A geostatistics-assisted approach to the deterministic approximation of climate data

Maria Lanfredi a, *, Rosa Coppola a, Mariagrazia D’Emilio a, Vito Imbrenda a, Maria Macchiati b, Tiziana Simoniello a

a Institute of Methodologies for Environmental Analysis, Italian Research Council, C.da S. Loja, Tito, PZ, I-85050, Italy
b Department of Physics, “Università Federico II”, via Cinthia, Napoli, I-80126, Italy

A R T I C L E   I N F O

Article history:
Received 12 August 2014
Received in revised form 7 December 2014
Accepted 11 December 2014
Available online

Keywords:
Geostatistics
Geographical model
Climatic surface

A B S T R A C T

We propose a nonconventional application of variogram analysis to support climate data modelling with analytical functions. This geostatistical technique is applied in the theoretical domain defined by each model variable to detect the systematic behaviours buried in the fluctuations determined by other driving factors and to verify the ability of candidate fits to remove correlations from the data. The climatic average of the atmospheric temperature measured at 387 European meteorological stations has been analysed as a function of geographical parameters by a step-wise procedure. Our final model accounts for non-linearity in latitude with a local-scale residual correlation that decays in approximately ten kilometres. The variance of the residuals from the fitted model (approximately 3% of the total) is mostly determined by local heterogeneity in transitional climates and by urban islands. Our approach is user-friendly, and the support of statistical inference makes the modelling self-consistent.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Recent satellite remote sensing technologies for Earth Observation (EO) have supplied a large amount of spatial data that are promising for improving our understanding of the climate system. Contextually, the sparse and uneven data provided by ground stations are still an essential source of information on many key variables characterizing climate dynamics. Currently, the collection of data obtained from meteorological networks, which are generally regarded as valid for spatial inferences of the state of the low atmosphere (Geiger et al., 2003), are also used within climatic studies at the planetary scale. As an example, the series of global datasets, HadCRUT, gridded on a $5^\circ \times 5^\circ$ latitude—longitude box grid, has been widely exploited for the evaluation and attribution of climate change (e.g., Brohan et al., 2006; Jones and Stott, 2011; Jones et al., 2012).

Multi-resolution, both in time and in space, provides the standard hierarchical framework for studying the dynamics of the climate system. Because details at different resolutions generally characterize different physical structures, a coarse-to-fine descriptive strategy is used to separate the broad scale context that is properly climatic from the local contexts of weather dynamics.

In regional studies, numerical models (Rummukainen, 2010; Feser et al., 2011) add details to global-scale climate models, thus improving simulations and forecasts. Within projects focused on long-term simulations or projections, Regional Climate Models (RCMs) currently operate at horizontal grid resolutions between 25 and 50 km [e.g., PRUDENCE (Christensen and Christensen, 2007), ENSEMBLES (Hewitt, 2005) and NARCCAP (http://www.narccap.ucar.edu)]. On the whole, there is an increasing demand of fine-scale data (e.g., Jeffrey et al., 2001; Huld et al., 2006; Hancock and Hutchinson, 2006; Daly et al., 2008; Tang et al., 2012) that can be useful to understand any environmental process linked to climate. Within the proper climatic context, a horizontal resolution of $7-10$ km is currently recognized as a good target (e.g., Suklitsch et al., 2011).

Our research activity (e.g., Lanfredi et al., 2004; Simoniello et al., 2008, 2011) examines complex processes linking climate and the land surface (Piao et al., 2006; Cleland et al., 2007; Prieto-Blanco et al., 2009). Such studies use remote sensing observations of land and require realistic and accurate climatic surfaces obtained by interpolating data from meteorological stations to be interfaced with remote information. In particular, we need to construct air temperature surfaces that can be linked to land surface maps to
better understand biosphere spatio-temporal patterns and
to characterize exchange processes (e.g., carbon emission—absorption) that actively involve climatic fluctuations (e.g.,
Cox et al., 2000; Yuan et al., 2010).

