Reproduction of extreme temperature and precipitation events by two stochastic weather generators

Martin Dubrovský and Jan Kyselý

Institute of Atmospheric Physics ASCR, Prague, Czechia
(dub@ufa.cas.cz, honza@ufa.cas.cz)

Acknowledgements: The present study is supported by the Grant Agency of the Czech Academy of Sciences (project IAA3017301) and the Grant Agency of the Czech Republic (205/06/1535)

Abstract. Stochastic weather generators are often used to provide synthetic weather series for hydrological and agricultural studies [3,5,9,10]. The quality of the weather generator may be measured by its ability to reproduce various climatic characteristics. These should include extreme events, which are commonly misreproduced by the present-day generators. This contribution assesses ability of two different stochastic daily weather generators to reproduce selected extreme temperature and precipitation events. The first generator is Met&Roll [1,8], which represents a parametric approach. The second weather generator, GeNNeR, is based on a nearest neighbours resampling technique, which is a non-parametric method making no assumption on the distribution of the variables being generated. In validating the generators, the characteristics of the extreme temperature and precipitation events derived from the synthetic weather series are compared with those derived from the observed weather series. The tests are made using observed weather series from 8 European stations.

Met&Roll

is a single-site 4-variate stochastic daily weather generator.

variables: - precipitation amount (PREC)
- daily temperature maximum (TMAX)
- daily temperature minimum (TMIN)
- global solar radiation (SRAD)

model: - precipitation occurrence ~ Markov chain
- precipitation amount ~ Gamma distribution
- (SRAD, TMAX, TMIN) ~ 3-variate 1st order AR model.

Various versions of Met&Roll were examined recently [8,11]. Three features were under focus, and are also tested in this experiment:

1) inclusion of the annual cycle of lag-0 and lag-1 correlations among SRAD, TMAX and TMIN (Table I: AC = YES/NO)
2) order of the Markov chain (Table I: Markov = 1 or 3)
3) daily weather generator is linked with 1st order autoregressive monthly generator (Table I: Monthly = Y/N)

Table I. Versions of Met&Roll examined in the experiments.
(those shown in figures are highlighted by yellow color)

acronym	AC	Markov	Monthly
MA *("basic" version)*	N	1	N
MB	Y	1	N
MC	N	3	N
MD	Y	3	N
ME	N	1	Y
MF	Y	1	Y
MG	N	3	Y
MH *(should be the best)*	Y	3	Y

GeNNeR

is a generator based on a nearest neighbours resampling procedure. Its development started last year. There are many features of the algorithm which might affect the quality of the generator and are presently a subject of optimisation. Generally, the **procedure is**:

(i) First term is selected randomly from those terms in the observed series whose Julian day (JD) is within ±Δ*D* days (Δ*D*=10 by default) from January 1st

(ii) having *n*-th term, ***X***(*n*), (n+1)-th term is selected:

- *K* nearest neighbors of ***X***(n) (using Mahalanobis distance to measure between-terms distance) are found in the learning series. These neighbours must satisfy conditions: (a) |JD-JD(n)|≤ Δ*D* days), (b) (optionally) the neighbours must be with the same precipitation category as ***X***(*n*)
- ***X***(*n*+1) is generated from this set. **Several algorithms** were considered, ***X***(*n*+1) =

  (**A**) average of *K* neighbours (= **AVG**)
  (**B**) randomly selected one of the *K* neighbours (= ***RND***)
  (**C**) **AVG** + C × ***STD*** × RANDOM(-1,1); where ***STD*** is a vector of standard deviations of individual variables, C is coefficient
  (**D**) ***RND*** + C × **STD** × RAND(-1,1)
  (**E**) ***AVG*** + C × ***STD*** × ERF^{-1}(RAND(-1,1))
  (**F**) ***RND*** + C × ***STD*** × ERF^{-1}(RAND(-1,1))
  (**G**) ***RND*** + (1+C) × (***RND'*** - ***RND***)

(note: The versions without the random term /i.e. *x*A and *x*B/ cannot produce extremes which are beyond the range of values found in observed data → problem: the extreme derived from arbitrarily long synthetic series will not exceed extreme of the observed series, from which the synthetic series is resampled).

In present experiments, we varied 4 parameters of the resampling algorithm (see Table II for the specifications of tested versions of GeNNeR):

PT: while finding neighbours, precipitation is considered either binary variable (PT=B) or a real-number variable (PT=B)
CP (conditioning on precipitation): while searching for the neighbours, only terms with precipitation parameter being the same as .... are considered
St: generation strategy
C: value of C coefficient

Table II. Versions of the resampling algorithm examined in the experiments. (those shown in figures are highlighted by yellow color)

acro-nym	PT	CP	St	C	acro nym	PT	CP	St	C	acro nym	PT	CP	St	C
AA	B	N	A	1	**BA**	B	Y	A	1	**CA**	R	N	A	1
AB	B	N	B	1	**BB**	B	Y	B	1	**CB**	R	N	B	1
AC	B	N	C	1	**BC**	B	Y	C	1	**CC**	R	N	C	1
AD	B	N	D	1	**BD**	B	Y	D	1	**CD**	R	N	D	1
AE	B	N	E	1	**BE**	B	Y	E	1	**CE**	R	N	E	1
AF	B	N	F	1	**BF**	B	Y	F	1	**CF**	R	N	F	1
AG	B	N	G	1	**BG**	B	Y	G	1	**CG**	R	N	G	1
DC	B	N	C	0.5	**EC**	B	Y	C	0.5	**FC**	R	N	C	0.5
DD	B	N	D	0.5	**ED**	B	Y	D	0.5	**FD**	R	N	D	0.5
DE	B	N	E	0.5	**EE**	B	Y	E	0.5	**FE**	R	N	E	0.5
DF	B	N	F	0.5	**EF**	B	Y	F	0.5	**FF**	R	N	F	0.5
DG	B	N	G	0.5	**EG**	B	Y	G	0.5	**FG**	R	N	G	0.5

other parameters of the algorithm, which might be subject of optimisation in future experiments:

- higher order of the algorithm (more recent terms might be used while searching for the neighbours)
- larger number of precipitation categories (for precipitation-conditional algorithm)
- ΔD (half-width of the date-based neighbourhood)
- K (number of nearest neighbours)
- possible transformation of some variables (solar radiation, precipitation; value→rank transformation might be considered)

Data

Table III. Daily weather series were taken from the "European Climate Assessment" project (http://eca.knmi.nl):

acronym	station	state	LAT	LONG	ALT [m]
BAMBE	BAMBERG	DE	+49:53:00	+10:53:00	282
HOHEN	Hohenpeissenberg	DE	+47:48:00	+11:01:00	977
PRAHA	Praha	CZ	+50:05:27	+14:25:09	191
SALAM	Salamanca	ES	+40:56:50	-05:28:19	790
SMOLE	Smolensk	RU	+54:45:00	+32:04:00	239
SODAN	Sodankyla	FI	+67:22:00	+26:39:00	179
VALEN	Valencia	ES	+39:28:48	-00:21:08	11
ZUGSP	Zugspitze	DE	+47:25:00	+10:59:00	2960

Present experiment

For each station, ten 30-years series was generated using 8 versions of Met&Roll (Table 1) and 36 versions of GeNNeR (Table 2). Statistics from the synthetic series (10×30 years) are then compared with those derived from the learning series.

validation characteristics

a) temperature

max(TMAX) = annual/30-year maximum of TMAX
min(TMIN) = annual/30-year minimum of TMIN
max(ΔTMAX↑) = annual/30-year max. of interdiurnal growth in TMAX
max(ΔTMIN↓) = annual/30-year max. of interdiurnal drop in TMIN
Lmax(Heat) = annual/30-year maximum length of heat spell*
Lmax(Cold) = annual/30-year maximum length of cold spell*

*where

heat spell = continuous period with TMAX>avg(TMAX) + 1.645×std(TMAX)
cold spell = continuous period with TMIN<avg(TMIN) − 1.645×std(TMIN)

(avg(*) and std(*) are climatological means and standard deviations of TMAX and TMIN for a given day of the year; smoothed by robust locally weighted regression. note: 1.646 correspond to 95th percentile of standard normal distribution

b) precipitation

max(PREC) = annual/30-year maximum of PREC
max(5-day PREC) = annual/30-year maximum of 5-day precipitation
Lmax(Dry) = annual/30-year maximum length of dry spell*
Lmax(Wet) = annual/30-year maximum length of wet spell*

where **dry spell** = continuous period with PREC<0.1 mm
wet spell = continuous period with PREC≥0.1 mm

Conclusions

a) GeNNeR

- Precipitation extremes (both 1-day and 5-day) are best reproduced by FD version (not shown in the figures), which employs the random term and uses precipitation amount as a predictor.

- Inclusion of the random term has not so definite effect on reproduction of the other extreme climatic characteristics; better specification of the random term will hopefully help.

- conditioning the search for the neighbours on precipitation occurrence (versions B*x* and E*x*) or using the precipitation amount (C*x* and F*x*) as a predictor perform similarly well.

b) Met&Roll

- effects of the three modifications (with respect to the basic version of Met&Roll) on reproduction of most extreme characteristics [e.g. max(TMAX) and min(TMIN)] is only slight.

- 3rd order Markov chain improves reproduction of the dry spells (but not the wet spells)

- inclusion of the monthly generator significantly affects precipitation. It improves reproduction of the 5-day precipitation extremes, but the effect on extreme 1day precipitation sum is not so definite.

c) most of the extreme characteristics analysed here are reproduced better by GeNNeR than by Met&Roll

future:

To assess performance of the two generators, we will use:

- quantitative characteristics for comparison of observed vs synthetic weather series
- more stations from ECA&D database

References

[1] Dubrovský, M., 1997: 'Creating Daily Weather Series with Use of the Weather Generator', *Environmetrics* **8**, 409-424.

