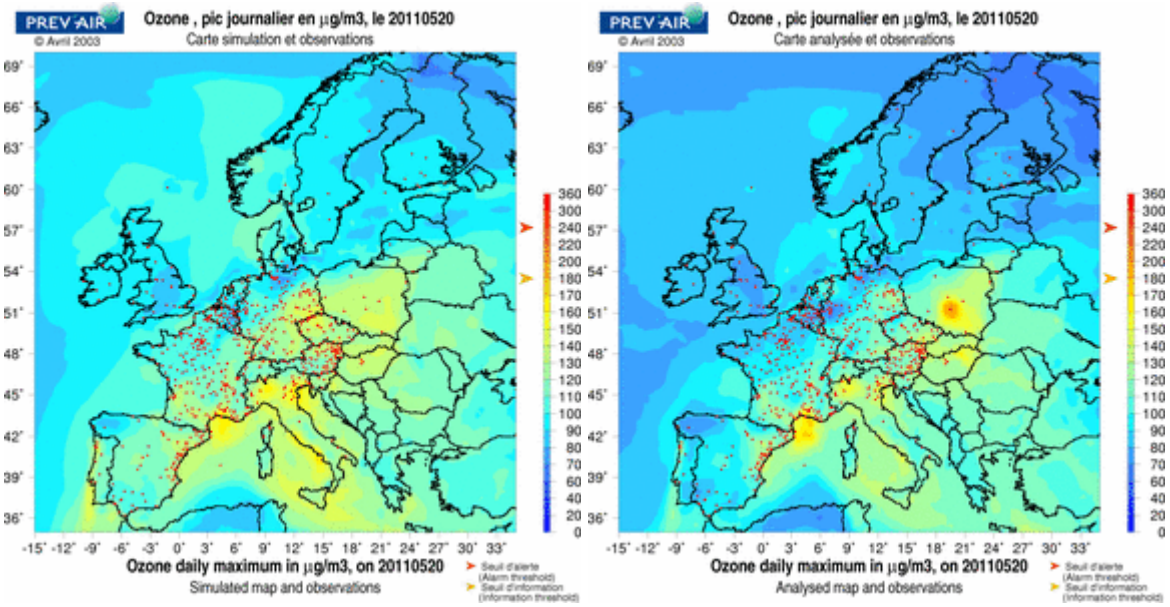


Figure 15 – Left: Map of the daily maximum concentration of ozone derived from the 24 simulated maps (CHIMERE raw outputs). Right: Map of the daily maximum concentration of ozone derived from the 24 hourly kriging maps (kriging with external drift). Input data: background observations from OzoneWeb (Europe) and BASTER (France) databases and CHIMERE model outputs (0.5° x 0.5°). 20 May, 2010.

❒ Cokriging with unknown mean

Model and hypotheses:

The linear model of coregionalization is a widely used multivariate model (Rivoirard, 2003). It proved appropriate to describe the links between in-situ measurements and chemistry-transport model outputs (Chilès *et al.*, 2008; De Fouquet *et al.*, 2011).

The concentration simulated by the CTM is seen as a regionalized variable. The observed (Z) and simulated (CTM) concentrations are decomposed into different scale components. For example, in a purely spatial framework:

$$\begin{cases} Z(x) = \varepsilon(x) + Z^1(x) + Z^2(x) \\ CTM(x) = CTM^1(x) + CTM^2(x) \end{cases}$$

Where:

- $\varepsilon(x)$ is the measurement error at point x
- Z^1 and CTM^1 describe local variability (short-range components)
- Z^2 and CTM^2 describe large-scale variability (long-range components)
- for each scale k, Z^k and CTM^k have similar spatial structures but can exhibit different amplitudes of variation.

This model makes it possible to differentiate the correlation between the observations and the CTM according to the spatial scale.

As in residual kriging, under the assumption of stationarity, the same relationship between both variables holds for the whole domain.

Estimation:

At an estimation point x_0:

$$Z^*(x_0) = \sum_{x_i \in V_0} \lambda_i^Z Z(x_i) + \sum_{x_j \in V_1} \lambda_j^{CTM} CTM(x_j)$$

V_0 and V_1 are the estimation neighbourhoods applied to the measuring data and the CTM data respectively. The x_i and x_j are the measurement sites and the model output points (grid cell centres) located within V_0 and V_1 respectively.

The λ_j and λ_j are the cokriging weights.

NB: when appropriate, other auxiliary variables can be added as external drift for *Z*. In that case, the linear model of coregionalization to be fitted should describe the relationship between the residual *R* and the CTM.

Kriging variance:

- It represents the estimation error variance of concentration *Z*.
- It does not include the uncertainty on the multivariate geostatistical model.

Examples:

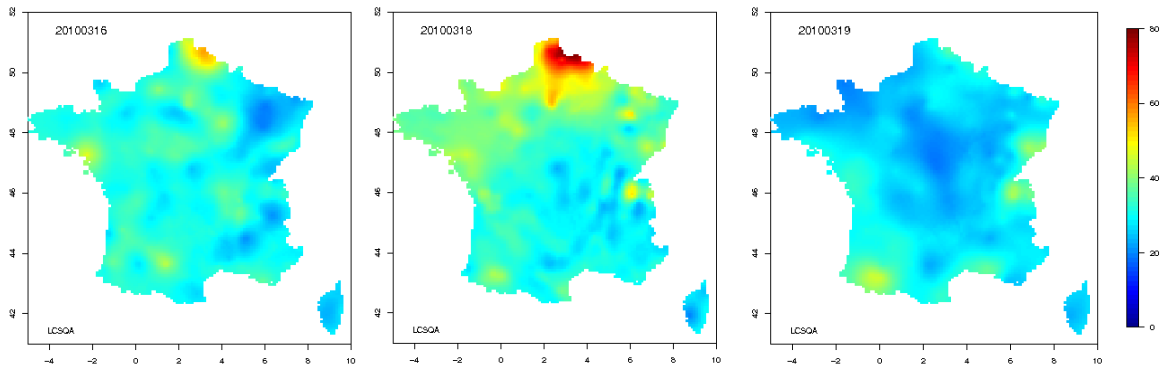


Figure 16 - Daily maps of PM_{10} concentrations produced by ordinary cokriging. Input data: daily background observations from the French monitoring network and CHIMERE model outputs (0.15° x 0.10°). 16, 18 and 19 March, 2010.

Remark:

In air quality mapping applications, an output grid is generally defined in agreement with the monitoring network density and the spatial resolution of the CTM and ancillary variables.

Current legislation contains no requirement about the estimation spatial support, i.e. **about the spatial unit over which concentrations should be computed on average** before comparison with air quality thresholds. A punctual support is usually considered: concentration is estimated at the centre of each grid cell (x_0 in the previous formulas) by point kriging. If the targeted variable is the mean concentration per cell, then block kriging (Chilès and Delfiner, 1999) can be performed (Denby, 2009).

