Transformation of CarpatClim datasets to grid-box average datasets

ORSZÁGOS METEOROLÓGIAI SZOLGÁLAT 1870-2020

Beatrix Izsák[1,2], Tamás Szentimrey[3], Mónika Lakatos[1], Zita Bihari,[1] Andrea Kircsi[1]

[1] Hungarian Meteorological Service, Unit of Climatology, Hungary
[2] ELTE Faculty of Science, Doctoral School of Earth Sciences Budapest, Hungary
[3] Varimax Limited Partnership Budapest, Hungary

CarpatClim database

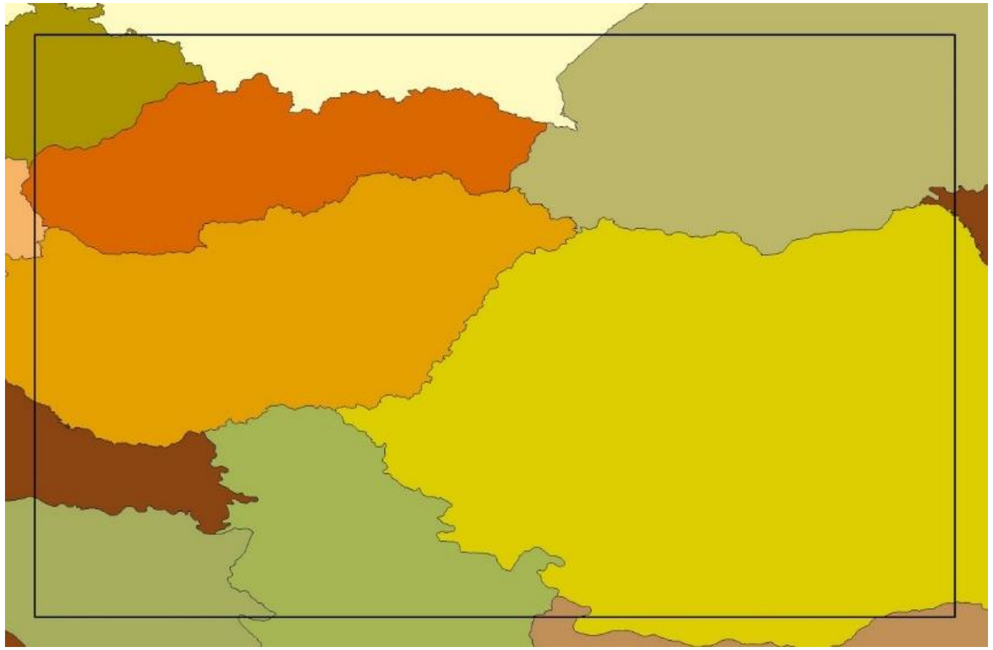
Daily gridded data series for basic meteorological variables in Carpathian Region (1961-2010)

Spatial resolution: 0,1°

Project of JRC (2011-2013) (10 participants)

Methodology: MASH for homogenization, MISH for gridding

CARPATCLIM
Climate of the Carpathian Region



ORSZÁGOS METEOROLÓGIAI SZOLGÁLAT 1870-2020

http://www.carpatclim-eu.org/pages/home/

THE COMMONLY USED METHODS AND SOFTWARE

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

MASHv3.03

(Multiple Analysis of Series for Homogenization; *Szentimrey, T.*)

For homogenization, quality control and missing value completion of station daily data series

MISHv1.03

(Meteorological Interpolation based on Surface Homogenized Data Basis; *Szentimrey, T.and Bihari, Z.*)

For gridding (interpolation) of homogenized daily data series

ORSZÁGOS METEOROLÓGIAI SZOLGÁLAT 1870-2020

The main features of MISHv1.03

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

- COPERNICUS C3S_311a_Lot4 project: Climate monitoring products for Europe based on Surface in-situ Observations- evaluation of new E-OBS data set
- *Transformation of CARPATCLIM dataset is necessary to comparisons*
- **MISH specialty**

- that the necessary statistical parameters - like spatial trend and correlation structure - are modelled for a very dense half minutes grid and saved.
- These statistical parameters were modelled during also the construction of CarpatClim datasets and they were also outputs of our MISH procedure applied for gridding.

-Using these saved parameters the transformation of CarpatClim datasets for grid-box average datasets is possible.

Additive model:

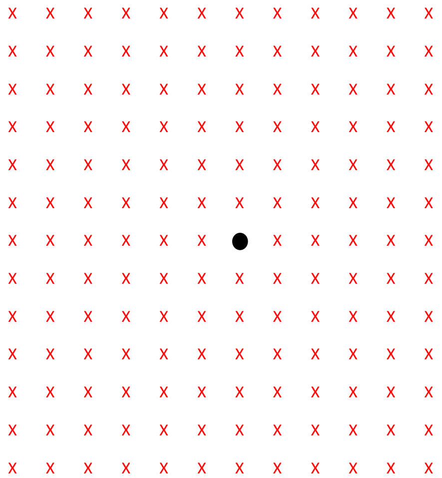
- The linear or additive model is appropriate in case of normal probability distribution.
- The spatial trend or median m(s)=E(s) were modelled for a very dense half minutes grid therefore we can transform the temperature CarpatClim gridpoint datasets for grid-box average datasets

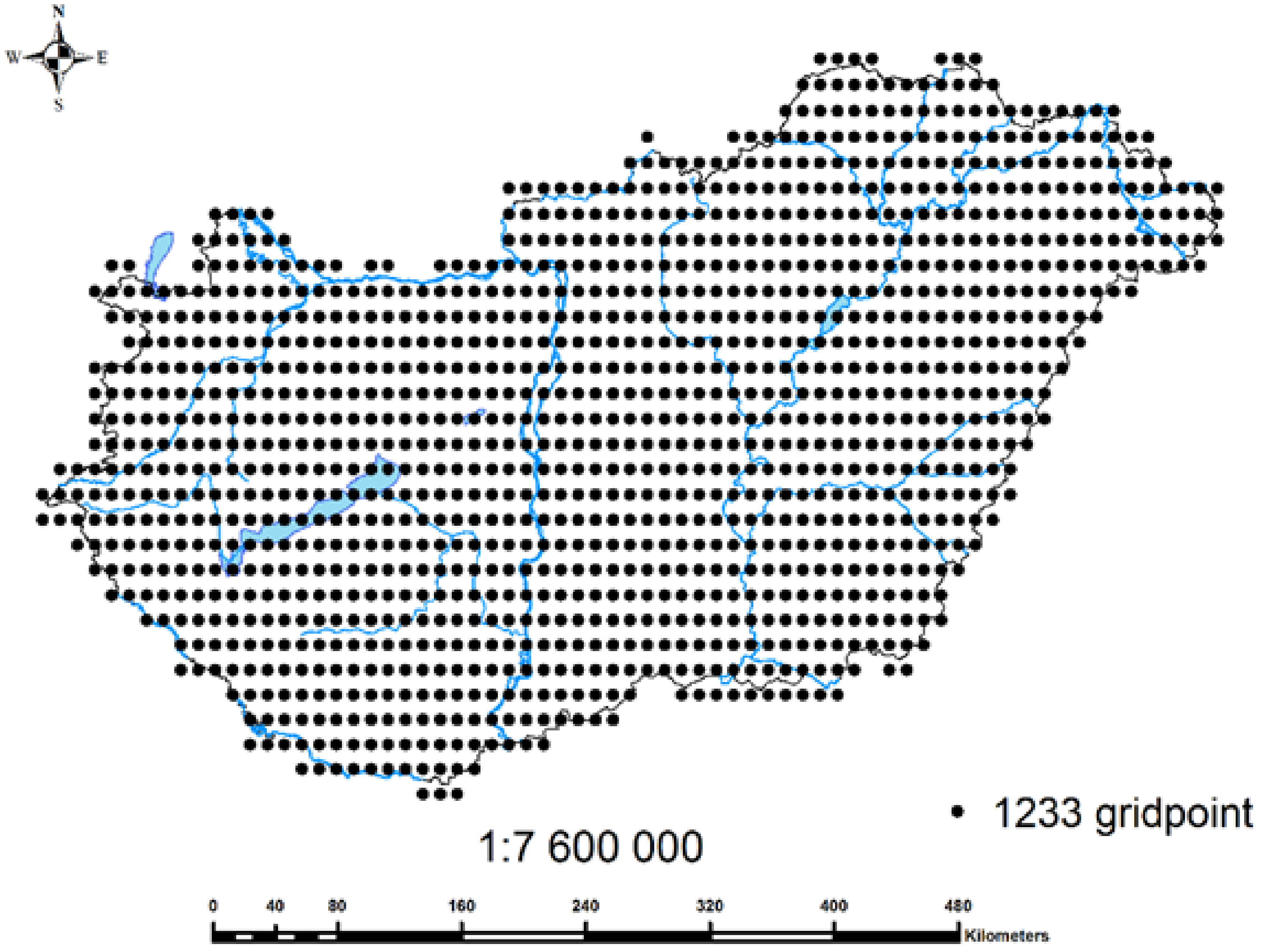
Multiplicative model:

- In case of a quasi lognormal distribution (e.g. precipitation sum) we deduced a mixed additive multiplicative formula which is used also in our MISH system.
- The spatial median m(s) were modelled for a very dense half minutes grid therefore we can transform the precipitation CarpatClim gridpoint datasets for grid-box average datasets

For Example the Hungarian grid

The resolution is 0.1 degrees, but the modeling is 0.5 minutes. So, around a grid point, 169 modeled medianes are saved in a box.





ORSZÁGOS METEOROLÓGIAI SZOLGÁLAT
1870-2020

Assuming the linear model – in case of normal distribution - the appropriate additive meteorological interpolation formula is as follows,

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

where Z($\mathbf{s}_0$) predictand, **s** are the element of the given space domain $\mathfrak{D}$

Z($\mathbf{s}_i$) (i=1,…,M) predictors

Where $\sum_{i=1}^{M}\lambda_i=1$ and the λ_i (i=1,…,M) minimize the root-mean-square error and these are known functions of some climate statistical parameters

$E(\mathbf{s}_i)$ (i=0,…,M) are the expected values or spatial trend values. (In case of normal distribution the expected value equals median.)

Mathematical background of transformation of gridpoint datasets for grid-box average datasets in case of additive model

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

Let $\mathbf{s}_{0,k} \in B(\mathbf{s}_0)$, $(k = 1, \dots, K)$ where B(s_0) **the gridbox round s_0.**

Then interpolated values within the gridbox are,

$$\hat{Z}(\mathbf{s}_{0,k}) = \sum_{i=1}^{M} \lambda_{i,k} \left(E(\mathbf{s}_{0,k}) - E(\mathbf{s}_i) \right) + \sum_{i=1}^{M} \lambda_{i,k} Z(\mathbf{s}_i) \qquad (k = 1, \dots, K)$$

Therefore the grid-box average is,

$$\hat{Z}_{Average}(\mathbf{s}_0) = \frac{1}{K}\sum_{k=1}^{K}\hat{Z}(\mathbf{s}_{0,k}) = \frac{1}{K}\sum_{k=1}^{K}\left(\sum_{i=1}^{M}\lambda_{i,k}\left(E(\mathbf{s}_{0,k}) - E(\mathbf{s}_i)\right) + \sum_{i=1}^{M}\lambda_{i,k}Z(\mathbf{s}_i)\right) \approx$$

as a consequence of the similar stochastic connection of the predictands within the grid-box with the predictors,

$$\approx \frac{1}{K}\sum_{k=1}^{K}\left(\sum_{i=1}^{M}\lambda_i\left(E(\mathbf{s}_{0,k}) - E(\mathbf{s}_i)\right) + \sum_{i=1}^{M}\lambda_i Z(\mathbf{s}_i)\right) = \sum_{i=1}^{M}\frac{1}{K}\sum_{k=1}^{K}\lambda_i\left(E(\mathbf{s}_{0,k}) - E(\mathbf{s}_0)\right) + \hat{Z}(\mathbf{s}_0) =$$

$$= \left(\bar{E}(\mathbf{s}_0) - E(\mathbf{s}_0)\right) + \hat{Z}(\mathbf{s}_0) \qquad = \left(\bar{\mathrm{m}}(\mathbf{s}_0) - \mathrm{m}(\mathbf{s}_0)\right) + \hat{\mathrm{Z}}(\mathbf{s}_0)$$

$\bar{\mathrm{m}}(\mathbf{s}_0)$: the average of the medians in a BOX around $\mathbf{s}_0$ point

$\mathrm{m}(\mathbf{s}_0)$: median in the $\mathbf{s}_0$ point

$\hat{\mathrm{Z}}(\mathbf{s}_0)$: the interpolated value from the CarpatClim gridpoint in the $\mathbf{s}_0$ point

Mathematical background of transformation of gridpoint datasets for grid-box average datasets in case of multiplicative model

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

$$\vartheta > 0 \,,\, q_i > 0, \quad \lambda_i \geq 0 \; (i = 1,\ldots,M) \qquad \sum_{i=1}^{M} \lambda_i = 1$$

During the construction of CarpatClim datasets we applied this interpolation formula with interpolation paramaters:

$$\vartheta = m(\mathbf{s}_0), \quad q_i = m(\mathbf{s}_0)/m(\mathbf{s}_i)$$

Where $m(\mathbf{s}_i), (i = 0, \ldots, M)$, *are the spatial median values.*

Let $\mathbf{s}_{0,k} \in B(\mathbf{s}_0)$, $(k = 1, \ldots, K)$ where B(s_0) **the gridbox round $\boldsymbol{s_0}$.**

Then interpolated values within the gridbox are,

$$\hat{Z}(\mathbf{s}_{0.k}) = m(\mathbf{s}_{0.k}) \cdot \left(\prod_{Z(\mathbf{s}_i) \geq m(\mathbf{s}_i)} \left(\frac{Z(\mathbf{s}_i)}{m(\mathbf{s}_i)} \right)^{\lambda_{i,k}} \right) \cdot \left(\sum_{Z(\mathbf{s}_i) \geq m(\mathbf{s}_i)} \lambda_{i,k} + \sum_{Z(\mathbf{s}_i) < m(\mathbf{s}_i)} \lambda_{i,k} \cdot \left(\frac{Z(\mathbf{s}_i)}{m(\mathbf{s}_i)} \right) \right)$$

Therefore the grid-box average is,

$$\hat{Z}_{Average}(\mathbf{s}_0) = \frac{1}{K}\sum_{k=1}^{K}\hat{Z}(\mathbf{s}_{0,k}) =$$

$$= \frac{1}{K}\sum_{k=1}^{K} m(\mathbf{s}_{0.k}) \cdot \left(\prod_{Z(\mathbf{s}_i) \geq m(\mathbf{s}_i)} \left(\frac{Z(\mathbf{s}_i)}{m(\mathbf{s}_i)} \right)^{\lambda_{i,k}} \right) \cdot \left(\sum_{Z(\mathbf{s}_i) \geq m(\mathbf{s}_i)} \lambda_{i,k} + \sum_{Z(\mathbf{s}_i) < m(\mathbf{s}_i)} \lambda_{i,k} \cdot \left(\frac{Z(\mathbf{s}_i)}{m(\mathbf{s}_i)} \right) \right) \approx$$

as a consequence of the similar stochastic connection of the predictands within the grid-box with the predictors,

$$= \left(\frac{1}{K}\sum_{k=1}^{K} m(\mathbf{s}_{0.k}) \right) \cdot \left(\prod_{Z(\mathbf{s}_i) \geq m(\mathbf{s}_i)} \left(\frac{Z(\mathbf{s}_i)}{m(\mathbf{s}_i)} \right)^{\lambda_i} \right) \cdot \left(\sum_{Z(\mathbf{s}_i) \geq m(\mathbf{s}_i)} \lambda_i + \sum_{Z(\mathbf{s}_i) < m(\mathbf{s}_i)} \lambda_i \cdot \left(\frac{Z(\mathbf{s}_i)}{m(\mathbf{s}_i)} \right) \right) =$$

$$= \frac{\frac{1}{K}\sum_{k=1}^{K} m(\mathbf{s}_{0.k})}{m(\mathbf{s}_0)} \cdot \hat{Z}(\mathbf{s}_0) = \frac{\overline{m}(\mathbf{s}_0)}{m(\mathbf{s}_0)} \cdot \hat{Z}(\mathbf{s}_0)$$