[2] Buchtele, J., Buchtelová, M., Fořtová, M., and Dubrovský, M., 1999: 'Runoff Changes in Czech River Basins - the Outputs of Rainfall-Runoff Simulations Using Different Climate Change Scenarios', *J. Hydrol. Hydromech.*, **47,** 180-194.

[3] Dubrovský M., Zalud Z. and Stastna M., 2000: "Sensitivity of CERES-Maize yields to statistical structure of daily weather series". *Climatic Change* **46**, 447- 472.

[4] Huth R., Kyselý J., Dubrovský M., 2001: "Time structure of observed, GCM-simulated, downscaled, and stochastically generated daily temperature series". *Journal of Climate*, **14**, 4047-4061.

[5] Žalud Z., Dubrovský M., 2002: "Modelling climate change impacts on maize growth and development in the Czech republic". *Theoretical and Applied Climatology*, **72**, 85-102.

[6] Huth R., Kyselý J., Dubrovský 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

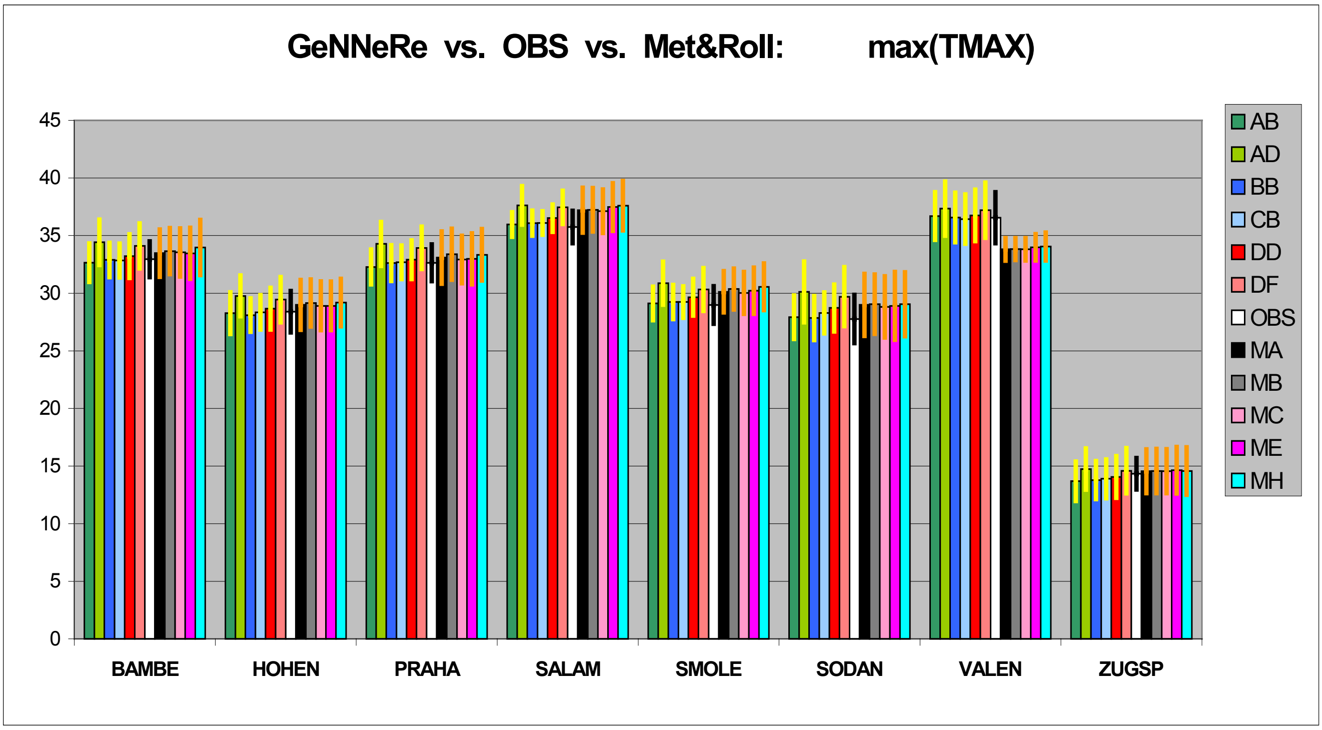
[7] Hejzlar J., Dubrovský M., Buchtele J. and Růžička M., 2003: "The effect of climate change on the concentration of dissolved organic matter in a temperate stream (the Malše River, South Bohemia)". *Science of Total Environment* **310** *143-152.*

[8] Dubrovský M., Buchtele J., Žalud Z., 2004: High-Frequency and Low-Frequency Variability in Stochastic Daily Weather Generator and Its Effect on Agricultural and Hydrologic Modelling. *Climatic Change* **63** (No.1-2), 145-179.

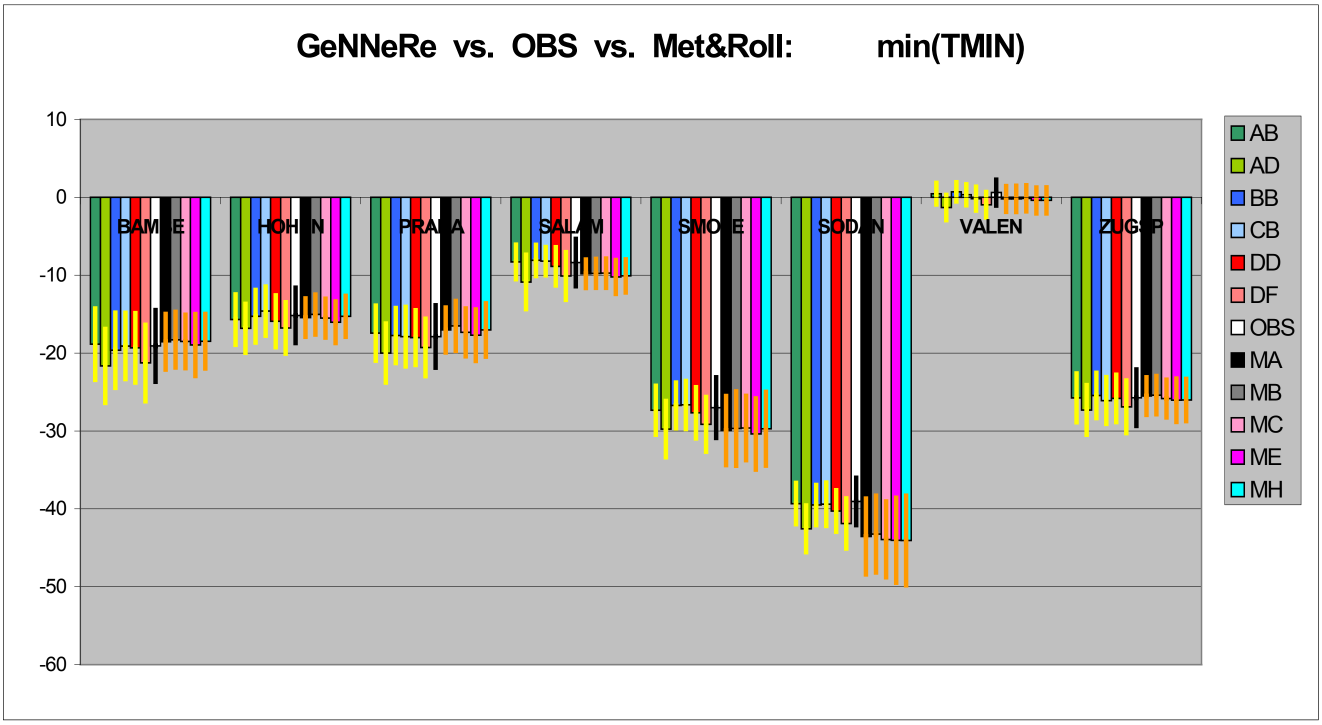
[9] Trnka M., Dubrovský M., Semerádova D., Žalud Z., 2004: Projections of uncertainties in climate change scenarios into expected winter wheat yields. *Theoretical and Applied Climatology*, **77**, 229-249.

[10] Trnka M., Dubrovský M., Žalud Z., 2004: Climate Change Impacts and Adaptation Strategies in Spring Barley Production in the Czech Republic. *Climatic Change* **64** (No. 1-2), 227-255.

