Bulgarian Geophysical Journal, 2014, Vol. 40
National Institute of Geophysics, Geodesy and Geography, Bulgarian Academy of Sciences

SIMPLE POSTPROCESSING METHOD FOR VERTICAL CORRECTION BASED ON STRATIFIED NEAR-SURFACE ATMOSPHERIC PARAMETERS

H. Chervenkov

National Institute of Meteorology and Hydrology, blvd Tsarigradsko Shose 66, Sofia 1784, Bulgaria,
e-mail: hristo.tchervenkov@meteo.bg

Abstract. The paper presents shortly a relatively simply and transparent, but meteorologically consistent method for physical interpolation with vertical correction to arbitrary points of interest, based on the distributed three dimensionally and stratified near-surface atmospheric parameters. Such procedure is needed relatively often in various environmental studies, when the output of some meteorological model has to be adjusted more precisely to location with known elevation, using one ore other postprocessing technique. The original authors' proposal is described and demonstrated briefly using as test dataset the output of regional climate model RegCM4 for the monthly mean temperature for the year 2000 over Bulgaria and records from the station observations in the NIMH-BAS network.

Key words: Postprocessing, Physical interpolation, Vertical correction, RegCM4, Near-surface atmospheric stratification

Introduction

Practically all spatially distributed hydrological and ecological models need certain meteorological information, most frequently formatted as initial dataset, containing the values of some input parameters. Thus, for instance, they use air temperature to drive processes such as evapotranspiration, snowmelt, soil decomposition, and plant productivity. Since most near-surface air-temperature data are collected at irregularly spaced point locations (for example the network of the measurement stations) rather than over continuous surfaces, the point-based temperatures must be accurately distributed over the landscape in order to be useful in spatially distributed modelling. On other hand the output of various atmospheric circulation models (ACM), either for numerical weather prediction, or for simulation of the global and regional climate, is presented most often with finite resolution, i.e. as distribution of the values in the gridpoints or gridcells of fixed mesh. Depending on the model an its concrete implementation, the spatial (horizontal and vertical) step is changed up to two orders, but typical values for the horizontal one of the contemporary hydrostatic numerical weather prediction and regional climate models is 8-10 km. The computational process of the ACM uses model representation of the topography, which more or less differs from the real one. Thus, generally speaking, due to the vertical displacement, even the perfect model can produce different outcome from the unbiased measurement (the objective “truth”) in horizontally collocated point. Key point in the model verification studies is to distinguish the differences, caused by such reasons, from those by the model physics weaknesses. Further, from practical point of view, in many local environmental studies, where the typical resolution is 1-2 km, to use directly meteorological input with (significantly) coarser resolution can leads to serious biases in the results. A pragmatic way to overcome this problem in the case when high-resolution data are not available, is to use different subgridding techniques, either incorporated in the model system as in Giorgi et al., 2003 and in Im et al., 2010 or in external form, as post processing procedures. Due mainly to the enormous complexity of the modern ACM, the second approach is much more easily feasible for the end-user, which in the common case utilizes this model as encapsulated meteorological driver for his specific tasks. Several methods exist for spatial interpolation of point-based data, including inverse-distance weighting (IDW), kriging, 2-dimensional splines, and trend-surface regression (Myers, 1994). They can be purely mathematical as the listed above, or combined - mathematical based on physical assumptions. As stated in Dodson and Marks, 1997, these methods often work well over relatively flat, homogeneous terrain. In mountainous terrain, however, the strong relationship between vertically stratified meteoparameters and elevation precludes a simple interpolation of point-based observations. Unless the effect of elevation on, for example, the temperature is explicitly accounted for, an interpolation can produce grossly inaccurate results. For example, in the case where a set of temperature observations or the model gridpoints are located around the base of a mountain, an interpolation which ignores elevation would seriously overestimate the temperature at the mountain top, as it would not account for the fact that temperature generally decreases with increasing elevation.

The presented work describes one proposed by the author physically based interpolation technique, which can be easily implemented in various applications. The scheme is very simple and transparent and can be used for any vertically stratified parameters (temperature, humidity characteristics).

The rest of the paper is organized as follows: The proposed innovative approach is described and briefly commented in Section 2. A numerical experiment, demonstrating the possibilities of the scheme, is described in Section 3. The last section contains consist discussion and concluding remarks.

