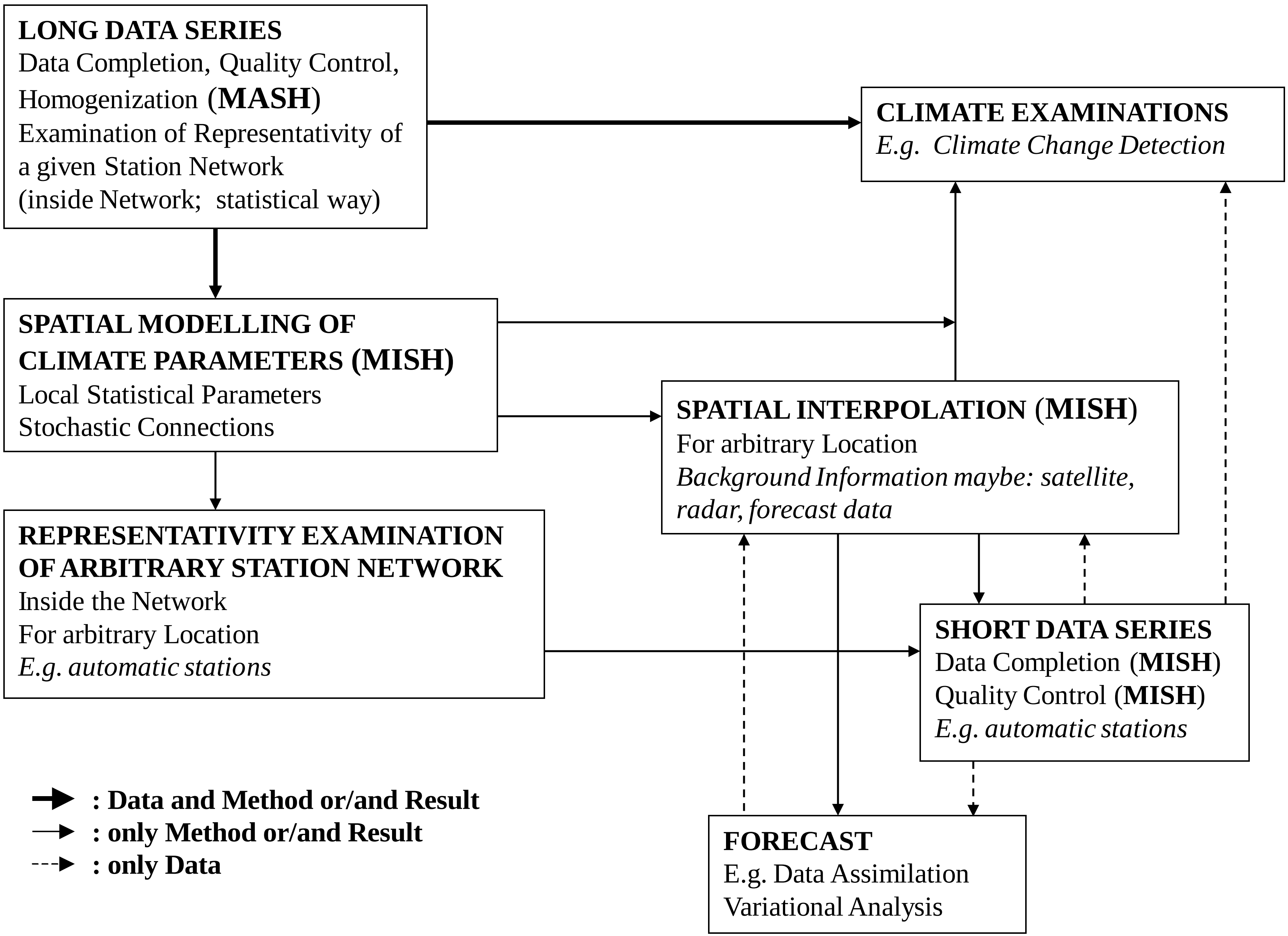
# New developments of interpolation method MISH: modelling of interpolation error RMSE, automated real time Quality Control

