DOI:10.28974/idojaras.2024.2.1

IDŐJÁRÁS

Quarterly Journal of the HungaroMet Hungarian Meteorological Service
Vol. 128, No. 2, April – June, 2024, pp. 143–154

Statistical modeling of the present climate by the interpolation method MISH – theoretical considerations

Tamás Szentimrey

Varimax Limited Partnership, Budapest, Hungary

Author E-mail: szentimrey.t@gmail.com

(Manuscript received in final form October 31, 2023)

Abstract— Our method MISH (Meteorological Interpolation based on Surface Homogenized Data Basis; *Szentimrey* and *Bihari*) was developed for spatial interpolation of meteorological elements. According to mathematical theorems, the optimal interpolation parameters are known functions of certain climate statistical parameters, which fact means we could interpolate optimally if we knew the climate. Furthermore, the data assimilation methods also need to know the climate if Bayesian estimation theory is to be correctly applied. Therefore, we have developed the MISH system also to model the climate statistical parameters, i.e. present climate, by using long data series. It is a nonsense that we try to model the future climate but we do not know the present climate.

Key-words: climate modeling, climate statistical parameters, data series, spatial interpolation, data assimilation, MISH, MASH

1. Introduction

In the statistical climatology, the climate can be formulated as the probability distribution of the meteorological events or variables. The purpose of the statistical climatology is to estimate or model the climate probability distribution or equivalently the climate statistical parameters. Furthermore, the meteorological data series make possible to estimate or model the climate statistical parameters in accordance with the establishments of statistical climatology principles.

Our method MISH (Meteorological Interpolation based on Surface Homogenized Data Basis; *Szentimrey* and *Bihari*, 2007, 2014; *Szentimrey*, 2017, 2021, 2023c) was developed for spatial interpolation of meteorological elements. According to the mathematical theorems, the optimal interpolation parameters are known functions of certain climate statistical parameters, which fact means we could interpolate optimally if we knew the climate. Consequently, according to the principles of climatology, the modeling part of software MISH is based on long meteorological data series. The main difference between MISH and the geostatistical interpolation methods built in GIS (Geographic Information System) is that the sample for modeling at GIS methods is only the predictors, which is a single realization in time, while at the MISH method we use spatiotemporal data for modeling, which form a sample in time and space alike.

We focus on the methodology of the modeling subsystem built in MISH. This subsystem was developed to model the following climate statistical parameters for half minutes grid: monthly, daily expected values, standard deviations, and the spatial and temporal correlations. Consequently, the modeling subsystem of MISH is completed for all the first two spatiotemporal moments. If the joint spatiotemporal probability distribution of a given meteorological element is normal (e.g., daily and monthly mean temperatures) then the spatiotemporal moments above uniquely determine this distribution, which is the mathematical model of the present climate for this meteorological element.

In our conception, the meteorological questions and topics cannot be treated separately. Therefore, we present a block diagram (*Fig. 1*) to illustrate the possible connection between various important meteorological topics. The software MISH (*Szentimrey* and *Bihari*, 2014) and MASH (Multiple Analysis of Series for Homogenization; *Szentimrey*, 2023a,b) were developed by us. These software were applied also in the CARPATCLIM project (*Szentimrey et al.*, 2012a,b; *Lakatos et al.*, 2013). The paper of *Izsák et al.* (2022) presents another application to create a representative database for Hungary.

