

# The CaliM&Ro Project: Calibration of Met&Roll Weather Generator for sites without or with incomplete meteorological observations

Martin Dubrovsky<sup>(1)</sup>, Ladislav Metelka<sup>(2)</sup>, Daniela Semeradova<sup>(3)</sup>, Miroslav Trnka<sup>(3)</sup>, Olga Halasova<sup>(2)</sup>, Martin Ruzicka<sup>(4)</sup>, Ivana Nemesova<sup>(1)</sup>, Stanislava Kliegrova<sup>(2)</sup>, Zdenek Zalud<sup>(3)</sup>

<sup>(1)</sup> Institute of Atmospheric Physics ASCR, Prague, Czechia (dub@ufa.cas.cz)

<sup>(2)</sup> Czech Hydrometeorological Institute, Hradec Kralove, Czechia

<sup>(3)</sup> Institute of Landscape Ecology MAFU, Brno, Czechia

<sup>(4)</sup> Institute for Hydrodynamics ASCR, Prague, Czechia

## Abstract.

The main aim of the 3-year (2005-2007) CaliM&Ro project ([www.ufa.cas.cz/dub/calimaro/calimaro.htm](http://www.ufa.cas.cz/dub/calimaro/calimaro.htm)) is interpolation of the single-site four-variable stochastic daily weather generator (WG) Met&Roll. The paper gives the sample results obtained during the first year of the project. The experiments are based on daily weather series from 45 Czech stations. WG parameters are interpolated using the nearest neighbours method with accounting for the effect of the altitude. Observed weather series and synthetic weather series produced by site-calibrated and interpolated WGs are (a) analysed in terms of the frequency of heat waves and cold waves, and (b) fed into the WOFOST crop growth model with model crop yields being analysed thereafter. Two validation tests are made: (a) Performance of WG is validated by comparing characteristics (heat and cold wave frequencies and the model crop yields) obtained with the site-calibrated WG with those obtained with observed weather series. (b) Performance of the interpolation method is assessed by comparing characteristics obtained with the interpolated WG with those obtained with the site-calibrated WG.

## 1. Introduction

Stochastic weather generators (WGs) are used to produce observed-like synthetic weather series, which may serve as an input to various weather-dependent models (e.g. crop models [Dubrovsky et al., 2000; Zalud and Dubrovsky, 2002; Trnka et al., 2004a, 2004b] and rainfall-runoff models [Buchtele et al., 1999]). Typically, WGs are used in assessing sensitivity of the modelled processes to variability and changes in climatic characteristics. CaliM&Ro project (2005–2007) focuses on calibrating a stochastic single-site four-variables daily weather generator Met&Roll for sites with non-existent or incomplete historical daily weather series (these are normally used to determine WG parameters). To calibrate WG for the ungauged location, WG parameters may be interpolated from the surrounding stations. In our project, we test applicability of various interpolation techniques, including common methods implemented in GIS (with a stress put on co-kriging), neural networks and nearest neighbours. The performance of the interpolation techniques is examined in three ways: (A) Assessment of accuracy of interpolation of individual WG parameters (= comparison of the interpolated WG parameters with those derived from the site-specific observed weather series). (B) Comparison of the climatic characteristics derived from synthetic series produced