ANOVA (Analysis Of Variance)

$Z(\mathbf{s}_j, t)$ $(j = 1,..,N;\ t = 1,...,n)$ – data series $(\mathbf{s}_j$: location; t : time $)$

$$\hat{E}(\mathbf{s}_j) = \frac{1}{n}\sum_{t=1}^{n} Z(\mathbf{s}_j, t) \quad (j = 1,..,N) \text{ – temporal mean at location } \mathbf{s}_j$$

$$\hat{D}^2(\mathbf{s}_j) = \frac{1}{n}\sum_{t=1}^{n} \left(Z(\mathbf{s}_j, t) - \hat{E}(\mathbf{s}_j)\right)^2 \quad (j = 1,..,N) \text{ – temporal variance at location } \mathbf{s}_j$$

$$\hat{E}(t) = \frac{1}{N}\sum_{j=1}^{N} Z(\mathbf{s}_j, t) \quad (t = 1,..,n) \text{ – spatial mean at moment } t$$

$$\hat{D}^2(t) = \frac{1}{N}\sum_{j=1}^{N} \left(Z(\mathbf{s}_j, t) - \hat{E}(t)\right)^2 \quad (t = 1,..,n) \text{ – spatial variance at moment } t$$

$$\hat{E} = \frac{1}{N \cdot n}\sum_{j=1}^{N}\sum_{t=1}^{n} Z(\mathbf{s}_j, t) = \frac{1}{N}\sum_{j=1}^{N} \hat{E}(\mathbf{s}_j) = \frac{1}{n}\sum_{t=1}^{n} \hat{E}(t) \text{ – total mean}$$

$$\hat{D}^2 = \frac{1}{N \cdot n}\sum_{j=1}^{N}\sum_{t=1}^{n} \left(Z(\mathbf{s}_j, t) - \hat{E}\right)^2 \text{ – total variance}$$

Partitioning of Total Variance (Theorem)

$$\hat{D}^2 = \frac{1}{N}\sum_{j=1}^{N}\left(\hat{E}(\mathbf{s}_j) - \hat{E}\right)^2 + \frac{1}{N}\sum_{j=1}^{N}\hat{D}^2(\mathbf{s}_j) = \frac{1}{n}\sum_{t=1}^{n}\left(\hat{E}(t) - \hat{E}\right)^2 + \frac{1}{n}\sum_{t=1}^{n}\hat{D}^2(t)$$

$\frac{1}{N}\sum_{j=1}^{N}\left(\hat{E}(\mathbf{s}_j) - \hat{E}\right)^2$ – spatial variance of temporal means

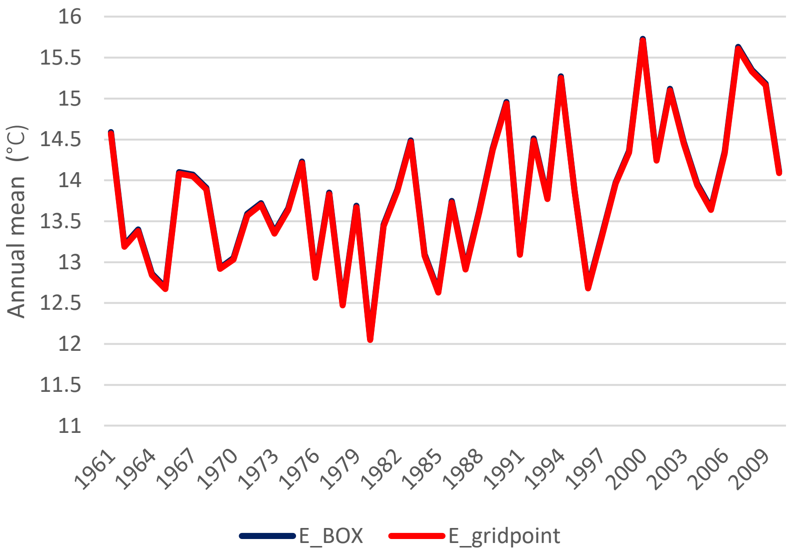
$\frac{1}{N}\sum_{j=1}^{N}\hat{D}^2(\mathbf{s}_j)$ – spatial mean of temporal variances

$\frac{1}{n}\sum_{t=1}^{n}\left(\hat{E}(t) - \hat{E}\right)^2$ – temporal variance of spatial means

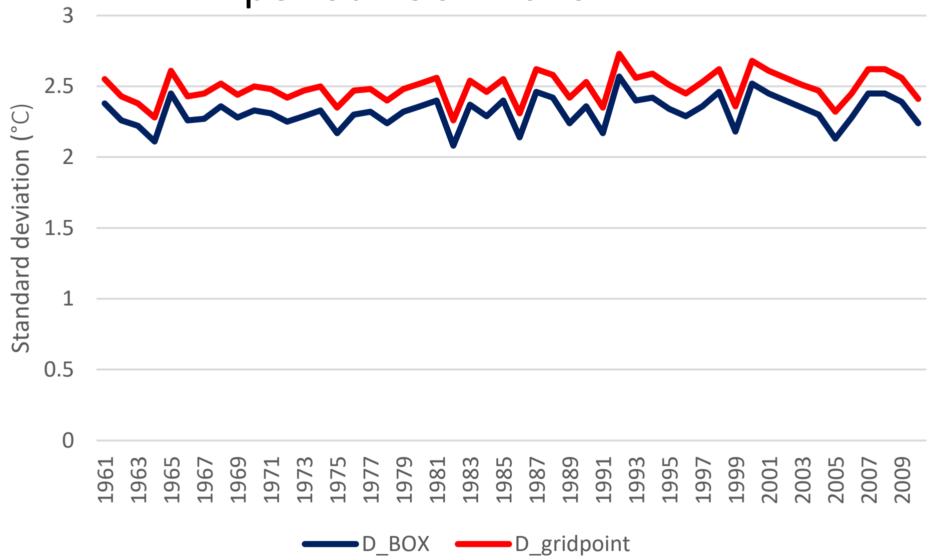
$\frac{1}{n}\sum_{t=1}^{n}\hat{D}^2(t)$ – temporal mean of spatial variances

Maximum temperature

Spatial mean series of annual mean maximum temperature for the period 1961-2010



Spatial standard deviation series of annual mean maximum temperature for the period 1961-2010