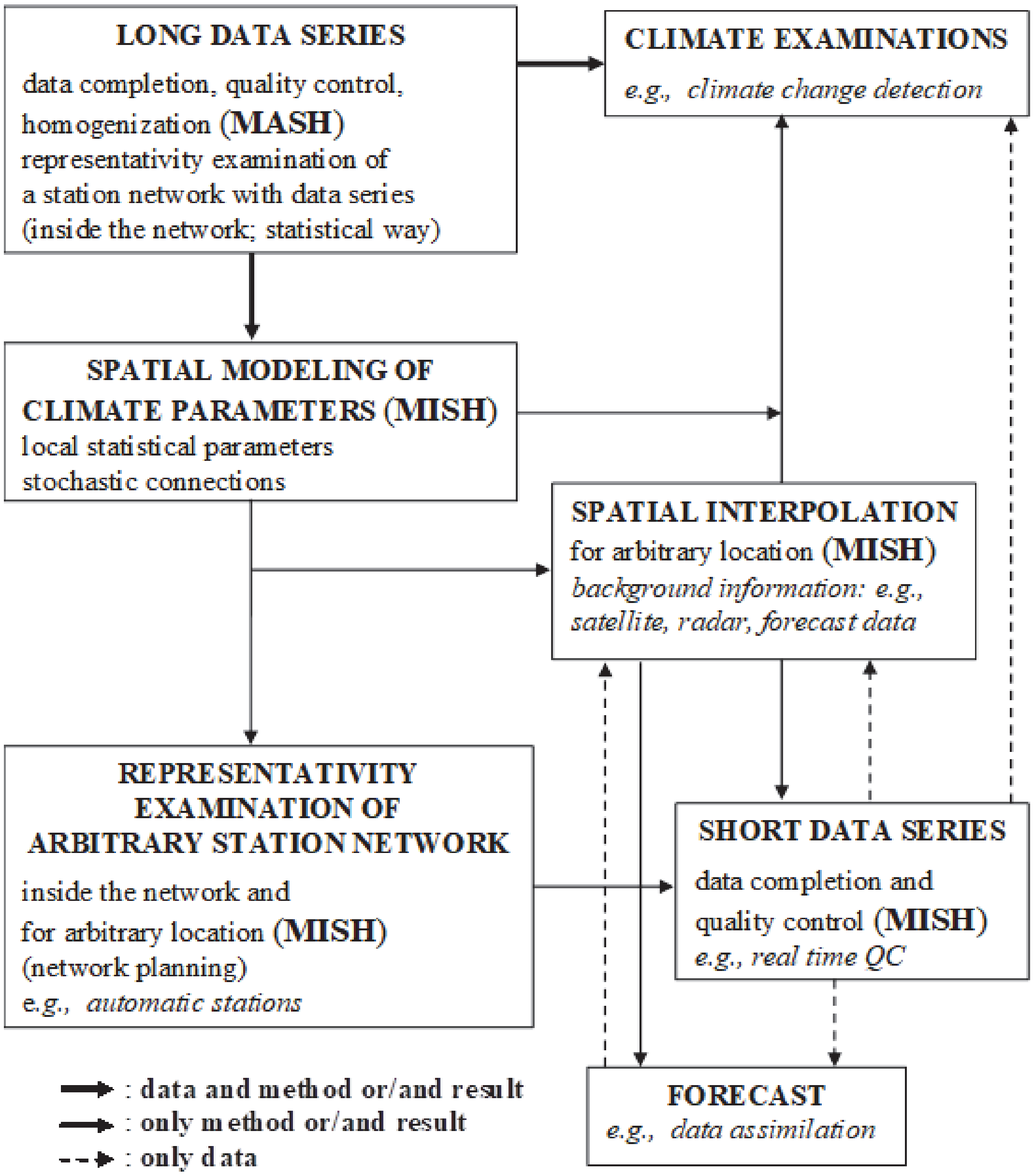

Fig. 1. Block diagram for the possible connection between various basic meteorological topics and systems.

2. Theoretical additive model of spatial interpolation

According to the interpolation problem for monthly or daily data, the unknown predictand $Z(\mathbf{s}_0, t)$ is estimated by use of the known predictors $Z(\mathbf{s}_i, t)$

$(i = 1, \ldots, M)$ where $\mathbf{s}$ is the location vector ($\mathbf{s} \in D$ spatialdomain) and t is the index of year. The type of the adequate interpolation formula depends on the probability distribution of the meteorological variable. Assuming normal distribution (e.g., temperature), the additive (linear) formula is adequate (*Szentimrey* and *Bihari*, 2007; *Szentimrey et al.*, 2011).