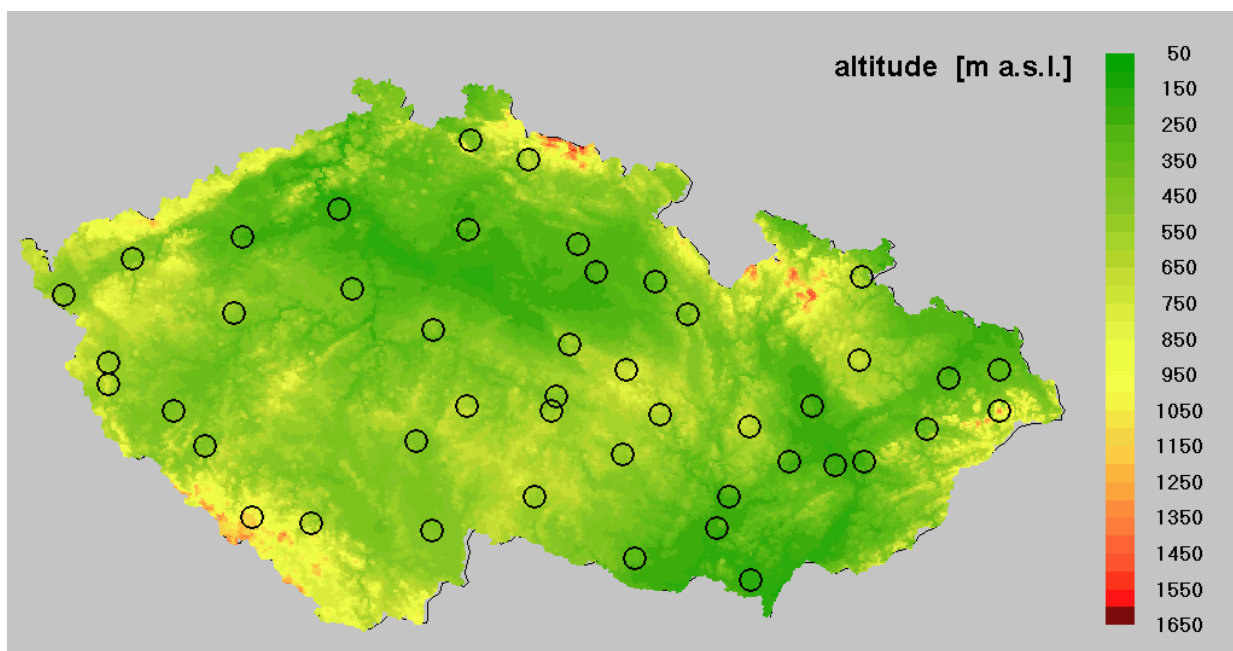
by the interpolated *vs.* site-calibrated WGs (interpolated WG = WG, whose parameters were interpolated; site-calibrated WG = WG, whose parameters were derived from the site-specific observed weather series). (C) Comparison of outputs from the crop growth models and rainfall-runoff models fed by synthetic series produced by interpolated *vs.* site-calibrated generators. In approach B, the weather series will be compared in terms of various climatic characteristics, including means, variability and extremes. In approach C, several crop growth models (WOFOST, CERES, STICS) and hydrological model SAC-SMA will be used. The experiments should show, what is the effect of the errors involved in WG interpolation process on the climatic characteristics derived from the synthetic weather series and on the output from the impact models fed by these series. In addition, the characteristics obtained using the synthetic weather series will be compared with those obtained with the observed series. Ideally, stochastic structure of observed and synthetic weather series should be the same.

The present paper brings the first results of the project. Nearest neighbours interpolator is used, and its performance is assessed in terms of heat and cold waves and wheat yields simulated by the WOFOST crop growth model.

## 2. Data and Methodology

### *Observed weather data*

The present analysis is based on 40-year daily weather series from 45 Czech stations (Fig. 1). The series consist of daily temperature extremes (*TMAX* and *TMIN*), daily precipitation amount (*PREC*) and daily sums of global solar radiation (*SRAD*). The former three variables are observed values, *SRAD* values have been estimated from the sunshine duration using the Angstrom-Prescott formula (Trnka et al., 2005).



**Figure 1** Topography of Czechia and location of the 45 stations used in the analysis.

### *Met&Roll weather generator*

Met&Roll is a parametric daily WG (Dubrovsky, 1997; Dubrovsky et al., 2004). It uses Markov chain (order = 1 to 3) to model precipitation occurrence, Gamma distribution to model precipitation amount and first-order autoregressive model to model solar radiation and daily extreme temperatures. Daily generator may be optionally conditioned on the monthly generator (Dubrovsky et al., 2004) which generates monthly means of the four variables using the 1<sup>st</sup> order autoregressive model. Parameters of the generator are allowed to exhibit annual cycle. Met&Roll was used mainly to provide synthetic series for crop growth model simulations – both in present and changed climates. It is also used to provide weather series for the hydrological SAC-SMA model (Buchtele et al., 1999). Met&Roll was also validated by Huth et al. (2001, 2003) and Kysely and Dubrovsky (2005) in terms of the extreme temperature characteristics.

