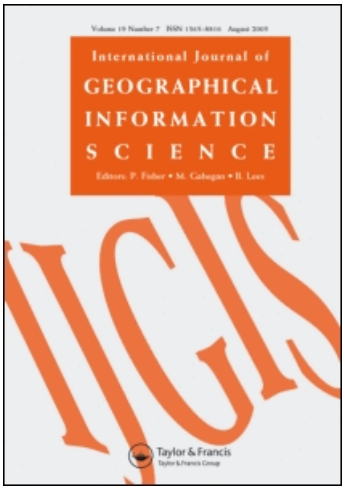
This article was downloaded by: *[Institut Fuer Informatik]*
On: *14 January 2010*
Access details: *Access Details: [subscription number 768490051]*
Publisher *Taylor & Francis*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



International Journal of Geographical Information Science
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713599799

Interactive visualization of uncertain spatial and spatio-temporal data under different scenarios: an air quality example

Edzer J. Pebesma [a]; Kor de Jong [a]; David Briggs [b]
[a] Faculty of Geosciences, Utrecht University, The Netherlands [b] Imperial College, London, UK

To cite this Article Pebesma, Edzer J., de Jong, Kor and Briggs, David(2007) 'Interactive visualization of uncertain spatial and spatio-temporal data under different scenarios: an air quality example', International Journal of Geographical Information Science, 21: 5, 515 — 527
To link to this Article: DOI: 10.1080/13658810601064009
URL: http://dx.doi.org/10.1080/13658810601064009

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

International Journal of Geographical Information Science
Vol. 21, No. 5, May 2007, 515–527

Taylor & Francis
Taylor & Francis Group

Interactive visualization of uncertain spatial and spatio-temporal data under different scenarios: an air quality example

EDZER J. PEBESMA†, KOR DE JONG† and DAVID BRIGGS‡
†Faculty of Geosciences, Utrecht University, The Netherlands
‡Imperial College, London, UK

(*Received 23 November 2005; in final form 13 October 2006*)

This paper introduces a method for visually exploring spatio-temporal data or predictions that come as probability density functions, e.g. output of statistical models or Monte Carlo simulations, under different scenarios. For a given moment in time, we can explore the probability dimension by looking at maps with cumulative or exceedance probability while varying the attribute level that is exceeded, or by looking at maps with quantiles while varying the probability value. Scenario comparison is done by arranging the maps in a lattice with each panel reacting identically to legend modification, zooming, panning, or map querying. The method is illustrated by comparing different modelling scenarios for yearly NO_2 levels in 2001 across the European Union.

Keywords: Dynamic graphics; Maps; Probability density function; Cumulative density function; Environmental modelling

Downloaded By: [Institut Fuer Informatik] At: 09:05 14 January 2010

1. Introduction

Spatio-temporal data, attributes that vary over space and time, may be measured, as for example with air quality monitoring network data, or they may be modelled, i.e. be estimated or predicted by a model. Except for small measurement error, measured data usually have no uncertainty associated with them. In the case where the data result from a model however, the associated uncertainties may be considerable. Possible sources for the uncertainties may be that spatial or temporal interpolation or extrapolation took place, or that a physical, chemical or otherwise mechanistic model was used to model the attribute for unmeasured locations and/or moments in time where initial conditions, boundary conditions, model parameters and/or model structure were subject to uncertainty (e.g. Heuvelink 1998).

In many applications the results of a modelling study are presented as maps with predicted or estimated values. There may be some report that uncertainties were known, were quantified, and there may even be a map with associated standard errors for the predicted value. The information available is however often a full probability distribution for the attribute under study for each spatial location and moment in time. This distribution may be arrived at using assumptions of some parametric distribution, like the normal or Gaussian distribution, or may be based on a set of simulated spatio-temporal fields resulting from a Monte Carlo experiment. Because this (approximation of the) spatio-temporal probability distribution is a *function* (curve) the full information we have for each point in space/time and for any possible single map view is necessarily reduced. The need to deal with uncertainties becomes more urgent when we model them or want to

International Journal of Geographical Information Science
ISSN 1365-8816 print/ISSN 1362-3087 online © 2007 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/13658810601064009

compare them for different scenarios. These scenarios may refer to the model structure used, or e.g. to different boundary conditions for models forecasting future developments of an attribute. When two scenarios come up with different predictions, the immediate question is whether the difference can be attributed to chance, resulting from uncertainty in one or both of the scenarios, or to *really* opposing outcomes. Comparing prediction *intervals* instead of predicted values (Pebesma and De Kwaadsteniet 1997) is in this context already a step forward, and is simply obtained by classifying the probability distributions.

In this paper we present a simple method to communicate the full spatio-temporal probability distribution under different scenarios to end-users, without needing to reduce information. We implemented this method, and have provided examples of the method.

The organization of the paper is as follows: we will first introduce the data used for the example; next we will explore the ideas implemented, and show a number of resulting maps and graphs. We conclude with a discussion.

To illustrate the visualization procedure, we use data from a study where different interpolation strategies for air quality variables across the European Union were applied and compared. Here, this interpolation study serves only the purpose of showing how we can end up with spatial probability distributions, and how we can explore them visually. This paper does not intend to defend either of the modelling strategies made there, nor does it provide all detailed modelling choices made. The point of this paper is to use its results to illustrate our visualization approach. We first briefly explain the modelling approach used.