2.1. Climate statistical parameters

The expected values are
$\mathrm{E}\left(Z(\mathbf{s}_i, t)\right) = \mu(t) + E(\mathbf{s}_i)(i = 0, .., M)$,
where $\mu(t)$ is the temporal trend or the climate change signal and $E(\boldsymbol{s})$ is the spatial trend. The standard deviations $D(\mathbf{s}_i) = \mathrm{D}\left(Z(\mathbf{s}_i, t)\right)$ $(i = 0, \ldots, M)$ and the correlation system is as, $\mathbf{r}$ is the predictand-predictors correlation vector, $\mathbf{R}$ is the predictors-predictors correlation matrix.

2.2. Additive (linear) interpolation formula

Assuming the normal distribution of the variables and the above model of expected values, the appropriate additive meteorological interpolation formula is
$\hat{Z}(\mathbf{s}_0, t) = \lambda_0 + \sum_{i=1}^{M} \lambda_i \cdot Z(\mathbf{s}_i, t)$,
where $\sum_{i=1}^{M} \lambda_i = 1$ because of unknown $\mu(t)$. The quality of interpolation can be characterized by the root mean square error $RMSE(\mathbf{s}_0)$ and by the representativity value: $REP(\mathbf{s}_0) = 1 - \frac{RMSE(\mathbf{s}_0)}{D(\mathbf{s}_0)}$.

Remark: Multiplicative model of spatial interpolation

In this paper only the linear or additive model is described in detail, which is appropriate in case of normal probability distribution. However, perhaps it is worthwhile to remark that for case of a quasi lognormal distribution (e.g., precipitation sum), we deduced a mixed multiplicative additive formula which is used also in our MISH system, and it can be written in the following form:
$\hat{Z}(\mathbf{s}_0, t) = \vartheta \cdot \left(\prod_{q_i \cdot Z(\mathbf{s}_i, t) \geq \vartheta} \left(\frac{q_i \cdot Z(\mathbf{s}_i, t)}{\vartheta}\right)^{\lambda_i}\right) \cdot \left(\sum_{q_i \cdot Z(\mathbf{s}_i, t) \geq \vartheta} \lambda_i + \sum_{q_i \cdot Z(\mathbf{s}_i, t) < \vartheta} \lambda_i \cdot \left(\frac{q_i \cdot Z(\mathbf{s}_i, t)}{\vartheta}\right)\right)$,
where the interpolation parameters are $\lambda_i \geq 0$ $(i = 1, \ldots, M)$, $\sum_{i=1}^{M} \lambda_i = 1$, and $\vartheta = m(\mathbf{s}_0)$, $q_i = \frac{m(\mathbf{s}_0)}{m(\mathbf{s}_i)}$, where $m(\mathbf{s}_i)$ $(i = 0, \ldots, M)$ are the spatial median values.

2.3. The optimal interpolation

The optimal interpolation parameters λ_0, λ_i $(i = 1, \ldots, M)$ minimize the root mean square error $RMSE(\mathbf{s}_0)$, and these are known functions of the climate statistical parameters! The optimal interpolation is when we use the optimal parameters.

The optimal constant term is $\lambda_0 = \sum_{i=1}^{M} \lambda_i \big(E(\mathbf{s}_0) - E(\mathbf{s}_i)\big)$.
The vector of optimal weighting factors $\boldsymbol{\lambda} = [\lambda_1, .., \lambda_M]^T$ and the optimal representativity $REP(\mathbf{s}_0)$ can be written as function of the parameters: $D(\mathbf{s}_0)/D(\mathbf{s}_i)$ $(i = 1, \dots, M)$, $\mathbf{r}$, $\mathbf{R}$. The expected values and these parameters are climate statistical parameters, consequently we could interpolate optimally if we knew the climate well. For example, let us see the next theorem.

Theorem 1

If $D(\mathbf{s}_0)/D(\mathbf{s}_i) = 1$ $\quad (i = 1, \dots, M)$ then:

(i) The vector of optimal weighting factors: $\boldsymbol{\lambda} = \mathbf{R}^{-1}\left(\mathbf{r} + \frac{(1-\mathbf{1}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{r})}{\mathbf{1}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{1}}\mathbf{1}\right)$.