### *Interpolation*

In this paper, we focus on the nearest neighbours interpolator. The interpolated value is defined as a weighted average of values related to the surrounding stations. The bell-shaped weight function and corrections for the zonal, meridional, and altitudinal trends are taken into account during the interpolation of WG parameter:

$$Y^{\wedge}(\lambda, \varphi, z) = \sum_{i=1..K} [Y_i + a.(\lambda - \lambda_i) + b.(\varphi - \varphi_i) + c.(z - z_i)] \times w_i(\lambda, \varphi, z) \quad (1)$$

where

- $Y^{\wedge}(\lambda, \varphi, z)$  is an interpolated value
- $Y_i$  is a value of the parameter derived from the  $i$ -th station's weather series
- $\lambda$ ,  $\varphi$ , and  $z$  are longitude, latitude and altitude
- $w_i(\lambda, \varphi, z)$  is a bell-shaped weight function accounting for the distance  $d_i(\lambda, \varphi, z)$  between the  $i$ -th station and the station defined by  $[\lambda, \varphi, z]$  coordinates:  
 $w_i(\lambda, \varphi, z) = [1 - (d_i(\lambda, \varphi, z)/D)^3]^3$  for  $d_i(\lambda, \varphi, z) < D$   
 $w_i(\lambda, \varphi, z) = 0$  for  $d_i(\lambda, \varphi, z) \geq D$
- $a$ ,  $b$ ,  $c$  are parameters of a tri-variate linear regression model ( $Y = a.\lambda + b.\varphi + c.z + d$ ), which are estimated using data from all available stations (except for the one, for which we interpolate – during the cross-validation test)
- $D$  defines the surroundings of the location for which we interpolate. The value of  $D$  is a subject of optimisation. We found, that  $D \sim 100$  km provides optimal results in the present experiment.

To show the importance of accounting for the altitude effect, we will also use simpler interpolator, which differs from the previous one by neglecting effect of altitude: in that case  $c$  is set to zero in eq.(1) and  $a$  and  $b$  are determined using bi-variate regression with  $\lambda$  and  $\varphi$  being independent variables. The two interpolators will be referred to as  $WG_{xy}$  and  $WG_{xyz}$  (Table 1).

### *Experiment*

1. In the first step, all WG parameters are derived for the 45 stations.
2. The WG parameters are interpolated for each of the 45 stations using data from the remaining 44 stations. Each parameter is interpolated independently of the others, using both interpolation approaches: with/without altitude effect.
3. 40-year synthetic series are generated using site-calibrated WG (thus producing  $WG_0$  series) and interpolated WGs (producing  $WG_{xy}$  and  $WG_{xyz}$  series).

**Table 1** Types of weather series used in the present experiments.

| acronym           | weather series  |
|-------------------|---|
| OBS               | observed weather series   |
| WG <sub>0</sub>   | synthetic series produced by WG calibrated with the site-specific observed data                                 |
| WG <sub>xy</sub>  | synthetic series produced by WG, whose parameters were interpolated (not accounting for the effect of altitude) |
| WG <sub>xyz</sub> | synthetic series produced by WG, whose parameters were interpolated (accounting for the effect of altitude)     |

- The four weather series (OBS, WG<sub>0</sub>, WG<sub>xy</sub> and WG<sub>xyz</sub>) available for each station are analysed in terms of the heat and cold waves. The heat wave (Huth et al., 2000) is defined here as the longest continuous period (i) during which the maximum daily temperature reached at least  $T_{H1}$  in at least three days, (ii) whose mean maximum daily temperature was at least  $T_{H1}$ , and (iii) during which the maximum daily temperature did not drop below  $T_{H2}$ . The threshold temperatures  $T_{H1}$  and  $T_{H2}$  are set to 30 and 25°C, respectively. Similarly, the cold wave is defined as the longest continuous period (i) during which the minimum daily temperature dropped to or below  $T_{C1}$  in at least three days, (ii) whose mean minimum daily temperature was  $T_{C1}$  or lower, and (iii) during which the minimum daily temperature did not exceed  $T_{C2}$ .  $T_{C1}$  and  $T_{C2}$  thresholds are set to -12 and -5°C, respectively.
- The four weather series are fed into the WOFOST model to simulate wheat yields. Note that the WOFOST model is used here as a “black box” which transforms weather series and site-specific information (cultivar characteristics, field attributes, soil characteristics, planting details, management factors) into the crop yields. The site-specific information is the same for all 45 stations. Although the latter setting is not realistic, it is justifiable in our experiment, which aims to validate the WG and the interpolation method.