5.1.3 Non fulfillment of the assumptions

In this section, hypotheses presented in section 4.1.2 and other assumptions commonly adopted for air quality mapping are reviewed. Their non-fulfillment is a source of uncertainty which is not reflected in the kriging variance map. Possible ways of handling it are indicated.

Normality

In theory, kriging requires no assumption on the data distribution other than on the first two moments. However geostatistical inferences using kriging techniques are more efficient when data are distributed normally. Difficulties caused by severely skewed distributions may be alleviated by preliminary transformation of the data. The most common process is the logarithmic transform which is best suited to lognormally distributed data. The kriged estimates, together with a corrective factor which ensures unbiasedness, are then back-transformed to obtain the lognormal kriging estimates (Roth, 1998). Several approaches actually exist (Rivoirard, 1990).

Air quality data frequently show a skewed, near lognormal distribution. After controlling that the data could be considered as lognormally distributed, Horálek *et al.* (2010) tested lognormal kriging to improve PM_{10} and ozone maps. The idea was to provide more optimal conditions for kriging and ensure positivity of the estimates. For PM_{10} annual indicators (annual average and 36th highest value) they noted a slight improvement whereas for ozone, they did not notice any quality enhancement of the estimation.

In practice, log-normal kriging may not bring the expected improvement. It is shown to respect some unbiasedness properties when considering the whole domain but significant deviations from the true values can be locally observed, as illustrated by Roth (1998). Application of log-normal kriging thus requires thorough validation before implementation. Lognormal kriging can be generalized to other types of data transformation.

Isotropy

To simplify the geostatistical modelling, spatial variability is often assumed to be independent of the direction (*isotropy*). However, in air quality, pollution transport and dispersion are influenced by directional patterns stemming from the environmental characteristics (emissions, land use) and the meteorological conditions (wind flows).

Anisotropy can be integrated in kriging through the definition of both an anisotropic variogram and an anisotropic estimation neighbourhood (selection of input stations depending on direction). However, including anisotropy in the variogram should only make sense within a certain geographical domain, affected in the same way by directional factors.

Stationarity

Usual implementation of the kriging methods described in section 5.1.2 rests on some strong hypotheses which are:

- the selected auxiliary variables (in addition to the CTM) are relevant predictors over the whole domain;
- for kriging of the residual and cokriging only, the relationship between the observations and the CTM (and any other auxiliary variable if relevant) is the same at any output grid point;
- the unique calculated variogram can be applied everywhere.

Those assumptions (no dependence on the geographical location, called *stationarity*) are questionable, especially when maps are produced over very large domains like Europe.

Different possible ways of addressing non-stationarity exist:

- Locally rescaling the parameters of the variogram.
  A proportional effect can be modelled to deal with non-stationarity in the variance. More recently, techniques based on local cross-validation processes have been developed to locally optimize other parameters such as the range and anisotropy (Magneron and Demongin, 2011).

- Splitting the domain into geographical subdomains.
  As described in section 2.2, the variability and performance of the chemistry-transport models depend on the geographical area. In MACC project different subregions could be identified in Europe. It would be interesting to divide the domain accordingly and define as many geostatistical models as subdomains. The main problem will be the connection between the output grids.

- Splitting the domain according to secondary variables.
  This solution has been proposed in previous ETC/ACC papers (Horálek *et al.*, 2010). Two maps are separately produced, a rural and an urban one, using the rural and urban stations respectively. Merging is performed following criteria based on population density and taking account of possible inconsistencies between urban and rural concentrations.
  Such methodology can be more particularly useful when the CTM performs differently according to the station classification (example: positive bias at rural stations and negative bias at urban sites as is the case for NO_2 with CHIMERE model.)

An extension would be the definition of a non-stationary space-time covariance (De Iaco *et al.*, 2005).

Station representativeness

The notion of station representativeness was largely discussed in chapter 2. The question here concerns representativeness with respect to the auxiliary variables. This issue is related to the spatial distribution of the monitoring network: do the observation points cover the whole range of values taken by each auxiliary variable?
If not, then multivariate kriging of concentrations in the entire domain means applying the relationship between the observations and the auxiliary variables in areas where it is not necessarily valid. Possible consequences of such extrapolation are incoherent and negative estimates, as pointed out by De Fouquet (2006). Some examples are estimation in elevated mountainous areas using monitoring stations located at low or medium altitudes and topographical data; or (as is often the case for NO_2) estimation in sparsely populated areas using monitoring stations mostly located in urbanized areas and population density data.
Extrapolation areas should thus be clearly identified to properly interpret the maps.

Linear relationship

Observed concentrations are supposed to be linearly related to the CTM and if the case arises to the other secondary variables (cf. section 5.1.2).

For seasonal or annual mapping based on average measurements, linearity is easy to verify during the exploratory data analysis and relevant auxiliary variables can be selected. When maps are produced at a shorter time step (hour or day), kriging has to be automatized. Linear relationship with the predefined auxiliary variables is no longer ensured on each date. This can occur for pollutants resulting from both local and regional time-varying contributions, like PM_{10}: during large scale episodes, PM_{10} concentrations can be poorly correlated to local emission data whereas for other days there is more evident link. Minimum automatic control is necessary to prevent inconsistent results (e.g. negative values due to a negative regression weight).

5.2 Use of the kriging variance

5.2.1 What does kriging variance represent?

Kriging variance quantifies the possible dispersion of the true value around the estimate. It depends on:

- the variogram model,
- the relative layout of the target (estimation point or cell) and the data points.

It does not directly depend on data values.

Kriging variance can be seen as a precision indicator (Chilès and Delfiner, 1999) reflecting:

- the variability of the studied phenomenon within the domain of interest. The more variable this phenomenon is (as described by the variogram parameters), the more difficult it is to yield a precise estimate, the higher the kriging variance is. This provides global information over the domain since in general, apart from local adjustment according to local variability (proportional effect or other techniques, see section 5.1.3), the variogram is an average representation of spatial variability.
- the uncertainty about station representativeness and the uncertainty of the monitoring measurement. Kriging variance integrates the average variance of the measurement errors and small-scale variability if those are modelled as a nugget effect: the higher this component is, the higher the kriging variance is. This point is further discussed in section 5.2.2.
- the uncertainty due to the absence of monitoring. The further a grid point is distant from the measuring stations, the less reliable the estimation is, the higher the kriging variance is. This provides local information depending on the grid point location.

Kriging variance can thus be helpful:

- for global comparison between different times, i.e. to compare the overall precision of maps produced at different periods (days or years) with the same monitoring network;
- for comparison in space, i.e. to distinguish between areas with dense monitoring coverage and more precise estimates and areas with sparse monitoring coverage and poor estimation precision.

Remarks:

- In a classical geostatistical framework, kriging variance does not include:
  - the uncertainties due to possible deviance from the hypotheses (stationarity, linearity, etc.);
  - the uncertainties on the variogram parameters;
  - the uncertainty on the drift estimation when the drift is modelled separately and kriging is performed on the residual.

  But for local rescaling of the variogram, it does not reflect the local variability of the phenomenon either.