Downloaded By: [Institut Fuer Informatik] At: 09:05 14 January 2010

2. Spatio-temporal probability distributions: an example

2.1 *Air quality data*

The example data used are European air quality data used from Airbase (http://airbase.eionet.eu.int). We chose the variable NO_2, and used annual averages from hourly recorded measurements, for 2001. We selected stations where at least 75% of the hourly measurements were valid. Figure 1 shows a map of Europe with the selected monitoring network stations.

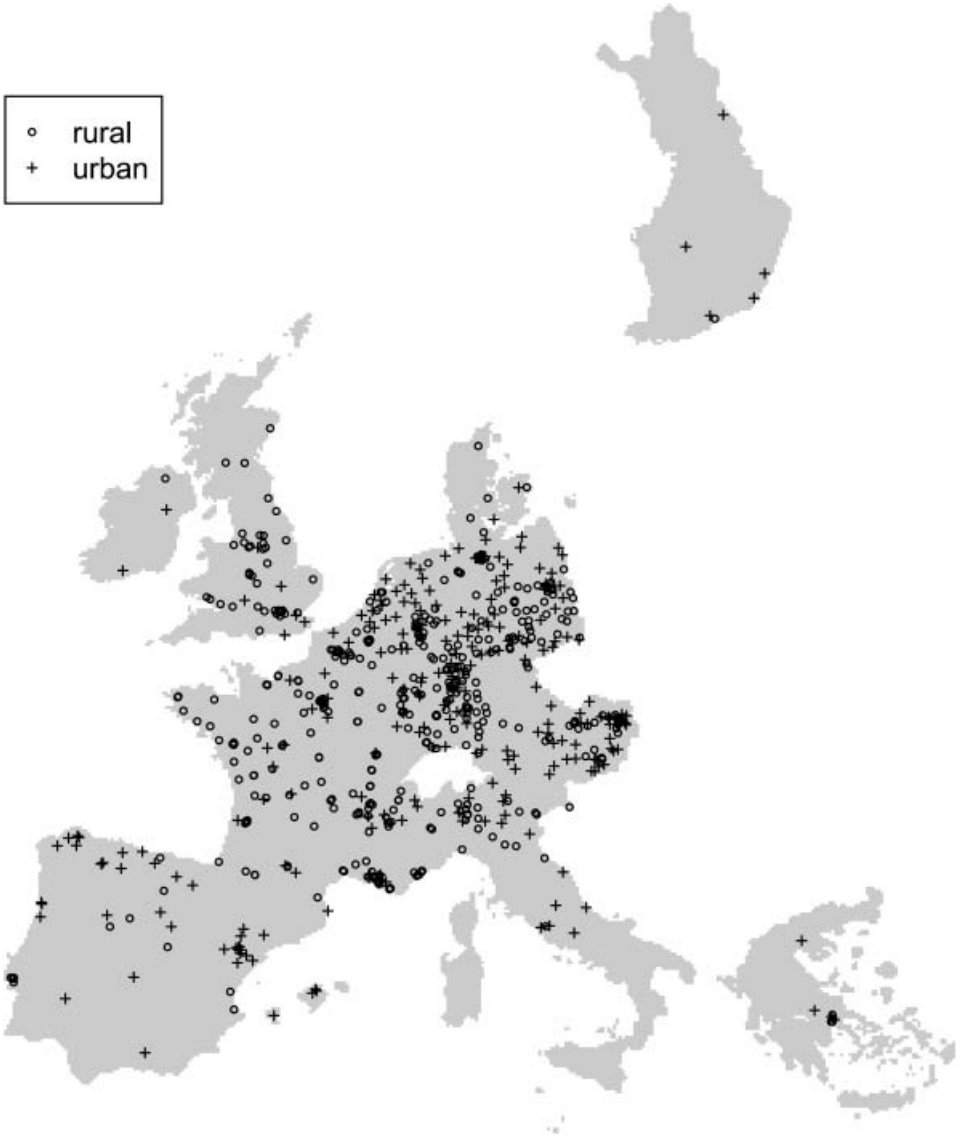
We will use three different approaches, or scenarios, to interpolate annual average NO_2 levels, which differ with respect to the site type and to the external covariate information used. First we will explain the general method used for all three, and how they lead to predictive probability density functions.

2.2 *Interpolating air quality*

Three different modelling (interpolation) scenarios are used to predict NO_2 concentration on a $1\,\text{km} \times 1\,\text{km}$ grid for Europe; all three use universal kriging (sometimes called kriging with external drift or regression kriging) for interpolation (Chilès and Delfiner 1999).

In the universal kriging model the annual average NO_2 concentration on the log-scale, $Z(s_i)$, at location s_i is modelled as the sum of a meaningful trend and a random residual:

$$Z(s_i) = \sum_{j=0}^{p} X_j(s_i)\beta_j + e(s_i), \quad i = 1, \ldots, n \tag{1}$$



Downloaded By: [Institut Fuer Informatik] At: 09:05 14 January 2010

Figure 1. Selected NO_2 monitoring stations from Airbase, used for the spatial modelling.

with $X_f(s_i)$ known and relevant predictors (covariates) at location s_i (i.e. no simple functions of coordinates, but layers in the GIS data base), with $X_0(s) \equiv 1$ to allow for an intercept β_0, and $e(s)$ an intrinsically stationary random function with zero expectation and known variogram (Chilès and Delfiner 1999). Residual variograms were modelled isotropically, by substituting ordinary least squares residuals for the true residuals.