### 3. Results

The station-specific frequencies of the heat and cold waves derived from the four weather series are displayed in Figures 2 and 3, respectively; the mean crop yields simulated by WOFOST crop growth model fed by the four types of the weather series are shown in figure 4. The displayed characteristics obtained using various input weather series are compared graphically in Figure 5 and quantitatively in Table 2 in terms of three measures: Mean Bias Error (*MBE*), Root Mean Square Error (*RMSE*) and Reduction of Variance (*RV*), all being expressed in percents:

$$MBE = 100 \times \{AVG(y_i - y_{i,ref}) / AVG(y_{i,ref})\}$$

$$RMSE = 100 \times \{[AVG(y_i - y_{i,ref})^2]^{1/2} / AVG(y_{i,ref})\}$$

$$RV = 100 \times \{1 - AVG(y_i - y_{i,ref})^2 / AVG[AVG(y_{i,ref}) - y_{i,ref}]^2\}$$

where

$y_i$  = characteristic of the  $i$ -th station (tested weather series is used)

$y_{i,ref}$  = characteristic of the  $i$ -th station (reference weather series is used)

AVG = average of the characteristic across the set of 45 stations

The perfect fit between the two samples would be indicated by  $MBE = 0$ ,  $RMSE = 0$  and  $RV = 100$ .

Having results related to the four different weather series, we focus on two comparisons: (i)  $WG_0$  vs. OBS, and (ii)  $WG_{xy}$  vs.  $WG_0$  and  $WG_{xyz}$  vs.  $WG_0$ . (Note that X-axis is used for  $WG_0$  characteristics in Figure 1;  $WG_0$  plays a role in both comparison tests). In the former test (panel B vs. panel A in Figures 2-4) the differences are due to the imperfections of WG.

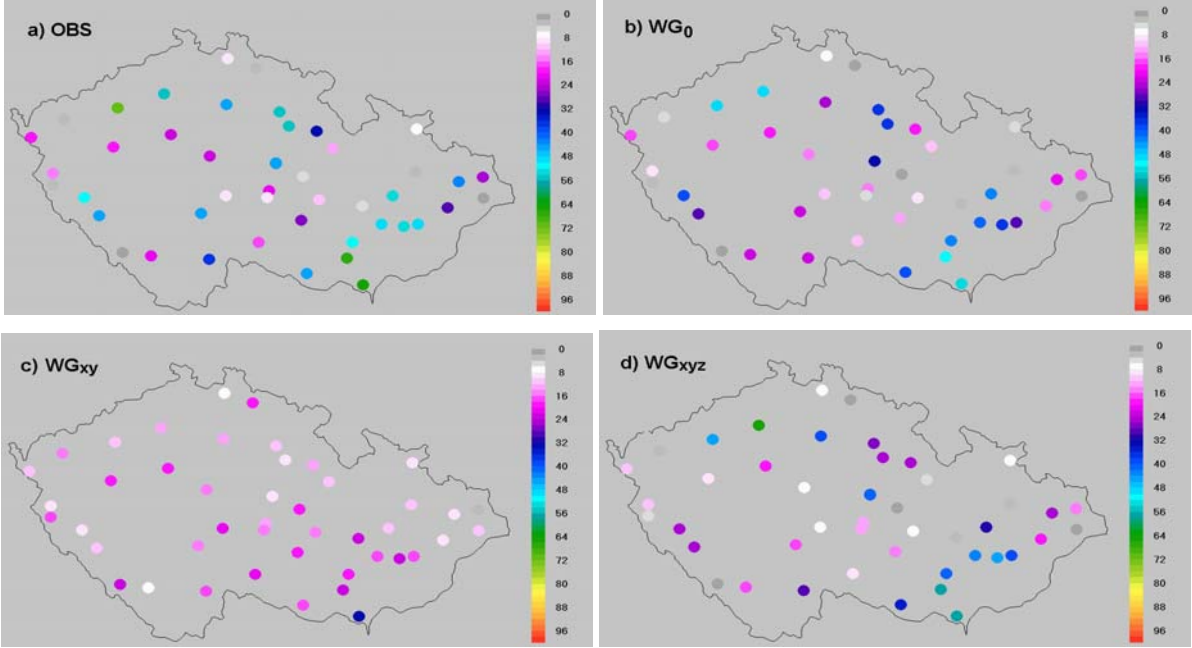


Figure 2 Number of heat waves in the 40-year weather series of four types (see Table 1 for specifications of the weather series).