CarpatClim gridBOX

Total mean: 13.85
Total variance: 6.10

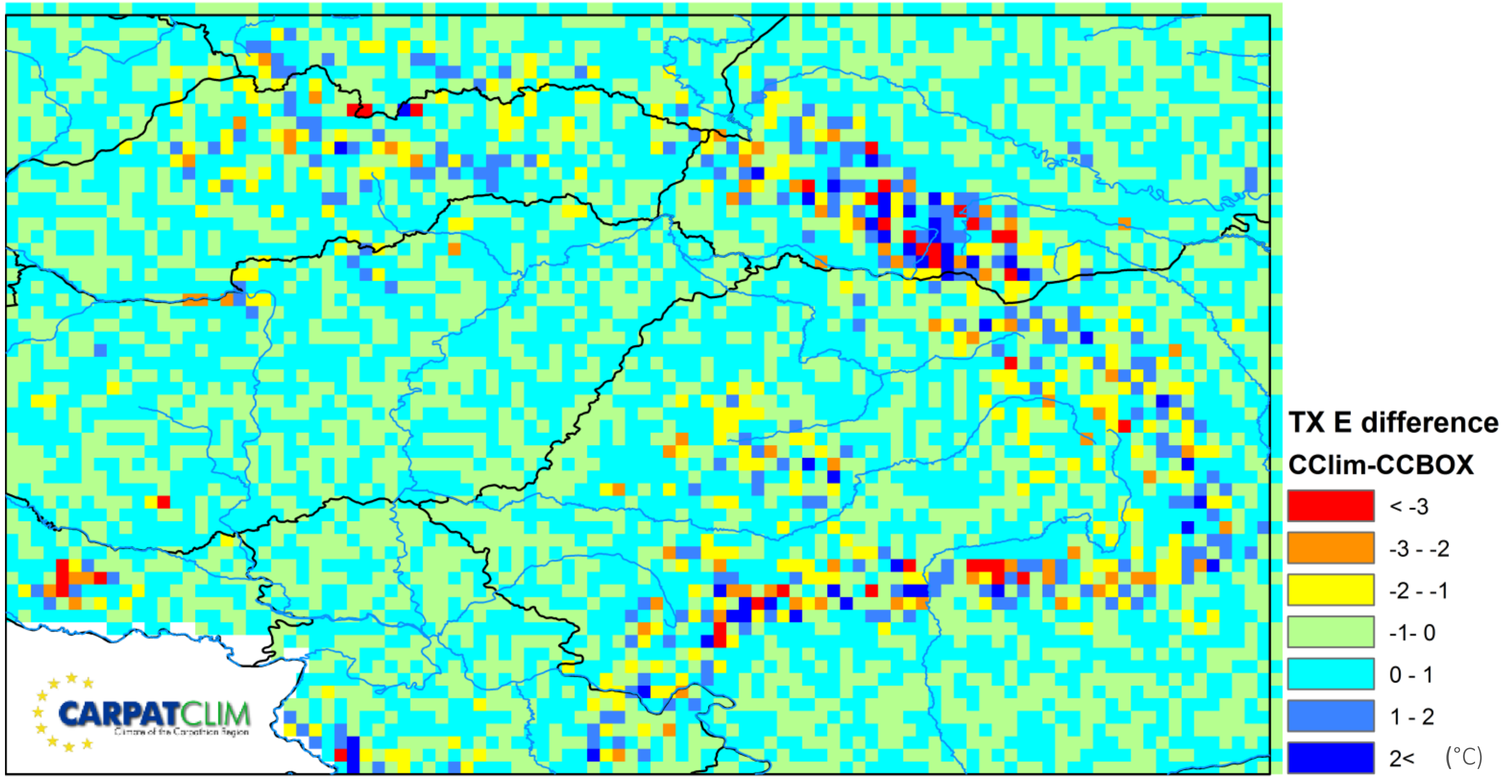
Spatial variance of temporal means: 5.32
Spatial mean of temporal variances: 0.78
Temporal variance of spatial means: 0.69
Temporal mean of spatial variances: 5.41

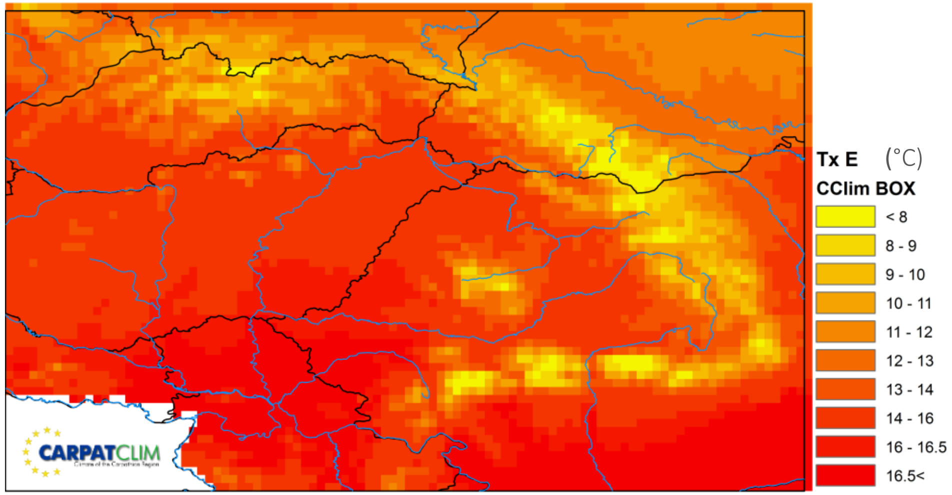
CarpatClim gridPoint

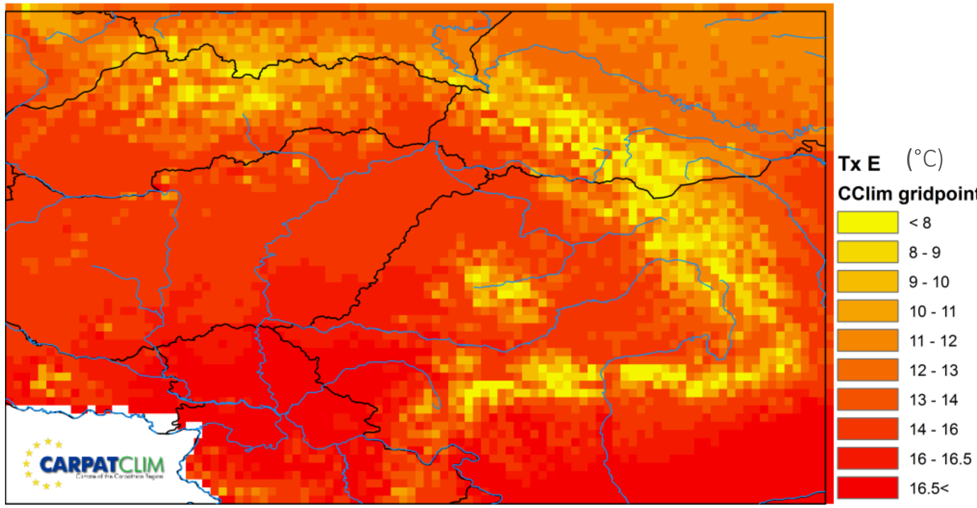
Total mean: 13.83
Total variance: 6.90

Spatial variance of temporal means: 6.13
Spatial mean of temporal variances: 0.78
Temporal variance of spatial means: 0.69
Temporal mean of spatial variances: 6.21

$\hat{E}\,(s_j)$ and the difference between the two CarpatClim Maximum temperature, 1961-2010