Predictors were chosen from a large set of available predictors using a forward stepwise procedure where candidate predictors entered the model only when (i) they added at least 1% to the adjusted R^2, and (ii) the sign of the regression coefficient agreed with the expected sign at the moment of inclusion.

2.3 *From prediction and prediction error to cumulative density functions*

We define the predictive cumulative probability density function (CDF) for scenario k and location s_0 as a function of concentration level c as

$$P_k(c, s_0) = \text{Prob}(Z_k(s_0) < c), \tag{2}$$

where Prob(x) refers to probability of event x, and with s_0 an arbitrary location on the spatial prediction grid. On locations where we measured $Z_k(s_i)$, we can say that $P_k(c, s_i)$ is either 0 or 1, as the measurement is below or above c, when measurement error is absent and c different from $Z(s_i)$. We are however concerned with predictions on a regular grid where we did not measure Z, and for these cases, depending on an interpolation approach (scenario) k, we can obtain a continuous probability $P_k(c, s_0)$ as follows.

If we are willing to assume that prediction errors for NO_2 on the log-scale are, for all three modelling scenarios, normally distributed, we can obtain the values for (2) from the cumulative density function of the standard normal distribution Φ, by $P_k(c, s_0) = \Phi\left(\left(c - \hat{Z}_k(s_0)\right)/\sigma_k(s_0)\right)$, with $\hat{Z}_k(s_0)$ and $\sigma_k(s_0)$ the predicted value and prediction standard error at location s_0, respectively, for scenario k. Function Φ is tabulated in basic statistics text books, present in hand calculators; the pnorm function in the R statistical program (Ihaka and Gentleman 1996) can also be used. Next, we can back-transform this distribution from the log-scale to the observation scale by taking the exponent of c in $P_k(c, s_0)$.

For spatio-temporal problems, $Z_k(s_0)$ generalizes to $Z_k(s_0, t_0)$ and the CDF becomes $P_k(c, s_0, t_0)$.

2.4 *The three interpolation scenarios*

The three modelling approaches (or scenarios) differ with respect to the regressors included, the residual variogram model, and the spatial extent (or scale) of the effects, or covariates included. The modelling scenarios for NO_2 concentrations are:

(1) *global scale*—at this scale global, background NO_2 concentration was modelled based on globally varying variables not influenced by human action. On the global scale for NO_2 the predictors selected were:

- square root of altitude,
- one climate factor (obtained from a factor analysis of a large number of climate variables),
- square root of distance from the sea.

For the modelling and prediction, background monitoring stations were selected.

(2) *rural scale*—the rural scale extends the global scale, tuning it to local and human influence variables. The observations used are identical to the global scale, but apart from the global scale predictors, for the rural scale the following predictors were additionally selected:

- square root of difference in altitude between the central grid square and the mean of the 24 surrounding grid squares in a 5×5 window on a $1\,\text{km} \times 1\,\text{km}$ grid,
- amount of high density residential area in a $5\,\text{km} \times 5\,\text{km}$ window,
- amount of minor roads in a $5\,\text{km} \times 5\,\text{km}$ window,

Downloaded By: [Institut Fuer Informatik] At: 09:05 14 January 2010

- amount of major roads in a 5 km × 5 km window,
- the fraction of non-residential urban land in a 21 km × 21 km window.

Major and minor roads were derived (reclassified) from relatively low resolution data on roads across the EU. These data were acquired from AND Ltd (http://www.and.com). Non-residential urban land was derived from the CORINE Land Cover 1990 map. Land cover data for 2000 was released in 2005, too late to be incorporated in this study. Window averages were calculated using approximately circular windows.

(3) *urban scale*—for a set of distinct measurements, directly influenced by urban activity, a regression model was made for (on the log-scale) residuals from the global model. This regression model selected the following variables:

- high density residential area in the 1 km × 1 km grid cell of the monitoring network station,
- major roads in the 1 km × 1 km grid cell of the monitoring network station,
- minor roads in the 1 km × 1 km grid cell of the monitoring network station.

For prediction, the urban scale residual predictions were added to the global scale predictions to get a prediction on the log-scale. The same was done for the prediction variances.

On the log-scale, we assumed a normal distribution of the prediction errors.

Figure 2(*a*) gives an example of a single cumulative probability density function obtained for the NO_2 data set at a specific location. From such a curve we can read:

- the median, which is the concentration value where this curve reaches cumulative probability value 0.5;
- the degree of uncertainty: the flatter the curve, the more uncertain, the steeper the curve the less uncertainty the data have; certainty about the concentration value would result in a step from 0 to 1 at that particular value;