**Tamás Szentimrey**

**Hungarian Meteorological Service**

# Possible Connection of Topics and Systems



**Additive model of spatial interpolation (normal distribution, temperature)**

**Daily or monthly data for a date**

Predictand: $Z(\mathbf{s}_0)$

Predictors (observations): $Z(\mathbf{s}_i)$ $\left( i = 1,\ldots,M;\ \mathbf{s}\ \text{location vector} \right)$

**Statistical Parameters**

Local Parameters:

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

$\Big($Temporal trend: $\mathrm{E}(Z(\mathbf{s}_i, t)) = E(\mathbf{s}_i) + \mu(t)$ $\left( \text{year}\ t;\ i = 0,\ldots,M \right)\Big)$

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

# Stochastic parameters

Correlations:

**r** : predictand-predictors correlation vector,

**R** : predictors-predictors correlation matrix.

Covariances (from correlations and st. deviations):

**c** : predictand-predictors covariance vector,

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

Representativity Value: $REP(\mathbf{s}_0) = 1 - \dfrac{RMSE(\mathbf{s}_0)}{D(\mathbf{s}_0)}$

Optimal Interpolation Parameters : $\lambda_i \ \left( i = 0, \ldots, M \right)$

minimize RMSE.

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

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

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

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

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

$D(\mathbf{s}_0)/D(\mathbf{s}_i)$ $\left( i = 1, \ldots, M \right)$, $\mathbf{r}$, $\mathbf{R}$.

**Conclusion**

The expected values (spatial trend), the standard deviations and the correlations (stochastic part) are climate statistical parameters in meteorology.

That means:

**We could interpolate optimally if we knew the climate well!**

## Modelling of climate statistical parameters

The obtained optimal interpolation formula:

$$\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^{M} \lambda_i \left(E(\mathbf{s}_0) - E(\mathbf{s}_i)\right) + \sum_{i=1}^{M} \lambda_i Z(\mathbf{s}_i) \quad ,$$

where the weighting factors: $\boldsymbol{\lambda} = \mathbf{C}^{-1}\left(\mathbf{c} + \frac{\left(1 - \mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{c}\right)}{\mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{1}}\mathbf{1}\right)$

Unknown statistical parameters:

$E(\mathbf{s}_0) - E(\mathbf{s}_i)$ (do not depend on temporal trend of climate change),

$D(\mathbf{s}_0)/D(\mathbf{s}_i)$ $(i = 1,\ldots,M)$, $\mathbf{r}$, $\mathbf{R}$.

**Modelling:** can be based on long station data series $Z(\mathbf{S}_k, t)$ $(t = 1,\ldots,n)$ belonging to the stations $\mathbf{S}_k$ $(k = 1,..,K)$. Sample in space and in time!

# Difference between Geostatistics and Meteorology

**Amount of information for modelling the statistical parameters.**

## Geostatistics

Information: only the actual predictors $Z(\mathbf{s}_i)$ $(i = 1,\ldots, M)$.

Single realization in time!

## Meteorology

Information: Stations with long data series. Sample in space and in time!

Consequently the climate statistical parameters in question (expectations, covariances) for the stations are essentially known.

**Much more information for modelling!**

**Theorem**

Let us assume for the daily values within a month:

i, Expected values and standard deviations:

$$E_t(\mathbf{s}_0) - E_t(\mathbf{s}_i) = e_{0i}, \quad D_t(\mathbf{s}_0)/D_t(\mathbf{s}_i) = d_{0i} \quad (i = 1,...,M; t = 1,..,30)$$

ii, Correlations:

$$\mathrm{corr}\left(Z_{t_1}(\mathbf{s}_{i_1}), Z_{t_2}(\mathbf{s}_{i_2})\right) = r^S_{i_1 i_2} \cdot r^T_{t_1 t_2} \quad (i_1, i_2 = 1,...,M;\ t_1, t_2 = 1,..,30)$$

$r^S_{i_1 i_2}$ : correlation structure in space, $r^T_{t_1 t_2}$ : correlation structure in time.