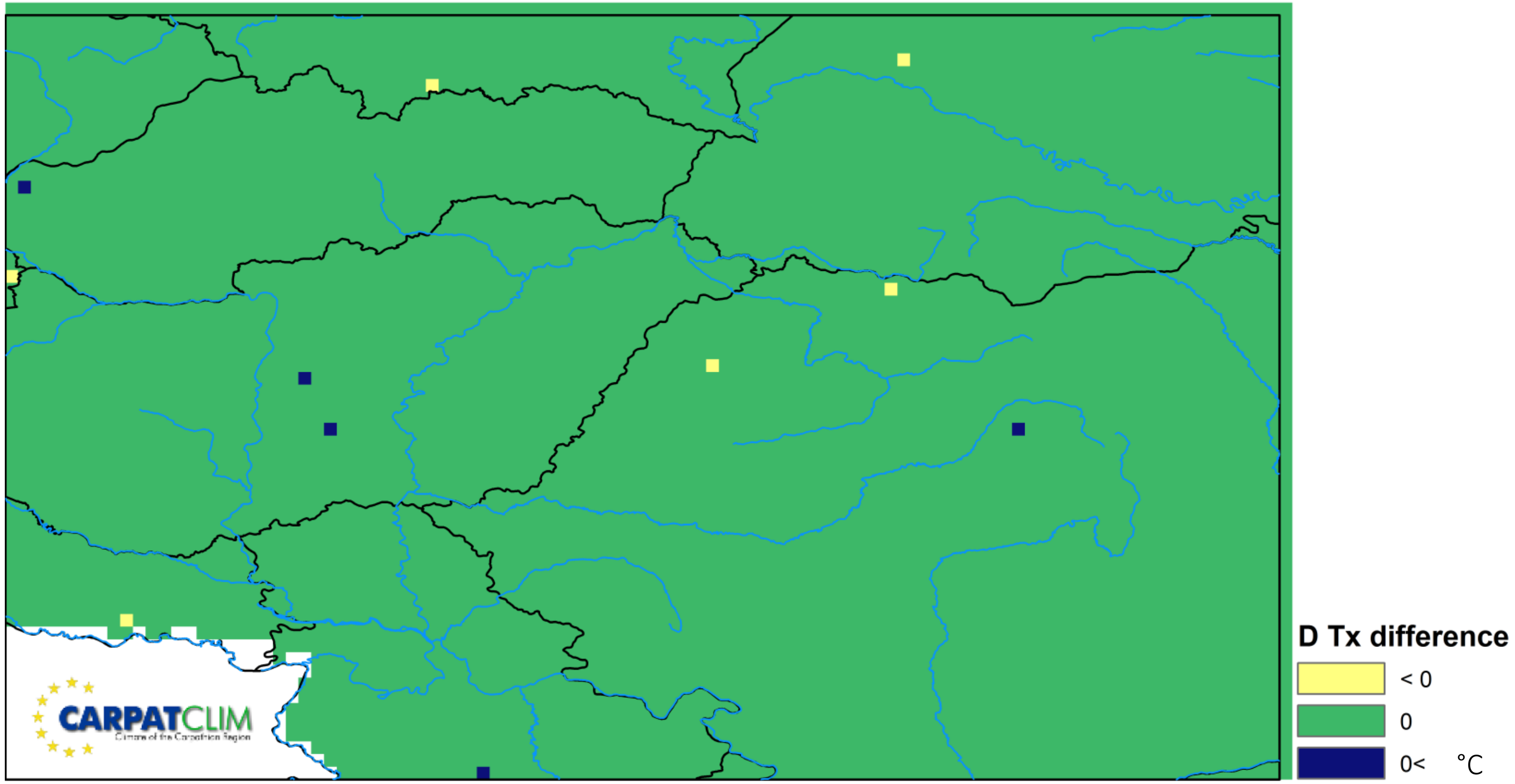


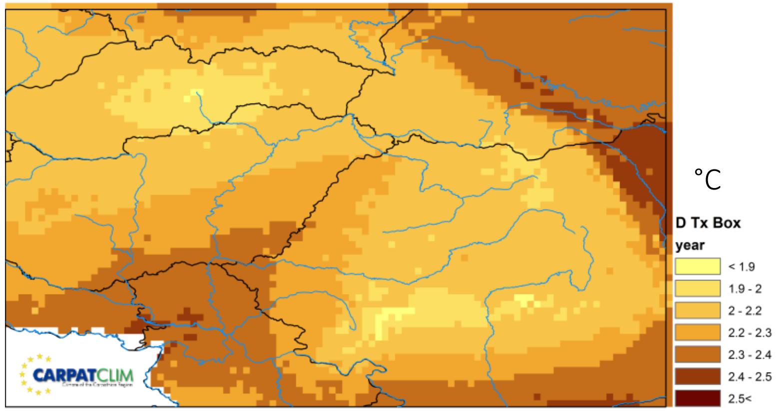


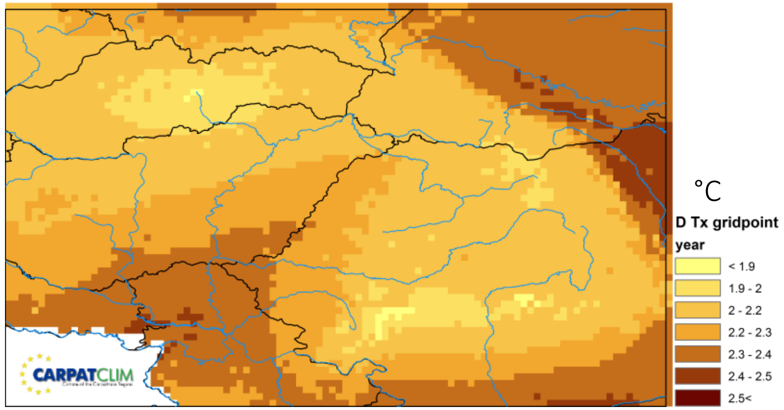
$$\hat{E}(\mathbf{s}_j) = \frac{1}{n}\sum_{t=1}^{n} Z(\mathbf{s}_j, t) \quad (j = 1,.., N) \; - \; \text{temporal mean at location } \mathbf{s}_j$$

$\widehat{D}\,(s_j)$ and the difference between the two CarpatClim

Maximum temperature, 1961-2010



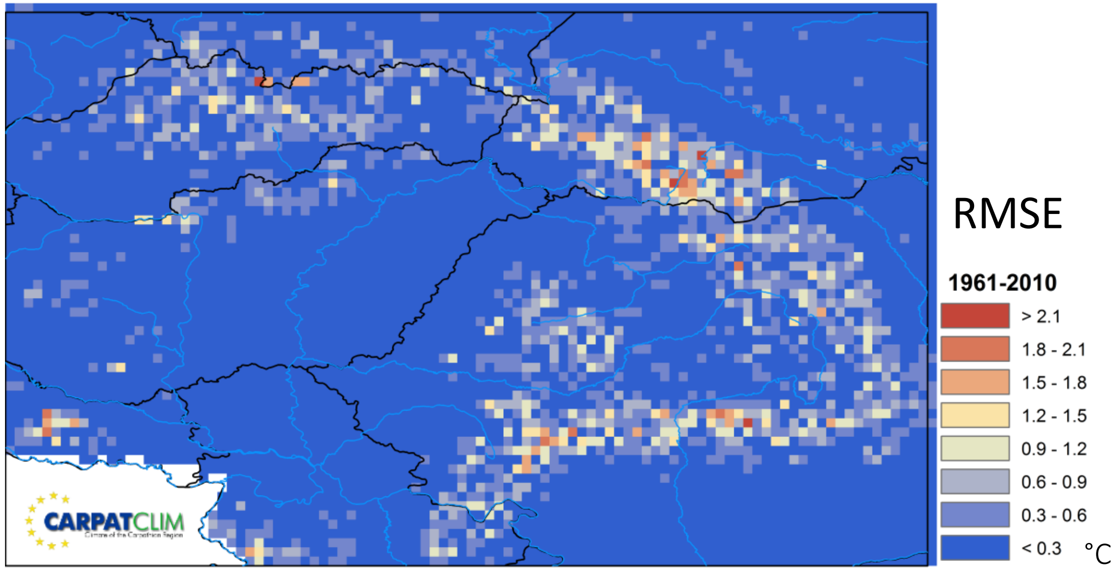




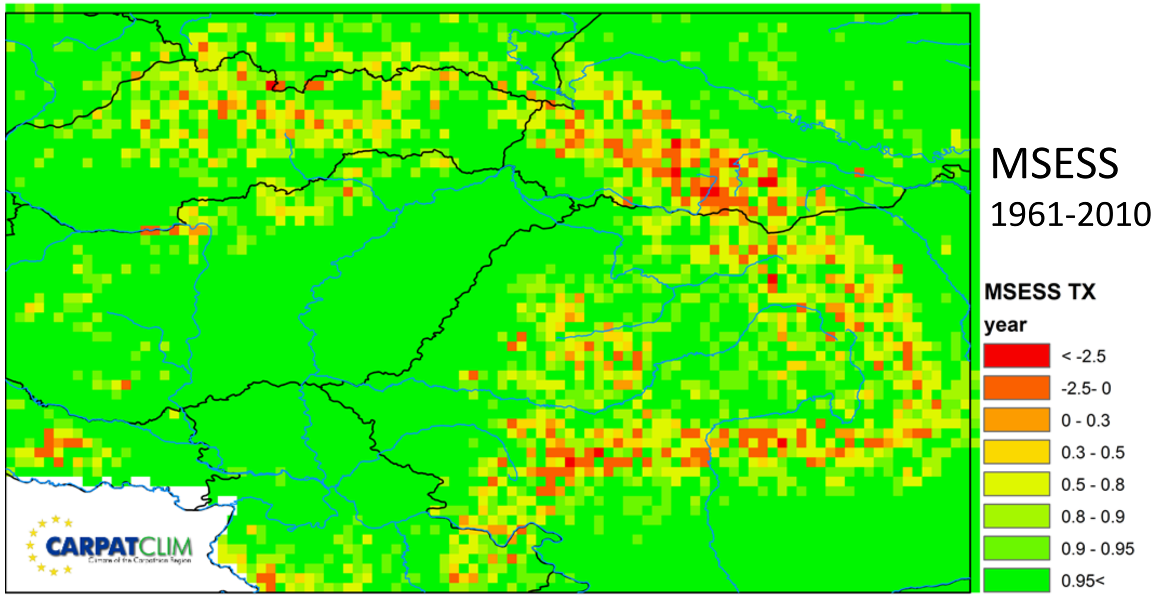
The difference is less than 0.01°C

$$\hat{D}^2(\mathbf{s}_j) = \frac{1}{n}\sum_{t=1}^{n}\left(Z(\mathbf{s}_j, t) - \hat{E}(\mathbf{s}_j)\right)^2 \quad (j = 1,.., N) \; - \; \text{temporal variance at location } \mathbf{s}_j$$