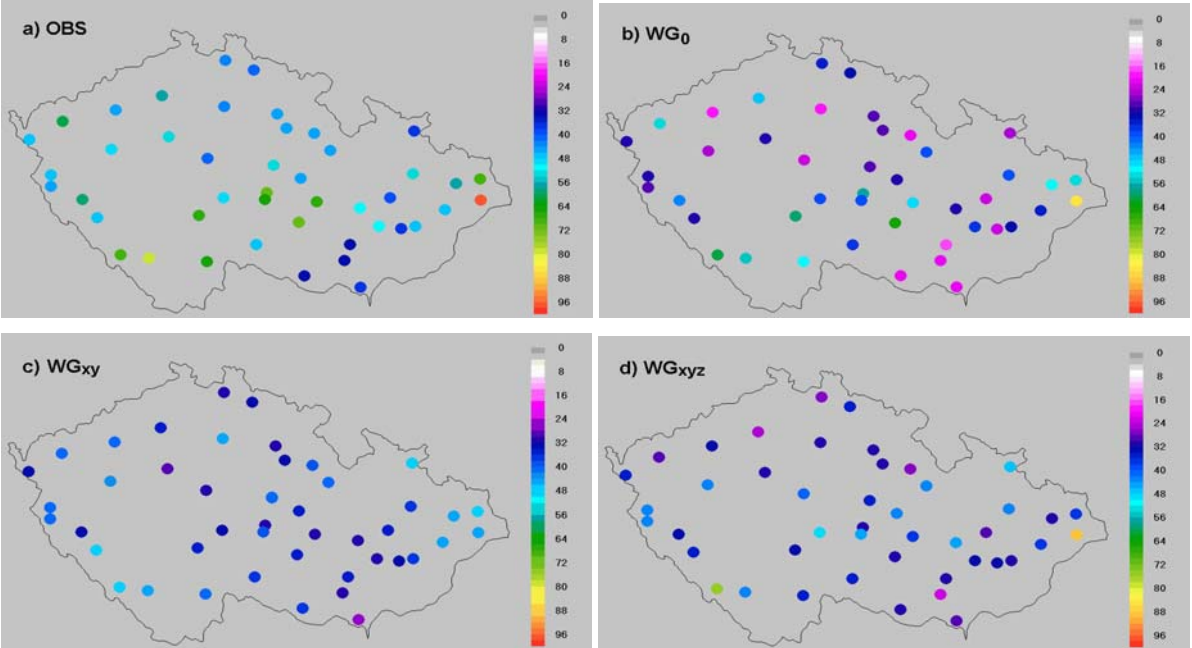
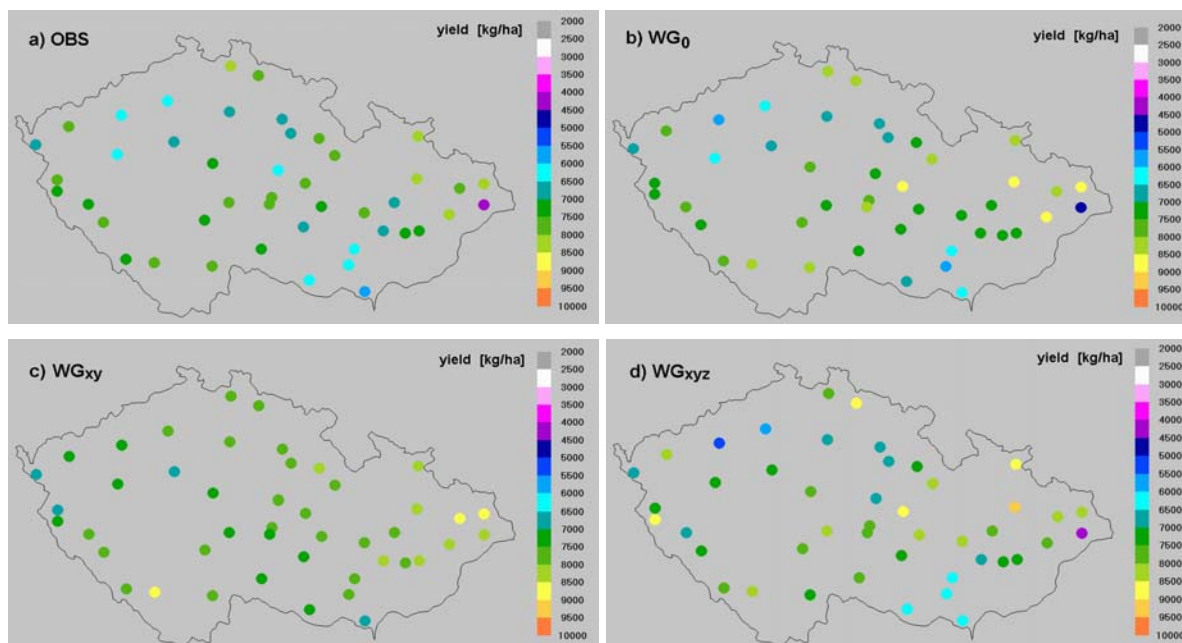