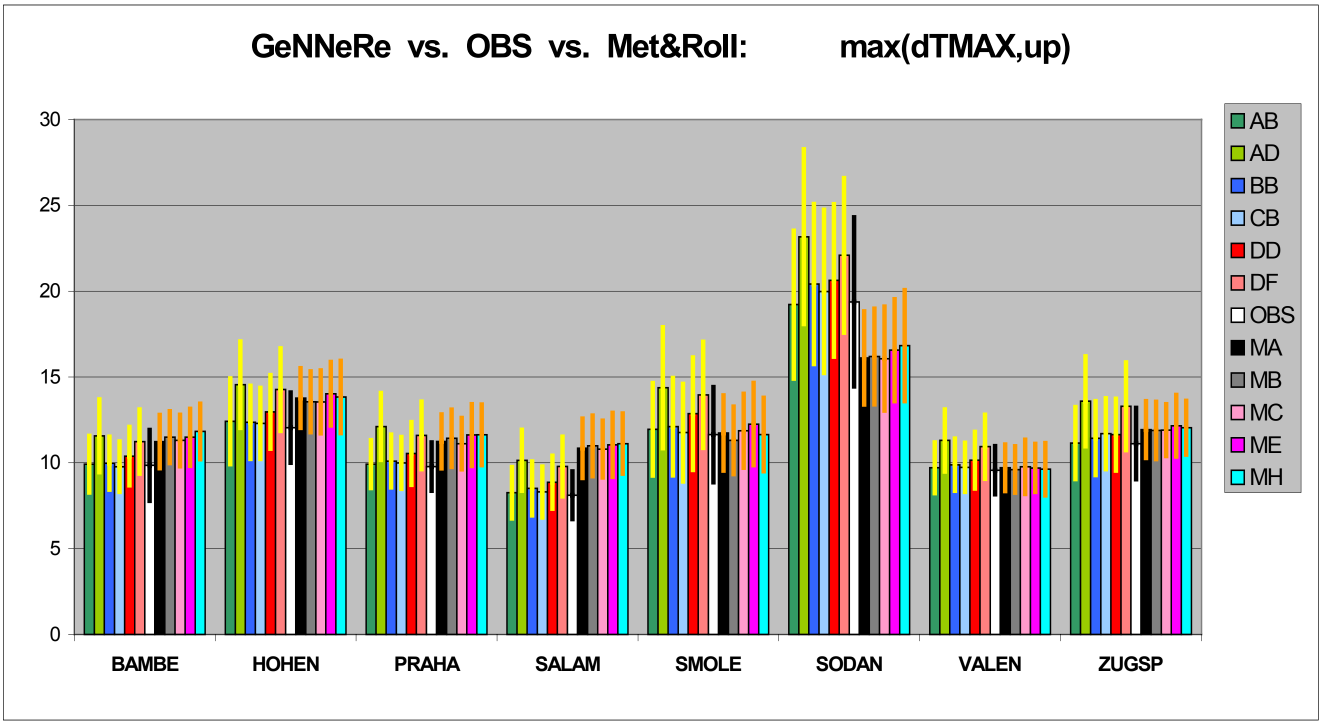
[11] Kyselý J., Dubrovský M., 2005: Simulation of extreme temperature events by a stochastic weather generator: effects of interdiurnal and interannual variability reproduction. *Int.J.Climatol.* **25**, 251-269.



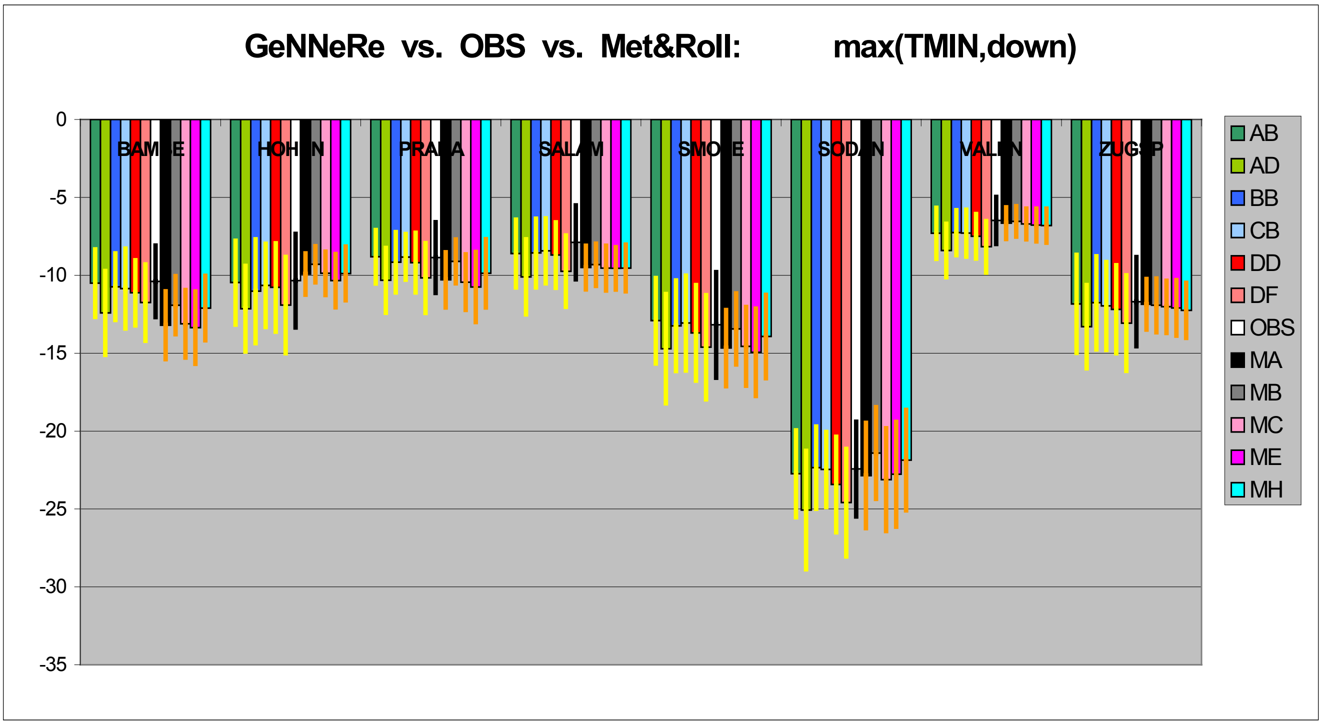
Annual maximum of TMAX (avg±std from 30 years /OBS/ or 300 years /GeNNeR and Met&Roll/)



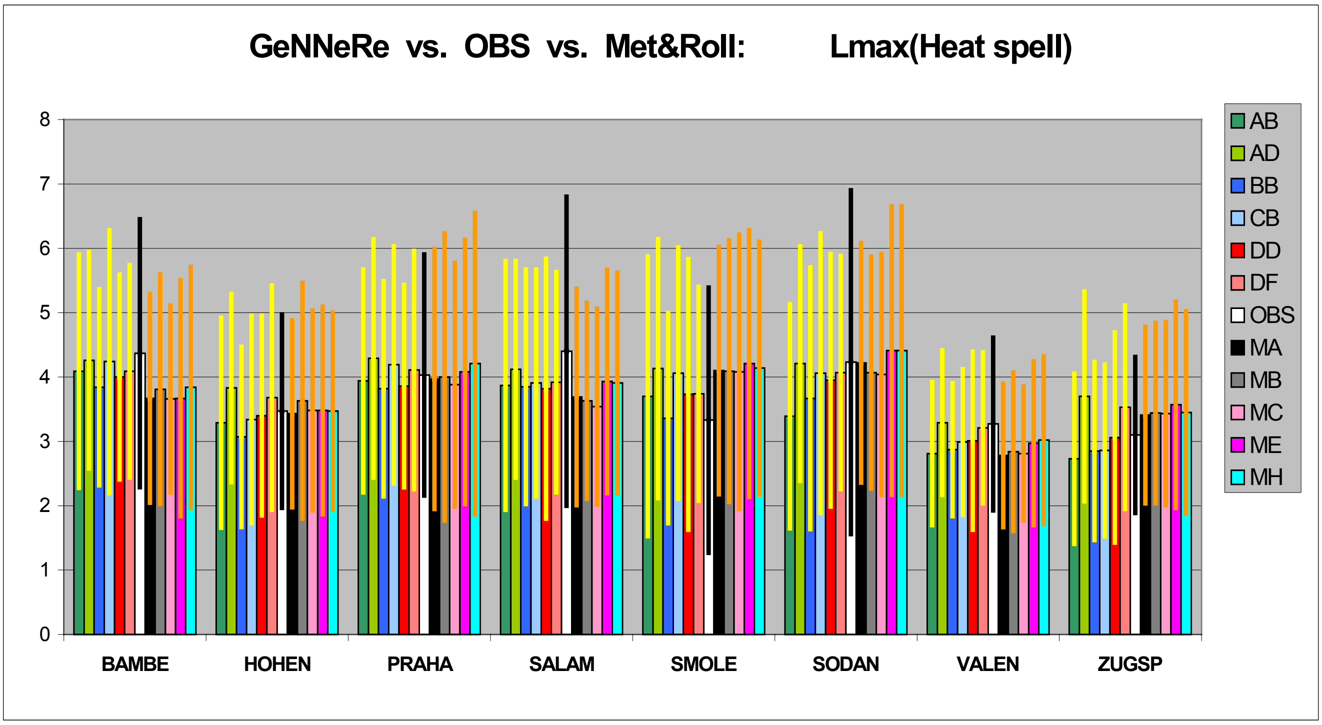
Annual minimum of TMIN (avg±std from 30 years /OBS/ or 300 years /GeNNeR and Met&Roll/)