Methodology

To test the capabilities of the novel scheme, it is forced to reproduce the value of some stratified meteoparameter in selected reference point, from some discrete 3D distribution of the same. A meteorological station is located in this point and the measurements there are accepted as objective truth. Thus, the closeness of the reproduced value to the measured one is treated as estimation of the methods quality.

The most widely used approach for assessing the departure of the model output to the measurements in the numerical simulation studies is to compare them with the obtained results in the nearest gridpoint. As stated above however, for many problems, like model performance evaluation and local high-resolution issues, to relay only on this approach is not sufficient. Obviously many other techniques based either on purely mathematical or mathematical and physical assumptions can be applied, but the testing of sophisticated schemes is out of the scope of our work. Instead, along with the "nearest gridpoint" approach, we will present one innovative technique. Basic fact in meteorology is that the heterogeneity of the atmosphere in vertical direction is significantly stronger than in horizontal one. That's why it is essential to account this effect, especially over a complex terrain, for the stratified variables like, for example, the temperature. The scheme on figure 1 illustrates, for simplicity in one dimension, the original authors' idea.

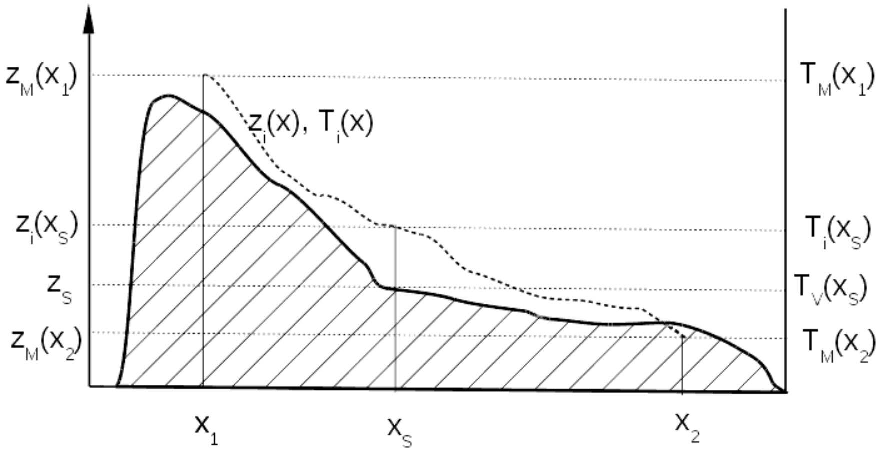

Fig. 1. Explanatory illustration of the proposed method

Let the dashed area sketches the terrain cross-section along the line between the gridpoints x_1 and x_2, which surrounds the point of interest with coordinate x_S. It is always possible to construct local continuous (with IDW or any other interpolation method) function $z_i(x)$ describing the dependence of the elevation from the horizontal coordinate, as depicted with dotted line on figure 1. The defined in equation (1) IDW is deterministic interpolator, where β is a positive real number, called the power parameter. The smoothness (if $\beta>1$) and computational feasibility makes the IDW preferable in many applications, as the presented here.

$$P_i(\beta) := \frac{\sum_{j=1}^{n} \frac{P_j}{r_j^{\beta}}}{\sum_{j=1}^{n} \frac{1}{r_j^{\beta}}} \tag{1}$$

In this implementation P is the surface temperature T, $n=4$ and, correspondingly, T_1, T_2, T_3, T_4. are the values in the nearest four (i.e. the surrounding) gridnodes, r_1, r_2, r_3, r_4.are the distances between them and the station location, and traditionally β is set to two. Thus, equation (1) can be rewritten as:

$$T_i = \frac{\sum_{j=1}^{4} \frac{T_j}{r_j^{2}}}{\sum_{j=1}^{4} \frac{1}{r_j^{2}}} \tag{2}$$

To find the IDW-value T_i. is necessary to locate the station in the grid first and, second, to find the distances. The first task is solved with efficient binary search, and the second – with procedure based on the haversine formula, keeping in mind that the geographical coordinates of the station and the gridpoints are known. The haversine formula gives the great-circle distance d between two points on a sphere (i.e. Earth) with radius R from their longitudes λ_1, λ_1 and latitudes φ_1, φ_2:

$$d = 2R \arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos(\varphi_1)\cos(\varphi_2)\sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)} \right) \tag{3}$$

