Dynamic visualisation of maps: planned developments

Edzer J. Pebesma, Kor de Jong
Dept. of Physical Geography
Utrecht University

January 2004

1 Introduction

Within the context of the EU project APMOSPHERE (http://www.apmosphere.org/), a capacity for 6 months computer programming has been reserved. This document describes the planned development in this period. The aim of development serves three aims: (i) (dynamic) analysis of maps showing probability density functions, (ii) map comparison, and (iii) interoperability of aguila with other map formats and statistical environments (R, S-PLUS). The starting point is an existing, very advanced and flexible, open source dynamic map visualisation program AGUILA (see http://pcraster.geog.uu.nl/).

2 Spatial probability density function maps

Quite often, spatial information is not deterministic but probabilistic. For example, if air quality maps are a result from spatial interpolation, regression modelling or Bayesian hierarchical modelling, then at each point we get a predicted concentration and a prediction standard error. Assuming a Gaussian distribution for the error, the mean and variance provide sufficient information for the probability density function of the attribute $Z(s)$ (e.g. PM_{10}) at each location on a map s,

$$Pr(Z(s) < c) = f(c, s).$$

The assumption of normality is not necessary: other ways to characterize $Pr(Z(s) < c)$ are (i) taking another parametric density function (e.g. log-normal, binomial etc.), (ii) by characterizing a set of density values such as the minimum, first quartile, median, third quartile and maximum, or the deciles (0, 10, ..., 90, 100-percentile), or (iii) by providing a random sample for $Z(s)$ of at least a reasonable size.

So, for each location s we have a function $f(c)$ to communicate to the user. Communicating only the predicted value of $Z(s)$ (its expectation) hides every notion of uncertainty. Providing only uncertainty measures such as the prediction variance or standard error hides information about levels: large uncertainties may not be “bad” as long as we know the values are either very large or very small.

2.1 Classifying prediction intervals

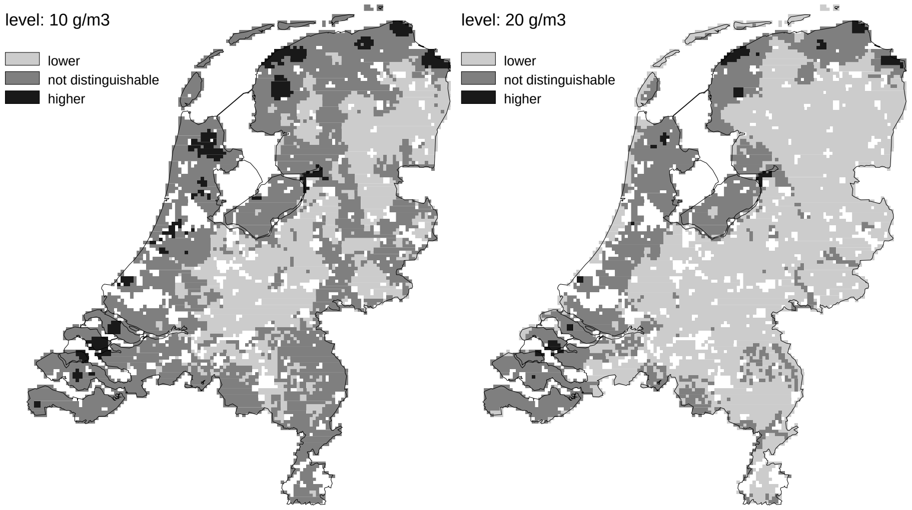
A simple way to present both uncertainty and predicted level is by constructing e.g. 95% prediction intervals, for which

$$Pr(c_l(s) < Z(s) < c_u(s)) = 1 - \alpha$$

with $\alpha = 0.05$ This leaves us to show two maps, that of $c_l(s)$ and of $c_u(s)$. A simple way to visualize this is to classify the intervals with respect to pre-chosen critical level c_{crit}, as follows:

condition	value	meaning
$c_u < c_{\text{crit}}$	lower	prediction is below critical level
$c_l > c_{\text{crit}}$	higher	prediction is above critical level
$c_l < c_{\text{crit}} < c_u$	not distinguishable	prediction and critical level cannot be distinguished, based on available information

An example of a map, thus classified is given here (potassium concentration in upper groundwater in the Netherlands; $c_{\text{crit}} = 10 g/m^3$ (Pebesma and De Kwaadsteniet, 1997):



For dynamic visualisation of such information, we need to provide $Pr(Z(s) < c)$ to aguila, and we need to be able to modify program such that we can interactively modify c_{crit} and/or α, and continuously see the updated classified prediction intervals.