Then the Optimum Interpolation Parameters for the daily values and monthly mean are identical: $\lambda_{i,t} = \lambda_{i,month} \quad (i = 0,...,M; t = 1,..,30)$.

Moreover the Representatvity Values for the daily values and monthly mean are also identical: $REP_t(\mathbf{s}_0) = REP_{month}(\mathbf{s}_0) \quad (t = 1,..,30)$.

**Remark about interpolation error RMSE**

(to charactarize quantitatively the uncertainties of interpolation)

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

$REP(\mathbf{s}_0) = 1 - \dfrac{RMSE(\mathbf{s}_0)}{D(\mathbf{s}_0)}$ depends on the parameters:

$D(\mathbf{s}_0)/D(\mathbf{s}_i)$ $(i = 1,\ldots,M)$, $\mathbf{r}$, $\mathbf{R}$.

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

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

## Modelling of monthly statistical parameters in MISH

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

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

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

**Remark:** Support program ANOVA (Analysis Of Variance) for modelling part, in order to evaluate the modelling results.

## Interpolation applications for monthly and daily data

$$\hat{Z}(\mathbf{s}_0) = \lambda_0 + \sum_{i=1}^{M} \lambda_i \cdot Z(\mathbf{s}_i) \ , \quad REP(\mathbf{s}_0) = 1 - \frac{RMSE(\mathbf{s}_0)}{D(\mathbf{s}_0)}$$

The Optimum Interpolation Parameters $\lambda_i$ $(i = 0,...,M)$ and Representativity Value $REP(\mathbf{s}_0)$ can be calculated from the above modelled monthly parameters.

# ANOVA (Analysis Of Variance) examination

## Partitioning of Total Variance of station data series

$$\hat{V} = \frac{1}{M \cdot n} \sum_{j=1}^{M} \sum_{t=1}^{n} \left( Z(\mathbf{s}_j, t) - \hat{E} \right)^2 = \frac{1}{M} \sum_{i=1}^{M} \left( \hat{E}(\mathbf{s}_i) - \hat{E} \right)^2 + \frac{1}{M} \sum_{i=1}^{M} \hat{D}^2(\mathbf{s}_i) = \hat{S}^2_{space} + \hat{D}^2_{time}$$

$\hat{S}^2_{space}$ is the variance of spatial trend, $\hat{D}^2_{time}$ is the mean temporal variance.

## CARPATCLIM ANOVA results for months in Hungary

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Tx** | | | | | | | | | | | | |
| $D_t$: | 2.67 | 3.24 | 2.69 | 1.87 | 1.96 | 1.64 | 1.71 | 1.98 | 1.96 | 1.83 | 2.43 | 2.11 |
| $S_s$: | 1.00 | 1.23 | 1.33 | 1.21 | 1.31 | 1.34 | 1.37 | 1.39 | 1.43 | 1.34 | 1.21 | 1.02 |
| **Tn** | | | | | | | | | | | | |
| $D_t$: | 2.76 | 2.88 | 1.86 | 1.35 | 1.20 | 1.12 | 1.21 | 1.18 | 1.29 | 1.67 | 1.97 | 2.12 |
| $S_s$: | 0.85 | 0.85 | 0.83 | 0.88 | 0.91 | 0.87 | 0.90 | 0.88 | 0.82 | 0.77 | 0.70 | 0.80 |
| **R** | | | | | | | | | | | | |
| $D_t$: | 22.5 | 22.9 | 21.3 | 25.6 | 36.2 | 39.0 | 39.3 | 40.7 | 36.5 | 35.7 | 33.3 | 27.9 |
| $S_s$: | 7.1 | 5.9 | 6.9 | 7.8 | 7.8 | 8.9 | 9.3 | 10.5 | 10.2 | 8.3 | 10.8 | 8.6 |