Surface data are mostly obtained by the pure interpolation of
sampled observations (e.g., by Thin Plate Smoothing Splines;
Hopkinson et al., 2012). More complex strategies combine human-
expert knowledge and statistical methods to satisfy the increasing
demand for spatial climate data sets in digital form (Daly et al.,
2008). All of these methodologies directly supply end-users with
gridded data; the underlying physical mechanisms that shape the
climatic surfaces are not singled out and thus remain encapsulated
within the complexity of the gridding algorithms. Nevertheless,
modelling the relationships between a given variable and the fac-
tors that generate its spatial patterns is crucial in many scientific
frameworks. In our case, we have to consider that the spatial
variability of both the land surface and low atmosphere variables is
influenced by geography and topography. Any study focussing on
fluctuations generated by mutual interactions between these two
environments needs to discriminate geographic-induced back-
ground patterns that could distort correlation analyses.

This requirement led us to work on the development of a
regression approach that can account for causal linkages between
geographic factors and temperature. General non-linear regression
implies that functional form selection, estimation of best-fit pa-
rameters, and evaluation of fit performances are rather difficult. In
contrast to linear regression, there is no closed-form expression for
the best-fitting parameters and departures from the optimal
approximation can occur, which could not be accounted for by
global cost functions and require weighty goodness-of-fit tests
(Caouder and Huet, 1997; Graineceau and Ruppero, 2004; Demidenko,
2006).

Here, we focus on a simpler approach by developing an additive
regression model that is non-linear in the explanatory variables.
The ability of such a model to generate random errors starting from
spatially structured patterns can be considered as an a posteriori
criterion to evaluate its performance. The main idea of our proposal
is that we can use variogram analysis (Cressie, 1993; Wackernagel,
2003) to characterize the scale properties of the response variable
along pseudo-directions that are defined by the explanatory vari-
ables of the model within an identification—estimation-checking
iterative approach to model building. This analysis can be particu-
larly useful in the diagnostic checking phase to verify the ability of
the fit to remove correlation structures from the data and thereby
randomize residuals from the fitted model ("whitening"). Efficient
best fits should flatten the variogram at the right variance level;
improper best fits should result instead in residual correlation be-
tween the response and explanatory variables over large scales.
This validation is also important because it enables us to evaluate if
the prediction error is actually the minimum allowed by the
intrinsic degree of randomness of the data. Rigorously speaking,
long-range correlation could also be observed in the case of fractal
data, but this peculiar circumstance is recognizable due to the
typical power law dependence that characterizes them (e.g., Brown
and Liebowitch, 2010). Thus we are limited to consider deterministic
against stationary randomness. Of course, differently from the
standard geostatistical applicative framework, the model variables
are not necessarily spatial coordinates.

We illustrate our strategy by building up a geographical model
for the climatic average of atmospheric temperature over Europe.
Data from 387 meteorological stations were recorded over the 30-
year period from 1961 to 1990, where the latest global "Normals"
are currently defined for climate reference (http://www.wmo.int/
pages/themes/climate/statistical_deviations_of_climate.php) ac-
cording to the World Meteorological Organization. Although air
surface temperature is one of the most continuous and studied
variables within climate analyses, not only its deep dynamical
features in time are still discussed (e.g., Lanfredi et al., 2009
and references therein) but also in truly applicative contexts there is no
single strategic approach to the modelling, as observed above for
climatic variables in general. We refer to the 8 km x 8 km resolu-
tion of the GIMP-S-AVHRR (Global Inventory Modelling and Mapp-
ing Studies-Advanced Very High Resolution Radiometer) data,
which are usually exploited for monitoring land cover in climatic
studies (e.g., Zeng et al., 2013). This resolution corresponds well to
the typical finest scales of RCMs (e.g., Suklitsch et al., 2011) and, as
it will be shown in the following, emerges naturally from scale ana-
lysers as a reasonable boundary between locality and globality. The
main variables shaping the basic structural part of the spatial
variability of near-surface temperature at that resolution in a cli-
matic context are latitude, longitude and elevation. We have also
included the distance from the coastline to illustrate our approxi-
mation process step by step. The final part of the paper concerns a
detailed discussion of the residuals from the fitted model and the
comparison between the performances of our model against a
standard multi-regressive linear model.