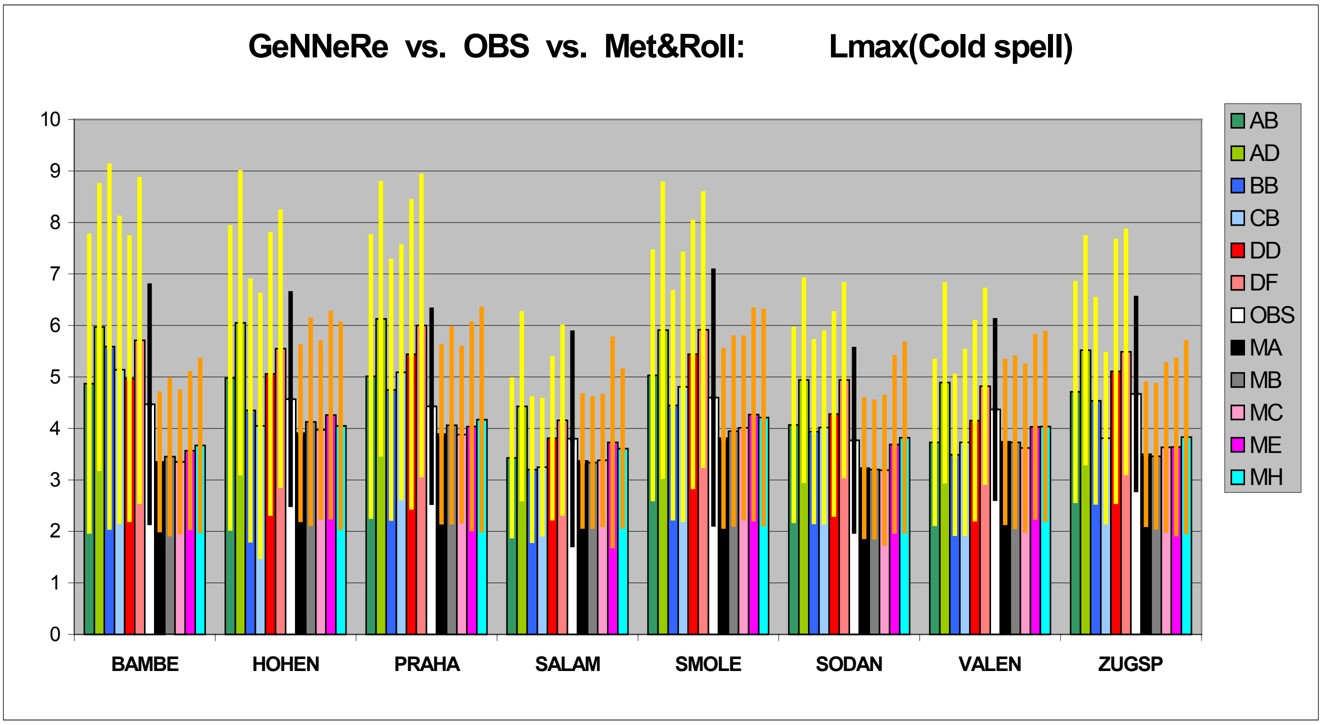
Annual maximum of interdiurnal growth in TMAX (avg±std from 30 years /OBS/ or 300 years /GeNNeR and Met&Roll/)



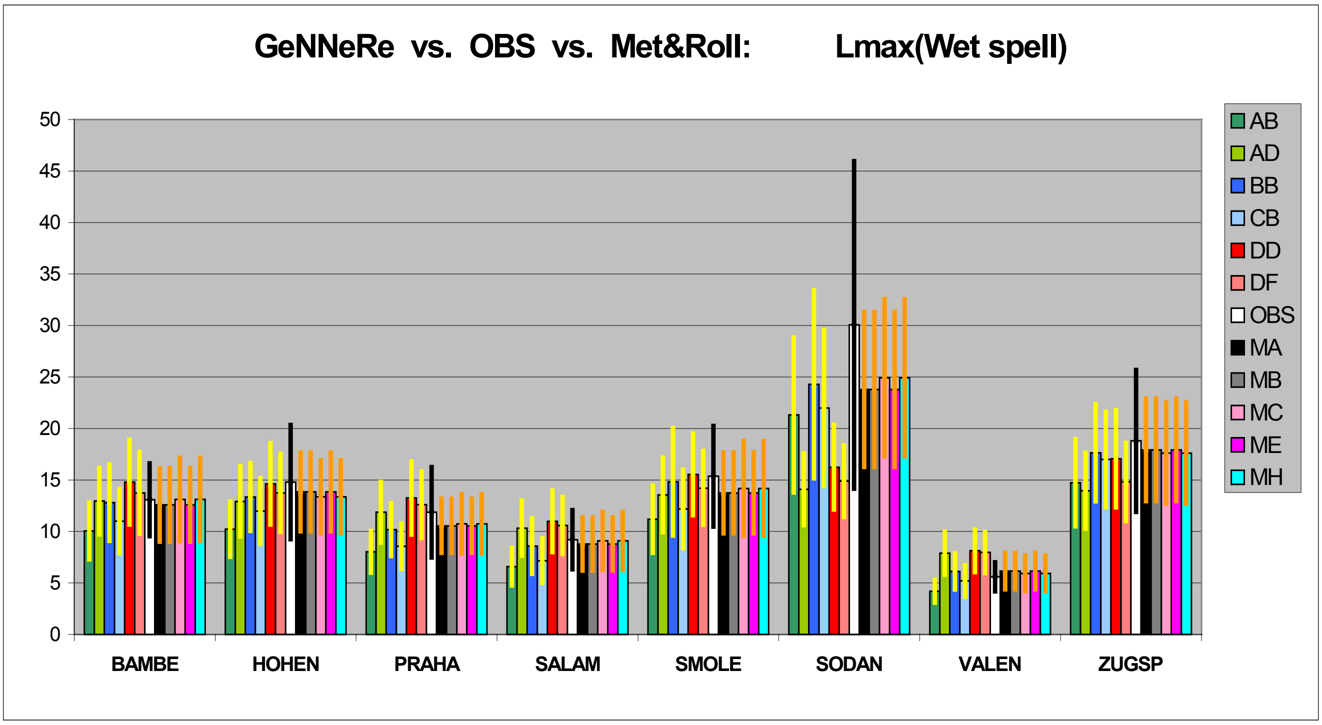
Annual maximum of interdiurnal drop in TMIN (avg±std from 30 years /OBS/ or 300 years /GeNNeR and Met&Roll/)



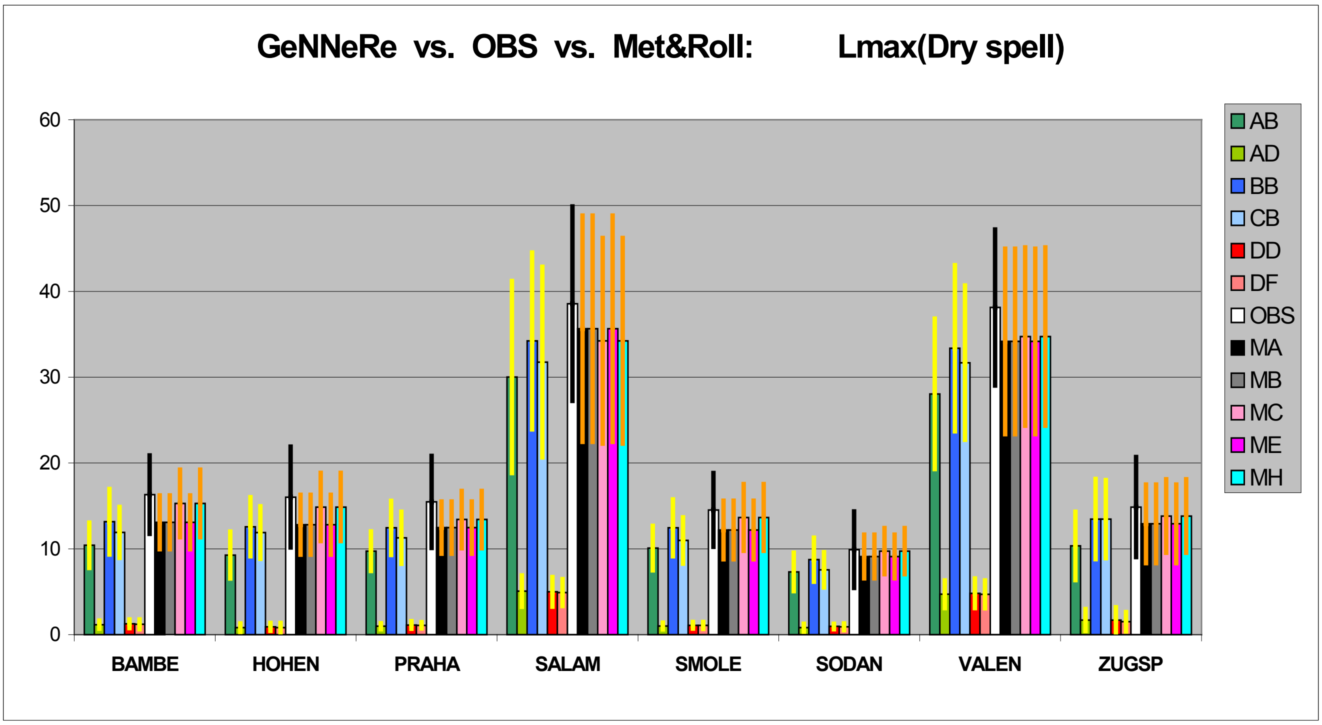
Annual maximum length of heat spell (avg±std from 30 years /OBS/ or 300 years /GeNNeR and Met&Roll/)