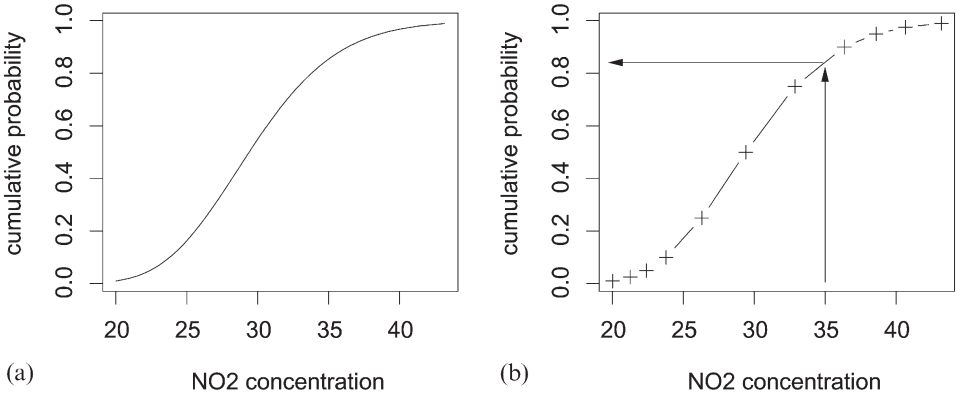


Figure 2. (*a*) cumulative probability plot for a given location s_0 and given scenario; values are obtained by assuming a normal distribution on the log-scale, and are drawn from P values ranging from 0.01 to 0.99; (*b*) deriving cumulative probability P from a discretized representation of the CDF, using linear interpolation.

Downloaded By: [Institut Fuer Informatik] At: 09:05 14 January 2010

- quantiles, by choosing a cumulative probability value (y) and reading the corresponding NO_2 concentration (x);
- cumulative probability, by choosing an NO_2 value (x) and reading the corresponding probability value (y).

3. Dimensions

We want to explore the spatio-temporal CDF $P_k(c, s_i, t_j)$. In our case space spans two dimensions ($s_i=(x_i, y_i)$), so the total number of dimensions spanned is six: $\{P, k, c, x, y, t\}$. Their characteristics are:

- space, s_i: either two- or three-dimensional Euclidean space, or possibly latitude/longitude locations on a sphere; space may be regularly discretized (gridded) or irregularly discretized (vector). Our implementation deals with regular grids;
- time, t_j: although we consider a non-temporal process in the example, the attribute Z may be modelled as varying over space and time: $Z(s_i, t_j)$ then has as CDF $P_k(c, s_i, t_j)$; time may be regularly or irregularly discretized;
- scenario, k: inevitably scenario is a discrete entity that may or may not have some natural order;
- attribute level c (concentration in our example); a continuous variable that may be discretized in some way, but we will approach it slightly differently;
- cumulative probability, P: a continuous variable ranging from 0 to 1.

4. Visualization

We will be looking at maps that plot an attribute as a function of 2D space. An obvious visualization of 3D space is to look at subsequent 2D slices through a 3D body. Time can also be examined by showing an animation of 2D maps (or slices).

The actual function we want to visualize is that of P as a function of c, and c as a function of P. We do this by making maps of either $P(c, s, t_j)$ and modifying c dynamically (follow the arrow in figure 2(b)), or by making maps of $c(P, s, t_j)=P^{-1}(c, s, t_j)$, the quantile function which inverts the cumulative probability function (revert the arrow direction in figure 2(b)).

4.1 *Scenarios*

We define scenarios as discrete, different entities, e.g. resulting from different modelling approaches and/or levels of information availability. A scenario can be identified with a short, descriptive string.

Building on the ideas from Cleveland (1993), implemented in the trellis library of S-Plus and the lattice package in R, we compare maps corresponding to different scenarios by (i) avoiding duplication where possible (e.g. of axes, legends) to optimally dedicate space to graphing data, (ii) make everything the same for each scenario except the attribute variability shown; this includes legend classes, legend class colour, plotting area, and (iii) give for each scenario an identical response to a query (zoom, pan, identify) on a single map.

4.2 *Space–time*

Although our approach is not limited to gridded data, the implementation of our method uses a regular discretization in space (i.e. gridded data in space). The

Downloaded By: [Institut Fuer Informatik] At: 09:05 14 January 2010

discretization chosen for time is regular but less rigid: the available time points do not have to be equidistant, but their distance should be a multiple of a basic time step. This allows for variability in time step, allowing periods with little variability to be stored with little loss of space, while still allowing for a simple, discrete index.

4.3 *Probability*

For each scenario k and each space–time point (s_i, t_j) we have a cumulative density function $P(c, \cdot)$. One way would be to discretize c and to store the P values associated with each attribute (concentration) level. When (part of) the cumulative density functions are rather steep, i.e. the P values increase fast between two values of c that are close together, we would need a large number of c values to characterize the full CDF for each space–time point. The alternative, which we used, is to discretize P, and store the corresponding c values for each space–time point. We can then approximate P for each c by (linearly) interpolating c between the two corresponding P values (figure 2(*b*)).

To characterize a cumulative probability density function we also used a semi-regular discretization: probability steps do not have to be equidistant but their distance should be a multiple of a basic probability step. We used for example the nine cumulative probabilities 0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95 and 0.99 in our example, resulting in a basic probability step of 0.01 without needing the full 99 maps for each moment in time, that would have been needed if discretization of P would have been regular.

4.4 *Filling the scenario slots differently*

Of course, it is not necessary for the scenario slot to contain scenarios. If we consider a single scenario, it can be used to compare future projections for different time steps, e.g. for 2010, 2015, 2020 and 2030, in order to compare their uncertainties.

5. Cross-sections: map views, graphs and interactive analysis

Recalling (2), which says that the spatial cumulative density function is a function of location s_i and attribute level c

$$P_k(c, s_i) = \text{Prob}(Z_k(s_i) < c),$$

the user can for a moment in time generate map views for:

- a quantile: given a level p, the maps show at each location the value c for which $P_k(c, s_i) = p$ (example: figure 3);
- the probability being below/above a threshold: given a level c, the maps show for each location the cumulative density function $P_k(c, s_i)$ (example: figure 4);
- the classified probabilities, expressed as $(1-\alpha)$-prediction intervals relative to an attribute level c—see table 1.
  An example of classified probabilities, for (approximate) 90% prediction intervals ($\alpha = 0.1$) is shown in figure 5.