Figure 3 The same as previous figure but for the frequency of cold waves.

Ideally, the climatic characteristics should be the same in both observed weather series and weather series produced by the site-calibrated weather generator, except for the deviations due to stochasticity of the generator. In case of the perfect fit, the grey circles in Figure should follow the 1:1 line. In the latter test (compare panels C vs. B and D vs. B in Figures 2-4), results obtained with the interpolated generator are compared with those obtained with the site-calibrated generator. In this case, the differences show how the interpolation imperfections affect results obtained with the synthetic weather series. The perfect fit, which would be indicated by triangles lying on 1:1 line in Figure would mean that the interpolation error is either zero or it does not affect results obtained with the interpolated generator. Altogether, in case of the perfect performance of WG and interpolator, the mean characteristics obtained with observed weather series, synthetic series produced by site-calibrated weather generator, and synthetic series produced by the interpolated generator should be the same. In reality, however, the results obtained by individual weather series differ. Figures 2-5 together with Table 2 show:

**Table 2.** Comparison of characteristics (number of heat waves, number of cold waves, and wheat yields simulated by WOFOST crop growth model) obtained with the four weather series (see Table 1 for specifications) in terms of mean bias error (*MBE*), root mean square error (*RMSE*) and reduction of variance (*RV*).

| input weather series       | with respect to OBS |                 |               | with respect to WG <sub>0</sub> |                 |               |
|----------------------------|---------------------|-----------------|---------------|---------------------------------|-----------------|---------------|
|                            | <i>MBE</i> [%]      | <i>RMSE</i> [%] | <i>RV</i> [%] | <i>MBE</i> [%]                  | <i>RMSE</i> [%] | <i>RV</i> [%] |
| a) frequency of heat waves |                     |                 |               |                                 |                 |               |
| WG <sub>0</sub>            | -29                 | 38              | 71            |                                 |                 |               |
| WG <sub>xy</sub>           | -52                 | 90              | -61           | -32                             | 84              | -28           |
| WG <sub>xyz</sub>          | -29                 | 43              | 64            | -1                              | 32              | 82            |
| b) frequency of cold waves |                     |                 |               |                                 |                 |               |
| WG <sub>0</sub>            | -29                 | 31              | -47           |                                 |                 |               |
| WG <sub>xy</sub>           | -28                 | 37              | -113          | 2                               | 39              | 2             |
| WG <sub>xyz</sub>          | -28                 | 37              | -112          | 1                               | 37              | 12            |
| c) model wheat yields      |                     |                 |               |                                 |                 |               |
| WG <sub>0</sub>            | 3                   | 5               | 79            |                                 |                 |               |
| WG <sub>xy</sub>           | 7                   | 13              | -46           | 4                               | 12              | -8            |
| WG <sub>xyz</sub>          | 3                   | 8               | 43            | 1                               | 7               | 58            |



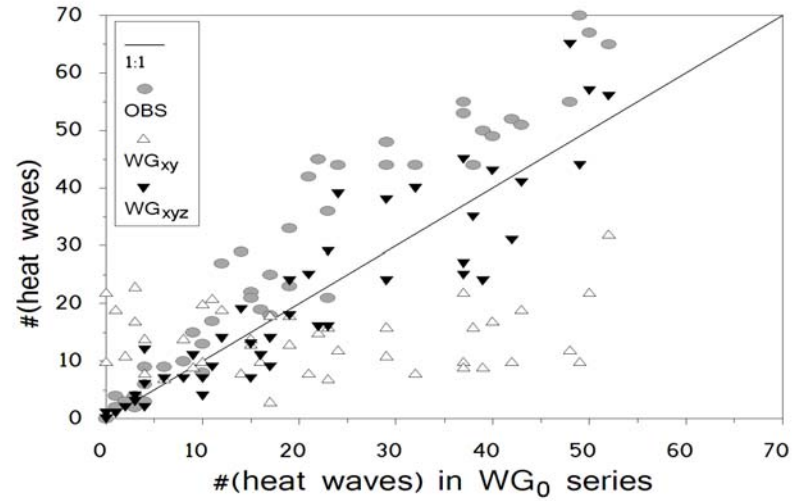
**Figure 4** Mean 39-year crop yields simulated by WOFOST crop growth model using the same weather series as those employed in Figures 2 and 3.

