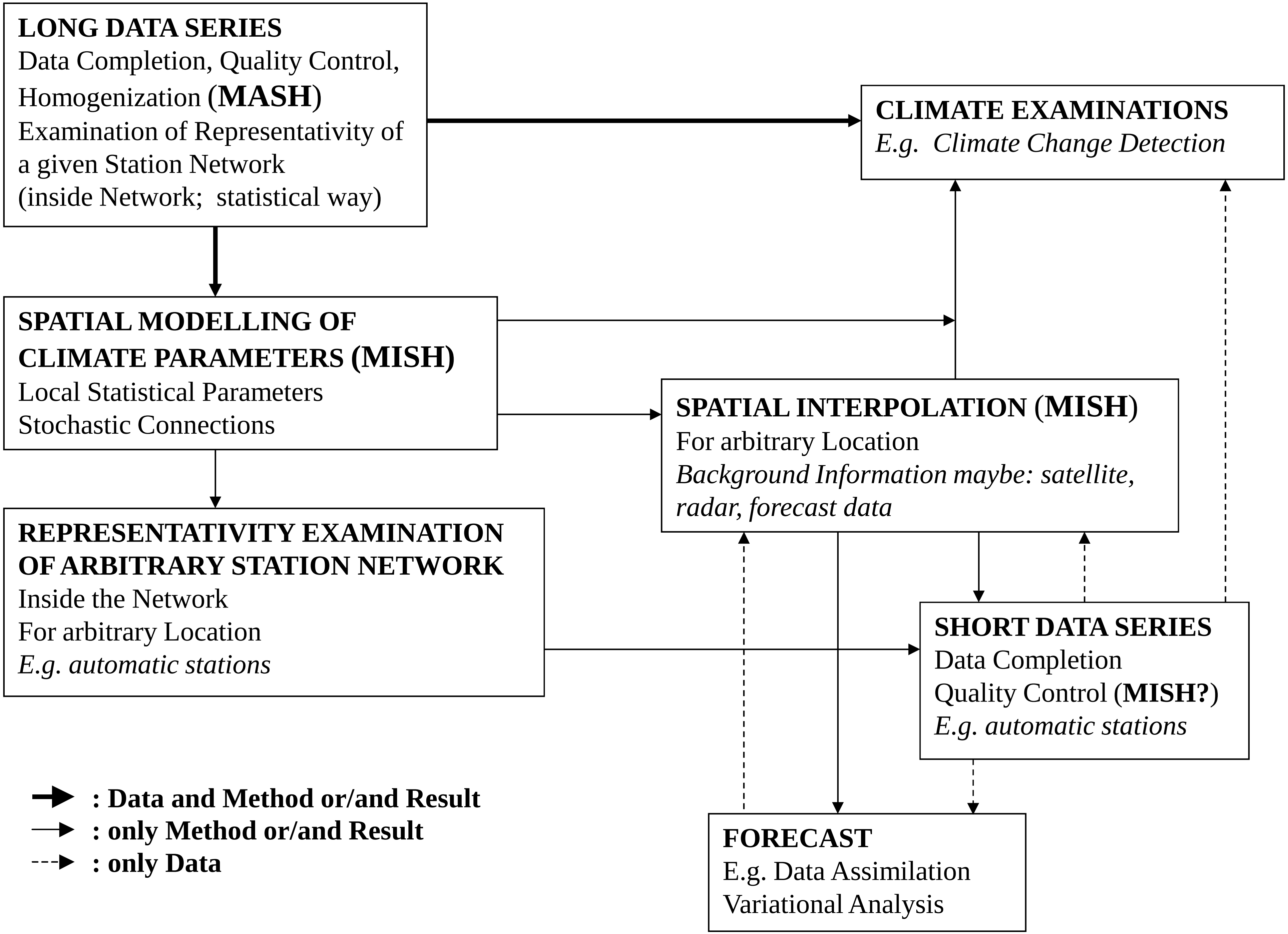
Mathematical questions of spatial interpolation of climate variables

Tamás Szentimrey, Zita Bihari, Mónika Lakatos

Hungarian Meteorological Service

Possible Connection of Topics and Systems



Schema of Meteorological Examinations

1. **Meteorology:** Qualitative formulation of the problem.

2. **Mathematics:** Quantitative formulation of the problem.

3. **Software:** Based on Mathematics.

4. **Meteorology:** Application of Software.

<u>John von Neumann:</u> Without quantitative formulation of the meteorological questions we are not able to answer the simplest qualitative questions either.