- Kriging variance is dependent on the spatial support used for kriging. If block kriging is performed instead of point kriging, in concrete terms if average concentrations per cell are estimated instead of punctual values at the grid cell centres, then kriging variance should be lower: the larger the spatial support, the less variable the quantity to estimate and the more precise the kriging estimates. However, for comparison with environmental objectives, and in the absence of any precise obligation, punctual concentrations are often calculated.

5.2.2 Uncertainty in the input data and kriging variance

On the one hand one has uncertainties involved in the measurement data, on the other hand uncertainties contained in the CTM and auxiliary variables that are used in the kriging. these two aspects of uncertainty in input data can be considered in more detail:

Measurement data

Observation data contain measurement errors:

$$Z(x) = Y(x) + \varepsilon(x)$$

with *Z(x):* observation at point *x*
Y(x): unobserved true process
ε*(x): measurement error*

Those errors result in an additional offset at the origin of the variogram, which is equal to their average variance. In most air quality applications, it is modelled by a nugget effect together with small-scale variability.
However, if measurement errors do not have the same variance and that variance is known at each data point, the kriging system can be modified to take account thereof.
In that approach, sometimes called *filtered kriging*, the variable to predict is the error-free value. The corresponding kriging variance has been compared to the one obtained by modelling a global nugget effect (Cárdenas and Malherbe, 2003). In point kriging, filtering the measurement error results in a reduction of the kriging variance as also pointed out by Christensen (2011).
Christensen (2011) proposes further developments for cases in which the observed process and the measurement error are correlated.

CTM outputs and other auxiliary variables

In section 5.1.2 different types of kriging relevant for combining measurement and auxiliary data - in particular CTM output fields - have been presented. The selection of one technique can be guided by investigating the relationships between concentration measurements and

auxiliary variables in view of the assumptions, possibilities and limitations of each method. In any case cross-validation and independent validation tests remain fundamental to support that choice. Depending on the selected technique, the CTM and auxiliary variables have different impact on the kriging variance.

In *kriging of the residual* a global relationship with the auxiliary variables (or a global relationship per station class, see Horálek *et al.*, 2010) is established, assuming that it is valid at any point of the domain. This may be a strong hypothesis on European scale but doing so allows the user to keep control over the multivariate relationship and be sure of its physical consistency.
Kriging variance reflects uncertainty in the residual interpolation and is not directly influenced by the CTM and auxiliary variables.

In *external drift kriging* a local relationship with the auxiliary variables is implicitly determined at each estimation point. It is established within a moving neighbourhood defined in terms of distance, sectors and/or number of stations. The user has no immediate control over those local relationships which may be sensitive to the location and characteristics of stations selected around each estimation point. As main advantage, he can take account of the fact that the CTM shows different behaviour (with respect to the observations) according to the geographical region.
Kriging variance is the sum of two parts corresponding to the estimation error variances of the residual and of the drift respectively. The values of the auxiliary variables at both measurement and estimation locations are directly involved in the calculation of the second term.

In *ordinary cokriging* between the measurements and the CTM, a global bivariate model is defined over the whole domain, with possibility of describing a different CTM behaviour (different variability relative to the observations) depending on the spatial scale.
Kriging variance is the sum of different terms based on the single and cross variograms of the observed concentrations and the CTM. It is then influenced by the spatial variability of the CTM.

5.2.3 Spatial representativeness of stations

With an assumption on the error distribution, kriging variance can be used to define areas where concentration levels fall within a given range from the values measured at the stations. Such application has been presented in section 2.2.1.

5.2.4 Probability of exceedance thresholds

Adopting the same probabilistic framework as for spatial representativeness, and making an assumption on the error distribution as well, kriging variance can be used to assess compliance with air quality limit values.
This issue has been more particularly investigated for PM_{10} (Horálek *et al.*, 2007; Denby *et al.*, 2008a, 2008b; Malherbe *et al.*, 2011). All those studies are based on the production of daily mean concentration maps by multivariate kriging. Different approaches are then proposed, reflecting the complexity of the problem and the search for methodologies suitable for operational and reporting purposes.

In Horálek *et al.* (2007):

- Daily estimates were respectively averaged over the year and summed after comparison with the 50 µg/m^3 threshold to calculate maps of annual mean concentration and number of daily exceedances (Figure 17, left);
- In a second stage, maps of uncertainty (standard deviation) were established for both indicators: the uncertainty on the annual mean was derived from the daily mean uncertainties, taking the temporal covariance into account. The uncertainty on the number of daily exceedances was calculated by perturbing the daily mean fields (Figure 17, right).

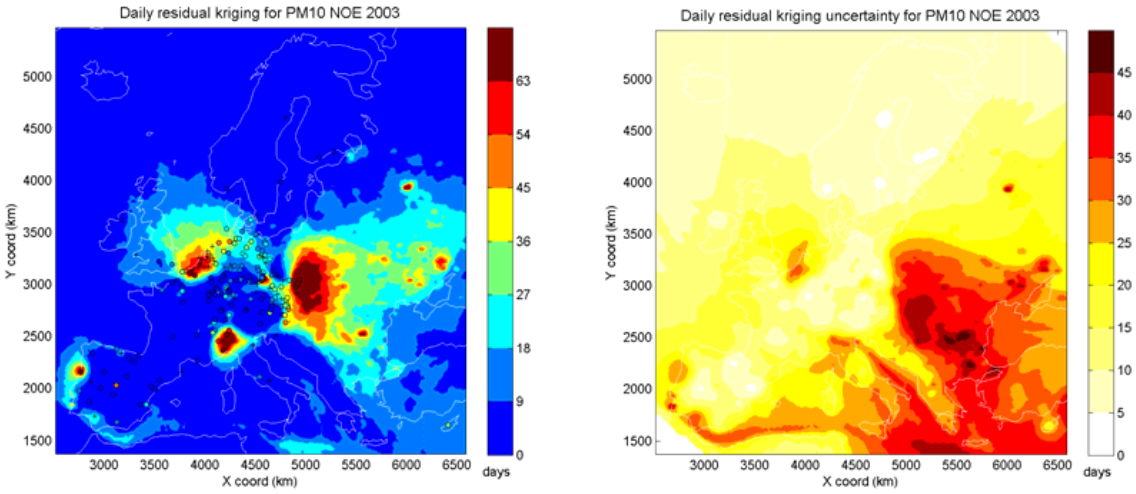


Figure 17 – Estimated number of exceedances (NOE) of the daily 50 µg/m^3 threshold (left) and associated estimated uncertainty (right) calculated over Europe for year 2003 by Horálek* et al. *(2007). AirBase data, EMEP outputs and ancillary variables were used as input in the estimation process.

In Denby *et al.* (2008a and 2008b):

- The map of the number of exceedances and the associated uncertainty map (Figure 18) were combined to map the probability of exceedance (>35) for the number of exceedance days (Figure 19). This was achieved by treating the uncertainty as having a Gaussian distribution.