(ii) The optimal representativity is

$$REP(\mathbf{s}_0) = 1 - \sqrt{(1 - \mathbf{r}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{r}) + (1 - \mathbf{1}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{r})^2 \cdot \frac{1}{\mathbf{1}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{1}}}\,.$$

Consequently the unknown climate statistical parameters are $E(\mathbf{s}_0) - E(\mathbf{s}_i)$ $(i = 1, \dots, M)$ (do not depend on temporal trend of climate change) and the correlations $\mathbf{r}$, $\mathbf{R}$.

3. Modeling of climate statistical parameters in MISH

The main difference between geostatistics and meteorology can be found in the amount of information being usable for modeling the statistical parameters. In geostatistics, the usable information or the sample for modeling is only the actual predictors $Z(\mathbf{s}_i, t)$ $(i = 1, \dots, M)$ which belong to a fixed instant of time, that is a single realization in time (*Cressie*, 1991), while in meteorology we have spatiotemporal data, namely long data series which form a sample in time and space as well and make possible to model the climate statistical parameters in question. If the meteorological station location $\mathbf{S}_k$ $(k = 1, .., K)$ have long data series $Z(\mathbf{S}_k, t)(t = 1, .., n)$ then the climate statistical parameters can be estimated statistically for the stations (*Szentimrey et al*, 2011).

3.1. Modeling of monthly climate statistical parameters for a half minutes grid

3.1.1. First step of modeling by using model variables

The monthly climate statistical parameters belonging to the stations $\mathbf{S}_k$ $(k = 1, .., K)$ can be used for modeling the correlation structure as well as the spatial variability of local statistical parameters (*Szentimrey* and *Bihari*, 2014). The basic principle of this neighborhood modelling is as follows. Let $P(\mathbf{s})$, $Q(\mathbf{s})$, $r(\mathbf{s}_1, \mathbf{s}_2)(\mathbf{s}, \mathbf{s}_1, \mathbf{s}_2 \in D)$ be certain model functions depending on different model variables with the following properties, within a neighborhood $\|\mathbf{S}_{j1} - \mathbf{S}_{j2}\| < d_0 \quad (j_1, j_2 = 1, .., K)$:

(a) Modeling of correlations is $r(\mathbf{S}_{j1}, \mathbf{S}_{j2}) \approx \operatorname{corr}(Z(\mathbf{S}_{j1}, t), Z(\mathbf{S}_{j2}, t))$;

(b) Modeling of difference of means (*E*) is $P(\mathbf{S}_{j1}) - P(\mathbf{S}_{j2}) \approx E(\mathbf{S}_{j1}) - E(\mathbf{S}_{j2})$;

(c) Modeling of ratio of standard deviations (*D*) is $\frac{Q(\mathbf{S}_{j1})}{Q(\mathbf{S}_{j2})} \approx \frac{D(\mathbf{S}_{j1})}{D(\mathbf{S}_{j2})}$.

The model variables may be distance, height, topography.

3.1.2. Second step of modeling by interpolation

Predictand location is $\mathbf{s}_0$, predictor station locations are $\mathbf{S}_{0i}$ $(i = 1, \ldots, M)$. The weighting factors can be calculated according to *Theorem 1*, where $\mathbf{r}$, $\mathbf{R}$ contain the modeled predictand-predictors, predictors-predictors correlations based on Section 3.1.1 (a).

Modeling of means, expected values (*E*) by additive interpolation is

$$E(\mathbf{s_0}) = \sum_{i=1}^{M} \lambda_i \big(P(\mathbf{s_0}) - P(\mathbf{S_{0i}})\big) + \sum_{i=1}^{M} \lambda_i E(\mathbf{S_{0i}})$$

Modeling of standard deviations (*D*) by multiplicative interpolation is

$$D(\mathbf{s}_0) = \prod_{i=1}^{M} \left(\frac{Q(\mathbf{s}_0)}{Q(\mathbf{S}_{0i})} \cdot D(\mathbf{S}_{0i}) \right)^{\lambda_i}.$$

3.2. Relation of daily and monthly data interpolation

Theorem 2

Let us assume the following properties for the daily values within a month:

(i) Expected values and standard deviations are

$$E_d(\mathbf{s}_0) - E_d(\mathbf{s}_i) = E_{0i}, \ \frac{D_d(\mathbf{s}_0)}{D_d(\mathbf{s}_i)} = D_{0i} \quad (i = 1, \ldots, M; \ \ d = 1, .., 30)$$