2. Data and study area

The annual mean air temperature data concerning the 30 years
climatic period from 1961 to 1990 were obtained from 387 mete-
orological stations located in the European part of the Eurasian
continent (Fig. 1) by averaging daily data. Most of the data were
provided by the European Climate Assessment & Dataset (ECAD)
project (Klein-Tank et al., 2002); few stations (~2%) were integrated
from local databases to introduce additional information in poorly
represented areas.

Differences in latitude and elevation are expected to play a
major role in determining the mean annual value of the air tem-
perature, but the longitude and distance to the sea could also be
significant parameters. In particular, from the point of view of
general atmospheric circulation, the investigated area falls in the
Ferrel cell of the Northern hemisphere where prevailing winds are
westerlies. Because the west coast of Europe is located on the
Atlantic Ocean, whereas the eastern part is continental, the west-
erlies move hot air masses inland from the sea in the direction of
increasing longitude during winter. As a consequence, non-
stationary behaviours are expected in the West(Sud)/East(Nord)
direction. This variability should prevalently concern annual ex-
cursions, but the annual mean values could also be affected.
Moreover, sea proximity, in general, modifies the minimum tem-
perature in coastal swaths, which is why this parameter is included
in the set of geographical parameters potentially involved in
determining air temperature spatial variability.

3. Method

3.1. Variogram analysis

In this Section, we provide some basic definitions and concepts concerning the
variogram analysis (for a detailed discussion see Cressie, 1993).

If Z(s) is a regionalized stationary variable with a constant mean \mu and variance \sigma^2 in a d-dimensional Euclidean space D, the quantity 2\gamma(s), which has been called a
variogram by Matheron (1962), is defined as:

\[ \gamma(s) = \frac{1}{2} \sum_{i=1}^{n} (Z(s_i) - Z(s_j))^2 \]  

Due to the stationary assumption, this is a function of the increments
\Delta s = s_2 - s_1 only and \gamma(\Delta s) = \sigma^2 for large values of \Delta s asymptotically.

When the mean is assumed to be a constant, this equality holds:

\[ \gamma(Z(s) + \Delta s) \leq \gamma(Z(s) + \Delta s) + \gamma(Z(s) - Z(s))^2 \]  

where \gamma(.) indicates the expected value, and we can estimate the variogram as:
Eq. (3) is often reported as the variogram definition in applicable contexts. Hereafter, we will use the word variogram to indicate the statistics defined in Eq. (3).

Let us consider now a non-stationary variable \( Y \) that can be represented as the superposition of a surface trend \( \mu(s) \) and a stochastic variable \( \sigma(s) \), where \( Z(s) \) is a normal variable, as is usually the case in many applicable fields, particularly in climate problems:

\[
Y(s) = \mu(s) + \sigma(s), \quad s \in D
\]

In this case, the second term of Eq. (3) can be written as:

\[
\sigma^2 = \frac{1}{2} \left[ \left( \sigma(s) - \sigma(s + \Delta s) \right)^2 \right]
\]

and thus, on scales larger than that characterizing stochastic correlation, we have:

\[
\sigma^2 = \frac{1}{2} \left[ \left( \sigma(s) - \sigma(s + \Delta s) \right)^2 \right]
\]

Possible systematic behaviors, resulting in large-scale correlation, appear within such empirical variograms as trends steadily climbing beyond the total variance of the data. In standard applications, these trends are preventively removed to study the small-scale properties of the process, but it is possible to use variogram analysis to examine such surface trends.