Further, for a given location s_i and moment in time t_j we can give the graph of the k curves of P as a function of c (lower-right quadrant of figures 3, 4 and 5), in which the curves should have been labelled or drawn in colour to identify the scenario, k.

Downloaded By: [Institut Fuer Informatik] At: 09:05 14 January 2010

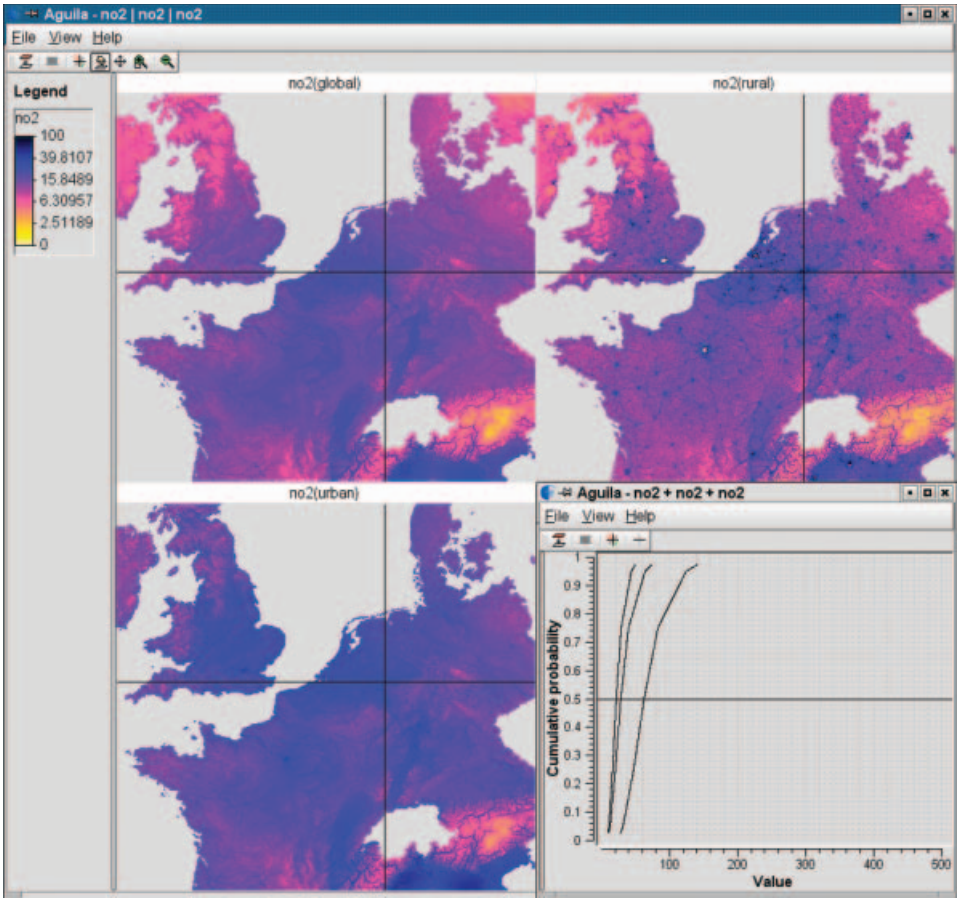


Figure 3. Map view for three scenarios showing the 0.5 quantile (median) of the NO_2 predictive probability distribution, along with a graph of the three cumulative probability functions under the map cursor (cross) and the probability cursor (horizontal line), adjustable for the quantile probability level.

We can now vary interactively:

- the map cursor by clicking/dragging over the map; the value under cursor changes in the cursor widget (not shown) and the graph of P as a function of c changes, as s_i changes;
- the area shown by zooming or dragging (panning) in one of the map views; zooming and panning affects all scenarios k;
- the time cursor by moving the time pointer in the time panel (not shown);
- the cumulative probability (P) level between 0 and 1, by moving (click + drag) the horizontal line in the CDF plot of figure 3;
- toggle whether we want to modify either the P level or the c level;
- the level cursor (i.e. level c) for which we want to show exceedance probabilities by moving (click + drag) the vertical line in the CDF plot of figures 4 and 5;
- the confidence *level* of confidence intervals by modifying $1-\alpha$ (default 0.95);
- the legend minimum, maximum, number of legend classes and colour ramp.

