Supplementary Information

Evaluation of CORDEX regional climate models in simulating extreme dry spells in Southwest China

Tao Feng[1,2], Zachary Tipton[3], Lan Xia[1,2], Youli Chang[1,2]

1. Department of Atmospheric Sciences, Yunnan University, Kunming 650091, China

2. Key laboratory of atmospheric environment and processes in the boundary layer over the low-latitude plateau region, Yunnan University, Kunming, 650091, China

3. Department of Biological Sciences, University of Arkansas, Fayetteville Arkansas 72701, United States

Revised

October 10, 2019

Corresponding author: Tao Feng, Email: taofeng@ynu.edu.cn

Youli Chang, Email: ylchang@ynu.edu.cn

Part A. Model evaluation metrics

To evaluate RCM performance in simulating extreme dry spells, full assessments of simulated AMDSL intensity, frequency, and spatial distribution are needed. In this study, we focused on tail distribution characteristics of the extreme AMDSL; the spatial correlation coefficient is mainly employed to interpret the five RCMs' ability to simulate the intensity of AMDSL and spatial consistency.

Spatial correlation – often measured as a pattern correlation coefficient (PCC) – indicates the strength and direction of a linear relationship between two fields (for example model simulation and observation). The best known being the Pearson correlation coefficient, which is obtained by dividing the covariance of the two variables by the product of their standard deviations. A larger positive value of a correlation coefficient indicates a stronger positive correlation between model simulation and observation.

$$R = \frac{\sum_{k=1}^{N}(M_k - \bar{M}) \cdot (O_k - \bar{O})}{\sqrt{\sum_{k=1}^{N}(M_k - \bar{M})^2 \sum_{k=1}^{N}(O_k - \bar{O})^2}} \tag{A1}$$

where M_k and O_k indicate the model pattern of interest and the corresponding observed pattern, respectively. N denotes the number of model grid cells or observation points used.

To summarize the spatial consistency and magnitude of RCM variance against observations in a simple way, we used Taylor diagrams to evaluate the performance of the individual extreme index in each RCM simulation (Taylor, 2001).

A Taylor diagram is a polar-style graph including the correlation coefficient (R)

between simulations and observations, the centered RMSE (E'), and the standard deviation of RCM simulation and observations, respectively.

$$\sigma_M = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(M_k - \bar{M})^2} \quad \text{(A2)}$$

$$\sigma_O = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(O_k - \bar{O})^2} \quad \text{(A3)}$$

The radial distance from the origin reflects STD, the cosine of the azimuth angle gives R, and the radial distance from the observed point is proportional to the centered RMSE difference between simulations and observations. Taylor diagrams are particularly beneficial in evaluating the relative accuracy of different complex models.

$$E'^2 = \sigma_M^2 + \sigma_O^2 - 2\sigma_O\sigma_M R \quad \text{(A4)}$$

Furthermore, standardized model variance is applied to avoid different units' cross variables. The comparison between RCM simulations and observations is finally quantified by spatial correlation (R), root-mean-square difference (E'/σ_O), and magnitude of variations represented by the ratios of modeled-to-observed standard deviation (σ_M/σ_O) across the grid cells of the observation data set. The closer the correlation is to the horizontal axis and the ratio nearer 1.0, the better the model reproduces observation values.

Part B. The generalized extreme value theory

Extreme weather and climate events are considered as a major source of risk for all human societies. The definition of an extreme weather event is an event that is rare at a particular place and time of year. Definitions of rare vary, and a statistical distribution or a sample distribution is required.