### 3.2. Rationale

Because trend and local variability are generally on rather different scales, we can separate the stochastic and deterministic terms in Eq. (6) by inspecting the sampling variogram. The deterministic trend does not contribute significantly to the variogram at the local scale due to its continuity properties \( \lim_{\Delta s \to 0} \sigma(s - \mu(s + \Delta s)^2) = 0 \), and by looking at such small scales, we can obtain an estimate of the variance \( \sigma^2 \) not explained by the trend. Thus the actual reference variance \( \sigma^2 \) for evaluating the adequacy of the fit is:

\[
\sigma^2 = \frac{1}{2} \left( \sigma(s) - \sigma(s + \Delta s) \right)^2
\]

This variance is obviously lower than that of the data. As an example, the maximum determination coefficient we can obtain is not \( R^2_{\text{max}} = 1 \) but lower:

\[
R^2_{\text{max}} = \frac{\sigma^2}{\sigma^2_0}
\]

Thus, Eq. (8) gives us a criterion to evaluate the explained variance. In addition, variogram analysis of residuals permits a quick and easy detection of the improper functional form, capturing subtle deviations from the flat behaviour expected for well-fitted surface trends. Generally, geographical coordinates are used in classical applications of geostatistics to extract spatial correlation structures. Here, we instead use the selected explanatory variables of the fitting model to infer the main systematic patterns in the long range. In practice, we define the vector of the explanatory variables \( x = (x_1, ... x_n) \) and search for an additive model \( Y(x) = X_i(x_1) + ... + x_0 + \varepsilon \), where \( Y(x) \) is the modelled variable; the terms \( X_i(x_1) \) are analytical, possibly non-linear functions, and \( \varepsilon \) are stochastic residuals. We estimate variograms in the \( n \) pseudo-directions defined by each one of the variables \( (x_1, ... x_n) \) within a step-wise procedure to evaluate the expected fit error and the actual ability of selected functional forms to reconstruct the actual long-range patterns. A final standard variogram analysis is performed on \( \varepsilon \) to assess the randomness of the residuals on the scales concerned thus making the method auto-consistent.

According to the scale separation approach, this analysis allows us to model broad-scale patterns. For continuous variables, a local scale correlation is expected to be present in the residuals. If the correlation range is comparable to the typical distances between the meteorological stations, the variability generated at these resolutions can be successively interpolated by using geostatistical tools (e.g., by Kriging, Cressie, 1993).

### 4. Results

#### 4.1. Development of the regressive model

The geographical temperature distribution is likely the main source of the large-scale coherence of our data. To express this regularity as an explicit function of geographical variables, we looked at the plots of \( Y \) versus the explanatory variables (latitude, elevation, longitude, and distance from the coastline) in Fig. 2.

In Fig. 2, the major roles of latitude (Fig. 2a) and elevation (Fig. 2b) are evident. Nevertheless, the non-uniform distribution of mountain chains implies an accumulation of low temperature values at latitudes between 37° and 50° N (Fig. 2a) that can distort the estimation of latitude dependence. Thus, the first problem was to convert the temperature to the temperature at sea level. The dependence of temperature on elevation has been widely discussed in the literature, which reports a linear average temperature decrease with a lapse rate around \( \Delta = 6 \text{ °C km}^{-1} \), with slight
seasonal differences (e.g., Linacre, 1992; Hudson and Wackernagel, 1994; Minder et al., 2010). If we would estimate such a lapse rate directly from the plot in Fig. 2b we would obtain an incorrect value due to the spread of temperature values in low elevation sites, which are mainly controlled by latitude. To minimize the influence of latitude, we sorted the sampling sites according to increasing latitude and focused on elevation (Δh) and temperature (ΔT) gradients. In such a way, we looked at elevation increments between sites located on the same or close latitudinal parallels. The variogram of ΔT against Δh (Fig. 3a) shows that the expected percentage of unexplained variance is approximately 12–14 % which implies $R^2_{\text{max}} = 0.88 – 0.86$.

Such a variance is prevalently due to other factors because temperature variations with elevation become significant at approximately $Δh = 100$ m, where the trend becomes evident (Fig. 3a). Over a larger scale, linearity seems to work rather well (Fig. 3b). The estimated lapse rate $ΔT_ι = 5.6$ °C km$^{-1}$ is similar to the values reported in the literature and the squared determination coefficient is $R^2 = 0.87$, which is consistent with $R^2_{\text{max}}$.

