UNIVERSITY OF CENTRAL ASIA
GRADUATE SCHOOL OF DEVELOPMENT
Mountain Societies Research Institute

Reconstruction of Hydrometeorological Data Using Dendrochronology and Machine Learning Approaches to Bias-Correct Climate Models in Northern Tien Shan, Kyrgyzstan

Erkin Isaev [1], Mariiash Ermanova [2], Roy C. Sidle [1], Vitalii Zaginaev [2], Maksim Kulikov [1] and Dogdurbek Chontoev [2]

[1]Mountain Societies Research Institute, UCA
[2]Institute of Water Problems and Hydropower, NAS KR

Bishkek
2022

Objective: to show the importance of GCM spatial resolution and the new parameterization of the physical processes in a region of complex orography, using historical simulations from the CMIP5, CMIP6, and CORDEX.

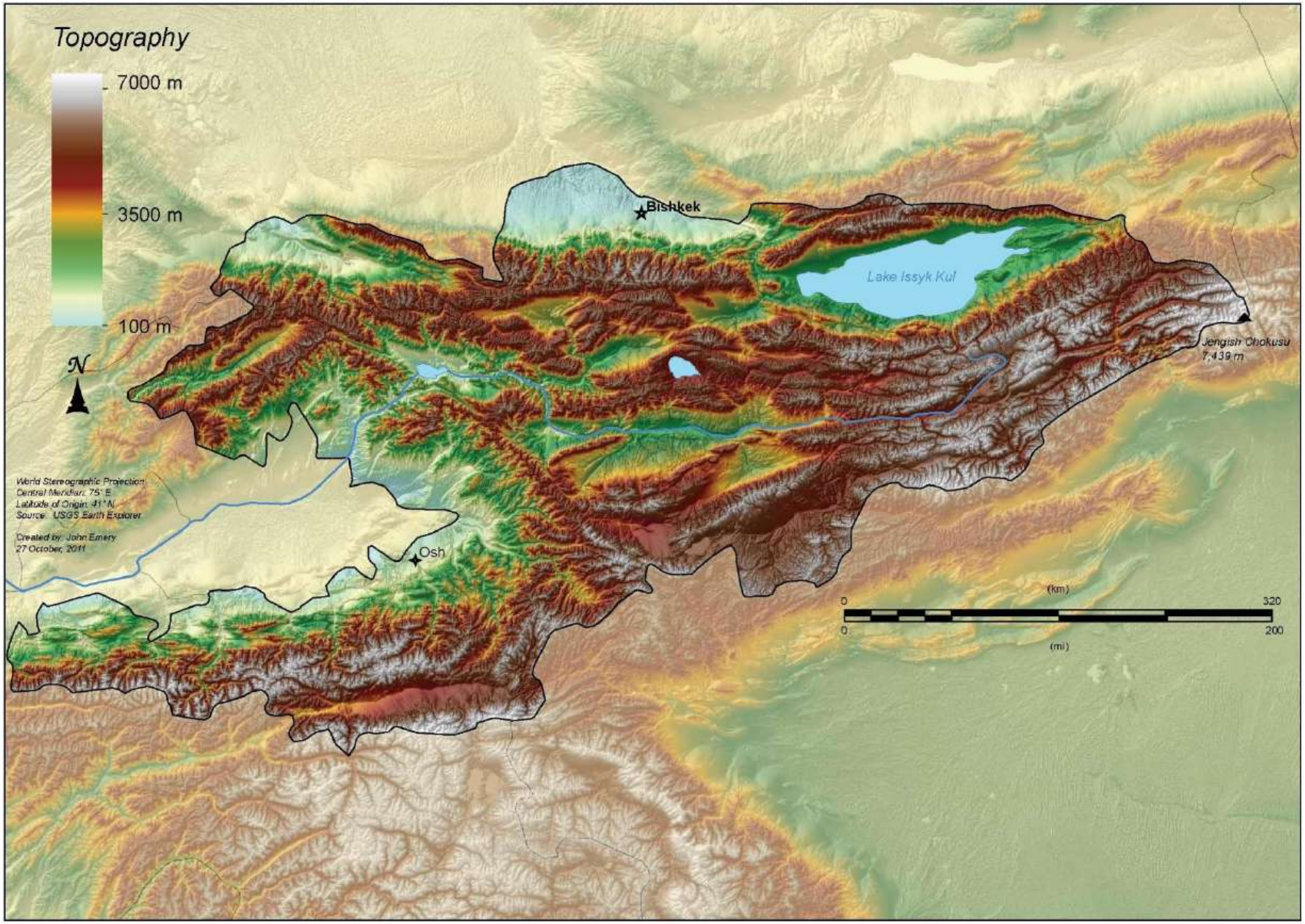
Task:

--Collection and statistical processing of data;

-reconstruction of hydrometeorological data using tree rings and Machine Leaning approaches;

- bias correction of GCM for mountainous region.