Monthly RMSE for the year



Monthly MSESS for the year



Monthly series: $\boldsymbol{Z(y,m)}$ (y: year, m: month)

$$V(Z(m)) = \frac{1}{n_y}\sum_{y=1}^{n_y}\left(Z(y,m) - \frac{1}{n_y}\sum_{y=1}^{n_y} Z(y,m)\right)^2$$

$$V(Z) = \frac{1}{12}\sum_{m=1}^{12} V(Z(m))$$

$$MSE\left(Z_{1,2}(m)\right) = \frac{1}{n_y}\sum_{y=1}^{n_y}\left(Z_1(y,m) - Z_2(y,m)\right)^2$$

$$MSE(Z_{1,2}) = \frac{1}{12}\sum_{m=1}^{12} MSE\left(Z_{1,2}(m)\right)$$

Monthly RMSE for the year:

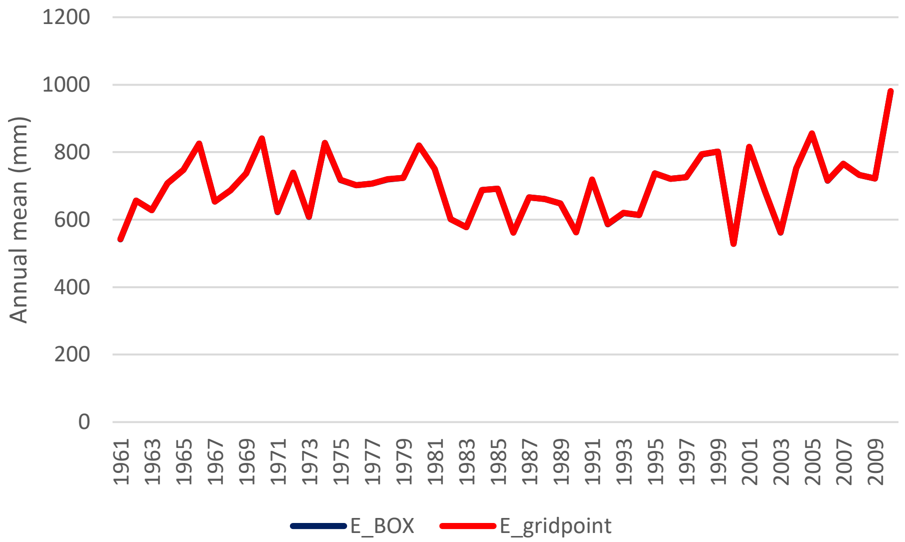
$$RMSE(Z_{1,2}) = \sqrt{MSE(Z_{1,2})}$$

$$MSESS(Z_{1,2}) = 1 - MSE\left(Z_{1,2}\right)/V\left(Z_2\right)$$

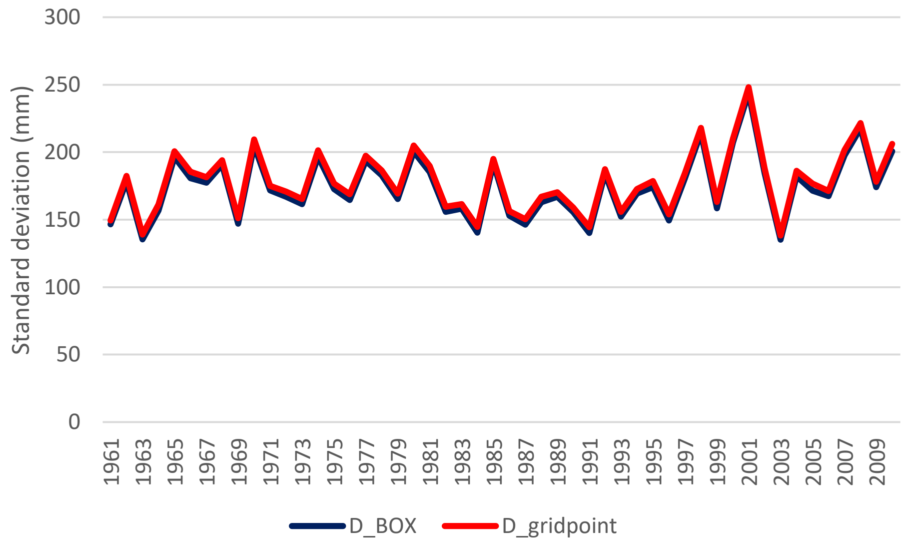
Maximum temperature

ANOVA results for precipitation

Spatial mean series of annual precipitation sum for the period 1961-2010



Spatial standard deviation series of annual precipitation sum for the period 1961-2010



CarpatClim gridBOX

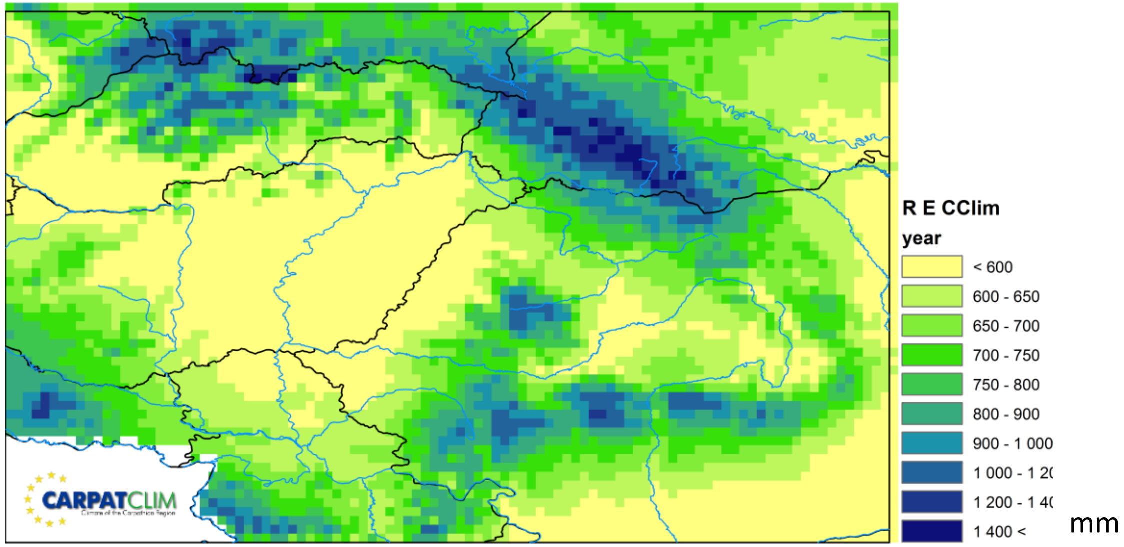
Total mean: 700.96
Total variance: 39178.73