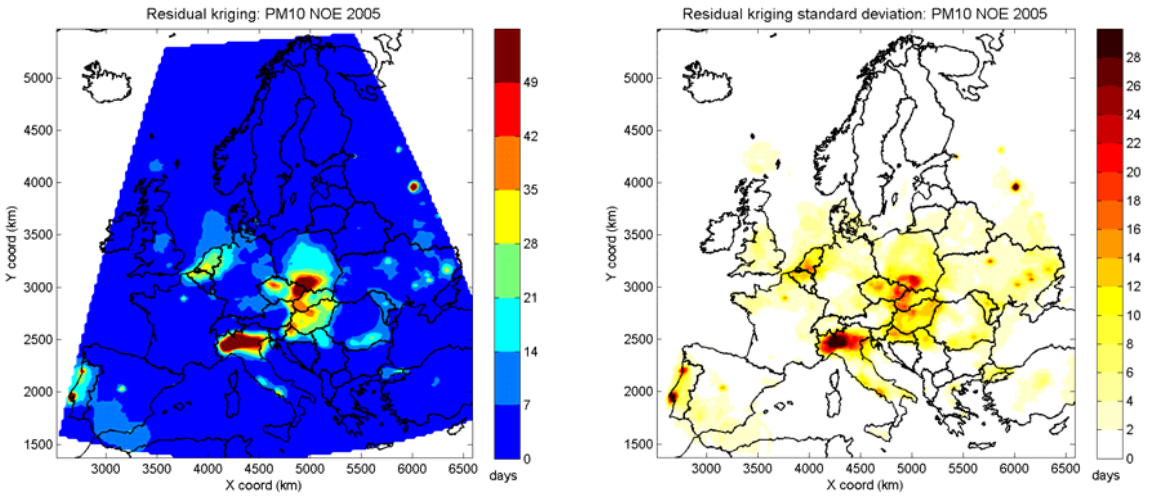


Figure 18 – Estimated number of exceedances of the daily 50 µg/m^3 threshold (NOE) (left) and associated estimated uncertainty (right) calculated over Europe for year 2005 by Denby et al. (2008a). In this example, AirBase data, LOTOS-EUROS outputs and ancillary variables were used as input.

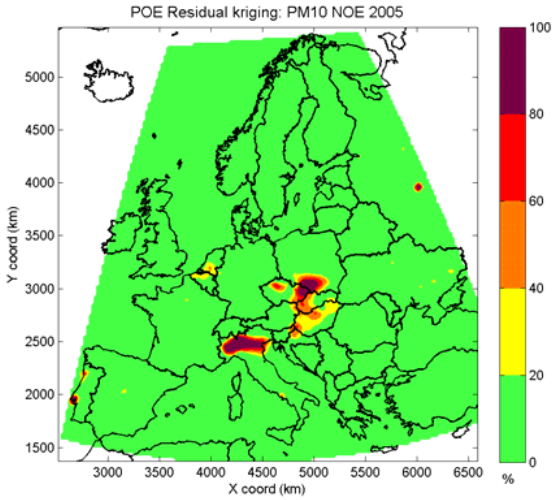


Figure 19 – Estimated probability (%) that the NOE be higher than 35. Map derived from Figure 18 estimates (Denby, 2008a).

In Malherbe *et al.* (2011):

- The daily PM_{10} concentration maps and associated kriging variance maps were used to produce daily probability of exceedance maps (probability of exceeding 50 µg/m^3), conventionally assuming that the estimation error is normally distributed[(*)] and ignoring temporal correlation as first approach (this will be accounted for in future developments).
  [(*)] $\varepsilon(x) = \sigma_K(x).T$, with $T \rightsquigarrow N(0,1)$ and $\sigma_K(x)$ is the kriging error standard-deviation.
- The translated Poisson distribution for the count of events was then used to map the probability that the annual number of exceedance days is over 35 (Figure 20).

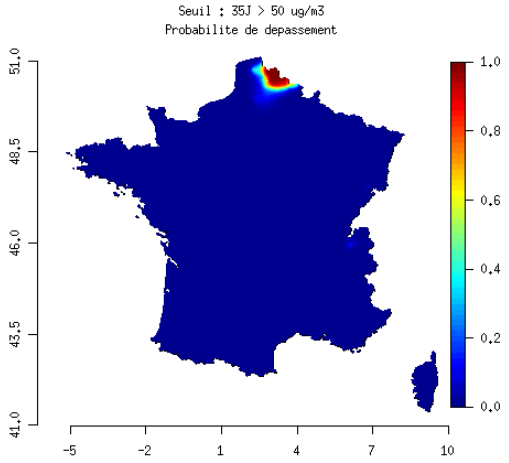


Figure 20 – Estimated probability that the annual number of daily exceedances is greater than 35.

Note that in theory, non-linear geostatistics offers a more rigorous framework for addressing the problem of exceedances (Deraisme *et al.*, 2011). In practice, this means heavier developments and less straightforward implementation which might not be suitable for day to day application. Further investigations need to be carried out.

5.3 Merging and other mapping methods

Next to the mapping methods and map merges on the basis of spatial related criteria discussed in previous sections, especially section 5.1.3 on non-stationarity, several alternative methods for air quality mapping exist, each with its specific methodological emphasis.

In the field of spatial statistics, spatio-temporal air quality mapping based on hierarchical Bayesian models has been developing for the last decade (e.g. Cretarola *et al.*, 2010; Fuentes and Raftery, 2005). Those models provide a flexible framework to incorporate spatio-temporal interactions at different hierarchical levels (Cameletti *et al.*, 2010) and combine uncertain in-situ data with uncertain numerical model outputs and other possible covariates. The result at each grid point is the posterior distribution of the concentration of interest.

In the field of computer graphics, geometric modelling and visualization, multi-resolution analysis is a powerful tool to represent a function at multiple levels-of-detail and produce a new mapping method. It makes use of the wavelet transform to decompose a data series in a cascade going from the smallest scales to the largest. For example in Beaulant *et al.* (2005), multi-resolution analysis and wavelet transform techniques are explored to combine point measurements and different data sources such as outputs of meteorological models. The method originates from the work of Beyer *et al.* (1997) who created maps for the new European Solar Radiation Atlas (ESRA) and is connected to the ARSIS concept (Improvement of the spatial resolution by injection of structures) developed in the field of image processing (Ranchin and Wald, 2000).

In the field of artificial intelligence, more and more studies have demonstrated the capacity of integrated artificial neural network model to forecast pollutant concentrations (Kolehmainen *et al.* 2000, Pérez and Reyes, 2006, Dutot *et al.* 2007, Ibarra-Berastegi *et al.* 2008). Once trained, the neural network can be applied to each raster grid cell to yield a two-dimensional map showing the interpolation of the field. Groselj *et al.* 2004 propose a new mapping based on a clustering methods and Kohonen network. Within the ETC/ACC some effort has been made on using neural network modelling to produce European annual average $PM_{2.5}$ maps on basis of a limited set of $PM_{2.5}$ measurements. Through a neural network technique pseudo PM2.5 stations are produced that support in the interpolations (Denby *et al.*, 2011a). However the technique has never been further employed within the ETC/ACM due to its complexity and non-transparency in relation to its quite limited improvement compared to the routinely applied multi-linear regression.