Once temperature is converted to the sea level temperature ($T_l$), we can model it according to latitude (Fig. 4a). The variogram in Fig. 4b reveals that approximately 7% of the variance can be considered independent of latitude, resulting in the reference determination coefficient of $R^2_{\text{max}} = 0.96$. In addition, it shows that the sea level mean temperature can be considered stationary over a latitude range $Δlat = 1°$. This scale signifies the first crossover (from independence to dependence) that characterizes the relationship between temperature and latitude.

To model the data in Fig. 4a, we consider three fitting functions that will be discussed comparatively: the linear trend, the second order polynomial trend, and the Gaussian trend.

4.1.1. Linear trend

Linear approximation is usually adopted when the target area is rather limited in latitude (e.g., Chuanyan et al., 2005; Shao et al., 2012).

\[ T_l(\text{lat}) = a_1 \text{lat} + b_1 \]  

Whatever the actual decay of temperature with latitude may be, $T_l$ should be suited when the Earth surface portion that we consider can be approximated with a tangential plan. To the best of our knowledge, generally explicit cross-over scales between linearity and non-linearity are not discussed.

4.1.2. Second order polynomial trend

Linacre and Geerts (Linacre, 1992; Linacre and Geerts, 2002) found that a second order power law well represented the latitude trend in data concerning the whole Northern Hemisphere.

According to the authors, this model is grounded on physical reasons because it is the consequence of the influence of the solar irradiance at the ground. The solar irradiance at the ground depends on the product of the irradiance at the top of the atmosphere and the attenuation of the solar beam through the atmosphere, both depending on latitude (Linacre, 1992). We tested a general second order polynomial (Adjusted Linacre and Geerts trend) assuming that the quadratic dependence due to solar irradiance has to be taken into account. However, the approximated model has to include some additional terms to account for some peculiar features of the European continent, in particular, the land–sea distribution (this influences temperature and is also expected to introduce a slight dependence between latitude and longitude that are mutually constrained within the ideal equation that would define the European land surface):

\[ T_{ALG}(\text{lat}) = a_{ALG}\text{lat}^2 + b_{ALG}\text{lat} + c_{ALG} \]  

4.1.3. Gaussian trend

Upon careful inspection of Fig. 4a (see red lines), three apparent regimes roughly coincide with the Mediterranean area, the continental area, and the northernmost maritime swath. The
superposition of two Gaussian functions could account for slight curvature changes and the Gaussian parameters could inform us about temperature variability features within these areas:

$$T_C(\text{lat}) = A_1e^{-\frac{(\text{lat} - \text{lat}_C)^2}{2s_1^2}} + A_2e^{-\frac{(\text{lat} - \text{lat}_C)^2}{2s_2^2}}$$  \hspace{1cm} (11)

Fig. 5 shows variograms for the three selected models. The value of $R^2$ associated to the linear fit is $R^2 = 0.89$ with approximately 11\% of unexplained variance. The linear fit regularizes data well in a range of approximately 2 or 3 latitude degrees, but is inefficient for larger areas. The estimated unexplained variance is approximately 8.4\% for the ALG filter, which regularizes better temperature data. The Gaussian filter with two components shows the best performance because the estimated unexplained variance (~7.4\%) is close to the expected one.

The Gaussian solution follows the curvature changes that we observe along the temperature pattern (see Fig. 6). It is interesting to note that the two Gaussian distributions are centred in the southern part of the two maritime swaths, whereas the continental pattern emerges from the superposition of these distributions that are weighted by the coefficients $A_1 = 2A_2$. Surely these patterns are grounded on general physical reasons, but nevertheless, we hypothesize that a small contribution could result from the different distributions of longitude of the three areas. The middle zone is the most continental, as it includes the main part of the inland eastern area, whereas the westernmost part of the grid in Fig. 1 is occupied by sea. As a consequence, we presume that the non-linear fit partially accounts for latitude—longitude constraints that are imposed by the specific land mass geography of the European peninsula.