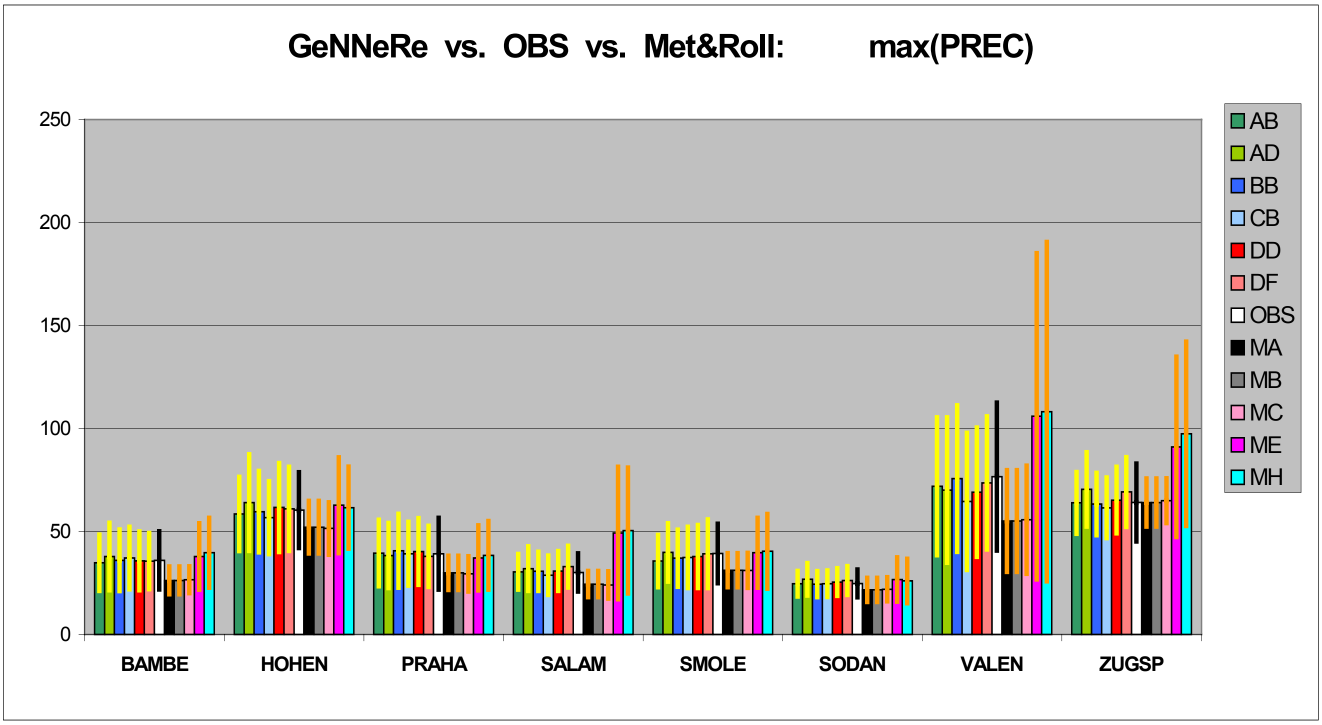
Annual maximum length of cold spell (avg±std from 30 years /OBS/ or 300 years /GeNNeR and Met&Roll/)



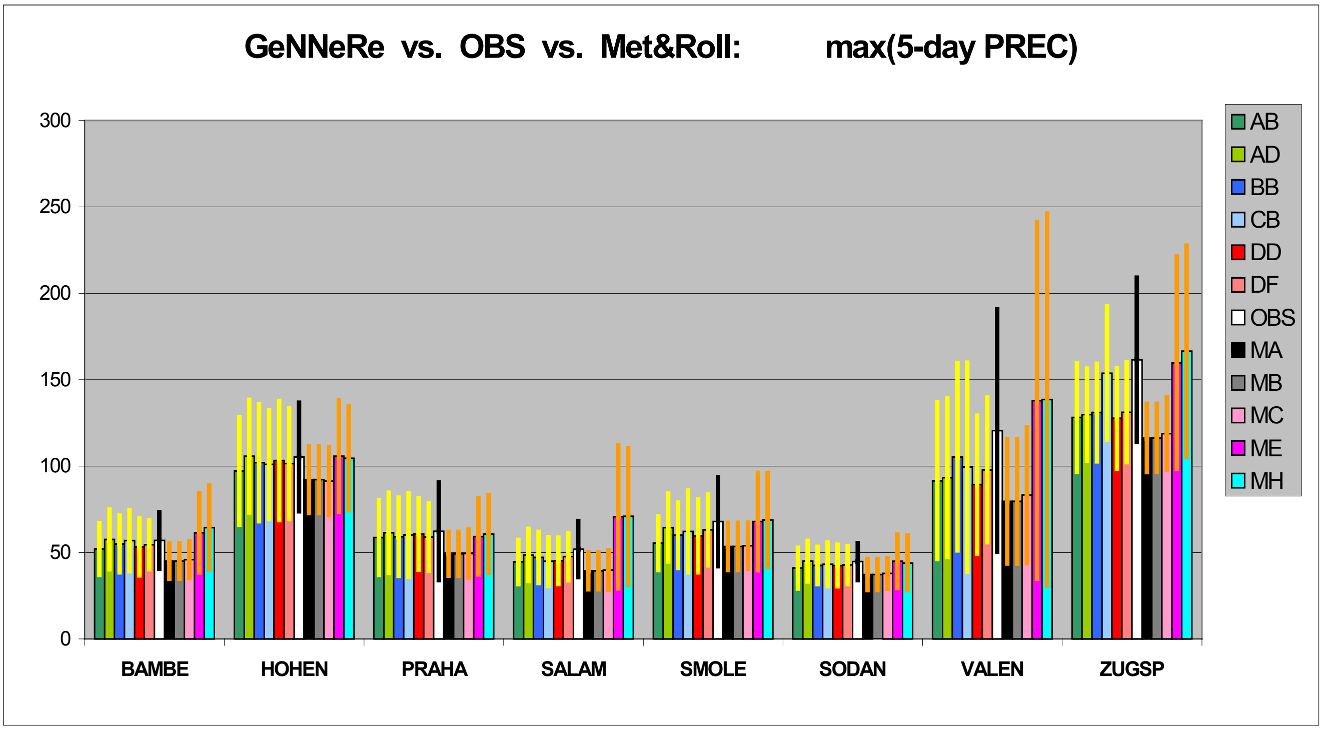
Annual maximum length of dry spell (avg±std from 30 years /OBS/ or 300 years /GeNNeR and Met&Roll/)



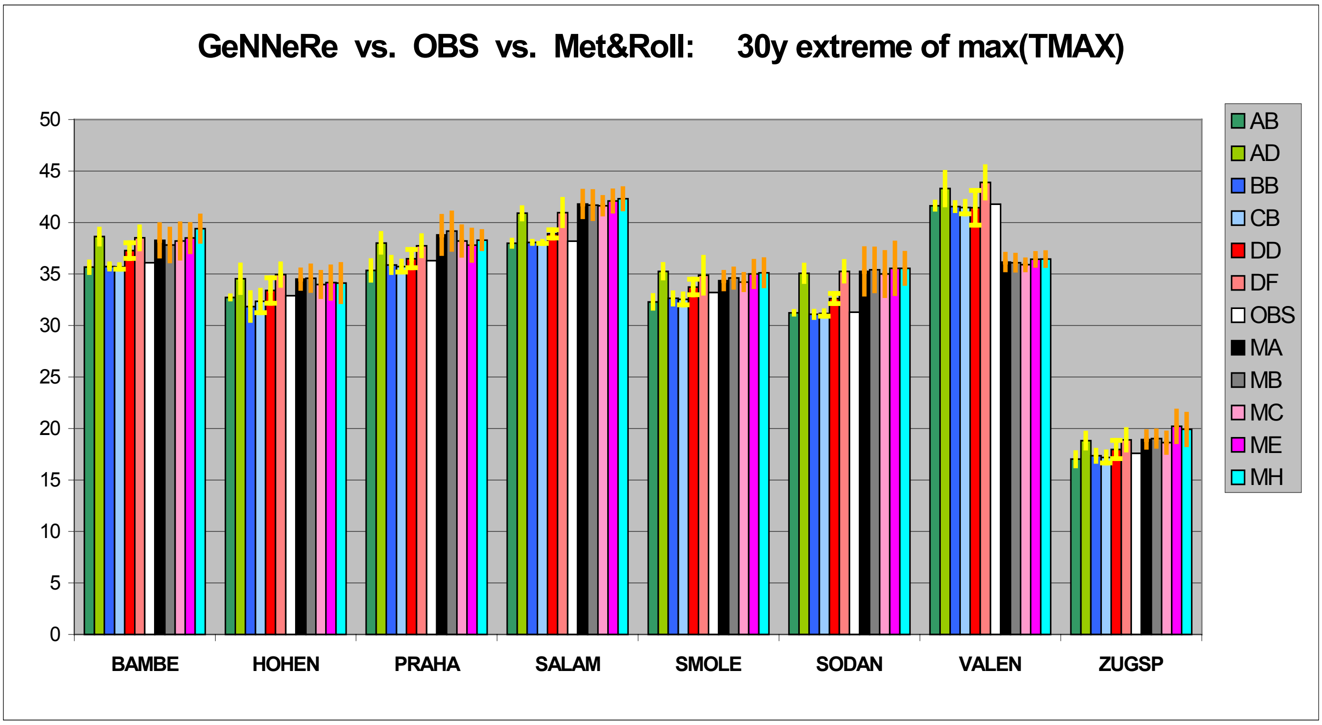
Annual maximum length of wet spell (avg±std from 30 years /OBS/ or 300 years /GeNNeR and Met&Roll/)