Downloaded By: [Institut Fuer Informatik] At: 09:05 14 January 2010

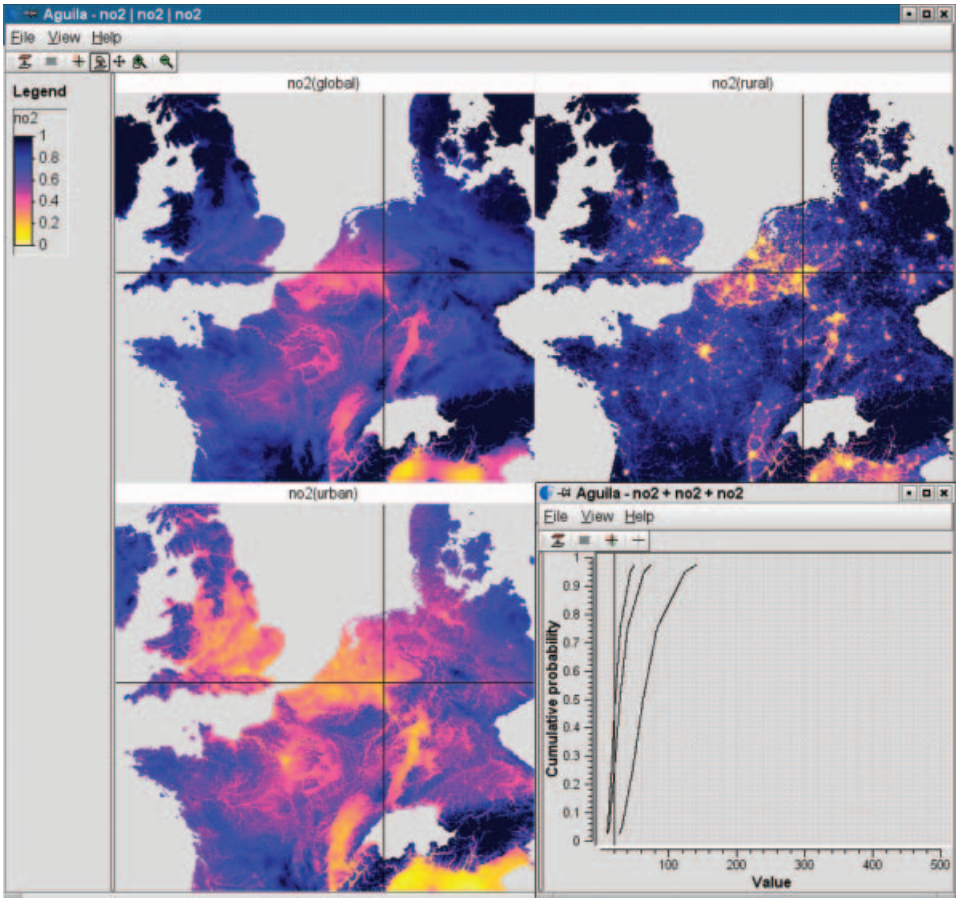


Figure 4. Map view for three scenarios showing probability of the true value of NO_2 being below 20 ppm, along with a graph of the three cumulative probability functions under the map cursor and the level cursor (vertical line), adjustable for the attribute (concentration) level below which we want probabilities. Note that the legend header should have been 'cumulative probability' instead of 'no2'.

6. Discussion

The method and tool presented here allows the interactive exploration of marginal cumulative probability density functions for each space–time point of a regularly discretized spatio-temporal random field. Thereby, it ignores spatial and/or temporal correlation among these functions. The method combines (i) ideas commonly implemented in interactive GIS viewers (e.g. ESRI 2004), (ii) dynamic graphics and linked windows (Cleveland and McGill 1988, Swayne *et al.* 1992), (iii)

Table 1.

value	legend class	condition
1.0	lower	$P_k(c, s_i) > 1-\alpha/2$
0.5	not distinguishable	$\alpha/2 < P_k(c, s_i) < 1-\alpha/2$
0.0	higher	$P_k(c, s_i) < \alpha/2$