*a) Observed weather series vs. site-calibrated WG*

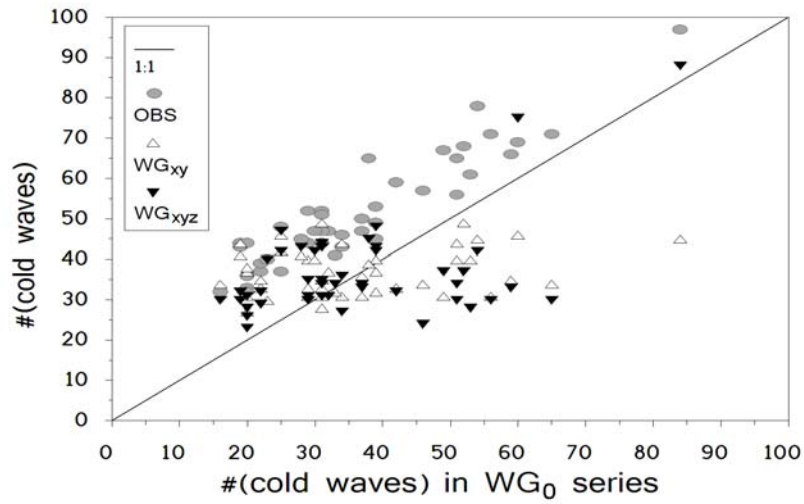
- Weather generator reproduces heat waves better than the cold waves (indicated by  $RV$  values in Table 1). However, frequencies of both heat waves and cold waves are underestimated by the generator (indicated by negative values of  $MBE$  in  $WG_0$  vs. OBS comparison).
- The good fit between mean model yields obtained with OBS and  $WG_0$  weather series (indicated by low values of  $MBE$  and  $RMSE$  and high value of  $RV$  in  $WG_0$  vs. OBS comparison) suggest applicability of WG in crop modelling: WOFOST model does not exhibit sensitivity to WG imperfections.

*b) Site-calibrated vs. interpolated WGs*

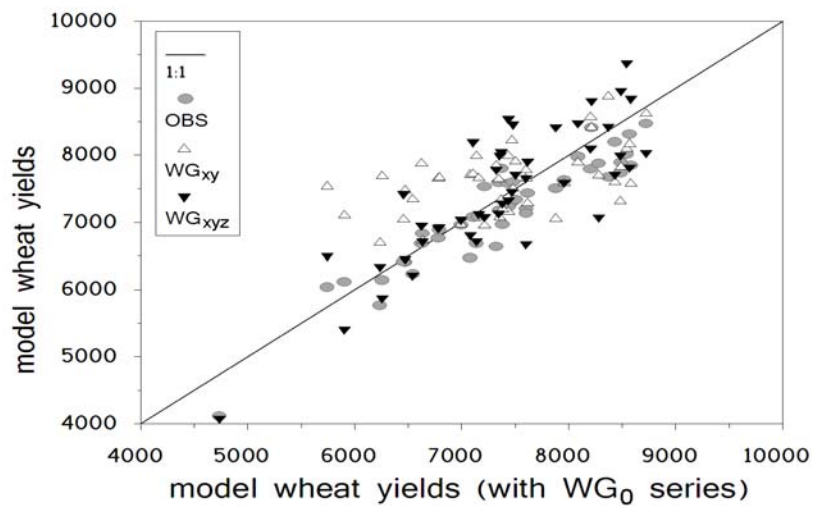
- The interpolator, which does not account for the altitude effect, fails. This is seen in figures (panels C vs. B in Figures 2-4; empty triangles in Figure 1) and indicated by low values of  $RV$  in  $WG_{xy}$  vs.  $WG_0$  comparison.
- In accounting for the altitude effect, frequencies of heat waves in the weather series produced by the interpolated generator ( $WG_{xyz}$  series) are comparable with those found in weather series produced by the site-calibrated generator. The values of  $RMSE$  indicate that the error due to interpolation ( $RMSE(WG_{xyz}$  vs.  $WG_0) = 32\%$ ) is slightly lower than the error due to weather generator imperfections ( $RMSE(WG_0$  vs. OBS) = 38%).
- Cold waves are poorly reproduced by the interpolated generator.
- Errors due to WG interpolation affect the model yields to a greater extent than the WG imperfections ( $RMSE(WG_{xyz}$  vs.  $WG_0) = 7\%$ ;  $RMSE(WG_0$  vs. OBS) = 5%).



