Mathematical questions of spatial interpolation and summary of MISH

Tamás Szentimrey

Varimax Limited Partnership

Budapest

Introduction

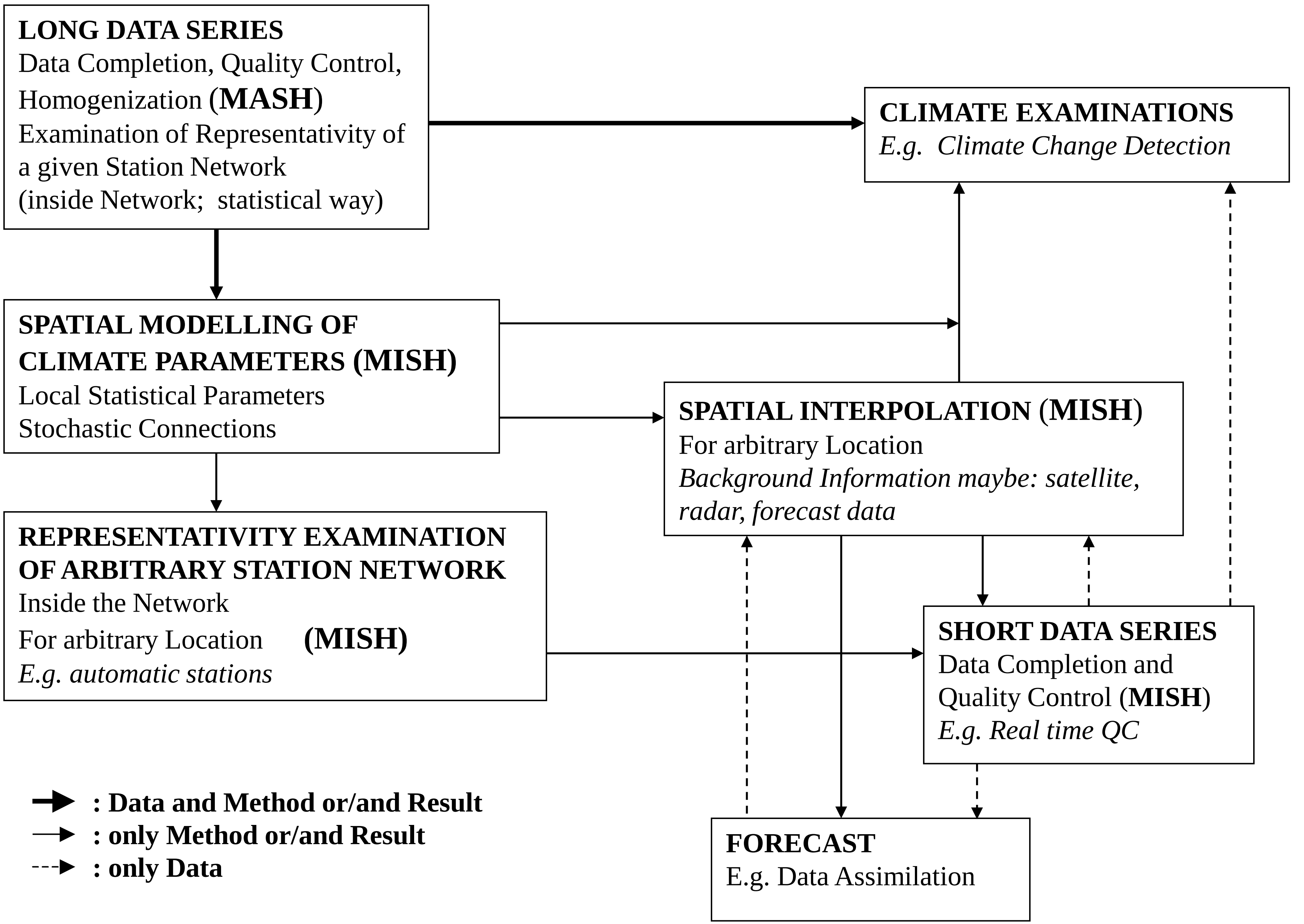
I retired from the Hungarian Meteorological Service.

I continue my activity in VARIMAX Limited Partnership.

Software: the new versions MASHv4.01 and MISHv2.01
are planned to share on the website of VARIMAX next year.

Presentation: summary of our conception on spatial interpolation.

Possible Connection of Topics and Systems



What is the Mathematics of Spatial Interpolation in Meteorology?

– Nowadays the geostatistical interpolation methods built in **GIS** are applied in meteorology, e.g. the various kriging methods.

– The geostatistical methods with correct mathematics are based on spatial data only and cannot efficiently use the spatiotemporal data like the meteorological data series.