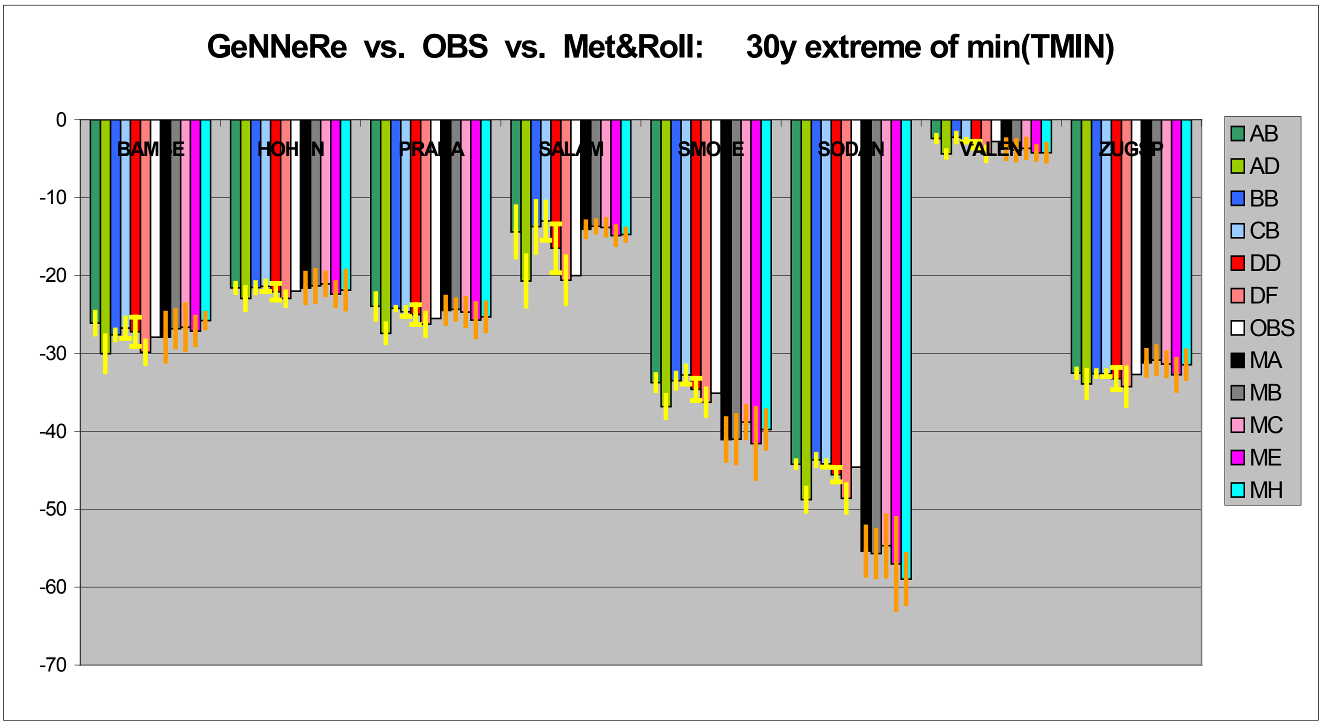
Annual maximum daily precipitation (avg±std from 30 years /OBS/ or 300 years /GeNNeR and Met&Roll/)



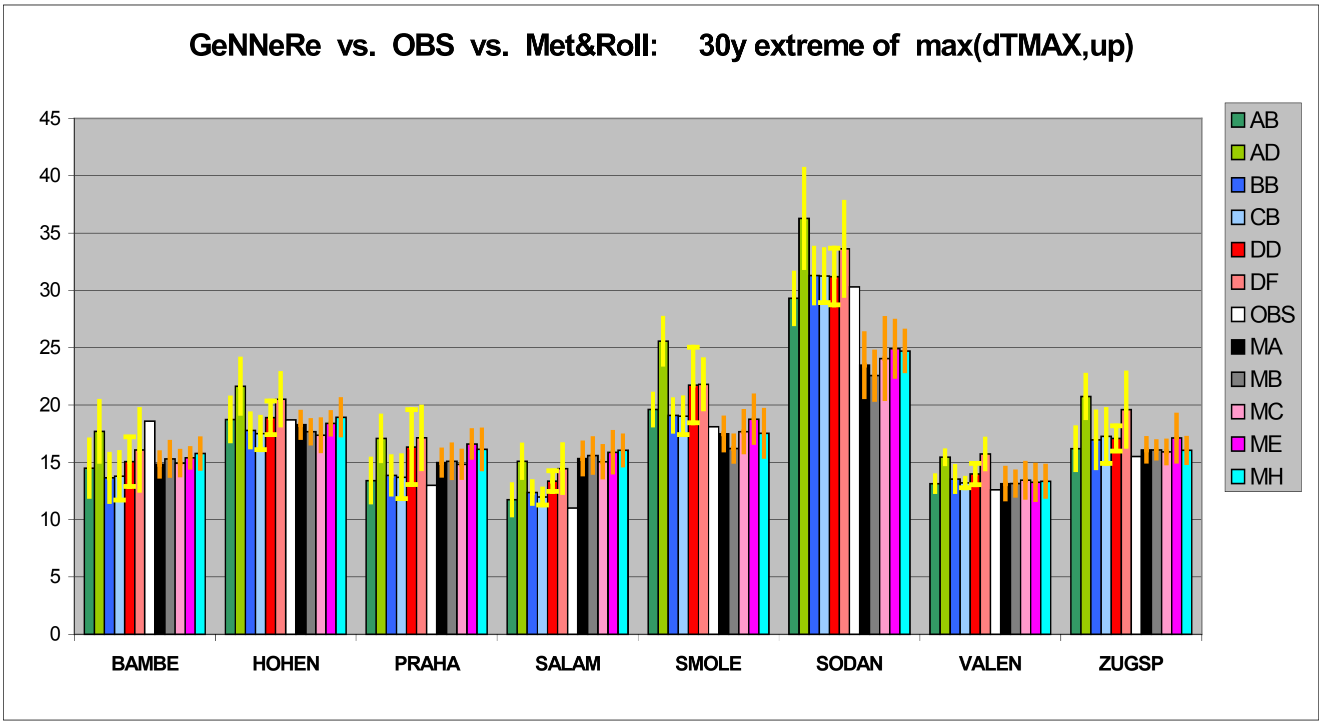
Annual maximum 5-day precipitation (avg±std from 30 years /OBS/ or 300 years /GeNNeR and Met&Roll/)



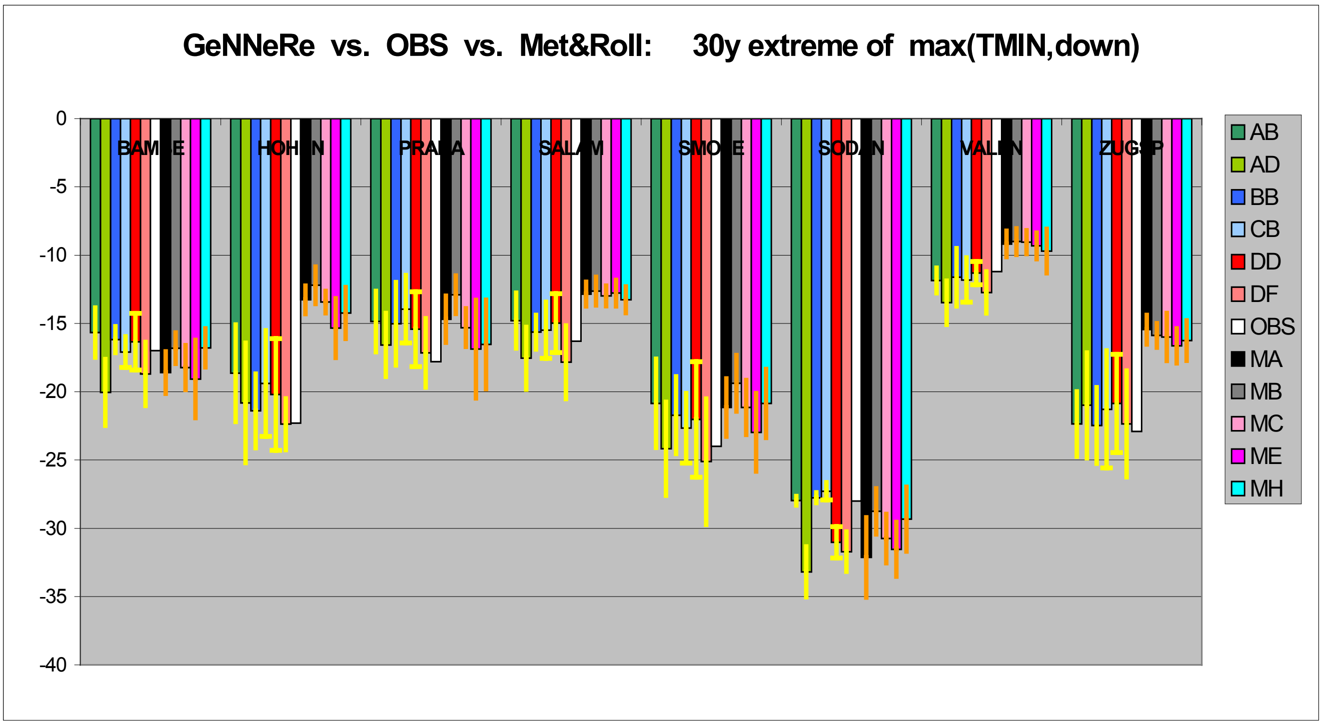
30-year extreme of TMAX (standard error is based on ten 30y series in case of synthetic weather series generated by GeNNeR and Met&Roll)