(ii) Correlations

$$\operatorname{corr}\left(Z_{d_1}(\mathbf{s}_{i_1}, t), \ Z_{d_2}(\mathbf{s}_{i_2}, t)\right) = r_{i_1 i_2}^{S} \cdot r_{d_1 d_2}^{T}$$

$$(i_1, i_2 = 1, \ldots, M; \ \ d_1, d_2 = 1, .., 30),$$

where $r_{i_1 i_2}^{S}$ is the correlation structure in space and $r_{d_1 d_2}^{T}$ is the correlation structure in time.

Then the optimum interpolation parameters for the daily values and the monthly mean are identical: $\lambda_{i,d} = \lambda_{i,month}$ $(i = 0,\ldots,M;\ \ d = 1,..,30)$.
Moreover, the representativity values for the daily values and the monthly mean are also identical: $REP_d(\mathbf{s}_0) = REP_{month}(\mathbf{s}_0)$ $(d = 1,..,30)$ in the case of optimal interpolation parameters.
Consequently, the monthly modeled climate statistical parameters and the modelling methodology can also be applied to model the daily climate statistical parameters.
Special modeling parts for daily data are modeling of temporal daily autocorrelations $\rho(\mathbf{s})$ and daily standard deviations $D_{daily}(\mathbf{s})$ per months. We assume that the daily data of a given month constitute an AR(1) process with common standard deviation $D_{daily}(\mathbf{s})$ and temporal first-order autocorrelation $\rho(\mathbf{s})$. Modeling of autocorrelation $\rho(\mathbf{s})$ by additive interpolation is

$$\rho(\mathbf{s}_0) = \sum_{i=1}^{M} \lambda_i \rho(\mathbf{S}_{0i}),$$

where autocorrelations $\rho(\mathbf{S}_{0i})$ belonging to the stations and the weighting factors are calculated according to Section 3.1.2. The daily standard deviation $D_{daily}(\mathbf{s})$ can be estimated by using the monthly standard deviation $D_{month}(\mathbf{s})$ and the first-order autocorrelation $\rho(\mathbf{s})$.
The *RMSE* can be calculated as follows: $RMSE(\mathbf{s}_0) = D(\mathbf{s}_0) \cdot (1 - REP(\mathbf{s}_0))$.

3.3. Modeled monthly and daily spatiotemporal climate statistical parameters in MISH

The necessary unknown climate statistical parameters can be modeled for a half minutes grid. These modeled monthly and daily spatiotemporal statistical parameters in MISH system are:

(i) spatial expected values (spatial trend) $E(\mathbf{s})$,

(ii) spatial standard deviations $D(\mathbf{s})$,

(iii) spatial correlations $r(\mathbf{s}_1, \mathbf{s}_2)$, and

(iv) temporal first-order autocorrelations $\rho(\mathbf{s})$.

Consequently, the first two spatiotemporal moments can be modeled for daily and monthly data by the MISH procedure! The normal distribution is uniquely determined by these moments. Some examples for the modeled climate statistical parameters are presented in *Fig. 2.*

Example

Mean temperature data in September for 10 arbitrary locations somewhere in Hungary.

Input: the location coordinates only without any temperature data.

Output: modelled climate statistical parameters

Location indices:

	1	2	3	4	5	6	7	8	9	10
Monthly Expected Values:	14.59	14.99	14.95	15.06	15.16	15.16	15.13	15.08	15.01	15.05
Daily Expected Values:	14.59	14.99	14.95	15.06	15.16	15.16	15.13	15.08	15.01	15.05
Monthly Standard Deviations:	1.34	1.62	1.68	1.67	1.68	1.66	1.72	1.66	1.61	1.64
Daily Standard Deviations:	2.84	3.44	3.47	3.46	3.47	3.60	3.73	3.58	3.48	3.46
Temporal Daily Autocorrelations:	0.74	0.74	0.75	0.75	0.75	0.73	0.73	0.73	0.73	0.74

Matrix of Spatial Autocorrelations:

