Application of MISH Method for Gridding of SPI Series

Tamás Szentimrey, Zita Bihari
Hungarian Meteorological Service
Budapest, Hungary

Abstract

The MISH (Meteorological Interpolation based on Surface Homogenized Data Basis; Szentimrey, Bihari) spatial interpolation method was developed at the Hungarian Meteorological Service. The main difference between MISH and the geostatistical interpolation methods built in GIS can be found in the amount of information used for modelling the necessary statistical parameters. In geostatistics the usable information or the sample for modelling is only the predictors, which are a single realization in time. While in meteorology we have spatiotemporal data, namely the long data series which form a sample in time and space as well. The SPI (Standardized Precipitation Index) series are often applied to characterize the drought variability within a region. The SPI values are certain transformed values of the precipitation sums assuming gamma distribution. After the transformation procedure the elements of the SPI series have expectedly standard normal distribution. However the SPI series can be calculated for station data series while there is the need to obtain gridded data series or mapping of SPI. In order to solve the above problem the following possible procedures are planned to compare during our presentation: - Gridding, interpolation of SPI series by geostatistical methods. - Calculation of SPI series based on the gridded precipitation series obtained by MISH. - Gridding, interpolation of SPI series by MISH.

Key words: SPI, interpolation, MISH, gridding

The MISH method

The MISH (Meteorological Interpolation based on Surface Homogenized Data Basis) method for the spatial interpolation of surface meteorological elements was developed at the Hungarian Meteorological Service (Szentimrey and Bihari, 2007). This is a meteorological system not only in respect of the aim but in respect of the tools as well. It means that using all the valuable meteorological information – climate and supplementary model or background information – is intended. For that purpose developing an adequate mathematical background was also necessary of course. In the practice many kinds of interpolation methods exist therefore the question is the difference between them. According to the interpolation problem the unknown predictand value is estimated by using the known predictor values. The type of the adequate interpolation formula depends on the probability distribution of the meteorological elements! Additive formula is appropriate for normal distribution (e.g. temperature) while some multiplicative formula can be applied for quasi lognormal distribution (e.g. precipitation). The expected interpolation error depends on certain interpolation parameters as for example the weighting factors. The optimum interpolation parameters minimize the expected interpolation error and these parameters are certain known functions of different climate statistical parameters e.g. expectations, deviations and correlations. Consequently the modelling of the climate statistical parameters is a key issue to the interpolation of meteorological elements. The various geostatistical kriging methods applied in GIS are also based on the above mathematical theory (Cressie, 1991). The main difference between MISH and the geostatistical interpolation methods can be found in the amount of information used for modelling the necessary statistical parameters. In geostatistics the usable information or the sample for modelling is only the predictors, which are a single realization in time. While in meteorology we have spatiotemporal data, namely the long data series which form a sample in time and space as well. The long data series is such a speciality of the meteorology that makes possible to model efficiently the statistical parameters in question.

The MISH method has been developed according to the above basic principles. The main steps of the interpolation procedure are as follows.

- To model the climate statistical parameters by using long homogenized data series.
- To calculate the modeled optimum interpolation parameters which are certain known functions of the modeled climate statistical parameters.
– To calculate the interpolation formula with the modeled optimum interpolation parameters and the predictor values.

The new software version MISHv1.02 consists of two units that are the modelling and the interpolation systems. The interpolation system can be operated on the results of the modelling system. In the following paragraphs we summarize briefly the most important facts about these two units of the developed software.

Modelling system for climate statistical (deterministic and stochastic) parameters:
- Based on long homogenized data series and supplementary deterministic model variables. The model variables may be as height, topography, distance from the sea etc.. Neighbourhood modelling, correlation model for each grid point.
- Benchmark study, cross-validation test for interpolation error or representativity.
- Modelling procedure must be executed only once before the interpolation applications!