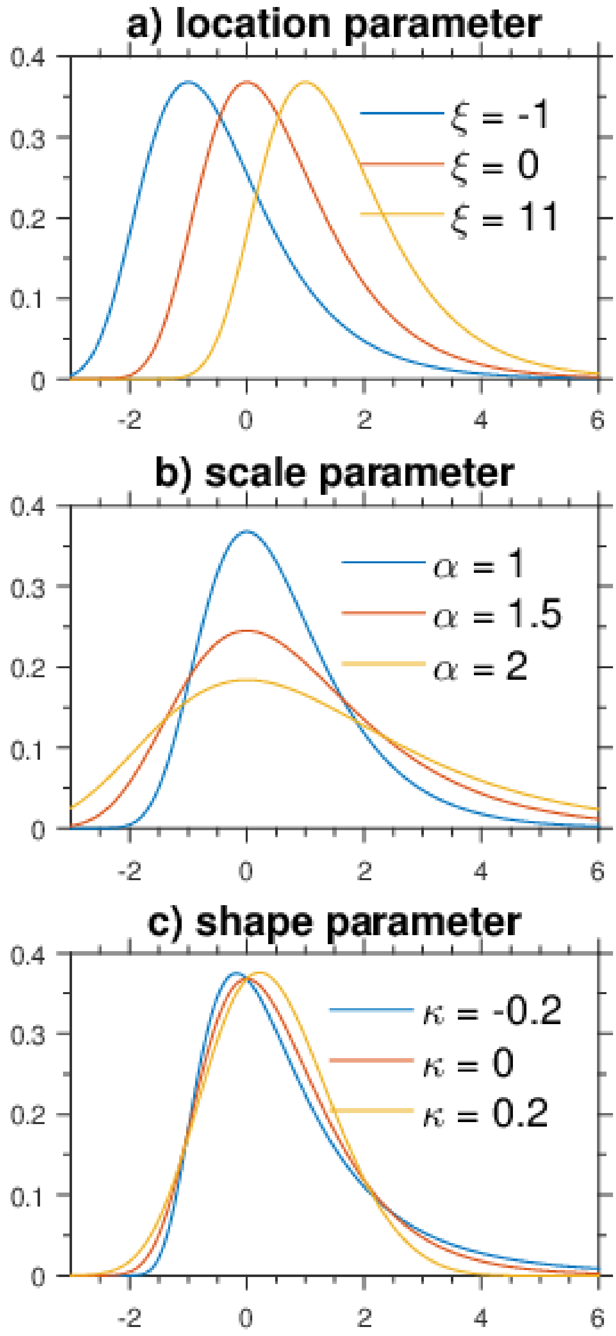
As mentioned in introduction section, the AMDSL was defined as annual maximum days without precipitation in this study. We assumed that the AMDSL is distributed according to a GEV distribution. The probability distribution function (PDF) and cumulated distribution function (CDF) for GEV distribution are given by

$$f(x;\ \xi,\alpha,k) = \frac{1}{\alpha}\left[\left(1+k\frac{x-\xi}{\alpha}\right)^{-1/k}\right]^{k+1} exp\left[-\left\{1+k\frac{x-\xi}{\alpha}\right\}^{-1/k}\right] \tag{B1}$$

$$F(x;\ \xi,\alpha,k) = exp\left[-\left\{1+k\frac{x-\xi}{\alpha}\right\}^{-1/k}\right] \quad k \neq 0 \tag{B2}$$

$$F(x;\ \xi,\alpha,k) = exp\left[-exp\left\{-\frac{x-\xi}{\alpha}\right\}\right] \quad k = 0 \tag{B3}$$

where x is the extracted block maxima and ξ $(-\infty < \xi < \infty)$is the location parameter determining position, α $(\alpha > 0)$ is the scale parameter present in the width of the PDF, k is the shape parameter describing the form of the distribution decay in the tail. When the shape parameter $k > 0$, the distribution is said to have a heavy tail with the density function decaying with a power law and is called Frechet distribution. When the shape parameter $k < 0$, the distribution has a bounded upper tail called a Weibull distribution. The $k = 0$ is referred to as exponential decay called a Gumbel distribution. Supplementary Figure 1 shows the physical meaning of the three GEV parameters. Review of recent developments in the statistical theory of extreme values is given in Coles et al (2001).



Supplementary Figure.1 A schematic of the generalized extreme value (GEV) theory.

After fitting the GEV distribution to the AMDSL, we can estimate how often the extreme quantiles occur with a certain return level. The *T*-year return value is defined as a value which is expected to be equaled or exceeded on average once every interval of T years, and $1/T$ is the return probability, respectively. X_T are estimated by inverting the GEV cumulative distribution function as follow:

$$F(X_T;\ \xi, \alpha, k) = 1 - 1/T \tag{B4}$$

$$X_T = \begin{cases} \xi - \alpha ln\left[-ln\left(1 - \frac{1}{T}\right)\right], & k = 0 \\ \xi - \frac{\alpha}{k}\left[1 - \left\{-ln\left(1 - \frac{1}{T}\right)\right\}^{-k}\right], & k \neq 0 \end{cases} \tag{B5}$$

In our current study, we calculated return values for a 20-yr return period of AMDSL at each model grid point and observation point.