1.00	0.99	0.99	0.98	0.97	0.96	0.97	0.97	0.98	0.98
0.99	1.00	0.99	0.99	0.98	0.95	0.96	0.96	0.97	0.98
0.99	0.99	1.00	0.99	0.99	0.94	0.95	0.95	0.96	0.97
0.98	0.99	0.99	1.00	0.99	0.91	0.93	0.93	0.95	0.96
0.97	0.98	0.99	0.99	1.00	0.90	0.91	0.91	0.93	0.94
0.96	0.95	0.94	0.91	0.90	1.00	0.99	0.99	0.98	0.98
0.97	0.96	0.95	0.93	0.91	0.99	1.00	0.99	0.99	0.98
0.97	0.96	0.95	0.93	0.91	0.99	0.99	1.00	0.99	0.99
0.98	0.97	0.96	0.95	0.93	0.98	0.99	0.99	1.00	0.99
0.98	0.98	0.97	0.96	0.94	0.98	0.98	0.99	0.99	1.00

Fig. 2. Example for modeling of present climate by MISH.

4. The main features of software MISHv2.01

The new software version of MISH method (*Szentimrey* and *Bihari*, 2007, 2014; *Szentimrey*, 2017, 2021, 2023c) is under development, and it is consisting of two units that are the modeling and the interpolation systems. The interpolation system can be operated on the results of the modeling system. We summarize briefly the most important facts about these two units of the software.

Modeling subsystem for climate statistical (local and stochastic) parameters:

– Modeling of all the first two spatiotemporal moments for daily and monthly data (expected values, standard deviations, spatiotemporal correlations).

– Based on long homogenized data series and supplementary deterministic model variables. The model variables may be such as height, topography, distance from the sea, etc. Neighborhood modeling, correlation model for each grid point, dense half minutes grid.

- Benchmark study, cross-validation test for interpolation error or representativity.
- Modeling procedure must be executed only once before the interpolation applications.

Totally different principle from the other methods!

Interpolation subsystem:

- 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 expected interpolation error RMSE is modelled too, representativity examination of arbitrary station network is performed.
- Real time quality control for daily and monthly data (additive model).
- Capability for application of supplementary background information (stochastic variables), e.g., satellite, radar, forecast data. (with QC: data assimilation)
- Data series completion that is missing value interpolation, completion for monthly or daily station data series.
- Capability for interpolation, gridding of monthly or daily station data series, as grid-point and grid-box average datasets alike.

The elder versions of MISH-MASH software can be downloaded from:
http://www.met.hu/en/omsz/rendezvenyek/homogenization_and_interpolation/software/
We plan to share the new version MISHv2.01 next year (2025).

5. The relationship between MISH and data assimilation

The MISH system is capable of interpolation with background information and quality control (see *Fig. 1*), which is essentially a data assimilation procedure (*Szentimrey*, 2016).

5.1. Interpolation with background information in MISH

The background information, e.g., forecast, satellite, radar data can be efficiently used to decrease the interpolation error. In this paper only the interpolation based on additive model or normal distribution is presented. According to Section 2, let

us assume that, $Z(\mathbf{s}_0, t)$ is the predictand, $\hat{Z}(\mathbf{s}_0, t) = \lambda_0 + \sum_{i=1}^{M} \lambda_i Z(\mathbf{s}_i, t)$ is the interpolated predictand, moreover, there is given the $\mathbf{G} = \{G(\mathbf{s}, t) \mid \mathbf{s} \in D\}$ background information on a dense grid.
Then the principle of interpolation with background information is that the interpolated predictand given $\mathbf{G}$ can be expressed as follows:

$$\hat{Z}_G(\mathbf{s}_0, t) = \hat{Z}(\mathbf{s}_0, t) + E\left(Z(\mathbf{s}_0, t) - \hat{Z}(\mathbf{s}_0, t) \mid \mathbf{G}\right), \quad (1)$$

where $\mathrm{E}\left(Z(\mathbf{s}_0, t) - \hat{Z}(\mathbf{s}_0, t) \mid \mathbf{G}\right)$ is the conditional expectation of $Z(\mathbf{s}_0, t) - \hat{Z}(\mathbf{s}_0, t)$, given $\mathbf{G}$.