We examined air temperature dependence on longitude and distance from the coastline by following the same approach. A slightly decreasing drift with longitude was detected in the residuals obtained by filtering dependence on latitude. Although the descriptive power of the linear best fit ($R^2 = 0.60$) was rather poor due to the high noise content, we decided to include this drift in the model because $R^2_{\text{max}}$ was estimated to be approximately 0.5—0.7. No significant correlation with the distance from the coastline was found.

The final model was:

$$T(h, \text{lat}, \text{lon}) = -0.0056h - 0.056\text{lon} + 16.94e^{-\frac{(\text{lat} - \text{lat}_C)^2}{10.52^2}} + 8.06e^{-\frac{(\text{lat} - \text{lat}_C)^2}{12.52^2}} + 0.68$$  \hspace{1cm} (12)

The parameterization of Eq. (12) was performed by integrating the contribution of all of the meteorological stations. Nevertheless, the actual predictive skill of the model was obtained by computing mean temperature in each of the 387 meteorological stations with parameters estimated on the basis of the remaining 386 stations (Cross-Validation). The unexplained variance was approximately 3\% with an RMSE = 0.7 °C.

The final varigrams (Fig. 7), estimated on the residuals of the model in Eq. (12), demonstrate no significant patterns, and Fig. 8 shows the synthetic map of the mean climatic air temperature. It is possible to identify latitude patterns and the significant influence of elevation that lowers temperature. Slight West—East behaviours are also detectable.

### 4.2. Analysis of the residuals. Comparison with a linear regression model

According to our results, the final representation of the temperature $T$ in a site $x$ is:
\[ T(x) = T(lat, lon, h) + \epsilon \]  

(13)

where \( T(lat, lon, h) \) is obtained from the model in Eq. (12) and the variable \( \epsilon \) represents the residual component, as explained in Section 3.2.

To estimate the local spatial correlation not explained by our model, we analysed these residuals in detail.

The box-plot in Fig. 9 summarizes the distribution of \( \epsilon \), which is symmetric and well centred around the value \( \epsilon = 0 \), with most of the values located in the range \((-0.5, 0.5)\). A dozen values are recognized as outliers. Incidentally, our procedure is not particularly sensitive to outliers because it is not a proper interpolation tool that tries to “pass” through all of the sample points. The largest residuals are obtained in the Mediterranean region, at the limit between Mediterranean and continental climates, and are likely related to the complex patterns of cold polar winds that are driven by mountainous barriers in a prevalently mild zone. Large positive residuals are also obtained in hot urban areas (e.g., Paris, Prague). These outliers are mainly responsible for the estimated RMSE = 0.7 °C. If we eliminate the Mediterranean area (up to 46°N), the value of this error drops down to RMSE = 0.5 °C.

These errors are comparable with errors reported in the literature and obtained in the analysis of temperature “Normals” through complex interpolation tools (see Hopkinson et al., 2012).

To bring out the added value of our procedure, we fitted a linear multi-regressive model (with the same variables \( lat, lon, \) and \( h \)) to the data. Fig. 10 shows the variogram of the residuals obtained by applying our model (black line) and the linear model (red line). The two variograms roughly coincide up to about a few hundreds of kilometres. Although the estimated RMSE = 0.85 °C of the linear model is not particularly worse than that of our model, it does not represent a measure of the adequacy of the fit because the linear model is not able to make the data stationary on large scales and provides ever more inefficient predictions. In this case, the mere deterministic approximation does not explain all of the variance generated by the broad scale coherence. To account for all of the variance, such residuals should be further interpolated, as an example through Kriging, thus splitting into two components the variability accounted for by the only model in Eq. (12).