6. Sub-grid variability and exposure assessment

Chemistry transport models (CTM) are usually implemented over regional domains with a spatial grid resolution of 10 km to 100 km. Simulated values represent grid-cell averaged concentrations. However, it is clear that there is a substantial amount of variability within those cells, known as *sub-grid variability*, which should be considered when population exposure is assessed from CTM outputs (Denby *et al.,* 2011b). Ignoring it may contribute to significant bias in exposure estimates which are provided by population weighted concentrations.

6.1 Exposure assessment based on mapping

Section 5 indicated how CTMs could supplement surface measurement data and be combined, where relevant, with ancillary variables that are well correlated to pollutant concentrations in order to produce air quality maps. Auxiliary variables are often known with high resolution so that the resolution of the final map can be increased compared to the CTM. In Horálek *et al.* (2010) rural and urban air quality maps were produced with spatial resolution of 10 km and 1 km, using AirBase measurement data and results from large scale EMEP simulations. EMEP data with initial resolution of 50 km were disaggregated to grid cells of 10x10km^2 and 1x1km^2. Other variables introduced into kriging were:

- ECMWF parameters: wind speed and surface solar data with initial resolution of approx. 27 km were disaggregated on both grids as well;
- altitude and land cover: data available with high spatial density were aggregated into the 10x10km^2 and 1x1km^2 resolution.

Rural and urban maps were then merged according to population density to provide final concentration estimates and calculate population weighted concentrations. At each stage of the process (interpolation/merging/exposure assessment) grid resolutions of 1 km and 10 km were tested. Cross-validation shows that applying a finer grid resolution for first estimating and second merging the rural and urban maps brings reduced uncertainty. Most important for an accurate estimate of exposure is the resolution of the merging: obviously, with a refined merging grid (1x1km^2), the population of small cities and towns is better accounted for.

6.2 Use of sub-grid covariance between population and concentration

Another approach for assessing exposure, based on CTM large scale outputs and population data is described in Denby *et al.* (2011b). In that study sub-grid variability and its impact on long term exposure estimates were assessed by investigating the covariance of concentration and population. This covariance proves to be the defining factor in determining the exposure estimation error due to sub-grid variability.

Let C_j be the average concentration over a given area A_j, P_j the mean population density within A_j and $cov_j(c,p)$ the covariance of concentration and population in A_j. The authors demonstrate that the population weighted concentration $C_{pw,j}$ in A_j can be written as:

$$C_{pw,j} = C_j \left(1 + \frac{cov_j(c,p)}{C_j P_j}\right)$$

This amounts to correcting C_j by an additive term which depends on the concentration-population covariance.

The methodology was developed for annual mean concentrations of NO_2 and PM_{10} and the ozone health indicator SOMO35. To allow exposure quantification over all of Europe a parameterized model of the covariance correction factor was built for each of those indicators. Based on multiple linear regression it makes use of the covariances between population and available proxy data (emissions, altitude). The methodology was successfully applied for annual PM_{10} and NO_2 in 2006, using EMEP model output fields (50x50 km^2 resolution). Results reveal that the error in exposure calculation using a CTM resolution of 50 km is largest for NO_2 (urban background exposure is underestimated by 44 ± 4%) and moderate for PM_{10} (exposure is underestimated by 15 ± 4%). For SOMO35 the covariance parameterization was deemed too uncertain for application over Europe but estimates based on observations alone show that population weighted SOMO35 concentrations are overestimated by around 13% when a CTM resolution of 50 km is used.
The methodology is promising for improving exposure assessment based on large scale modelling data. It requires that mean grid concentrations be well represented by the CTM. Limitations lie in the availability of high resolution proxy data to parameterize sub-grid variability and in the possibility of deriving accurate parameterization for each targeted air quality indicator.

7. Conclusion

The mapping activities of ETC/ACC have been described in detail in previous technical papers. Those documents provide some ways of addressing the various sources of uncertainty in the mapping methodology implemented by ETC/ACC and EEA for evaluating air quality over Europe and assessing population and ecosystem exposure. The purpose of this report is to further discuss and analyze the mapping processes in the light of recent papers or studies conducted within European projects or through national initiatives like the work carried out for the implementation of the French operational air quality forecasting and mapping system PREV'AIR.

Section 2 is dedicated to the definition of the representativeness of a monitoring station depending on the context. Information about representativeness is an added value that can be used to make a selection of stations which has itself a direct impact on the results and output variability. The definition and quantification of representativeness differ according to the use: reporting, mapping, assimilation or validation purpose.

Section 3 provides a brief overview of the assessment of measurement uncertainty.

Section 4 is about the assessment of uncertainty and variability in air quality modelling with chemistry transport models (CTMs). Recent works on this subject have been conducted and their results are presented: application of Monte-Carlo methods in EC4MACS project and ensemble methods in MACC project. Those approaches are difficult to compare. The first one examines the effects of perturbing input data and physical parameters on model results while the second one is more a methodology to take benefits from all chemistry-transport models with their specific parameterization. In Monte-Carlo methods output variability is expressed in terms of pollutant concentration range (synthesized by the standard deviation). In ensemble methods, it is characterized by the spread of the members and still has to be linked with the skill scores of the individual models calculated with AirBase database. Knowing the potential and the limitations of existing models, the last part of this section demonstrates the capacity of chemistry-transport models to capture the observed trends of major atmospheric pollutants.

Section 5 is dedicated to the mapping methods. More than a theoretical review of the kriging techniques and their assumptions, it is a review of difficulties faced by the operator during a mapping exercise.

Section 6 addresses the problem of sub-grid variability when chemistry transport models are run on large scale to assess population exposure. With resolution of approx. 50 km, exposure estimates, given as population weighted concentrations, can be significantly biased. More accurate assessment can be performed if the CTM is not used alone but included in a mapping process as described in Section 4. More recently a methodology based on the covariance of concentration and population has been developed. A covariance correction factor is applied to the concentration field simulated by the CTM to compensate for the underestimation (in the case of PM_{10} and NO_2) or overestimation (in the case of ozone SOMO35) due the grid resolution of the model. The possibility of representing that

covariance in a parameterized form, using proxy data, offers interesting potential for improving exposure quantification over Europe. The ETC/ACM aims to evaluate in 2012 non-stationary covariance models and anisotropic kriging in order to reduce the sub-grid variability currently present in its mapping results. On this basis and from primarily the AirBase measurement data point-of-view, it will evaluate the use of comprehensive probabilistic indicators (i.e. uncertainty, accuracy, confidence statements) suitable for communication within the policy field. Once deemed acceptable, the probabilistic indicators will be applied routinely in future (i.e., after 2012) mapping and assessment activities based on measurement data primarily. Another activity of the ETC/ACM will focus on similar issues but from primarily the modelling point-of-view.

Considering the range of difficulties encountered, several recommendations and propositions can be made for a more efficient mapping:

- Studies on station classification need to be continued to get a better match between the scale of the pollution process being mapped, the selection of monitoring sites used in input, the auxiliary data source variables and their resolution, the CTM spatial resolution and the targeted level of detail in the final map. In particular more robust classifications based on the properties of several-year time series could be developed as in Joly and Peuch (2012).