High-altitude relief from 350 to 7439 meters above sea level and 94% of the territory at an altitude of more than 1 km



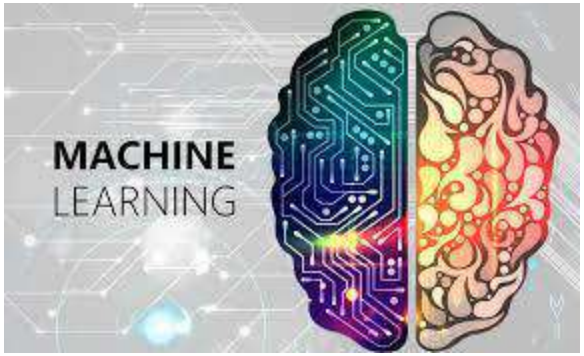
Methods

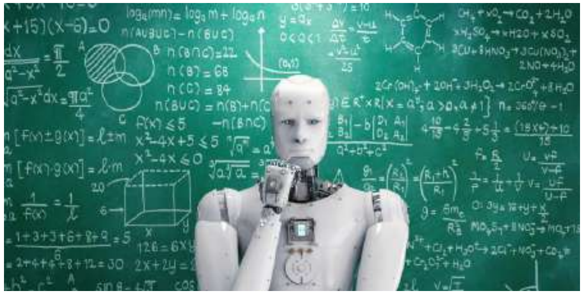
-Correlation analysis,
-Mann-Kendall Trend Test,
- ML approaches,
-BAIS-correction:

  -*method Bruyère et al. (2014),*
  -*ML RF*

Numerical experiments were conducted to verify machine learning (ML) approaches. We used :

- **Artificial Neural Network (ANN),**
- **Hyperparameter Tuning (HPT),**
- **Xgboost Regressor (XgbR),**
- **Random Forest Regressor (RFR),**
- **K-Nearest Neighbors Regressor (KNN),**
- **Decision Tree Regressor (DTR),**
- **Lasso Regression (LaR),**
- **Linear Regression (LR),**





Metrics

- root-mean-square error(RMSE)
- mean absolute error (MAE)
- The Kling–Gupta efficiency (KGE)