a)



b)



c)

**Figure 5** Comparison of the station-specific (45 stations) characteristics obtained with the four types of weather series: a) number of heat waves, b) number of cold waves, c) wheat yields simulated by WOFOST crop growth model. ( $WG_0$  series is used as a reference).

#### 4. Conclusions

The results shown in this paper originate from the experiments made in the first year (2005) of the CaliM&Ro project. These have shown that the interpolated WG performs reasonably well for some climatic characteristics (heat waves, in our case) as well as for the wheat yields modelled by WOFOST crop model. The new experiments, which follow the first ones, are now being made using updated experiment settings:

- (i) 125 Czech stations with observed daily weather series are now available.
- (ii) More interpolation techniques is tested (including co-kriging and neural networks).
- (iii) More impact models (crop growth models and rainfall-runoff models) are fed by the observed and synthetic weather series.
- (iv) More characteristics (mean, variability, extremes) will be employed in validating WG and interpolators. In addition, statistical significance of differences between results obtained with different weather series will be assessed.


**Acknowledgement:** CaliM&Ro project is supported by the Grant Agency of the Czech Republic, project 205/05/2265.

#### References

- Buchtele J., Buchtelova M., Fortova, M., Dubrovsky M., 1999: Runoff changes in Czech river Basins - the outputs of rainfall - runoff simulations using different climate change scenarios. *Journal of Hydrology and Hydromechanics* **47**, 180-194.
- Dubrovsky M., 1997: Creating Daily Weather Series With Use of the Weather Generator. *Environmetrics* **8**, 409-424.
- Dubrovsky M., Zalud Z. and Stastna M., 2000: Sensitivity of CERES-Maize yields to statistical structure of daily weather series. *Climatic Change* **46**, 447- 472.
- Dubrovsky M., Buchtele J., Zalud Z., 2004: High-Frequency and Low-Frequency Variability in Stochastic Daily Weather Generator and Its Effect on Agricultural and Hydrologic Modelling. *Climatic Change* **63**, 145-179.
- Huth R., Kysely J., Pokorna L., 2000: A GCM simulation of heatwaves, dry spells, and their relationships to circulation. *Climatic Change* **46**, 29–60.
- Huth R., Kysely J., Dubrovsky M., 2001: Time structure of observed, GCM-simulated, downscaled, and stochastically generated daily temperature series. *Journal of Climate* **14**, 4047-4061.
- Huth R., Kysely J., Dubrovsky M., 2003: Simulation of Surface Air Temperature by GCMs, Statistical Downscaling and Weather Generator: Higher-Order Statistical Moments. *Studia Geophysica et Geodaetica* **47**, 203-216.
- Kysely J., Dubrovsky M., 2005: Simulation of extreme temperature events by a stochastic weather generator: effects of interdiurnal and interannual variability reproduction. *Int.J.Climatol.* **25**, 251-269.
- Trnka M., Dubrovsky M., Semeradova D., Zalud Z., 2004: Projections of uncertainties in climate change scenarios into expected winter wheat yields. *Theoretical and Applied Climatology* **77**, 229-249.

Trnka M., Dubrovsky M., Zalud Z., 2004: Climate Change Impacts and Adaptation Strategies in Spring Barley Production in the Czech Republic. *Climatic Change* **64**, 227-255.

Trnka M., Zalud Z., Eitzinger J., Dubrovsky M., 2005: Global solar radiation in Central European lowland estimated by various empirical formulae. *Agricultural and Forest Meteorology* **131**, 54-76.

Zalud Z., Dubrovsky M., 2002: Modelling climate change impacts on maize growth and development in the Czech  ublic. *Theoretical and Applied Climatology* **72**, 85-102.