– While the meteorological data series make possible to obtain the necessary climate information i.e. to model the climate statistical parameters for spatial interpolation.

Spatial interpolation (normal distribution, e.g. temperature)

Daily or monthly data for a date ($\mathbf{s}$ location vector)

Predictand: $Z(\mathbf{s}_0)$ Predictors: $Z(\mathbf{s}_i)$ $(i = 1,\ldots, M)$

Statistical Parameters

Expected values (spatial trend): $E(\mathbf{s}_i) = \mathrm{E}(Z(\mathbf{s}_i))$ $(i = 0,\ldots, M)$

Standard deviations: $D(\mathbf{s}_i) = \mathrm{D}(Z(\mathbf{s}_i))$ $(i = 0,\ldots, M)$

$\mathbf{r}$: predictand-predictors correlation vector,

$\mathbf{R}$: predictors-predictors correlation matrix.

$\mathbf{c}$: predictand-predictors covariance vector,

$\mathbf{C}$: predictors-predictors covariance matrix.

Additive (Linear) Interpolation

Interpolation Formula:

$$\hat{Z}(\mathbf{s}_0) = \lambda_0 + \sum_{i=1}^{M} \lambda_i \cdot Z(\mathbf{s}_i) \ , \quad \text{where } \sum_{i=1}^{M} \lambda_i = 1 \ .$$

Root Mean Square Error: $RMSE(\mathbf{s}_0) = \sqrt{\mathrm{E}\left(\left(Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0)\right)^2\right)}$

Optimal Interpolation Parameters: $\lambda_i \ (i = 0,..., M)$ minimize RMSE

Optimal Interpolation with the above parameters: $\hat{Z}_{opt}(\mathbf{s}_0)$

The Optimal Interpolation Parameters are known functions of the statistical parameters!

<u>Optimal constant term:</u> $\lambda_0 = \sum_{i=1}^{M} \lambda_i \left(E(\mathbf{s}_0) - E(\mathbf{s}_i) \right)$

<u>Vector of optimal weighting factors:</u> $\boldsymbol{\lambda} = [\lambda_1, .., \lambda_M]^{\mathrm{T}}$

$$\boldsymbol{\lambda} = \mathbf{C}^{-1}\left(\mathbf{c} + \frac{(1 - \mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{c})}{\mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{1}} \mathbf{1} \right) \qquad \text{(covariance form)}$$

and vector $\boldsymbol{\lambda}$ can be written as function of parameters:

$D(\mathbf{s}_i)\ (i = 0, ..., M),\ \mathbf{r},\ \mathbf{R}$.

Optimal, minimal RMSE with optimal parameters
(to charactarize quantitatively the uncertainties of interpolation)

$$RMSE_{opt}(\mathbf{s}_0) = \sqrt{\left(D^2(\mathbf{s}_0) - \mathbf{c}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{c}\right) + \left(1 - \mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{c}\right)^2 \cdot \frac{1}{\mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{1}}}$$

and it can be written as function of parameters: $D(\mathbf{s}_i)(i = 0,..,M), \mathbf{r}, \mathbf{R}$.

Summary

The necessary statistical parameters for calculation of $\hat{Z}_{opt}(\mathbf{s}_0)$ and

$RMSE_{opt}(\mathbf{s}_0)$ are: $E(\mathbf{s}_i),\ D(\mathbf{s}_i)\ (i = 0,...,M),\ \mathbf{r},\ \mathbf{R}$

How can we know them?

Modelling of monthly, daily spatiotemporal statistical parameters in MISH

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

ii. Spatial standard deviations $D(\mathbf{s})$

iii. Spatial correlations $r(\mathbf{s}_1, \mathbf{s}_2)$

iv. Temporal autocorrelations $r(t_1, t_2)$

Modelling: can be based on long station data series. Sample in space and in time!

There is a substantial difference between Geostatistics and Meteorology

that is the amount of information for modelling the statistical parameters.

Conclusion:

We should know the present climate well! (not the future climate only)

Special advanced MATHEMATICS is needed!

The main features of MISHv2.01

(https://www.met.hu/en/omsz/rendezvenyek/homogenization_and_interpolation/software/)

I. Modelling system for climate statistical parameters in space

(expected values, standard deviations, spatiotemporal correlations)

- Based on long homogenized data series and model variables.
- Modelling procedure must be executed only once before the interpolation applications. **Totally different from the other methods!**

II. Spatial 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, annual values and many years' means can be interpolated.
- The expected interpolation error RMSE is modelled too, representativity examination of arbitrary station network.
- Real time Quality Control for daily and monthly data (additive model).
- Capability for application of background information such as satellite, radar forecast data. (with QC: data assimilation)
- Capability for gridding of data series, as grid point and grid-box average alike.