Similarly, knowing the values of the considered variable T, say for definiteness the temperature, in the gridpoints x_1 and x_2, namely $T_M(x_1)$ and $T_M(x_2)$, we can obtain also the temperature interpolation function $T_i(x)$. In the common case, however, the interpolated elevation at point x_S $z_i(x_S)$ differs from the actual z_S, which, as x_S, is a priori known. Assuming, that the atmosphere in this layer is polytropic, i.e. linearly stratified and according the definition for the vertical gradient (lapse rate), we can compute it as follows:

$$\gamma = -\frac{\partial T}{\partial z} \approx -\frac{\Delta T}{\Delta z} = -\frac{T(z_2) - T(z_1)}{z_2 - z_1}, \ z_2 \phi z_1 \tag{4}$$

Equation (4) can be generalized, where έ is small positive number:

$$\gamma = \begin{cases} -\dfrac{T(z_2) - T(z_1)}{z_2 - z_1}, \; z_2 - z_1 \phi \; \varepsilon \\ \\ 0, \qquad z_2 - z_1 \le \varepsilon \end{cases} \tag{5}$$

or for the used in figure 1 notation

$$\gamma_M = \begin{cases} -\dfrac{T_M(x_1) - T_M(x_2)}{z_M(x_1) - z_M(x_2)}, \; z_M(x_1) - z_M(x_2) \phi \; \varepsilon \\ \\ 0, \qquad z_M(x_1) - z_M(x_2) \le \varepsilon \end{cases} \tag{6}$$

Such generalization is needed obviously to avoid the division by zero when the gridpoints are with equal altitudes. The threshold έ can be set, for instance, to 0.1 m as in our implementation.

The basic idea of the proposed approach is to interpolate the altitude *z* and the temperature *T* to the coordinate x_S yielding $z_i(x_s)$ and $T_i(x_s)$ correspondingly. Then, using the calculated with equation (3) lapse rate, we can adjust $T_i(x_s)$ to the real elevation adding vertical correction factor, proportional to the difference between the interpolated $z_i(x_s)$ and the real altitude z_s in the following manner:

$$T_v(x_s) = T_i(x_s) - \gamma_M (z_i(x_s) - z_s) \tag{7}$$

In the case of the two-dimensional generalization, six different gradients can be defined between the values of the searched variable, with respect to the elevations of the four surrounding gridpoints. Intending to bound possibly thicker (and thus more representative) layer, the gradient between the lowest and highest gridpoint is taken under consideration in the current implementation.

The obvious merits of the proposed scheme are its clear physical sense and computational simplicity. The method is based on the linear stratification assumption, which, on one hand, is significantly smaller constrain than some in other techniques (for comprehensive review see – Dodson and Marks, 1997), where the gradient is prescribed to a constant over the whole domain. On the other hand, this stratification is confirmed by many experimental studies of lower-level atmosphere. Another strong point of the method is its locality – the gradient is calculated separately for every gridcell, which is more close to the physical reality, where the vertical distribution of the atmospheric parameters can vary greatly from point to point. As a drawback of the method could be pointed the fact, that in some cases the station can be not situated between the gridpoints in vertical direction. So, for instance, if the station is on the top/bottom of a convex/concave terrain segment and the gridpoints on its periphery, the method cannot retrieve the actual gradient at the altitude of the point of interest.

Numerical experiment

To demonstrate the possibilities of the proposed approach, the obtained results during sensitivity study of the regional climate model RegCM version 4 (for description see Pal et al., 2007 and references therein) over the territory of Bulgaria are used. The considered part of the model output consist of data for the monthly mean surface temperature in grid with 10×10 km resolution which are compared with the averages from the measurements in 30 stations of the network of NIMH-BAS. The model outcomes are compared with the observations using three methods: observation – nearest gridpoint (noted further as “mode 1”), observation – IDW value (“mode 2”) and observation – vertically corrected value with the above described procedure (“mode 3”). According equation (7) the magnitude of the vertical correction is proportional to the vertical displacement between the interpolated and real altitude, which, itself, depends from the mesh properties, but generally decreases by higher resolutions. Thus, to emphasize the impact of this correction, two stations, Kyustendil and Kasanlak, with significant displacement from the mesh are used in the performed tests. In most practical tasks however, such postpocessing procedures are applied on the whole dataset in the domain, rather than on selected stations. Further, such test can be treated as much more consistent way for overall method performance evaluation. Table 1 shows the positional parameters for the selected two stations – the distance to the nearest gridpoint r_{min}, the real altitude z_s, the altitudes of the four surrounding gridpoints $z_M(x_1)$, $z_M(x_2)$, $z_M(x_3)$, and $z_M(x_4)$.