# New modelling parts in MISH

**iv,** Modelling of temporal daily autocorrelations $\rho(\mathbf{s})$ per months.

**v,** Modelling of daily standard deviations $D_{daily}(\mathbf{s})$ per months.

This development is based on the modelled autocorrelation $\rho(\mathbf{s})$. Let us assume the daily data of a given month constitute an AR(1) process with common standard deviation $D_{daily}(\mathbf{s})$ and temporal autocorrelation $\rho(\mathbf{s})$. Then $D_{daily}(\mathbf{s})$ can be estimated by using the monthly standard deviation $D_{month}(\mathbf{s})$:

$$D_{daily}(\mathbf{s}) \approx \sqrt{30 \cdot \frac{1-\rho}{1+\rho}} \cdot D_{month}(\mathbf{s})$$

**Consequently the first two spatiotemporal moments can be modelled for daily and monthly data by MISH!**

# Modelling of Present Climate in MISH

## Example

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

| index | lambda | fi |
|---|---|---|
| 1 | 17.30 | 47.00 |
| 2 | 17.30 | 47.20 |
| 3 | 17.30 | 47.40 |
| 4 | 17.30 | 47.60 |
| 5 | 17.30 | 47.80 |
| 6 | 17.60 | 45.80 |
| 7 | 17.60 | 46.00 |
| 8 | 17.60 | 46.20 |
| 9 | 17.60 | 46.40 |
| 10 | 17.60 | 46.60 |

**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 |

## Modelling of interpolation error RMSE

The modelled Representativity Value is the same for monthly and daily data: $REP(\mathbf{s}_0) = 1 - \frac{RMSE(\mathbf{s}_0)}{D(\mathbf{s}_0)}$

Consequently the modelled RMSE can be calculated as,

i, for monthly data: $RMSE_{month}(\mathbf{s}_0) = D_{month}(\mathbf{s}_0) \cdot (1 - REP(\mathbf{s}_0))$

ii, for daily data: $RMSE_{daily}(\mathbf{s}_0) = D_{daily}(\mathbf{s}_0) \cdot (1 - REP(\mathbf{s}_0))$

```
Example MISH output: Predictand location: 17.60  47.40

Predictor Indexes :   12   13   10   30   16   11   14
 Weighting Factors: 0.20 0.07 0.20 0.18 0.11 0.13 0.11

Interpolation without Background Information:
 Predictand Value:     14.77
 Representativity:     0.814    RMSE:    0.57

Interpolation with Background Information:
 Predictand Value:     14.74
 Representativity:     0.822    RMSE:    0.54
```

## Automated real time Quality Control for daily and monthly data

Test schema of QC procedure at additive, normal model is:

$$\frac{Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0)}{RMSE(\mathbf{s}_0)} \in N(0,1),$$

where $Z(\mathbf{s}_0)$ is the predictand to be controlled, $\hat{Z}(\mathbf{s}_0)$ 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 similarly to the QC procedure built in our MASH method for station data series.

**Multiplicative Interpolation Formula of MISH**

Optimum Interpolation Formula depends on the probability distribution.

Multiplicative Formula based on lognormal distribution for precipitation sum:

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

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

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.

## Software

http://www.met.hu/en/omsz/rendezvenyek/homogenizationand_interpolation/software/

## MASHv3.03

Multiple Analysis of Series for Homogenization;

*Szentimrey, T.*

## MISHv2.01

Meteorological Interpolation based on Surface Homogenized Data Basis;

*Szentimrey, T.and Bihari, Z.*

# The main features of MISHv2.01

## 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.

## 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.
- 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.

**There is no royal road!**

**Thank you for your attention!**