Part C. Theory of *L*-Moment

The L-moments method is introduced by Hosking (1990) and is based on probability weighted moments (PWMs). Compared to the conventional moments, being linear functions of the data, L-moments are less sensitive to sampling variability or measurement errors in extreme data values, and therefore may be expected to yield more accurate and robust estimates of the characteristics or parameters of an underlying probability distribution (Hosking 1990). Thus, the estimated parameters via L-moment methods are more robust. Furthermore, L-moments have less bias in estimation and their asymptotes are closer to the normal distribution in finite samples. L-moments are defined as follows (Hosking and Wallis 2005):

Assuming $F(x)$ is the distribution function of random variable $X, X_{1:n} \leq \cdots \leq X_{n:n}$ are the order statistics of the samples. The L-moments of r is

$$\lambda_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} E(X_{r-k:r}), \quad r = 1,2 \ldots \tag{C1}$$

where $X_{k:n}$ denotes the k^{th}order statistic (k^{th} smallest value) in an independent sample of size n from the distribution of X while E denotes the expected value. In particular, the first four population L-moments are:

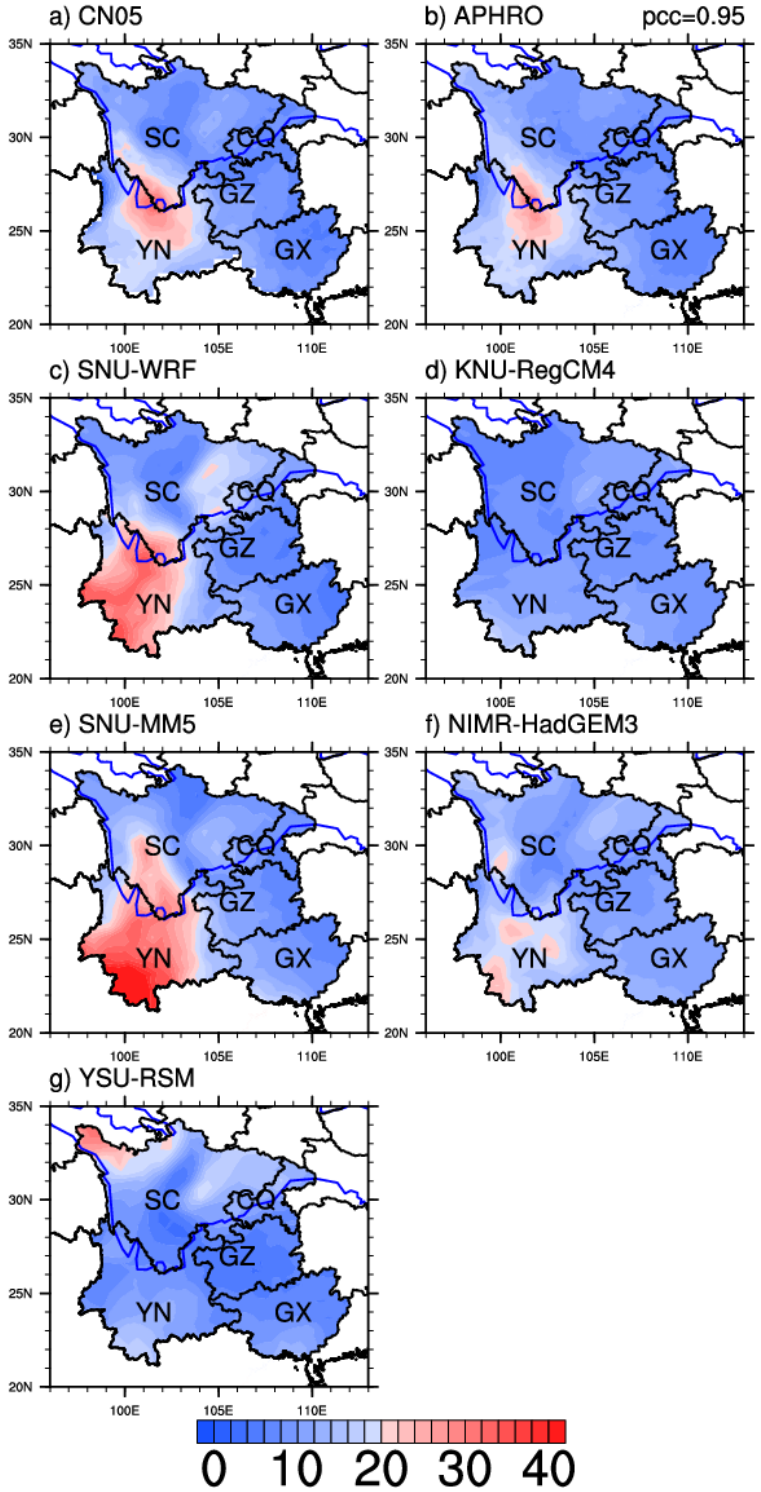
$$\begin{cases} \lambda_1 = E(X_{1:1}) \\ \lambda_2 = \dfrac{E(X_{2:2} - X_{1:2})}{2} \\ \lambda_3 = \dfrac{E(X_{3:3} - 2X_{2:3} + X_{1:3})}{3} \\ \lambda_4 = \dfrac{E(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4})}{4} \end{cases} \tag{C2}$$