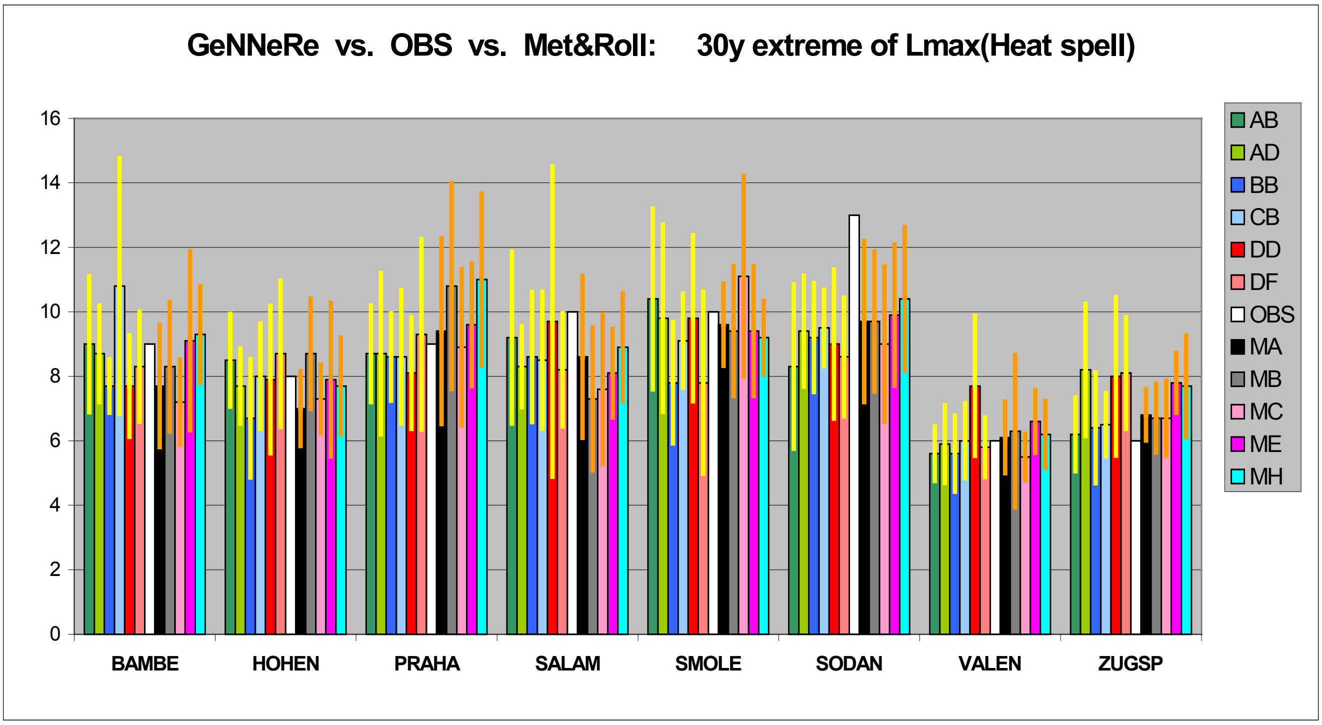
30-year extreme of TMIN (standard error is based on ten 30y series in case of synthetic weather series generated by GeNNeR and Met&Roll)



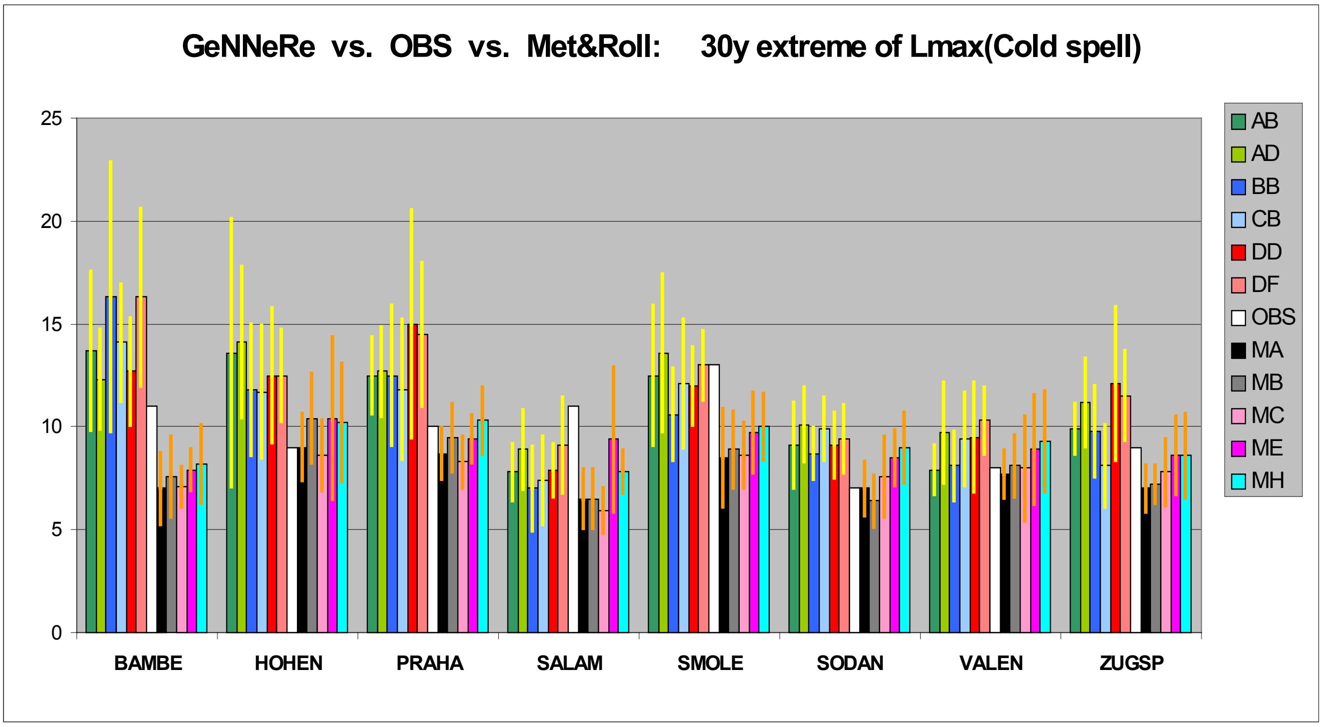
30-year maximum of interdiurnal growth in TMAX (std. error is based on ten 30y series in case of synthetic weather series generated by GeNNeR and Met&Roll)



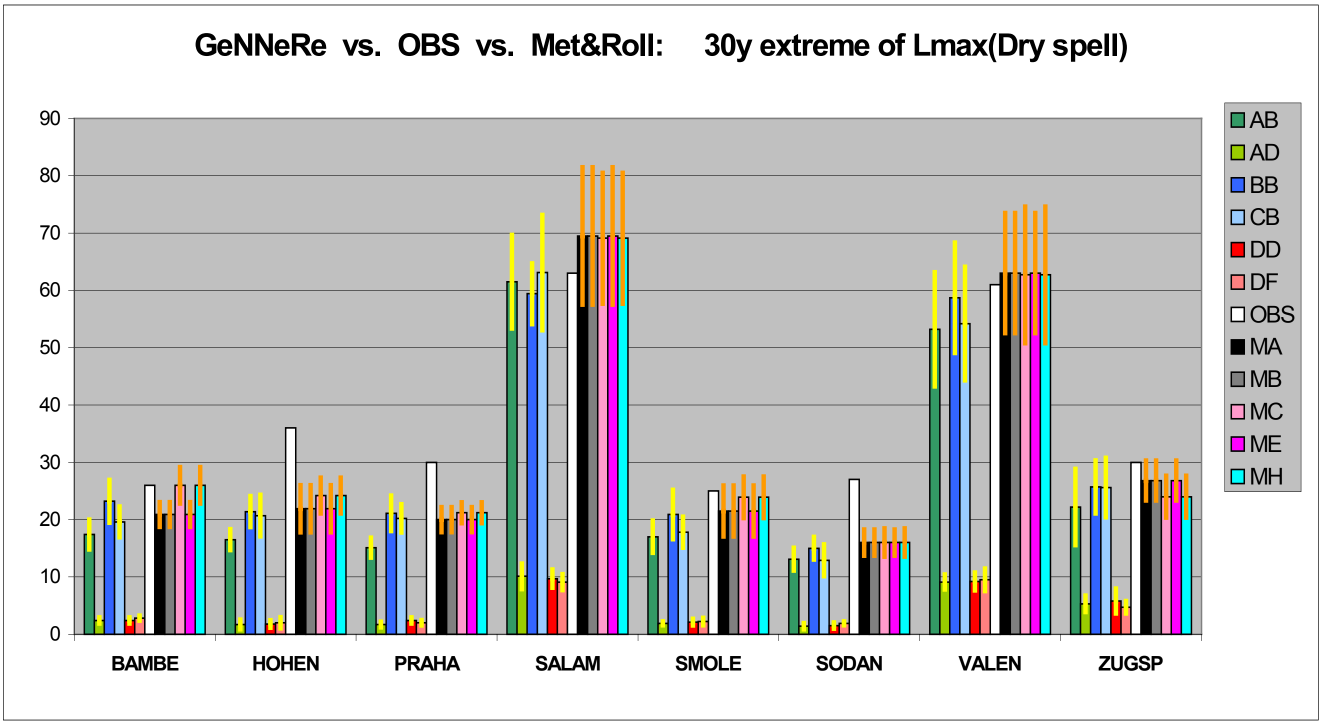
30-year extreme of interdiurnal drop in TMIN (std. error is based on ten 30y series in case of synthetic weather series generated by GeNNeR and Met&Roll)