Interpolation system:
- Additive (e.g. temperature) or multiplicative (e.g. precipitation) model and interpolation formula can be used depending on the climate elements.
- Daily, monthly values and many years’ means can be interpolated.
- Few predictors are also sufficient for the interpolation and no problem if the greater part of daily precipitation predictors is equal to 0.
- The interpolation error or representativity is modelled too.
- Capability for application of supplementary background information (stochastic variables) e.g. satellite, radar, forecast data.
- Data series complementing that is missing value interpolation, completion for monthly or daily station data series.
- Interpolation, gridding of monthly or daily station data series for given predictand locations. In case of gridding the predictand locations are the nodes of a relatively dense grid.

As it can be seen modelling of the climate statistical parameters is a key issue to the interpolation of meteorological elements and that modelling can be based on the long homogenized data series. The necessary homogenized data series can be obtained by our homogenization software MASHv3.02 (Multiple Analysis of Series for Homogenization; Szentimrey, 1999, 2007).

The Standardized Precipitation Index (SPI)
The Standardized Precipitation Index (SPI) is a drought index developed by McKee et al. (1993). The SPI is based only on precipitation and can be used to monitor conditions on a variety of time scales (Hayes, 2002). The SPI calculation for any location is based on long-term precipitation record for a desired period. This long-term record is fitted to a gamma (Γ) probability distribution, which is then transformed into the standard normal distribution. The mathematical details are as follows.

Let the precipitation sum series for a given period be \( X(s, t) \) \((t = 1, 2, ..., n)\), where \( s \) is the location vector and \( t \) is the time.

The assumed mixed distribution function of \( X(s, t) \) for \( x \geq 0 \) is: \( p_0 + G(x) \cdot (1 - p_0) \), where \( p_0 \) is the probability of 0 value, i.e. \( p_0 = P(X(s, t) = 0) \), furthermore function \( G(x) \) is a gamma \((\Gamma(p, \lambda))\) distribution function for \( x \geq 0 \): \( G(x) = \frac{\lambda^p}{\Gamma(p)} \int_0^x u^{p-1} e^{-\lambda u} du \).

The unknown parameters to be estimated are: \( p_0, p, \lambda \). After the estimation procedure based on the series \( X(s, t) \) \((t = 1, 2, ..., n)\), we can obtain the SPI series by the following transformation formula,

\[
Z(s, t) = \Phi^{-1}(p_0 + G(X(s, t)) \cdot (1 - p_0)) \quad (t = 1, .., n),
\]

where \( \Phi^{-1} \) is the inverse of the standard normal distribution function \( \Phi(z) \). As a consequence of the transformation construction the SPI values are approximately standard normal distributed, i.e. \( Z(s, t) \sim N(0, 1) \) \((t = 1, .., n)\).
Gridding, interpolation of SPI series (mapping)

Notations
According to the interpolation problem the unknown value belonging to the predictand location \( s_0 \) is estimated by use of the known values of predictor locations \( s_i \) \( (i = 1, ..., M) \) where the location vectors \( s \) are within the same climate region.

The precipitation sum series for a given period are:

\[
(t) \quad X(s_i, t) \quad (i = 0, \ldots, M; \quad t = 1, \ldots, n)
\]

The SPI series obtained by transformation:

\[
(t) \quad Z(s_i, t) = \Phi^{-1}\left(p_{0,i} + G_i(X(s_i, t)) \cdot (1 - p_{0,i}) \right)
\]

Since the SPI values are calculated from the precipitation values there are two basic possibilities for their interpolation.