Downloaded By: [Institut Fuer Informatik] At: 09:05 14 January 2010

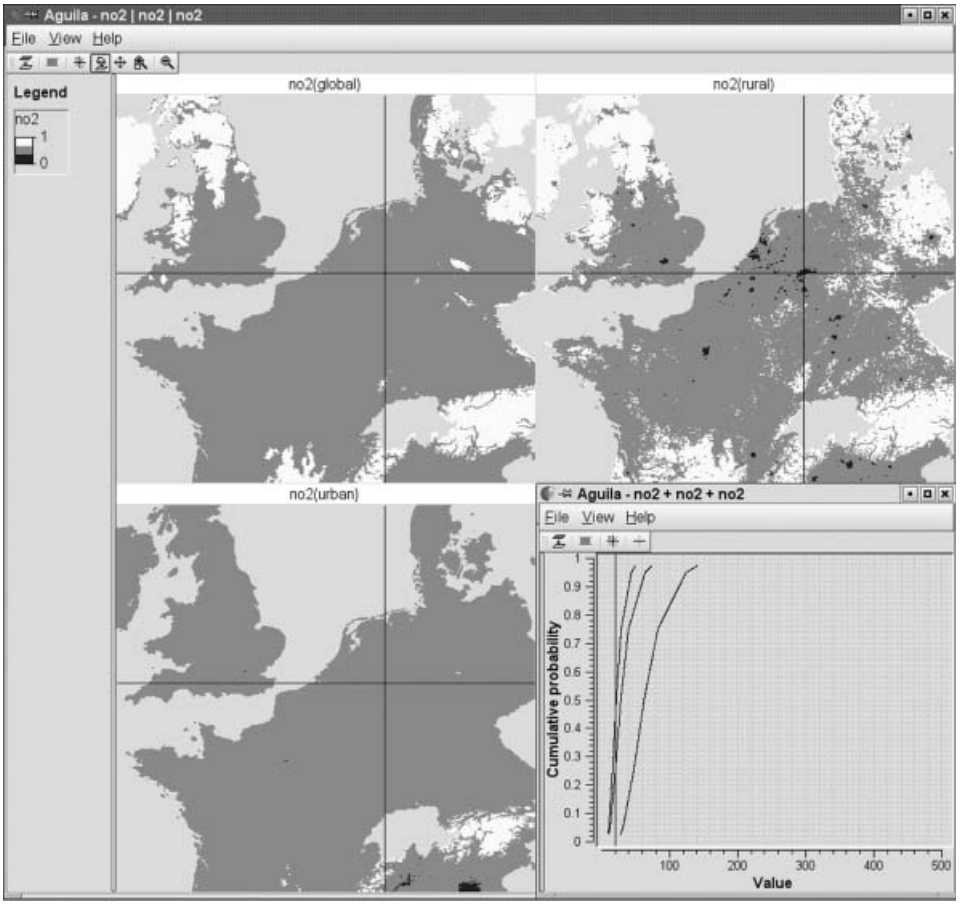


Figure 5. Map view for three scenarios showing classified cumulative density curves; with respect to the attribute (NO_2 concentration) level 20 ppm, we classify CDF curves as 'lower' (1, if $P>0.95$, dark grey), 'higher' (0, if $P<0.05$, light grey) 'not distinguishable' (0.5, if $0.05<P<0.95$). The fourth panel shows the level cursor (vertical line), which is dynamically adjustable for the attribute (concentration) with respect to which we want to classify probabilities.

the concept of conditioning plots, or trellis plots (Cleveland 1993), and (iv) previous attempts to show maps with confidence intervals (Pebesma and de Kwaadsteniet 1997) and the idea that (cumulative) probability is yet another dimension in GIS (Pebesma *et al.* 2000). We focus on continuous variables. When the emphasis is to visually explore observations, a dynamic graphics tool like xgobi (Swayne *et al.* 1992) or ggobi (http://www.ggobi.org/) may suffice.

As the actual method involves interaction with a computer program, we strongly encourage interested readers to try the real application (see the next section): a printed paper can never approximate the overwhelming experience of a dynamic graphics application. Dynamic graphics systems answer questions and generate new hypotheses faster than we can formulate them.

Many people are reluctant to present exceedance probabilities for a chosen threshold because thresholds unavoidably have an arbitrary component to them: why do we focus on the critical level of 35 and not 40? Given that cumulative or

Downloaded By: [Institut Fuer Informatik] At: 09:05 14 January 2010

exceedance probabilities are one way of exploring probability density functions, the need for a dynamic tool becomes imminent: after considering the probabilities of being above a chosen threshold, questions immediately following about how these probabilities change with threshold can be answered at the instant they are posed.

Cleveland (1993) laid out the basic ideas underlying many of the visualizations presented here. The two systems resulted from his work (trellis in S-Plus, lattice in R) do provide sufficient plotting capabilities to make the type of composite maps we present here, but they lack enough interactive functionality with the data plotted. Therefore, and for reasons of computational efficiency, our application was built from scratch.

6.1 *Other methods for visualizing uncertain data*

Downloaded By: [Institut Fuer Informatik] At: 09:05 14 January 2010

It is sometimes advocated (e.g. Goovaerts 1997) that simulation is the solution when there is a need to deal with uncertainty. Understanding uncertainty by looking at simulated random fields may help but is often not easy: it is hard to get an idea of exceedance probabilities by looking at (many of) them.

However, the tool presented here can be used to visualize sets of simulations in three ways:

(1) by computing a set of quantiles from a (large) set of simulations, and visualizing these using the tool we present here (note that under many conditions, like the examples used in this paper, simulation is not needed to obtain these quantiles);
(2) by using the time dimension to show simulations in an animated fashion (this will show sudden, large transitions, which can be avoided by smoothing transitions, see Ehlschlaeger *et al*. (1997));
(3) by using the scenario slot to show and compare realizations side by side.

Although simulations could be shown as a spatial time series, their order is usually of no value; it may also be useful to order them according to some posterior likelihood (Switzer 2000). Simulated random fields do show the spatial variation and correlation present in the attribute, whereas quantiles lose this information, and will be more smooth than the variable modelled.

Another, static alternative to visualizing data and their (predictive) uncertainty was given by Hengl *et al*. (2003) and Hengl *et al*. (2004). It uses a two-dimensional colour legend where one dimension (a colour scale, or hue range) represents the predicted value ($\hat{Z}(s_0)$) and a second dimension (greyness, whiteness, or saturation) increases with prediction error ($\sigma(s_0)$). Compared to our method, this approach has the obvious advantage of being static, but also a number of drawbacks:

- by using two dimensions, it still *separates* first order (Z, mean) from second order (σ, uncertainty) effects; our approach does not separate them, because it addresses the probability density function that unites them;
- on two-dimensional legends, matching between colours on the map and values in the legend is difficult;
- when $\sigma(s_0)$ increases, colour vanishes, and it becomes impossible to distinguish between values actually predicted;
- the method needs arbitrarily chosen threshold values for $\sigma(s_0)$, which play an important role in the resulting map.

Simple interpolation algorithms, such as ordinary kriging, show predicted surfaces where hot spots coincide with the data locations. Taking the predictive probability into account, it may turn out that the locations with the largest probability being above a given high critical level may *not* coincide with data locations, but rather with locations where uncertainty is large.