- When it comes to mapping air quality over Europe, usual application of kriging reaches its limits. Several common assumptions do not hold any longer, such as a unique variogram for describing variability over the whole domain or hypothesis of isotropy. The use of auxiliary data and their spatial characteristics and correlation with concentrations do implicitly play a role. Developments are still required to better adapt the methodologies to the dimensions of the study domain and to the actual variability of concentrations. The definition of non-stationary covariance models and anisotropic kriging offer interesting prospects. (Similar recommendations will be made by Gräler *et al.* (2011)).

- Assessing exceedances is a complex issue which cannot be addressed without simplifications and assumptions. Promising studies have been undertaken in that field and are worth pursuing. In particular kriging variance is a very helpful uncertainty indicator which has been used so far through different approaches to delimit exceedance areas. However it does not integrate all sources of uncertainty and only provides partial information on the estimation error. Further works could be dedicated to the elaboration of a more comprehensive uncertainty indicator. They could include the development of harmonized methodologies aimed at calculating probabilities of exceedance for different air quality statistics. This involves a reflection on the way of interpreting probability maps for the fulfilment of reporting requirements (assessment of the area and population exposed to exceedances).

- Sub-grid variability analysis and the estimation of population weighted concentrations as recently proposed by Denby et al. (2011b) have been carried out so far with raw CTM outputs. The capacity of that methodology to better quantify population exposure could be examined as well when it is applied to combined maps - i.e.

concentration maps based on measurement, modelling and other supplementary data - which have been produced with a resolution of appr. 10 km x 10 km.

- Finally, innovative ideas could be taken from the current and the alternative mapping and merging approaches in the field of spatial statistics, computer graphics and artificial intelligence. Data fusion techniques using wavelet transform could be applied to combine regional mapping with global modelling. Neural network classifiers might also be used to define typical weather or pollution situations for which more appropriate mapping processes could be defined. For example, in case of a kriging-based mapping approach, some parameters of the geostatistical model as the number and type of structures could be selected according to the meteorological situation.

References

AFNOR technical documents for assessing uncertainties in measurements: FD X43-070-1 (2007, generalities), FD X43-070-2 (2007, in-situ automatic measurements of SO_2, NO, NO_2, NO_x, O_3 et CO), FD X43-070-3 (2008, benzene measurement by passive sampling), FD X43-070-4 (2008, NO_2 measurement by passive sampling), FD X43-070-5 (2008, benzene measurement by active sampling), FD X43-070-6 (2011, automatic measurements of particulate matter), FD X43-070-7 (2011, measurement of HAP in PM_{10}), FD X43-070-8 (2011, measurement of lead, cadmium, arsenic and nickel in PM_{10}).

AirBase database, http://acm.eionet.europa.eu/databases/airbase/

AIR4EU reports: http://www.air4eu.nl/reports_products.html

Beauchamp M, de Fouquet Ch, Malherbe L, 2011. Spatial representativeness of an air quality monitoring station. Application to NO_2 in urban areas. Proceedings of the Spatial2 Conference: Spatial Data Methods for Environmental and Ecological Processes, Foggia, Italy, 1-2 September 2011. Available at http://dspace-unibg.cilea.it/handle/10446/898/?locale=en

Beaulant A, Wald L, Weber C, Perron G and Kleinpeter J, 2005. Multiresolution analysis for air pollution mapping over a city – Typical fields methods. In: Proceedings of the 5th International Conference on Urban Air Quality (UAQ), Valencia, Spain, March 29th – 31st, 2005.

Belis CA, Lagler F, Barbiere M, Mücke H-G, Wirtz K and Stummer V, 2009a. The evaluation of the Interlaboratory Comparison Exercise for SO_2, O_3, NO and NO_2, Langen 20th-25th September 2009. JRC Scientific and Technical Reports, doi:10.2788/94507.

Belis CA, Kapus M, Barbiere M and Lagler F, 2009b. The evaluation of the Interlaboratory comparison Exercise for SO2, CO, O3, NO and NO2 19.- 22. October 2009. JRC Scientific and Technical Reports, doi: 10.2788/47601.

Beyer H G, Czeplak G, Terzenbach U, Wald L, 1997. Assessment of the method used to construct clearness index maps for the new European Solar Radiation Atlas ESRA. Solar Energy, May, pp.17.

Bobbia M, Cori A, de Fouquet Ch, 2008. Représentativité spatiale d'une station de mesure de la pollution atmosphérique. Pollution Atmosphérique N°197.

Cameletti M, Ignaccolo R, Bande S, 2010. Comparing air quality statistical models. Graspa working paper, available at http://dspace-unibg.cilea.it/handle/10446/898/?locale=en

Cárdenas G, Malherbe L, 2003. Evaluation des incertitudes associées aux méthodes géostatistiques. Technical report, DRC-03-45599-2IEN-MECO-GCa-LMa, Laboratoire de surveillance de la qualité de l'air, www.lcsqa.org

Cárdenas G, Malherbe L, 2007. Application des méthodes géostatistiques à l'évaluation de la représentativité spatiale des stations. Technical report, DRC-08-85146-01036A, Laboratoire de surveillance de la qualité de l'air, www.lcsqa.org

Chilès J-P, Delfiner P, 1999. Geostatistics: modelling spatial uncertainty. Ed. Wiley, New-York, 695 p.

Chilès J-P, Séguret S, Riboud P-M, 2008. Geostatistical analysis of validation data of an air pollution simulator. Proceedings of the VIII International Geostatistics Congress, GEOSTATS 2008, Santiago, Chile, 1-5 December 2008, Vol. 2, pp 861-870.

Christensen WF, 2011. Filtered kriging for spatial data with heterogeneous measurement error variances. Biometrics, Vol. 67, Issue 3, pp. 947–957.

Colette A, Granier C, Hodnebrog Ø, Jakobs H, Maurizi A, Nyiri A, Bessagnet B, D'Angiola A, D'Isidoro M, Gauss M, Meleux F, Memmesheimer M, Mieville A, Rouil L, Russo F, Solberg S, Stordal F, Tampieri F, 2011. Air quality trends in Europe over the past decade: a first multi-model assessment. Atmos. Chem. Phys. Discuss., 11, 19029-19087, doi:10.5194/acpd-11-19029-2011 (final paper in press).

Cretarola L, Lasinio G, Arima S, Pollice A, 2010. Bayesian univariate space-time hierarchical models for mapping pollutant concentrations in the municipal area of Taranto. Graspa working paper, available at http://dspace-unibg.cilea.it/handle/10446/898/?locale=en

Denby B, Gola G, de Leeuw F, de Smet P, Horálek J, 2011a. Calculation of pseudo PM2.5 annual mean concentrations in Europe based on annual mean PM10 concentrations and other supplementary data. ETC/ACC Technical Paper 2010/9.
http://acm.eionet.europa.eu/reports/ETCACC_TP_2010_9_pseudo_PM2.5_stations

Denby B, Cassiani M, de Smet P, de Leeuw F, Horálek J, 2011b. Sub-grid variability and its impact in European wide air quality exposure assessment. Atmopsheric Environment, Vol. 45, 4220-4229.

