

Comments on White Paper 11 by Smith et al: "Spatial and temporal interpolation of environmental data"

I would like to highlight a couple of issues relating to interpolation that I think are of importance for informing the user community. In particular, I am referring to the large community of researchers who make use of spatially interpolated near-surface observed climatological data to analyse impacts of weather and climate on natural and managed ecosystems (e.g. biodiversity, agriculture, forestry) and on other conditions/activities of importance to society (e.g. water availability and quality, human health, tourism and recreation).

The two issues are: (i) offering best practice guidelines on the uncertainties of interpolated climatological data sets at different spatial resolutions and their potential applicability for estimating impacts; and (ii) the potential importance of elevation for determining the characteristics of interpolated grid box climate.

1. Guidelines on uncertainties in interpolated data sets for user applications

There are many different spatially interpolated datasets available for researchers to download from multiple sources. Some have been developed using standard statistical procedures that are recommended by meteorologists and climatologists, are based on quality-controlled station observations, include extensive documentation of uncertainties, have been peer-reviewed and are well documented. Others may fall short on one or more of these criteria. Unfortunately, there are no common standards available for users to judge the provenance and reliability of such data sets. More worrying perhaps is an increasing trend towards often uncritical application of data at ever-finer resolutions for impact assessment. The highest resolution interpolated data sets are often developed with specific users in mind, sometimes by user groups themselves (with or without meteorological and statistical expertise), to plug a perceived information gap using software that can be readily deployed to interpolate to any resolution desired (within the limits of computer power and storage). In some cases, it is questionable whether such high resolution representation of certain climatological variables can be justified on the basis of the station coverage and the inherent errors and uncertainties associated with the original data and the interpolation procedures.

In summary, I would urge the developers of any global repository that includes spatially interpolated observational data sets to make a serious effort to develop guidance on the limitations of such data sets. Rather than emphasising only its spatial resolution, it would be desirable to associate with each data set a presentation of standard uncertainty metrics as well as prominent advice on how to interpret these metrics in selecting a data set appropriate to the region and topic of interest.

2. Importance of elevation in gridded data sets

Most spatial interpolation procedures for near-surface climatological variables account for variations in topography, because these can have a significant influence on local climate. Topography is usually represented by elevation, commonly with reference to a digital elevation model. Sometimes other characteristics are also considered, such as slope or aspect, but I don't consider these here.

By far the most common approach, for a given grid box resolution, is to interpolate climate to the average (usually mean, but occasionally modal) elevation in a grid box. However, for many applications the climatic conditions at the average elevation aren't necessarily (indeed, one could argue, is rarely) the most relevant for the systems or activities affected by climate, especially in mountainous or hilly terrain. Arable agriculture is commonly conducted away from steep gradients, often in valley bottoms or on plateaux; settlements tend to be found in valleys; the limits of natural vegetation in a grid box are likely to be defined by the highest or lowest elevation rather than the average.

Interpolation to the maximum or minimum elevation in a grid box can be more challenging than interpolation to the average, simply because the vertical lapse rates of climate may be less well defined at the extremes of the altitudinal range, may be reversed (e.g. in inversion situations) and sometimes requiring extrapolation from observations at weather stations that are rarely found at these altitudinal extremes. Nevertheless, as long as the inherent errors of interpolation can be indicated, it would still be extremely valuable to represent the range of elevation as well as the average.

Obviously, the finer the resolution of interpolation, the more topographic detail will be represented, but I suggest that rather than trying to represent ALL fine scale elevations defined by a digital elevation model (i.e. by interpolating to ultra-fine resolution, without good justification based on the station coverage – see point 1, above), it is more defensible to retain a coarse spatial resolution in keeping with the station coverage, but to interpolate to a small number of representative elevations (e.g. mean, max, min, and/or percentiles) in each grid box, indicating for each elevation the uncertainties associated with the interpolation.

In summary, to account for the altitudinal distribution of systems and activities sensitive to climate, it is desirable for many users to be provided with spatially interpolated observed near-surface climatological data sets representing a range of grid box elevations in addition to data sets representing the grid box average. Uncertainties associated with such estimates should also be provided.