To standardize the high order L-moments, the L-moments ratio is defined as

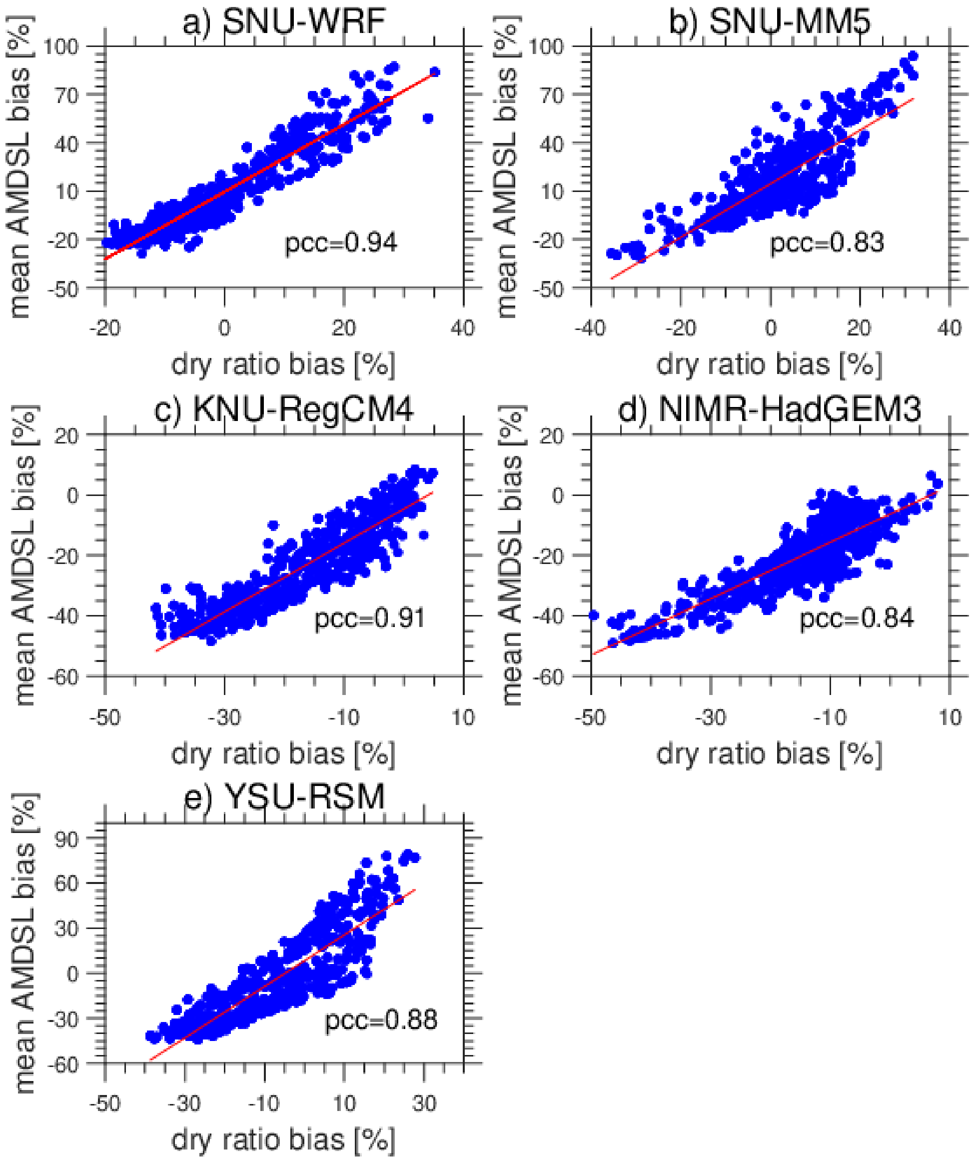
$$\tau_r = \frac{\lambda_r}{\lambda_2}, r = 3,4, \ldots \tag{C3}$$

τ_3 and τ_4 are the *L*-skewness and *L*-kurtosis, respectively, which are powerful tools to describe extreme precipitation distribution characteristics.