Spatial Interpolation Mathematics for Meteorology?

– Nowadays the geostatistical interpolation methods built in **GIS** are applied in meteorology.

– The mathematical basis of the geostatistical interpolation methods: **Geostatistics**

– The geostatistical methods can not efficiently use the meteorological data series.

– While the data series make possible to obtain the necessary climate information.

Additive model of spatial interpolation
(normal distribution, temperature)

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

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

(**s**: space, *t*: time)

Statistical Parameters

Deterministic Parameters:

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

Linear meteorological model for expected values:

$$\mathrm{E}(Z(\mathbf{s}_i, t)) = \mu(t) + E(\mathbf{s}_i) \quad (i = 0,..,M)$$

Temporal trend (unknown climate change): $\mu(t)$, Spatial trend: $E(\mathbf{s})$

Stochastic parameters

Covariance preferred in mathematical statistics and meteorology:

c : predictand-predictors covariance vector

C : predictors-predictors covariance matrix

Variogram preferred in geostatistics:

$\boldsymbol{\gamma}$: predictand-predictors variogram vector

$\boldsymbol{\Gamma}$: predictors-predictors variogram matrix

Additive (Linear) Interpolation

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

where $\sum_{i=1}^{M} \lambda_i = 1$, because of unknown $\mu(t)$.

Mean Square Error (MSE): $\mathrm{E}\left(\left(Z(\mathbf{s}_0, t) - \hat{Z}(\mathbf{s}_0, t)\right)^2\right)$

Optimal Interpolation Parameters :

λ_0, λ_i $(i = 1, \ldots, M)$ minimize MSE.

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: $\boldsymbol{\lambda} = [\lambda_1, .., \lambda_M]^{\mathrm{T}}$

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

ii, $$\boldsymbol{\lambda} = \boldsymbol{\Gamma}^{-1} \left(\boldsymbol{\gamma} + \frac{\left(1 - \mathbf{1}^{\mathrm{T}} \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma}\right)}{\mathbf{1}^{\mathrm{T}} \boldsymbol{\Gamma}^{-1} \mathbf{1}} \mathbf{1} \right)$$ (variogram form)

Conclusion

The spatial trend (deterministic part)

and the covariances (stochastic part)

are climate statistical parameters in meteorology.

It means that:

We could interpolate optimally

if we knew the climate well!

Remark

Inadequate formulas:

- Inverse Distance Weighting (IDW), $\lambda_0 = 0$, $\lambda_i \left(i = 1,\dots, M \right)$ not optimal
- Ordinary kriging, $\lambda_0 = 0$

Adequate formulas:

- Universal kriging,
- Regression (residual, detrended) kriging

But in geostatistics: modelling of statistical parameters is based on only the actual predictors

Modelling of climate statistical parameters

The obtained optimal interpolation formula:

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

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

Unknown statistical parameters: $E(\mathbf{s}_0)-E(\mathbf{s}_i)\left(i=1,\ldots,M\right)$, $\mathbf{c}$, $\mathbf{C}$

Modelling: can be based on long station data series $Z(\mathbf{S}_k,t)$ $\left(t=1,\ldots,n\right)$

belonging to the stations $\mathbf{S}_k$ $\left(k=1,..,K\right)$. 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 $Z(\mathbf{s}_i)$ $(i = 1,..., M)$.

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!

Remark 1

A single realization in time – i.e. only the actual predictors

$Z(\mathbf{s}_i)$ $(i = 1,..., M)$ at the geostatistical methods –

is insufficient climate information.

The data series make possible to know the climate!

It is the fundament of Climatology!

Remark 2

Nowadays however the geostatistical methods
– built in GIS – are applied in meteorology!

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

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

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

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, variational analysis

Minimization of the variational cost function:

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

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

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

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

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

In essence:
Interpolation with background information + Quality control

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: $\mathrm{E}\left(\mathbf{z} \mid \mathbf{g}\right) = \mathbf{g}$?

Importance of gridded databases with good quality!

For example: CARPATCLIM project

Software used at CARPATCLIM project

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

MASHv3.03

Multiple Analysis of Series for Homogenization;

Szentimrey, T.

MISHv1.03

Meteorological Interpolation based on Surface Homogenized Data Basis;

Szentimrey, T.and Bihari, Z.

The main features of MISHv1.03

I. Modelling system for climate statistical parameters in space

- 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 is modelled too.
- Capability for application of background information such as satellite, radar forecast data.
- Capability for gridding of data series.

There is no royal road!

Thank you for your attention!