5.2. Data assimilation and reanalysis data

The forecast and the reanalysis data are based on the data assimilation which procedure is in strong relationship with the methodology of interpolation with background information. The Bayes estimation theory is the mathematical background of the data assimilation methods in meteorology. The purpose of data assimilation is to determine a best possible atmospheric state using observations and short range forecasts. The typical way applied in practice to estimate the true atmospheric state is the minimization of the following variational cost function:

$$J(\mathbf{z}) = (\mathbf{z} - \mathbf{g})^{\mathrm{T}} \mathbf{Q}^{-1} (\mathbf{z} - \mathbf{g}) + (\mathbf{y}_0 - \mathbf{Fz})^{\mathrm{T}} \mathbf{P}^{-1} (\mathbf{y}_0 - \mathbf{Fz}), \quad (2)$$

where $\mathbf{z}$ is the analysis field, predictand (grid), $\mathbf{g}$ is the given background field (forecast), $\mathbf{y}_0$ is the given observations, predictors; $\mathbf{Fz} = \mathrm{E}(\mathbf{y}_0|\mathbf{z})$, $\mathbf{Q}$ is the background error covariance matrix, and $\mathbf{P}$ is the observation error covariance matrix.
It can be proved that this procedure is essentially an interpolation with background information including a quality control part for the predictors. The cost function (Eq. (2)) is known and referred by the forecasting community, as it is based on the Bayesian estimation theory. However, there are some mathematical omissions and simplifications at this cost function (*Szentimrey*, 2016).
This mathematical derivation of the Bayesian cost function is not correct, therefore, the decision according to Eq. (2) is not a real Bayes decision. For example this formula includes implicitly the assumption that the conditional expectation of $\mathbf{z}$, given $\mathbf{g}$ is identical with $\mathbf{g}$, i.e., $\mathrm{E}(\mathbf{z}|\mathbf{g}) = \mathbf{g}$, that means the conditional expectation does not depend on climate, and the forecast is always optimal. Or the relation of the background error covariance matrix $\mathbf{Q}$ and the climate is absolutely not known by the forecasting community. Consequently, the necessary climate statistical parameters are also neglected at the data assimilation procedures applied in practice.

The data assimilation technique is used also to produce reanalysis data series in order to monitor the climate change based on past observation series. However, beside the inadequacies mentioned above there are further sources of errors for reanalysis data. One of them is the inhomogeneity of the used station data series i.e., these series are often affected by artificial shifts due to changes in the measurement conditions (relocations, instrumentation). Another problem may be the little spatial representativity, i.e., relatively few station data series are used for production of reanalysis data series as a consequence of the data policy between the countries (*Lakatos et al.*, 2021; *Bandhauer et al.*, 2022).

6. Conclusion

It is a nonsense that we try to model the future climate but we do not know the present climate. We should know the present climate well, if want an efficient methodology for spatial interpolation and data assimilation. For this purpose, special advanced mathematics is needed of course. Originally, we have developed the MISH system also to model the climate statistical parameters, i.e., climate, by using long station data series, since the optimal spatial interpolation needs modeled climate statistical parameters. But the question can be turned back. The climate modeling can be based on spatial interpolation of station climate statistical parameters.

References

Bandhauer, M., Isotta, F., Lakatos, M., Lussana, C., Båserud, L., Izsák, B., Szentes, O., Tveito, O.E., and *Frei, C.*, 2022: Evaluation of daily precipitation analyses in E-OBS (v19.0e) and ERA5 by comparison to regional high-resolution datasets in European regions. *Int. J. Climatol. 42*, 727–747. https://doi.org/10.1002/joc.7269

Cressie, N., 1991: Statistics for Spatial Data., Wiley, New York.

Izsák, B., Szentimrey, T., Lakatos, M., Pongrácz, R., and *Szentes, O.*, 2022: Creation of a representative climatological database for Hungary from 1870 to 2020. *Időjárás 126*, 1–26. https://doi.org.10.28974/idojaras.2022.1.1

Lakatos, M., Szentimrey, T., Bihari, Z., and *Szalai, S.*, 2013: Creation of a homogenized climate database for the Carpathian region by applying the MASH procedure and the preliminary analysis of the data. *Időjárás 117*, 143–158.