2.2 Tail probability and quantiles

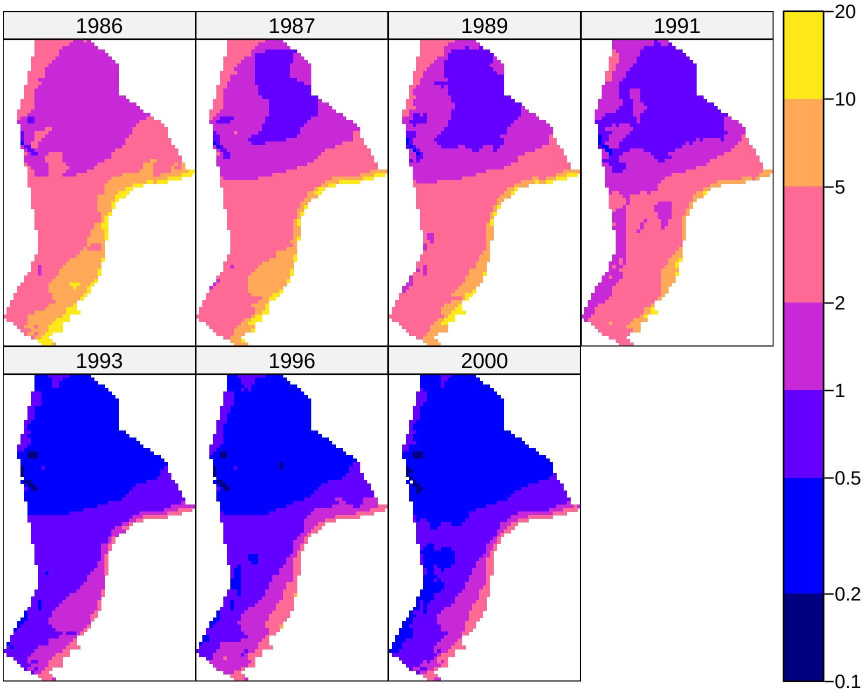
Two simpler, but related visualisations do not concentrate on *symmetric* prediction intervals (where α is divided equally over both tails) but on one-sided intervals, where α is only on one side. The density function

$$Pr(Z(s) < c) = f(c, s)$$

can be visualized dynamically by choosing (and manipulating) c, and showing $f(c, s)$ as the map, or alternatively the tail probability $1 - f(c, s)$. The other way around, we can choose (i.e., fix, and manipulate) $f(c, s)$ and show the corresponding $Z(s)$ value, which gives quantile maps, such as maps with the first quartile, the median, the 95-percentile etc.

3 Map comparison

In the apmosphere project, air quality indicators are spatially predicted using different prediction methods. One way to compare prediction methods is by (cross) validation, another way is to compare the resulting prediction maps. In order to compare maps, the maps need to be displayed in an organized way. Consider the following display of 7 maps:



which is a conditioning plot: PCB-138 predicted values are shown as a function of their spatial coordinates, conditioned on year (sea floor sediment concentration on the Dutch part of the North Sea). The key feature for this figure are:

- each panel has the same x- and y-axis (not shown here, to safe space); the aspect ratio is such that a unit in x-direction equals a unit in y-direction (i.e., it is a map).
- a single legend is used for the information in all panels (colour, grey tones)
- panels are arranged such that a minimum amount of space is wasted

Plots like this are routinely made with the Trellis system (Cleveland, 1993), available in R and S-PLUS, however, their information is not interactive: we cannot manipulate or query the plot. The proposed extension of aguila is to include a visual mode reflecting the one shown above, but in addition allow the following actions:

- identify cells, zoom or pan in one of the maps, affecting all others
- re-arrange the order (and maybe layout) of panels (sub-maps)
- modify legend (colour ramps, legend class boundaries), affecting all maps
- manipulate prediction intervals, probabilities or quantiles as described in the previous section

4 Interoperability

One of the former major advantages of the PCRaster GIS (`http://pcraster.geog.uu.nl/`—a well-defined, portable binary map format—now slowly becomes it's major disadvantage: computing now mainly concentrates on being larger systems, e.g. R, matlab, ArcGIS, GRASS, and having to convert to and from an exotic format, although not harder than it used to be, may seem prohibitively cumbersome. The planned development here is to loosen the dependency of aguila on the PCRaster CSF map format. Two steps are planned:

1. The first step has already been realized: exclusive reading and writing to CSF has been replaced by reading and writing through the GEOSPATIAL DATA ABSTRACTION LAYER (GDAL, `http://www.remotesensing.org/gdal/`), a library that supports some 40 different raster map formats, and to which the CSF format has been added.
2. The second step is to develop aguila as a library such that it can be called as a C function with a set of raster maps (in some format) passed in memory. In this form, it would be trivial to write e.g. an R (or S-PLUS) function that launches aguila with a stack of raster maps, without having to convert data. As aguila is an interactive tool, we still need to establish whether this function should *return* information, e.g. simple queried points, or the full state information such that a chosen visualisation can be replicated in a later session.

References

Cleveland, W.S., 1993, Visualizing Data. Hobart Press, Summit, New Jersey.

E.J. Pebesma and J.W. de Kwaadsteniet, 1997, Mapping Groundwater Quality in the Netherlands. Journal of Hydrology 200, pp. 364-386.