Denby B, de Leeuw F, de Smet P, Horálek J, 2009. Sources of uncertainty and their assessment in spatial mapping. ETC/ACC Technical Paper 2008/20.
http://acm.eionet.europa.eu/reports/ETCACC_TP_2008_20_spatialAQ_uncertainties

Denby B, Horálek J, de Smet P, de Leeuw F, Kurfürst P, 2008a. European scale exceedance mapping for PM10 and ozone based on daily interpolation fields. ETC/ACC Technical Paper 2007/8.
http://acm.eionet.europa.eu/reports/ETCACC_TP_2007_8_spatAQmaps_dly_interpol

Denby B, Schaap M, Segers A, Builtjes P, Horálek J, 2008b. Comparison of two data assimilation methods for assessing PM_{10} exceedances on the European scale. Atmospheric Environment, Vol. 42, pp 7122-7134.

Debry E, Malherbe L, Honoré C, Bessagnet B and Rouïl L, 2010. BCRD Incertitudes : Intégration probabiliste pour l'assimilation de données d'observation dans les modèles de qualité de l'air. Rapport final. Technical report, DRC-10-71769-02091A, INERIS, 2010.

De Fouquet Ch, Malherbe L, Ung A, 2011. Geostatistical analysis of the temporal variability of ozone concentrations. Comparison between CHIMERE model and surface observations. Atmospheric Environment, Vol 45, n°20, 3434-3446.

De Fouquet Ch, 2006. La modélisation géostatistique des milieux anthropisés. Habilitation à diriger des recherches. Mémoire des Sciences de la Terre N°2006-13, Académie de Paris, Université Pierre et Marie Curie, France, http://tel.archives-ouvertes.fr/tel-00536237/fr/

De Iaco S, Palma M, Posa D, 2005. Modeling and prediction of multivariate space–time random fields. Computational Statistics & Data Analysis, Vol 48, pp 525 – 547.

De Smet P, Horálek J, Coňková M, Kurfürst P, de Leeuw F, Denby B, 2010. European air quality maps of ozone and PM_{10} for 2008 and their uncertainty analysis. ETC/ACC Technical Paper 2010/10.
http://acm.eionet.europa.eu/reports/ETCACC_TP_2010_10_spatAQmaps_2008

De Smet P, Horálek J, Coňková M, Kurfürst P, de Leeuw F, Denby B, 2009. European air quality maps of ozone and PM10 for 2007 and their uncertainty analysis. ETC/ACC Technical Paper 2009/9.
http://acm.eionet.europa.eu/reports/ETCACC_TP_2009_9_spatAQmaps_2007

Deraisme J, Bobbia M, de Fouquet Ch, 2011. Contribution of geostatistics to the study of risks related to air pollution. Advanced Air Pollution, InTech.
http://www.intechopen.com/articles/show/title/contribution-of-geostatistics-to-the-study-of-risks-related-to-air-pollution

Dutot A L, Rynkiewicz J, Frédy E, Steiner 1, Rude J, 2007. A 24-h forecast of ozone peaks and exceedance levels using neural classifiers and weather predictions. Environmental Modelling and Software 22(9), pp 1261-1269.

Elbern H, Schmidt H, 2001. Ozone episode analysis by four dimensional variational chemistry data assimilation, J. Geophys. Res., 106, D4, pp 3569-3590, doi:10.1029/2000JD900448.

EC, 2011. Commission Staff Working Paper establishing guidelines for the agreements on setting up common measuring stations for PM2.5 under Directive 2008/50/EC on ambient air quality and cleaner air for Europe.
http://ec.europa.eu/environment/air/quality/legislation/pdf/sec_2011_77.pdf

EU, 2008. Directive 2008/50/EC of the European Parliament and of the Council of 21 may 2008 on ambient air quality and cleaner air for Europe. Official Journal of the European Union, 11.06.2008, L152/1-44.
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:152:0001:0044:EN:PDF

EU, 2005. Directive 2004/107/EC of the European Parliament and of the Council of 15 December 2004 relating to arsenic, cadmium, mercury, nickel and polycyclic aromatic hydrocarbons in ambient air. Official Journal of the European Union, 26.01.2005, L23/3-16.
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2005:023:0003:0016:EN:PDF

Fuentes M, Raftery AE, 2005. Model Evaluation and Spatial Interpolation by Bayesian combination of Observations with Outputs from Numerical Models. Biometrics, 61, 36-45.

Garaud D, Mallet V, 2011. Automatic calibration of an ensemble for uncertainty estimation and probabilistic forecast: Application to air quality, Journal of Geophysical Research, 116, D19304, doi:10.1029/2011JD015780.

Garaud D, Mallet V, 2010. Automatic generation of large ensembles for air quality forecasting using the Polyphemus system. *Geoscientific Model Development*, 3(1), pp 69-85.

GEMS Project, 2010. A Monitoring and Forecasting System for Atmospheric Composition. Final report of the GEMS project.
http://gems.ecmwf.int/do/get/PublicDocuments/1529/1434?showfile=true

GEOMON project. Report on station representativeness for key sites (D2.1.2). http://www.geomon.eu/science/act2/SciAct2_representativeness.php

Gerboles M, Buzica D, Brown RJC., Yardley RE, Hanus-Illnar A, Salfinger M, Vallant B, Adriaenssens E, Claeys N, Roekens E, Sega K, Jurasovic J, Rychlik S, Rabinak E, Tanet G, Passarella R, Pedroni V, Karlsson V, Alleman L, Pfeffer U, Gladtke D, Olschewski A, O'Leary B, O'Dwyer M, Pockeviciute D, Biel-Cwikowska J, TursicJ, 2011. Interlaboratory comparison exercise for the determination of As, Cd, Ni and Pb in PM_{10} in Europe. Atmospheric Environment, Vol. 45, pp. 3488-3499.

Gerboles M, Reuter HI, 2010. Estimation of the measurement uncertainty of ambient air pollution datasets using geostatistical analysis. JRC Scientific and Technical Reports, doi 10.2788/44902.

Groselj N, Zupan J, Reich S, Dawidowski L, Gomez D, Magallanes J, 2004. 2D mapping by Kohonen Networks of the air quality data from a large city. J. Chem. Inf. Comput. Sci. 44, 339–346.