Example for Modelling by MISH

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

Input: the location coordinates only without any temperature data.

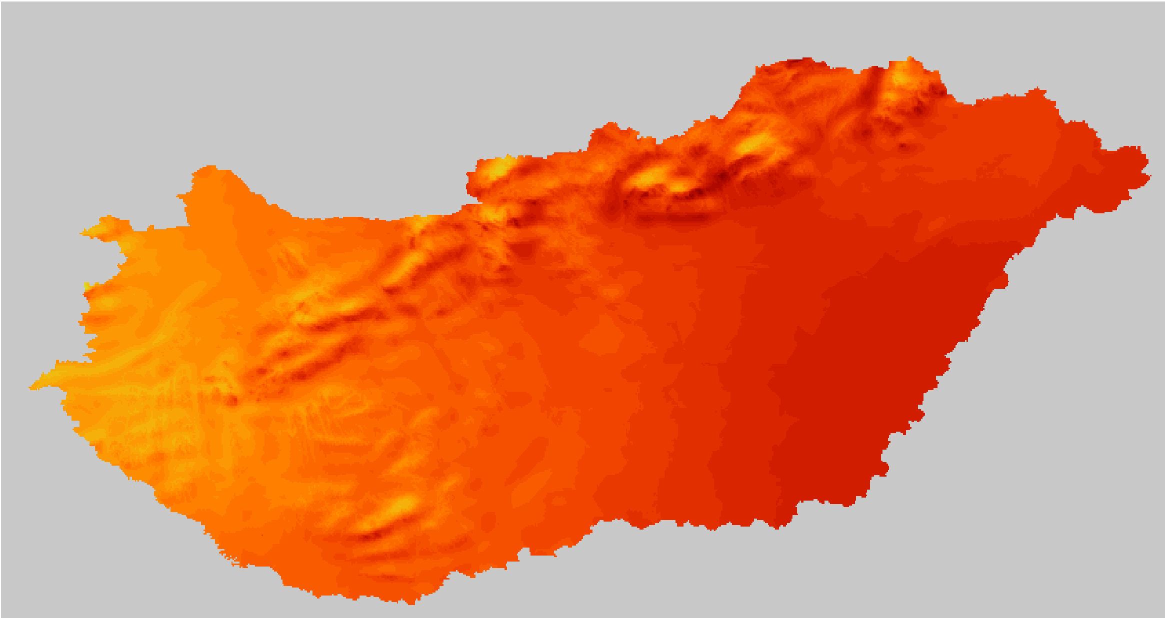
Output: modelled climate statistical parameters

	1	2	3	4	5	6	7	8	9	10
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

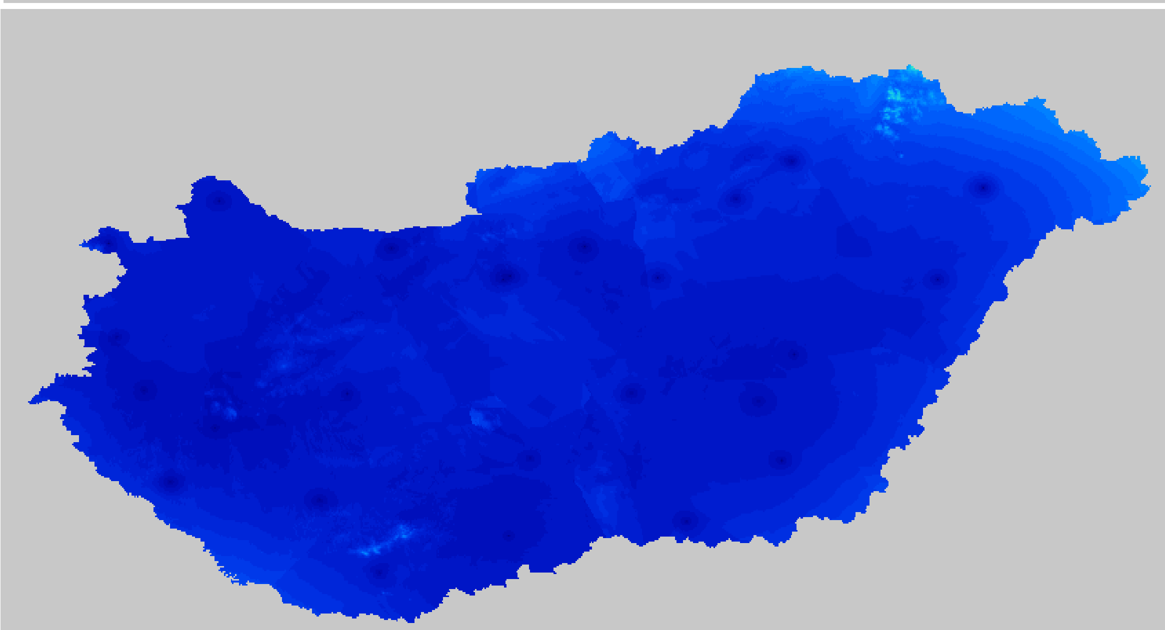
Matrix of Spatial Autocorrelations:

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

Modelled Maps for Monthly Mean Temperature in June (Hungary)



Modelled
St. Deviation (ºC)
Min: 0.87
Mean: 1.14
Max: 1.38



Modelled RMSE (ºC)
with 25 **arbitrary**
predictand locations
Min: 0
Mean: 0.16
Max: 0.58

Real time Quality Control for daily and monthly data

The principle of the test Statistics (*TS*) of Quality Control procedure at additive, normal model is as follows.

If the predictand $Z(\mathbf{s}_0)$ to be controlled and the predictors are correct, then

$$TS = \frac{Z(\mathbf{s}_0) - \hat{Z}_{opt}(\mathbf{s}_0)}{RMSE_{opt}(\mathbf{s}_0)} \in N(0,1) \quad (\text{=standard normal distribution})$$

where $\hat{Z}_{opt}(\mathbf{s}_0)$ is the interpolated value using modelled optimal parameters and $RMSE(\mathbf{s}_0)$ is the modelled interpolation error.

During the procedure multiple spatial comparison is tested.

Multiplicative Interpolation Formula of MISH

Optimum Interpolation Formula depends on the probability distribution.

Multiplicative Formula based on lognormal distribution for precipitation sum:

Predictand: $Z(\mathbf{s}_0, t)$ Predictors: $Z(\mathbf{s}_i, t)$ $(i = 1,.., M)$

$$\hat{Z}(\mathbf{s}_0, t) = \vartheta \cdot \left(\prod_{q_i \cdot Z(\mathbf{s}_i, t) \geq \vartheta} \left(\frac{q_i \cdot Z(\mathbf{s}_i, t)}{\vartheta} \right)^{\lambda_i} \right) \cdot \left(\sum_{q_i \cdot Z(\mathbf{s}_i, t) \geq \vartheta} \lambda_i + \sum_{q_i \cdot Z(\mathbf{s}_i, t) < \vartheta} \lambda_i \cdot \left(\frac{q_i \cdot Z(\mathbf{s}_i, t)}{\vartheta} \right) \right)$$

where $\vartheta > 0$, $q_i > 0$, $\sum_{i=1}^{M} \lambda_i = 1$ and $\lambda_i \geq 0$ $(i = 1,..., M)$,

are the interpolation parameters.

The optimum interpolation parameters are uniquely determined by certain climate statistical parameters.

Interpolation with Background Information

Background information can decrease the interpolation error.
For example: forecast, satellite, radar data

$Z(\mathbf{s}_0, t)$: predictand

$\hat{Z}(\mathbf{s}_0, t) = \lambda_0 + \sum_{i=1}^{M} \lambda_i Z(\mathbf{s}_i, t)$: interpolation

$\mathbf{G} = \{ G(\mathbf{s}, t) \mid \mathbf{s} \in D \}$: background information on a dense grid

<u>Principle of interpolation with Background Information</u>

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

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 $\mathbf{G}$.

Reanalysis data

Based on Data Assimilation (normality is assumed)

Minimization of the variational cost function:

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

$\mathbf{z}$: analysis field, predictand (grid),

$\mathbf{g}$: background field (forecast), assumption $\mathrm{E}(\mathbf{z}|\mathbf{g}) = \mathbf{g}$,

$\mathbf{y}_0$: observations, predictors; $\mathbf{Fz} = \mathrm{E}(\mathbf{y}_0|\mathbf{z})$,

$\mathbf{Q}$, $\mathbf{P}$: covariance marices

In essence:
Interpolation with background information + Quality control

Szentimrey, T. (2016): Analysis of the data assimilation methods from the mathematical point of view. In: *Mathematical Problems in Meteorological Modelling*, Springer International Publishing, Switzerland, 193–205

Problem with Reanalysis data

i, Inhomogeneous predictor station data series

ii, Few stations, little spatial representativity

iii, Problem with the data assimilation formula:

- Lack of good climate statistical parameters in matrix **Q**
- Assumption: $\mathrm{E}(\mathbf{z} \mid \mathbf{g}) = \mathbf{g}$?

Importance of gridded databases with good quality!

For example: CARPATCLIM project

There is no royal road!

(Archimedes)

Thank you for your attention!

There is no royal road! (Archimedes)

Question from Victor:

Do we not have a royal road or is there no royal road even in theory?

My answer:

We have not royal road in theory and in science.

Thank you for your attention!