Lakatos, M., Szentimrey, T., Izsák, B., Szentes, O., Hoffmann, L., Kircsi, A., and *Bihari, Z.,* 2021: Comparative study of CARPATCLIM, E-OBS and ERA5 dataset, Proceedings of the 10th Seminar for Homogenization and Quality Control in Climatological Databases and 5th Conference on Spatial Interpolation Techniques in Climatology and Meteorology (Eds. *Lakatos M, Hoffmann L, Kircsi A, Szentimrey T*), Budapest, Hungary, 2020, WCDMP-No. 86, pp. 84-101

Szentimrey, T. and *Bihari, Z.*, 2007: Mathematical background of the spatial interpolation methods and the software MISH (Meteorological Interpolation based on Surface Homogenized Data Basis), Proceedings of the Conference on Spatial Interpolation in Climatology and Meteorology, Budapest, Hungary, 2004, COST Action 719, COST Office, 2007, 17–27

Szentimrey, T, Bihari, Z., Lakatos, M., and *Szalai, S*., 2011: Mathematical, methodological questions concerning the spatial interpolation of climate elements. Proceedings of the Second Conference on Spatial Interpolation in Climatology and Meteorology, Budapest, Hungary, 2009. *Időjárás 115*, 1–11.

Szentimrey T. et al., 2012a: Final report on quality control and data homogenization measures applied per country, including QC protocols and measures to determine the achieved increase in data quality. Carpatclim Project, Deliverable D1.12. http://www.carpatclim-eu.org/docs/deliverables/D1_12.pdf

Szentimrey T. et al., 2012b: Final report on the creation of national gridded datasets, per country. Carpatclim Project, Deliverable D2.9. http://www.carpatclim-eu.org/docs/deliverables/D2_9.pdf

Szentimrey, T. and *Bihari, Z.,* 2014: Manual of interpolation software MISHv1.03, Hungarian Meteorological Service.

Szentimrey, T., 2016: Analysis of the data assimilation methods from the mathematical point of view. (pp. 193-205), In (eds. *Bátkai, A., Csomós, P., Faragó, I., Horányi, A., Szépszó, G*.) Mathematical Problems in Meteorological Modelling. Springer International Publishing, Switzerland, 193–205, https://doi.org/10.1007/978-3-319-40157-7_10

Szentimrey, T., 2017: New developments of interpolation method MISH: modelling of interpolation error RMSE, automated real time quality control, Proceedings of the 9th Seminar for Homogenization and Quality Control in Climatological Databases and 4th Conference on Spatial Interpolation Techniques in Climatology and Meteorology (Eds. *Szentimrey T, Lakatos M, Hoffmann L*), Budapest, Hungary, 2017, WCDMP-No. 85, 115–124.

Szentimrey, T., 2021: Mathematical questions of spatial interpolation and summary of MISH, Proceedings of the 10**th** Seminar for Homogenization and Quality Control in Climatological Databases and **5**th Conference on Spatial Interpolation Techniques in Climatology and Meteorology (Eds. *Lakatos M, Hoffmann L, Kircsi A, Szentimrey T*), Budapest, Hungary, 2020, WCDMP-No. 86, 59–68.

Szentimrey, T., 2023a: Overview of mathematical background of homogenization, summary of method MASH and comments on benchmark validation. *Int. J. Climatol.43*, 6314–6329. https://doi.org/10.1002/joc.8207

Szentimrey, T., 2023b: Manual of homogenization software MASHv4.01, Varimax Limited Partnership.

Szentimrey, T., 2023c: Statistical modelling of the present climate by the interpolation method MISH – theoretical considerations, (extended abstract), Proceedings of the 11**th** Seminar for Homogenization and Quality Control in Climatological Databases and 6th Conference on Spatial Interpolation Techniques in Climatology and Meteorology (Eds. *Lakatos M, Puskás M, Szentimrey T*), Budapest, Hungary, 2023, WCDMP-No. 87, 45–50. https://library.wmo.int/idurl/4/68452