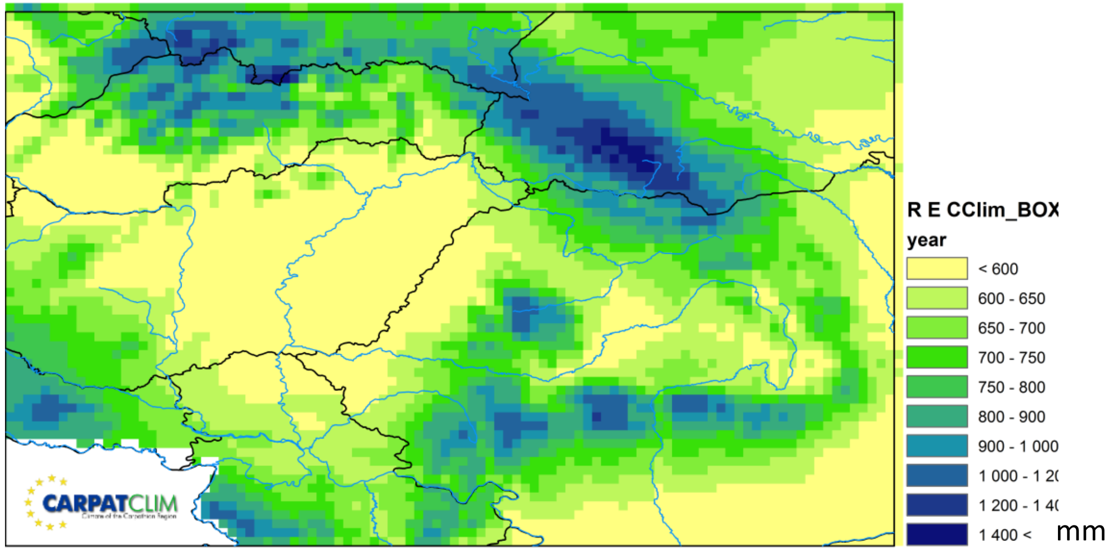
Spatial variance of temporal means: 23225.56
Spatial mean of temporal variances: 15953.17
Temporal variance of spatial means: 8293.03
Temporal mean of spatial variances: 30885.27

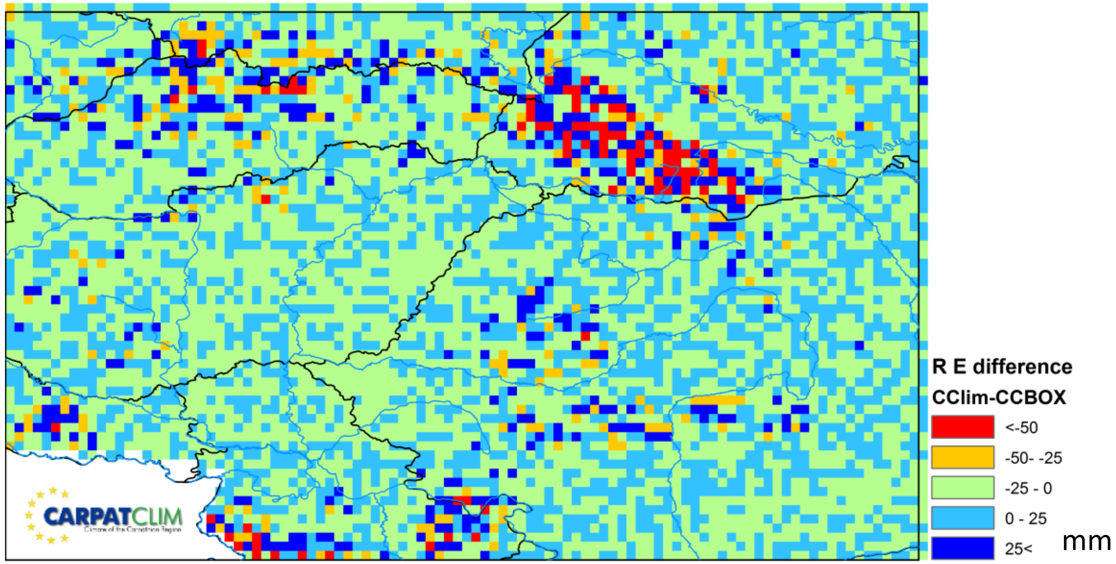
CarpatClim gridpoint

Total mean: 701.21
Total variance: 40565.66

Spatial variance of temporal means: 24571.22
Spatial mean of temporal variances: 15994.44
Temporal variance of spatial means: 8294.19
Temporal mean of spatial variances: 32271.01