$$\mathrm{KGE} = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (1)$$

$$\alpha = \sigma_s / \sigma_o \quad (2)$$

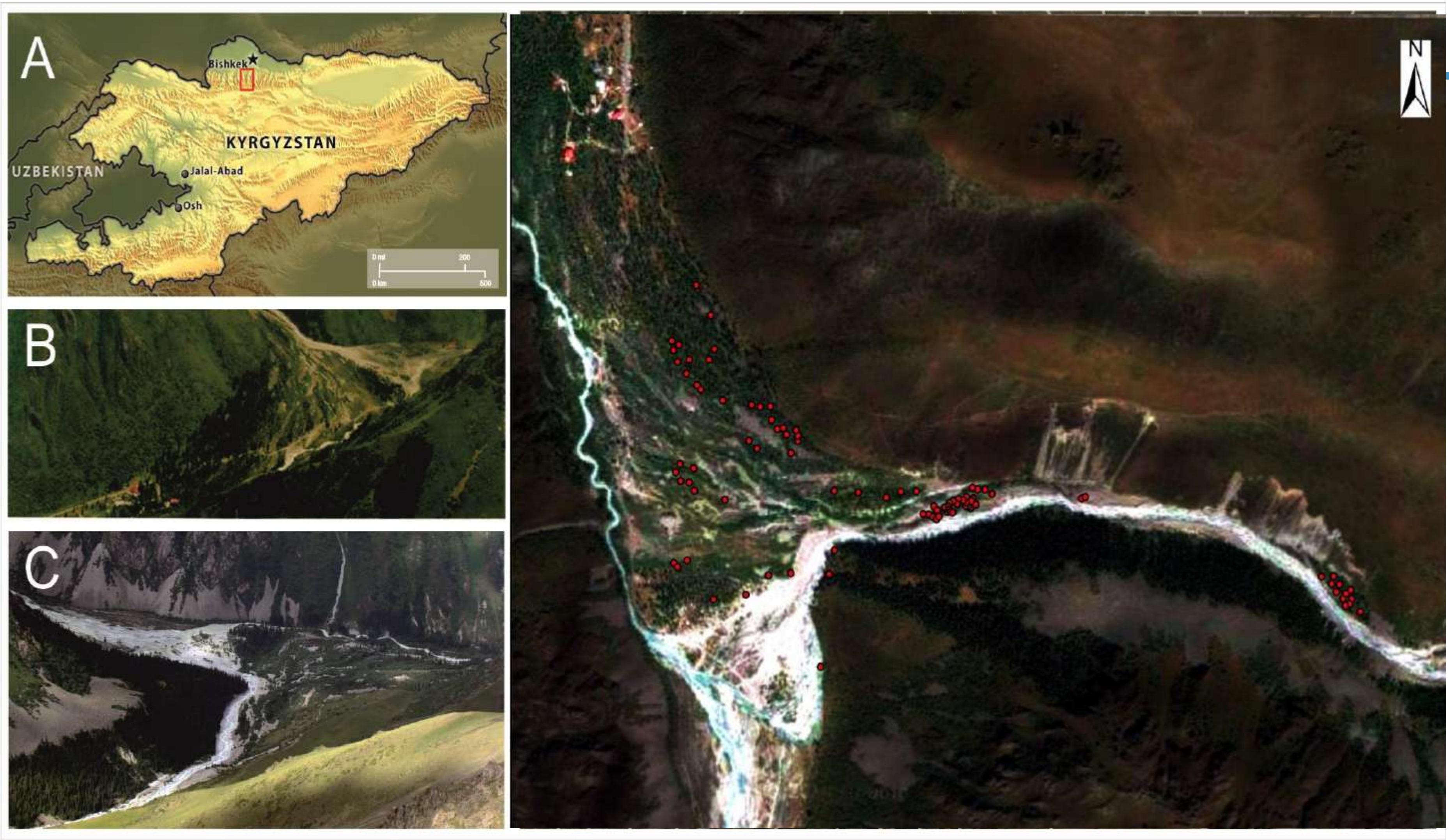
$$\beta = \mu_s / \mu_o \quad (3)$$

where r is the linear correlation between simulated and observed data, α is the ratio of standard deviation of observed and simulated data, β is the ratio of mean values of simulations and observations, s subscript represents the simulations, and o represents the observations.

the Taylor diagram was then applied to compare the climate model

6

Study site– Aksai valley



DATA

MATERIALS	Period	Parameters
Meteorological station at Baytik	1915-2013	Precipitation and temperature
Meteorological station Ala-Archa,	1979 - 2013	Precipitation and temperature
Hydrological station at Baytik,	1928-2013	Discharge
PICEA abies L. Karst tree	1886-2013	31 trees