Station	r_{min}, km	z_s, m	$z_M(x_1)$, $z_M(x_2)$, $z_M(x_3)$, $z_M(x_4)$, m	$z_i(x_s)$, m	z_s-$z_i(x_s)$, m
Kyustendil	1.699	520.0	1017.4, 919.2, 814.5, 859.1	862.8	-342.8
Kasanlak	5.350	392.0	561.6, 677.3, 662.3, 520.7	623.8	-231.8

Table 1. Positional parameters of the selected stations, identified with their World Meteorological Organisation (WMO) code.

Results and comments

The (calculated from the) observed values of the monthly mean temperature for the year 2000 are compared with the model output using the three commented above methods.

Common approach to assess the degree of agreement of the observed (measured) values O_i and their modelled correspondents M_i (in the presented study in the three modes) is to calculate certain statistical measures, among which most widely used are the root mean square error (*RMSE*), the correlation coefficient (also termed the Pearson correlation coefficient, *R*), the index of agreement (*IA*) and the mean bias (*BIAS*). Explicit formulas for the first two will not be given due to their popularity, and the last two are equal accordingly to:

$$IA = 1 - \frac{\sum_{i=1}^{N} \left(O_i - M_i\right)^2}{\sum_{i=1}^{N} \left(\left|M_i - \overline{O}\right| + \left|O_i - \overline{O}\right|\right)^2} \tag{8}$$

$$BIAS = \frac{1}{N} \sum_{i=1}^{N} (O_i - M_i) \tag{9}$$

The summation is along the number of comparisons *N* and the overlines notes averaging. The (dimensionless) index of agreement condenses the differences between observed and modeled values into one statistical quantity. It provides a measure of the match between the departure of each prediction from the observed mean and the departure of each observation from the observed mean. The index of agreement has a theoretical range of 0 to 1, with a value of 1 suggesting “perfect” agreement. The mean bias is simply the average bias between the observed and modeled values.

The calculated statistical quantities are summarized in Table 2.

	RMSE, ^{o}C	R	IA	BIAS, ^{o}C
Mode 1	1.112	0.992	0.995	0.197
Mode 2	1.144	0.991	0.995	0.225
Mode 3	0.923	0.995	0.997	-0.224

Table 2. Main features of the selected cyclones according column caption

Similar statistical treatment for the selected two stations is hampered by the insufficient length of the time series – only 12 values are available. That’s why the results for this comparison are presented only graphically, as shown on figure 2.

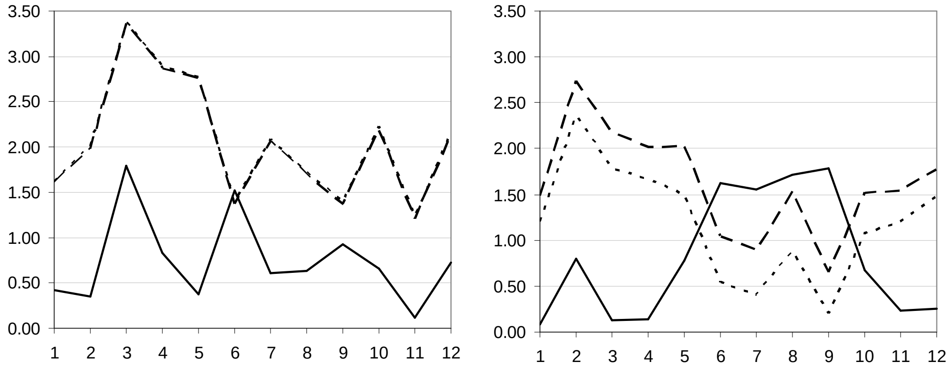

Fig. 2 Absolute biases (i.e. |obs-mode n| of mode 1 (dashed line), mode 2 (dotted line) and mode 3 (normal line) for station Kyustendil (Left) and Kasanlak (Right).