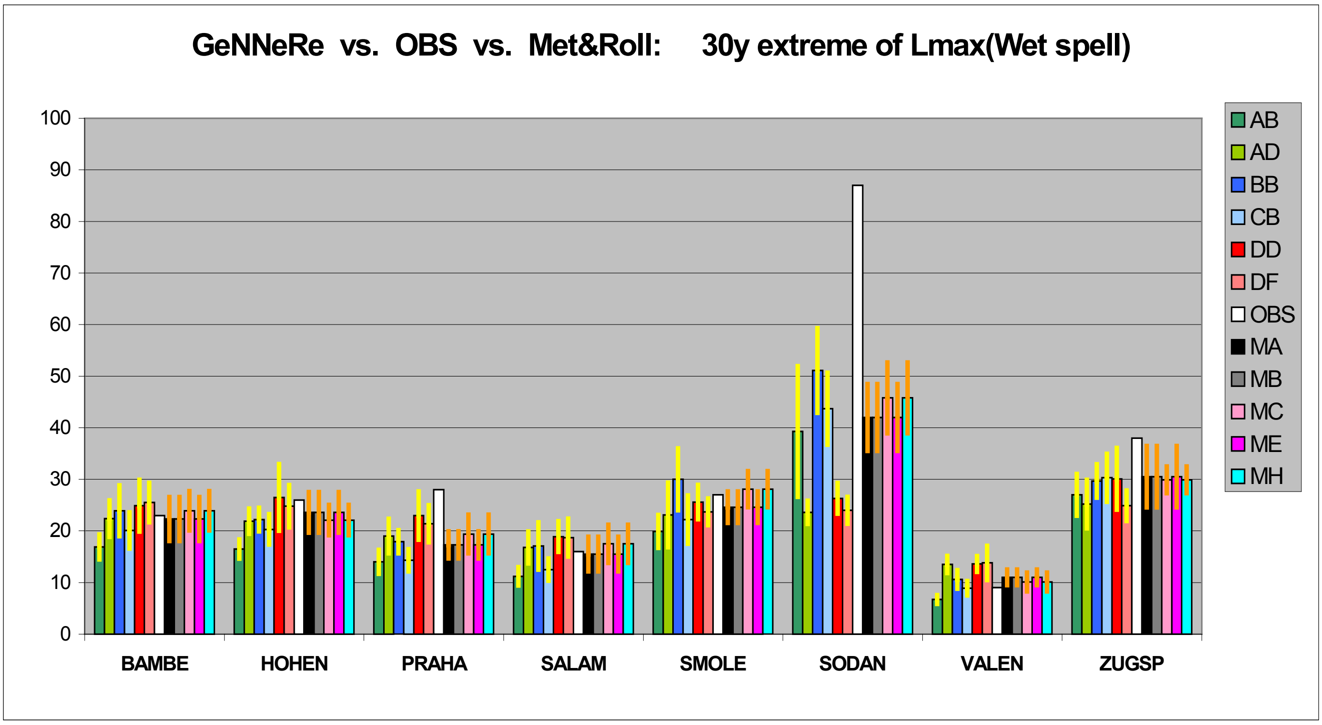
30-year maximum length of heat spell (std. error is based on ten 30y series in case of synthetic weather series generated by GeNNeR and Met&Roll)



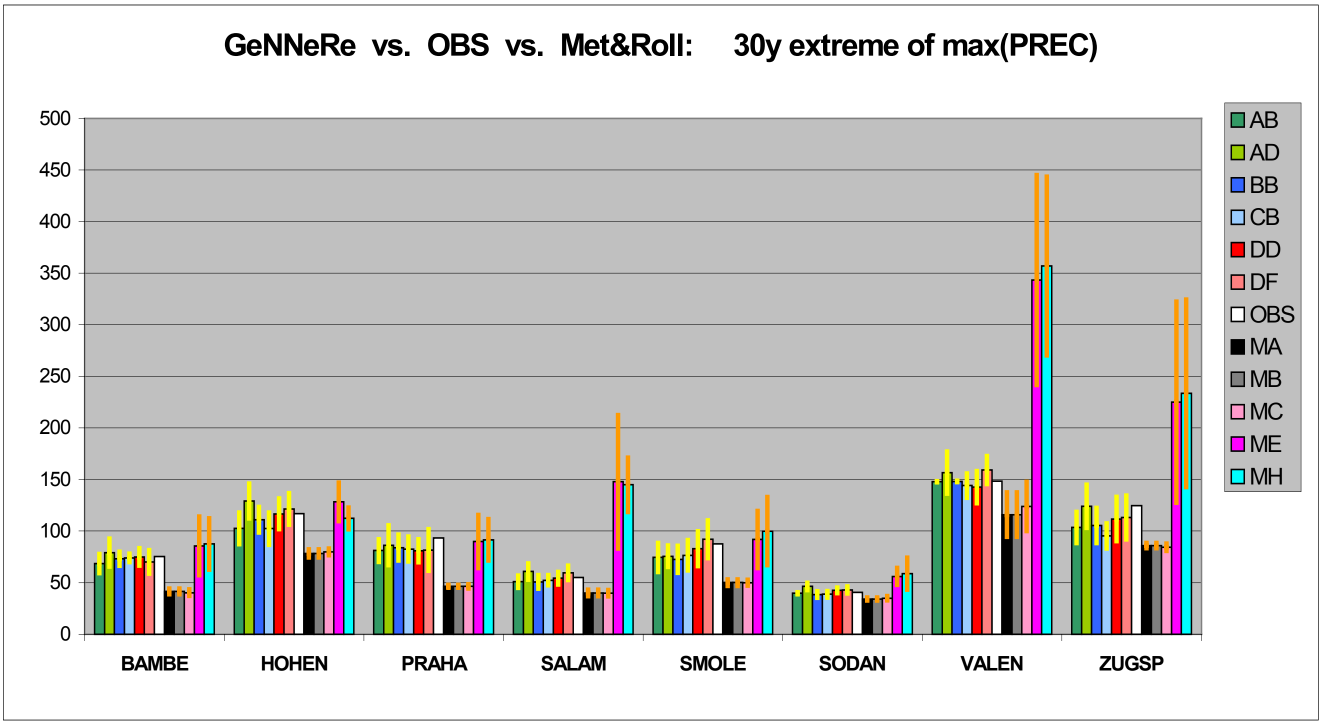
30-year maximum length of cold spell (std. error is based on ten 30y series in case of synthetic weather series generated by GeNNeR and Met&Roll)



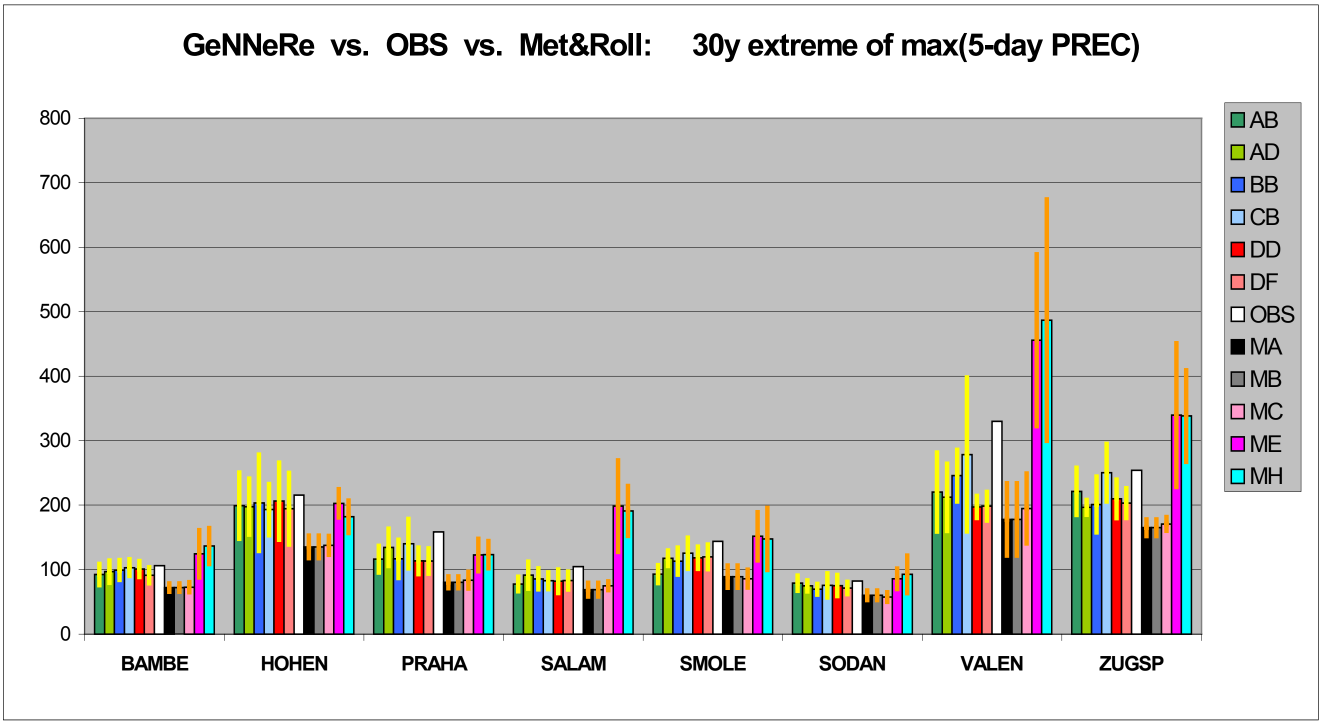
30-year maximum length of dry spell (std. error is based on ten 30y series in case of synthetic weather series generated by GeNNeR and Met&Roll)



30-year maximum length of cold spell (std. error is based on ten 30y series in case of synthetic weather series generated by GeNNeR and Met&Roll)



30-year maximum daily precipitation (std. error is based on ten 30y series in case of synthetic weather series generated by GeNNeR and Met&Roll)



30-year maximum 5-day precipitation (std. error is based on ten 30y series in case of synthetic weather series generated by GeNNeR and Met&Roll)