Gräler, B, Gerharz, L, Pebesma E, 2011. Spatio-temporal analysis and interpolation of PM10 measurements in Europe. ETC/ACM Techncial Paper 2011/10. To be published at http://acm.eionet.europa.eu/reports/#tp

GUM (ISO/IEC Guide 98-3:2008). Uncertainty of measurement -- Part 3: Guide to the expression of uncertainty in measurement (GUM:1995).
http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=50461

Henne S, Brunner D, Folini D, Solberg S, Klausen J, Buchmann B, 2010. Assessment of parameters describing representativeness of air quality in-situ measurement sites. Atmos. Chem. Phys., 10, 3561-3581, 2010. www.atmos-chem-phys.net/10/3561/2010/ doi:10.5194/acp-10-3561-2010

Honoré C, Rouïl L, Vautard R, Beekmann M, Bessagnet B, Dufour A, Elichegaray C, Flaud J-M, Malherbe L, Meleux F, Menut L, Martin D, Peuch A, Peuch V-H, Poisson N, 2008. Predictability of European air quality: The assessment of three years of operational forecasts and analyses by the PREV'AIR system. Journal of Geophysical Research –Atmospheres, 113, D04301, doi :10.1029/2007JD008761.

Horálek J, de Smet P, de Leeuw F, Coňková M, Denby B, Kurfürst P, 2010. Methodological improvements on interpolating European air quality maps. ETC/ACC Technical Paper 2009/16.
http://acm.eionet.europa.eu/reports/ETCACC_TP_2009_16_Improv_SpatAQmapping

Horálek J, de Smet P, de Leeuw F, Denby B, Kurfürst P, Swart R, 2008. European air quality maps for 2005 including uncertainty analysis. ETC/ACC Technical paper 2007/7.
http://air-climate.eionet.europa.eu/reports/ETCACC_TP_2007_7_spatAQmaps_ann_interpol

Horálek J, Denby B, de Smet P, de Leeuw F, Kurfürst P, Swart R, de Noije T, 2007. Spatial mapping of air quality for European scale assessment. ETC/ACC Technical paper 2006/6.
http://acm.eionet.europa.eu/reports/ETCACC_TechnPaper_2006_6_Spat_AQ

Ibarra-Berastegi G, Elias A, Barona A, Saenz J, Ezcurra A, Diaz de Argandona J, 2008. From diagnosis to prognosis for forecasting air pollution using neural networks: Air pollution monitoring in Bilbao, Environmental Modelling & Software, Vol 23, Issue 5, pp 622-637, ISSN 1364-8152, DOI: 10.1016/j.envsoft.2007.09.003.

Joly M, Peuch V-H, 2012. Objective classification of air quality monitoring sites over Europe. Atmospheric Environment, Vol. 47, pp. 111-123.

Kolehmainen M, Martikainen H, Hiltunen T, Ruuskanen J, 2000. Forecasting Air Quality Parameters Using Hybrid Neural Network Modelling. Environmental Monitoring and Assessment, 2000, Vol 65, No 1-2, pp 277-286.

MACC project. Monitoring atmospheric composition & climate. http://www.gmes-atmosphere.eu/

Macé T, Raventos C, Mathé F, 2009a. Rédaction de guides pratiques de calcul d'incertitude et formation des AASQA (Rapport 1/5). LCSQA Technical report, DRC-10-103351-01139A, www.lcsqa.org

Macé T, Raventos C, Mathé F, 2009b. Estimation des incertitudes sur les concentrations massiques de particules mesurées en automatique (Rapport 2/5). LCSQA Technical report, DRC-10-103321-00122A, www.lcsqa.org

Magneron C, Demongin T, 2011. Moving-geostatistics for automated interpolation - Application to environmental data. Les Journées de géostatistique, Fontainebleau, France, 15-16 September 2011.

Malherbe L, Beauchamp M, de Fouquet Ch, Létinois L, Ung A, 2011. Estimation of the areas of air quality limit value exceedances on national and local scales. A geostatistical approach. Proceedings of the Spatial2 Conference: Spatial Data Methods for Environmental and Ecological Processes, Foggia, Italy, 1-2 September 2011. Available at http://dspace-unibg.cilea.it/handle/10446/898/?locale=en

Malherbe L, Ung A, 2009. Travaux relatifs à la plate-forme nationale de modélisation PREV'AIR : Réalisation de cartes analysées d'ozone (Study related to the national modelling platform PREV'AIR : production of analysed maps). LCSQA Technical report, DRC-10-103351-01139A, www.lcsqa.org

Moussiopoulos N, Isaksen I (eds), 2007. Proceedings of the ACCENT Workshop on Model Benchmarking and Quality Assurance 29/30 May 2006, Thessaloniki, Greece.

Pérez Ballesta P and collaborators, 2008. EC Intercomparison of VOC measurements between National Reference Laboratories (AQUILA network). JRC Technical and Scientific Reports, doi 10.2788/11627.

Pérez P, Reyes J, 2006. An integrated neural network model for PM_{10} forecasting. Atmospheric Environment. Vol 40, Issue 16, May 2006, pp 2845-2851.

Peuch V-H, Joly M, 2011. Objective classification of air quality observation sites over Europe. International Workshop on Air Quality Forecasting Research, Québec City, Canada, 16-18 November 2010.

Ranchin T, Wald L, 2000. Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation. Photogrammetric Engineering and Remote Sensing, Vol 66(1), pp 49-61.

Rivoirard J, 1990. A review of lognormal estimators for in situ reserves (teacher's aid). Math. Geol. 22 (2), pp 213-221. DOI: 10.1007/BF00891825
http://www.springerlink.com/content/g06u46xxk3676711/

Rivoirard J, 2003. Course on multivariate geostatistics. Cours C-17 3. Ecole des Mines de Paris. http://cg.ensmp.fr/bibliotheque/public/RIVOIRARD_Cours_00608.pdf

Roth C, 1998. Is lognormal kriging suitable for local estimation? Math. Geol. 30 (8), pp 999-1009.

Rouïl L (ed), 2011a. Evaluation Report of the Air quality assessments in Europe for 2007. MACC R-EVA project deliverables.
http://www.gmes-atmosphere.eu/documents/deliverables/r-eva/. Ref: D-R-EVA_2.2.

Rouïl L 2011b. Assessment Report: Air quality in Europe for 2007. MACC R-EVA project deliverables. http://www.gmes-atmosphere.eu/documents/deliverables/r-eva/. Ref: D-R-EVA_3.1.

Singh K, Jardak M, Sandu A, Bowman K, Lee M, Jones D, 2011. Construction of non-diagonal background error covariance matrices for global chemical data assimilation. Geoscientific Model Development, 4, pp 299–316.

Tombette M, Mallet V, Sportisse B, 2009. PM10 data assimilation over Europe with the optimal interpolation method. Atmospheric Chemistry and Physics, 9:1, pp 57-70.

UBA (Spangl W, Schneider J, Moosmann L, Nagl C), 2007. Representativeness and classification of air quality monitoring stations – final report. Service contract to the European Commission – DG Environment Contract No. 07.0402/2005/419392/MAR/C1. Reports Bd. REP-0121. Umweltbundesamt (UBA), Wien.

Ward JH, 1963. Hierarchical Grouping to optimize an objective function. Journal of American Statistical Association, *58*(301), 236-244. http://iv.slis.indiana.edu/sw/data/ward.pdf

Wu L, Mallet V, Bocquet M, Sportisse B, 2008. A comparison study of data assimilation algorithms for ozone forecasts. J. Geophys. Res., 113, D20310, doi:10.1029/2008JD009991