Part D. Model bias diagnosis



Supplementary Figure.2 Spatial distributions of the GEV location parameter of the AMDSL. (a) CN05, (b) APHRO; and five RCMs: (c) SNU-WRF, (d) KNU-RegCM4, (e) SNU-MM5, (f) NIMR-HadGEM3, and (g) YSM-RSM.



Supplementary Figure.3 Scatter plots of relative model bias of the climatological AMDSL against relative model bias of the dry day's occurrence. (a) SNU-WRF, (b) SNU-MM5, (c) KNU-RegCM4, (d) NIMR-HadGEM3, and (e) YSM-RSM.

Supplementary Table.1 Name, acronym, and equation of the cumulative distribution function (CDF) of four extreme models used in the assessment.

Distribution function	Acronym	CDF Equation
Generalized extreme value	GEV	$F(x) = exp(-exp(-y))$, where: $y = -k^{-1}ln\left(\frac{1-k(x-\xi)}{\alpha}\right)$ if $k \neq 0$; $y = (x-\xi)/\alpha$ if $k = 0$
Pearson type III (PE3)	PE3	$F(x) = frac(e^{\frac{ky-y^2}{2}})(\alpha(2\pi^{-2}))$, where: $y = -k^{-1}ln\left(\frac{1-k(x-\xi)}{\alpha}\right)$ if $k \neq 0$; $y = (x-\xi)/\alpha$ if $k = 0$
Three-parameter lognormal	LN3	$F(x) = \Phi\left(\frac{ln(x-\gamma)-\mu}{\alpha}\right)$
Generalized Pareto	GPD	$F(x) = 1 - exp(-y)$, where: $y = -k^{-1}ln\left(\frac{1-k(x-\xi)}{\alpha}\right)$ if $k \neq 0$; $y = (x-\xi)/\alpha$ if $k = 0$

Note: In the above equations, ξ is the location parameter, α is the scale parameter, and k represents the shape parameter of the cumulative distribution function, where Φ is the Laplace Integral for three-parameter lognormal distributions.

Supplementary Table.2 Name, acronym, and *L*-kurtosis estimator of four extreme models used in the assessment.

Distribution function	Acronym	L-kurtosis estimator
Generalized extreme value	GEV	$\tau_4 = 0.10701 + 0.11093\tau_3 + 0.84838\tau_3^2 - 0.06669\tau_3^3 + 0.00567\tau_3^4 - 0.04208\tau_3^5 + 0.03763\tau_3^6$
Pearson Type III	PE3	$\tau_4 = 0.1224 + 0.30115\tau_3^2 + 0.95812\tau_3^4 - 0.57488\tau_3^6 + 0.19383\tau_3^8$
Three-parameter lognormal	LN3	$\tau_4 = 0.12282 + 0.77518\tau_3^2 + 0.122779\tau_3^4 - 0.13638\tau_3^6 + 0.11368\tau_3^8$
Generalized Pareto	GPD	$\tau_4 = 0.20196 + 0.95924\tau_3^2 - 0.20096\tau_3^3 + 0.04061\tau_3^4$

Supplementary Table. 3 Abbreviations

Abbreviation	Standard Name
AMDSL	Annual Maximum Dry Spell Length
CQ	Chongqing direct-controlled municipality, China
EASM	East Asian Summer Monsoon
ESM	Earth System Model
GCM	Global Climate Model
GX	Guangxi Province, China
GZ	Guizhou Province, China
ISM	Indian Summer Monsoon
LLH	Low-latitude highlands
MLE	Maximum-Likelihood Estimator
MRB	Mekong River Basin
PCC	Pattern Correlation Coefficient
PDF	Probability Distribution Function
PWM	Probability Weighted Moments
RCM	Regional Climate Model
SB	Sichuan Basin
SC	Sichuan Province
SETP	southeastern Tibetan Plateau
SCV	Southwest China Vortex
TMC	Traverse Mountain Chain
YN	Yunnan Province, China

Supplementary References

Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). An introduction to statistical modeling of extreme values (Vol. 208). London: Springer.

Hosking, J. R. M. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. *J. R. Statist. Soc. B*, 52(1), 105-124. doi:10.1111/j.2517-6161.1990.tb01775.x

Hosking, J. R. M., Wallis, J. R. (2005). Regional frequency analysis: an approach based on L-moments. Cambridge University Press.

Taylor, K.E. (2001). Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res*., 106, D7: 7183-7192, doi: 10.1029/2000JD900719.