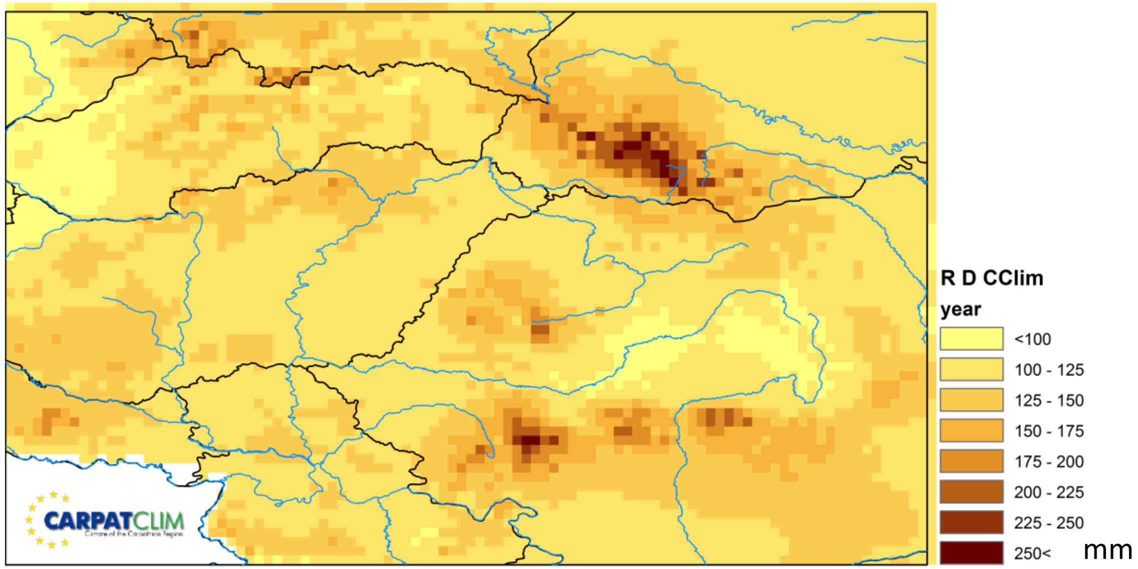


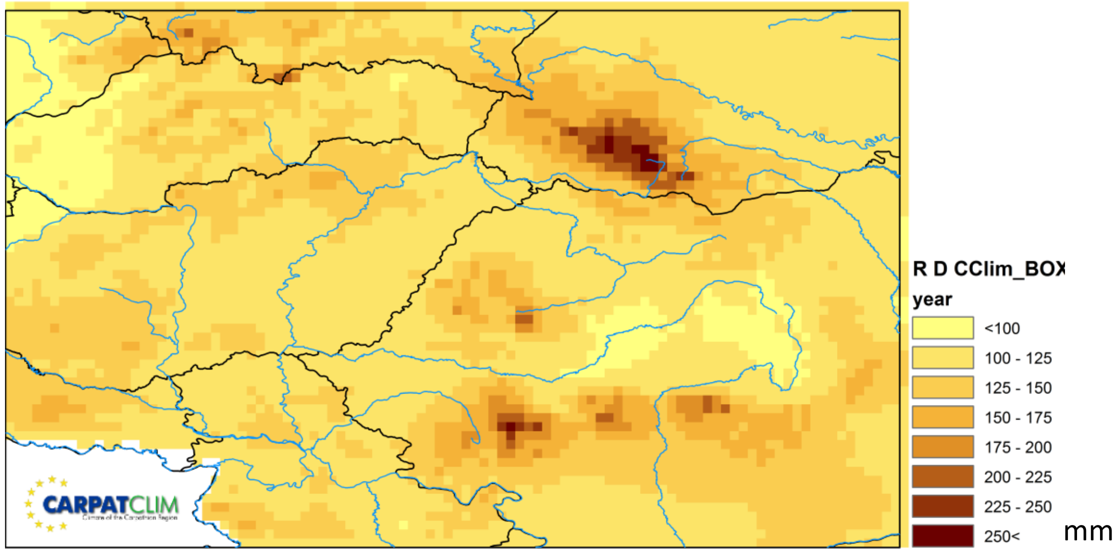


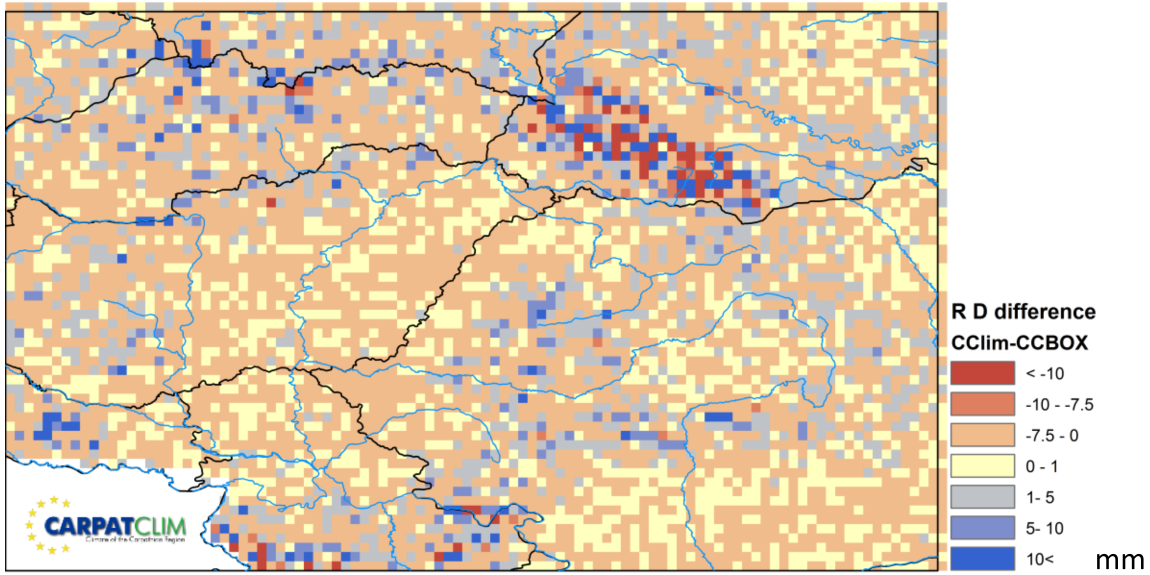


$\hat{E}\,(s_j)$ and the difference between the two CarpatClim (precipitation)
1961-2010

$$\hat{E}(\mathbf{s}_j) = \frac{1}{n}\sum_{t=1}^{n} Z(\mathbf{s}_j, t) \quad (j = 1,.., N) \; - \; \text{temporal mean at location } \mathbf{s}_j$$





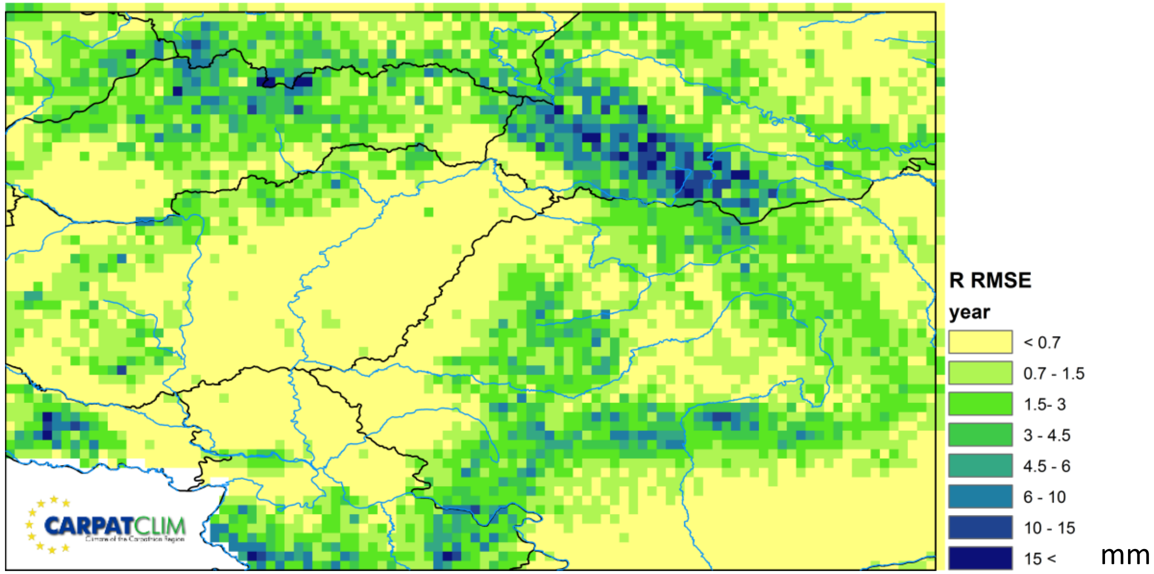


$\hat{D}$ (s_j) and the difference between the two CarpatClim, precipitation

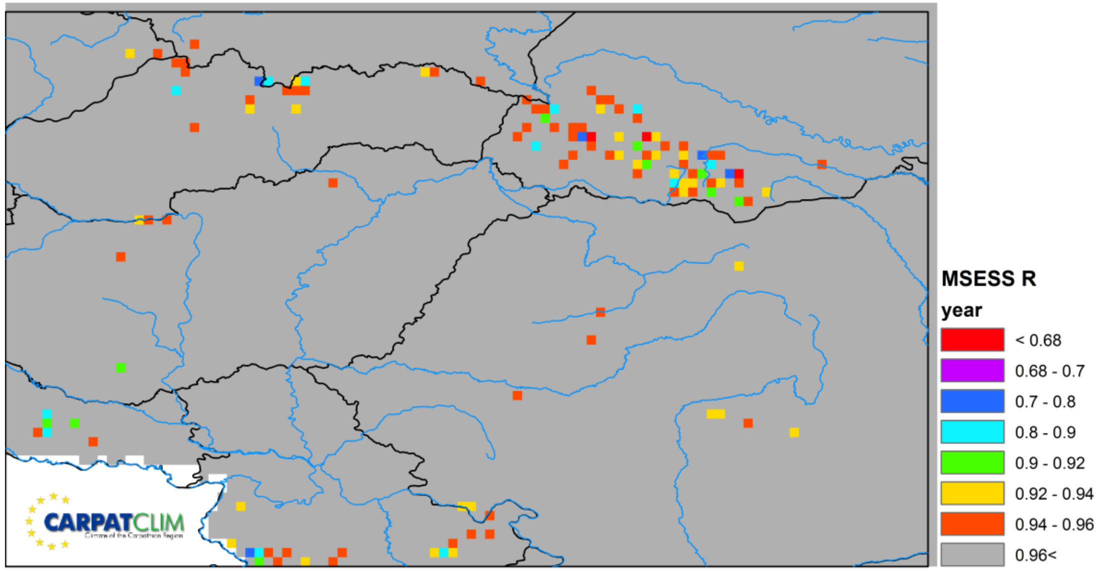
1961-2010

$$\hat{D}^2(\mathbf{s}_j)=\frac{1}{n}\sum_{t=1}^{n}\left(Z(\mathbf{s}_j,t)-\hat{E}(\mathbf{s}_j)\right)^2 \quad (j=1,..,N) \quad - \text{ temporal variance at location } \mathbf{s}_j$$

Precipitation



Monthly RMSE for the year 1961-2010



Monthly MSESS for the year 1961-2010

CarpatClim has two versions:
-gridpoint database (climate studies)
-box average database (climate modelling)

Climate studies: extremes
Climate modelling: smoothed data

There is a difference between the two CarpatClim where the surface is more complex: in the mountains, in the valleys, slopes and other surface forms.

ORSZÁGOS METEOROLÓGIAI SZOLGÁLAT 1870-2020

Thank you for your attention!

ORSZÁGOS
METEOROLÓGIAI
SZOLGÁLAT
1870-2020