Indirect method based on interpolation of precipitation
In this case first we interpolate the precipitation sum series \( X(s_i, t) \) and the SPI values \( Z(s_i, t) \) are after calculated. The type of the adequate interpolation formula depends on the probability distribution of the meteorological element in question. In practice linear or additive (kriging) formula is applied in general which is appropriate in case of normal probability distribution, however for precipitation sum we deduced a mixed additive multiplicative formula built in MISH system and it can be written in the following form,

\[
\hat{X}(s_0, t) = \theta \cdot \prod_{q_i X(s_i, t) \geq \theta} \left( q_i \cdot X(s_i, t) \right) \lambda_i \left( \sum_{q_i X(s_i, t) \geq \theta} \lambda_i + \sum_{q_i X(s_i, t) < \theta} \lambda_i \right)
\]

where \( X(s_0, t), \quad X(s_i, t) \) \( (i = 1, ..., M) \) are the predictand and predictors respectively, furthermore \( \theta > 0, \quad q_i > 0, \quad \sum_{i=1}^{M} \lambda_i = 1 \) and \( \lambda_i \geq 0 \) \( (i = 1, ..., M) \) are some interpolation parameters.

The optimum interpolation parameters are uniquely determined by certain climate statistical parameters which can be modelled by using the predictor data series \( X(s_i, t) \) \( (i = 1, ..., M; \quad t = 1, \ldots, n) \).

After this the estimated SPI values \( \hat{Z}(s_0, t) \) can be calculated on the basis of the interpolated series \( \hat{X}(s_0, t) \) \( (t = 1, \ldots, n) \), i.e.

\[
\hat{Z}(s_0, t) = \Phi^{-1}\left(\hat{p}_{0,0} + \hat{G}_0(\hat{X}(s_0, t)) \cdot (1 - \hat{p}_{0,0}) \right) \quad (t = 1, \ldots, n).
\]

Direct methods with ordinary kriging formula
Another way is to interpolate the predictand \( Z(s_0, t) \) by the predictors \( Z(s_i, t) \) \( (i = 1, ..., M) \) directly. Taking into account that the variables \( Z(s_i, t) \in N(0, 1) \) \( (i = 0, \ldots, M) \) i.e. their common distribution is about normal with expected values \( \text{E}(Z(s_i, t)) \approx 0 \) and st. deviations \( \text{D}(Z(s_i, t)) \approx 1 \) \( (i = 0, \ldots, M) \), the application of ordinary kriging interpolation formula is evident, i.e.

\[
\hat{Z}(s_0, t) = \sum_{i=1}^{M} \lambda_i \cdot Z(s_i, t) \quad , \quad \text{where} \quad \sum_{i=1}^{M} \lambda_i = 1.
\]

The vector of optimal weighting factors \( \lambda^T = [\lambda_1, ..., \lambda_M] \) can be written in covariance form,

\[
\lambda^T = \left(c^T + 1^T \frac{1 - 1^T C^{-1} c}{1^T C^{-1} 1} \right) C^{-1}, \quad \text{where} \quad c \text{ is the predictand-predictors covariance vector, and } C \text{ is the predictors-predictors covariance matrix, or equivalently in variogram form, preferred in}
\]
geostatistics, $\lambda^T = \left( \gamma^T + \Gamma^T \left( \frac{1 - I^T \Gamma^{-1} \gamma}{I^T \Gamma^{-1} I} \right) \Gamma^{-1} \right)^T$, where $\gamma$ is the predictand-predictors variogram vector, and $\Gamma$ is the predictors-predictors variogram matrix. As a consequence of $D(Z(s_i,t)) \approx 1$ ($i = 0, .., M$), the covariance form is identical with the correlation form.

The possible modelling procedures for the statistical parameters in question are as follows.

i, Geostatistical methods for variogram modelling. However these methods are based on only the momentary predictors $Z(s_i, t_0)$ ($i = 1, .., M$) in case of predictand $Z(s_0, t_0)$ at given time $t_0$. It means only a single realization in time for modelling, therefore, consequently less efficiency may be expected.

ii, MISH method for modelling of covariance (correlation) structure is based on the series of predictors $Z(s_i, t)$ ($i = 1, .., M$; $t = 1, .., n$). Moreover a spatial neighbourhood modelling part was developed and built in the MISH system.

Acknowledgements

This study was supported by the project of South East Europe Transnational Cooperation Programme DMCSEE (Drought Management Centre for South East Europe).

References