On the contrary, the variogram of the residuals of our model reaches the expected percentage of total variance (0.03) in a range of about ten kilometres (our target resolution). Our model is an accurate descriptor of the climatic field in grid cells of 10 km × 10 km and the short correlation range of the local fluctuations are a direct consequence of the good trend removal. This
scale, which naturally emerges from our analysis, as a reasonable boundary between locality and globality, is consistent with the typical finest scale RCMs (e.g., Suklitsch et al., 2011). The short correlation range that marks the representative domain of a single station (on grid cells wider than 10 km resolution are non-correlated) implies that a dense sampling dataset would be necessary to increase resolution. A rough estimation for the European continent (~10,000,000 km$^2$) recommends approximately 100,000 stations.

We also evaluated the possibility of introducing parameters accounting for very local topographic details, which have provided important predictive improvements in mountainous areas (Minder et al., 2010). Nevertheless, the parameters we can obtain from a DEM, such as the slope and aspect, are not able to improve the model due to the presence of heterogeneity, poor high-elevation sampling over Europe and the lack of detailed local-scale information.

**Fig. 8.** Synthetic map of the mean climatic air temperature estimated through model in Eq. (12) using an 8 km × 8 km grid.

**Fig. 9.** Box-plot of the model residuals in Eq. (12).

**Fig. 10.** Sample variograms of the residuals of two deterministic models: the model in Eq. (12) (black line); the linear regression model $T = -0.0057h - 0.53\text{lat} - 0.063\text{lon} + 37.73$ (red line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
5. Conclusions

We have proposed a regressive method for modelling broad-scale climate fields in a general non-linear context. Model building is achieved by a three-stage procedure based on identification, estimation, and diagnostic checking. In both the first and last stages, we use variogram analysis in the domains defined by each explanatory variable. In the exploration phase, the variogram informs us about the correlation between a response variable and a candidate explanatory variable. In particular, the variogram allows us to infer the scales involved and the level of the variance that cannot be explained. In the diagnostic phase, we analyse the same scale properties after the removal of the candidate fit. A model is acceptable if the residuals from the fitted model are consistent with a random field on the scales concerned and the variance is reduced to the previously estimated level. Because scales are separated, this procedure is also useful for identifying scale ranges where simple approximations may work well.

The application of the method to the climatic mean temperature recorded over Europe shows good results, especially considering the low number of explanatory variables (elevation, latitude, and longitude). A particularly interesting result was found about the latitude-dependence of temperature. While it is common knowledge that solar radiation is unevenly distributed and that intensity varies from one location to another depending upon the latitude, it is more difficult to understand how these patterns are altered by the land mass distribution. The best fit we designed describes three different climatic sub-zones in the European area: the Mediterranean region, the continental area, and the northern maritime swatch. Such a partition, which accounts for two relative maxima in the mean zone, the continental area, and the northern maritime swatch, reflects the synergic effect of radiances variability against latitude and land–sea distribution, which is characterized by the presence of a massive area of land in the continental zone. Over land areas that do not exceed 2–3 latitude degrees, a linear approximation is adequate.

A range of approximately 10 km marks the decay of the local correlation of the residuals that cannot be further interpolated due to the lack of information at such fine scales. The unexplained variance of our model is approximately 3% of the total variance; an estimate of the error by cross validation is RMSE = 0.7 °C (RMSE = 0.5 °C for the continental area). Such a prediction error is only due to the peculiar microclimatic variability, especially in the Mediterranean region and comes from an intrinsic descriptive limit that solar radiation is unevenly distributed and that intensity varies from one location to another depending upon the latitude, it is more difficult to understand how these patterns are altered by the land mass distribution. The best fit we designed describes three different climatic sub-zones in the European area: the Mediterranean region, the continental area, and the northern maritime swatch. Such a partition, which accounts for two relative maxima in the mean zone, the continental area, and the northern maritime swatch, reflects the synergic effect of radiances variability against latitude and land–sea distribution, which is characterized by the presence of a massive area of land in the continental zone.

Acknowledgements

This work is inserted in the framework of the project ORIENTATE – A network for the integration of climate knowledge into policy and planning-funded by the South East Europe Transnational Cooperation Programme (CCI 2007CB163P0069).

References