6.2 *User's perspective*

We did not evaluate extensively how happy prospective users (scientists, policy makers) are with the visualization opportunities offered here (Evans 1997, Aerts 2004). Although this would be very valuable to evaluate, especially in a context where several visualization alternatives were compared, it is simply beyond the scope of the current study, which is about presenting a new alternative.

Using our visualization approach, users may not actually get a better understanding of the actual phenomena studied, but they definitely get a better, not only qualitative but also quantitative understanding of the limits of our knowledge about the phenomena. Compared to studying maps with predicted (interpolated) values only, we believe that showing probability density functions better represent what we know, and what we do not know. The tool we present here allows scientists to explore these functions. Policy makers, although they may not like it, should be aware that predictions are different from reality. We are convinced that policy makers are aware that a modelled value is different from the true value, and that scientists should reveal the limits of their knowledge.

6.3 *Future plans*

In the probability view, the current implementation gives, strictly speaking, no exceedance probabilities but one minus the exceedance probability. An option to choose between these will be added. Another view that we may add is that of the actual probability *density* functions, with (shaded) the area under the density curve that is selected (or results from a chosen quantile). Although this poses a scaling problem as probability densities may vary by many orders of magnitude, their notion and connection to histograms may give an alternative and sometimes easier understanding of probabilities.

A foreseen future development is to extend the spatial visualization, which is currently limited to gridded data, with polygon data. Also, work is in progress to integrate this tool with the R environment for statistical computing.

7. Availability and applicability

The tool developed that implements the ideas presented here is called *Aguila*, which is Latin for eagle. The software is available in binary and source code form from http://pcraster.geo.uu.nl/projects/aguila/index.html

The source code is distributed under the GNU General Public License (GPL), available from http://www.fsf.org/licensing/licenses/gpl.html. Aguila uses the GDAL library (http://www.gdal.org/) to read the most commonly used grid formats.

As the tool is available in source code form, it can easily be and may get integrated in a modelling environment. Because the modelling environment can be anything that yields spatial or spatio-temporal probability distributions (or just time series, or scenarios of maps, or just only grid maps) it can range from e.g. flood

Downloaded By: [Institut Fuer Informatik] At: 09:05 14 January 2010

forecasting systems based on hydrological models using short term weather forecast scenarios as input to spatial prediction model building tools.

Acknowledgements

Much of the work for the development of the visualization tool presented here was supported by the EU-project APMOSPHERE (EVK2-2002-00577), one of the preliminary GMES projects under the 5th framework. Derek Karssenberg provided valuable suggestions. Tomislav Hengl and an anonymous reviewer are acknowledged for their helpful reviews.

References

AERTS, J.C.J.H., CLARKE, K.C. and KEUPER, A.D., 2003, Testing popular visualization techniques for representing model uncertainty. *Cartography and Geographic Information Science*, **30**, pp. 249–261.

CHILÈS, J.-P. and DELFINER, P., 1999, *Geostatistics, Modeling Spatial Uncertainty* (New York: Wiley).

CLEVELAND, W.S., 1993, *Visualizing Data* (Summit, NJ: Hobart Press).

CLEVELAND, W.S. and MCGILL, M.E., 1988, *Dynamic Graphics for Statistics* (California: Wadsworth & Brooks Cole).

EHLSCHLAEGER, C.R., SHORTRIDGE, A.M. and GOODCHILD, M.F., 1997, Visualizing spatial data uncertainty using animation. *Computers & Geosciences*, **23**, pp. 387–395.

ESRI, 2004, *Getting Started With ArcGIS* (Redlands, CA: ESRI).

EVANS, B.J., 1997, Dynamic display of spatial data reliability: does it benefit the map user? *Computers & Geosciences*, **23**, pp. 409–422.

GOOVAERTS, P., 1997, *Geostatistics for Natural Resources Evaluation* (USA: Oxford University Press).

HENGL, T., 2003, Visualisation of uncertainty using the HSI colour model: computations with colours. In *Proceedings of the 7th International Conference on GeoComputation*, University of Southampton, 8–10 September 2003 (CD-ROM).

HENGL, T., HEUVELINK, G.M.B. and STEIN, A., 2004, A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, **122**, pp. 75–93.

HEUVELINK, G.B.M., 1998, *Error Propagation in Environmental Modelling with GIS* (London: Taylor & Francis).

IHAKA, R. and GENTLEMAN, R., 1996, R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, **5**, pp. 299–314.

PEBESMA, E.J. and DE KWAADSTENIET J.W., 1997, Mapping groundwater quality in the Netherlands. *Journal of Hydrology*, **200**, pp. 364–386.

PEBESMA, E.J., KARSSENBERG, D. and DE JONG K., 2000, The stochastic dimension in a dynamic GIS. In *Compstat 2000, Proceedings in Computational Statistics*, J.G. Bethlehem and P.G.M. van der Heijden (Eds), pp. 379–384 (Heidelberg: Physica-Verlag).

SWAYNE, D.F., TEMPLE LANG, D., BUJA, A. and COOK, D., 2003, Evolving from XGobi into an Extensible Framework for Interactive Data Visualization, *Journal of Computational Statistics and Data Analysis*, **43**, pp. 423–444.

SWITZER, P., 2000, Multiple simulation of spatial fields. In *Accuracy 2000: Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, G.B.M. Heuvelink and MJ.P.M. Lemmens (Eds), pp. 521–528 (Delft: Delft University Press).

Downloaded By: [Institut Fuer Informatik] At: 09:05 14 January 2010