According the listed in table 2 values of the statistical quantities can be stated, that the overall (i.e. for the whole dataset of stations) performance in the three modes is practically equal. Probably the influence of the good behavior for some stations is masked by the bad one for others. The interpretation of figure 2 leads to the same conclusion. For the selected stations, however, the influence of the proposed method for vertical correction is discernible and significant. In all months, except one for the first station and except four for the second one, the vertically corrected values are apparently closer to the measurements. Reason for the opposite situation during the summer months can be weaknesses in the planetary boundary layer (PBL) parameterization in hot days with prevailing unstable stratification. In this case the correction adds biases to the modelled value. The differences between mode 1 and mode 2 are smaller, in case of the first station practically indiscernible, due most probably to the relatively high horizontal resolution. Deeper conclusions with significant confidence, however, can be obtained after further statistical treatment of relevantly longer time series.

Conclusion

The presented short study can be observed as concise demonstration of the proposed by the author postprocessing method for vertical correction to the point of interest of model output containing values of stratified parameters, in particular the temperature. The method is very simple, and, utilizing explicitly the three-dimensional structure of the model output, physically consistent and transparent. Keeping in mind that almost all ACMs uses terrain-following coordinate systems, such 3D-distribution is the common case. The presence of data for the considered variable at different altitude practically in all gridcells allows the calculation of the local lapse rate and as consequence the vertical correction. The discussed examples for two stations shows, that in the most cases this method produces noticeable closer to the measurements results, even when the point of interest is outside the layer, in which the lapse rate is calculated. In other cases, however, the results are worse, and thus the method has to be applied carefully, preferably point-by-point of interest. Statistical processing of longer datasets and records can be performed aiming to reveal a priori criterion when the method is applicable or not. Nevertheless the proposed technique illustrates a possible pragmatic way to adapt the available model output physical consistently to the specific needs of the end-user who looks for concrete value in the arbitrary point of interest.

References

Giorgi, F., Francisco, R., Pal, J., 2003 Effects of a subgrid-scale topography and land use scheme on the simulation of surface climate and hydrology. Part I: Effects of temperature and water vapor disaggregation. J. Hydrometeor., 4(2): 317-333.

Dodson, R., Marks, D., 1997 Daily air temperature interpolated at high spatial resolution over a large mountainous region Clim Res Vol. 8: 1-20.

Im, E.-S., Coppola, Giorgi, F., Bi, X. 2010 High Resolution Regional Climate Model for the Alpine Region and Effects of a Subgrid-Scale Topography and Land Use Representation, J. Climate vol. 23 1854-1873.

Myers DE 1994 Spatial interpolation. An overview. Geoderma 62(1):17-28

Pal JS, Giorgi F, Bi X, Elguindi N and others 2007 Regional climate modelling for the developing world: The ICTP RegCM3 and RegCNET. Bull Am Meteorol Soc 88: 1395-1409

Прост постпроцесорен метод за вертикална корекция основан на тримерно разпределени и стратифицирани атмосферни параметри в близост до подложната повърхност

Хр. Червенков

Резюме: Статията представя накратко един сравнително прост и обозрим, но метеорологически съдържателен метод за физическа интерполация с вертикална корекция към конкретна точка на интерес, основана на тримерно разпределени и стратифицирани атмосферни параметри в близост до подложната повърхност. Подобна процедура е необходима сравнително често в различни изследвания на околната среда, в които изходните данни от числен атмосферен модел трябва да се сведат по-прецизно до определено място с известна надморска височина, посредством една или друга постпроцесорна схема. Оригиналното авторско предложение е описано и демонстрирано в синтезиран вид, използвайки за тестов набор данни изхода на регионалния климатичен модел RegCM4 за средномесечната температура за 2000 година и записи от станционните наблюдения на НИМХ-БАН

Ключови думи: Постпроцесорна процедура, Физическа интерполация, Вертикална корекция, RegCM4, Атмосферна стратификация в близост до подложната повърхност