Восстановленные

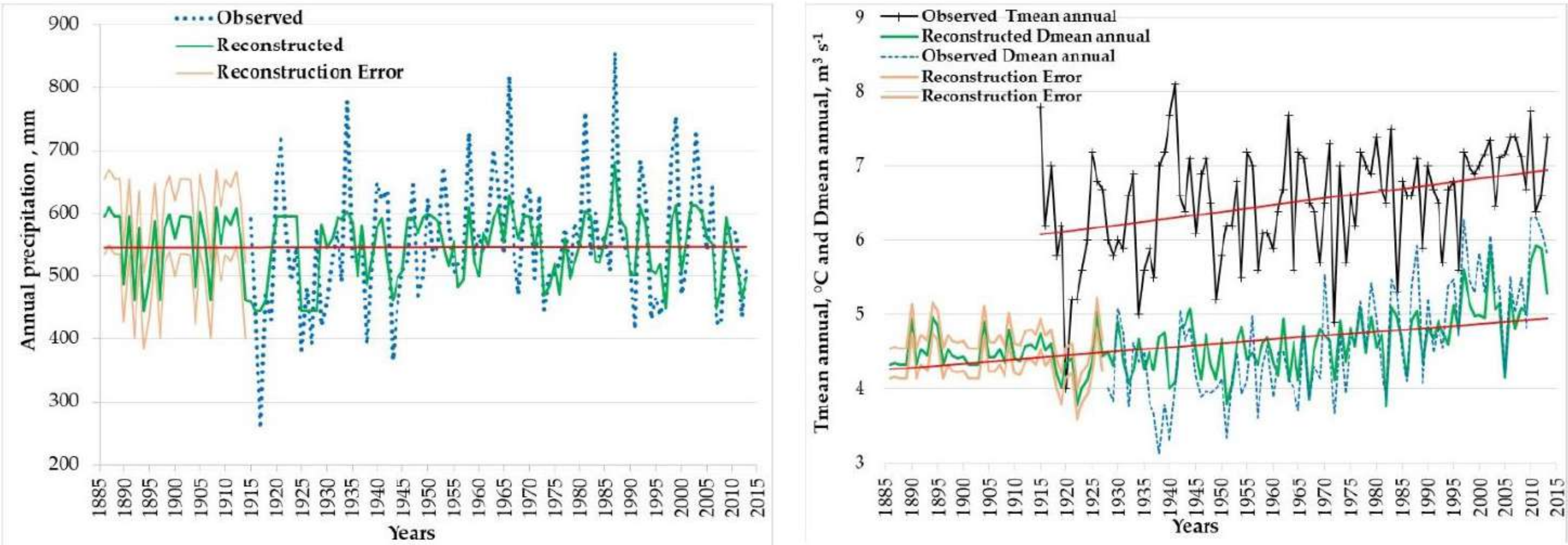
MATERIALS	Period	Parameters
Meteorological station at Baytik	1886-1915(29)	Precipitation and temperature
Meteorological station Ala-Archa,	1886-1979(93)	Precipitation and temperature
Hydrological station at Baytik,	1886-1928(42)	Discharge

Table 2. Correlation coefficients for comparisons between the hydrometeorological observations and tree-ring widths. Coefficients in red bold are significant at *p*-level 0.05.

Station Type Name		**Meteorological, Baytik**		**Meteorological, Ala-Archa**						**Hydro-Logical Gauging, Baytik**	**-**
	Variable	**P_a**	**$Tmean_a$**	**P_a**	**$Tmean_a$**	**$Tmax_a$**	**$Tmin_a$**	**$Tmax_{ma}$**	**$Tmin_{ma}$**	**$Dmean_a$**	**Tree-Ring Width**
MeteorologicalBaytik	P_a	1	−0.07	**0.83**	−0.21	**−0.36**	**0.44**	−0.2	0.22	0.04	**0.47**
	$Tmean_a$	−0.07	1	−0.05	−0.05	−0.01	0.19	**0.57**	**0.48**	**0.36**	−0.26
MeteorologicalAla-Archa	P_a	**0.83**	−0.05	1	−0.21	**−0.44**	**0.45**	−0.3	0.09	−0.09	**0.6**
	$Tmean_a$	−0.21	−0.05	−0.21	1	0.19	−0.12	0.02	−0.08	0.24	−0.26
	$Tmax_a$	**−0.36**	−0.01	**−0.44**	0.19	1	−0.13	0.22	−0.01	0.16	**−0.46**
	$Tmin_a$	**0.44**	0.19	**0.45**	−0.12	−0.13	1	−0.24	**0.39**	−0.24	**0.46**
	$Tmax_{ma}$	−0.2	**0.57**	**−0.3**	0.02	0.22	−0.24	1	**0.3**	**0.34**	**−0.39**
	$Tmin_{ma}$	0.22	**0.48**	0.09	−0.08	−0.01	**0.39**	**0.3**	1	**0.29**	−0.02
Hydrological Gauging Baytik	$Dmean_a$	0.04	**0.36**	−0.09	0.24	0.16	−0.24	**0.34**	**0.29**	1	**0.14**
-	Tree-ring width	**0.47**	−0.26	**0.60**	−0.26	**−0.46**	**0.46**	**−0.39**	−0.02	**0.14**	1

Table 3. Evaluation metrics amongst selected hydrometeorological reconstructed variables. Coefficients in red bold are significant at p-level 0.05.

Station, Variable, Unit	Metrics	LR	XgbR with HPT	RFR with HPT	KNN with HPT	LaR	DTR with HPT	ANN
Meteorological, Baytik, P_a, mm	MAE RMSE	70.8 89.2	**60.9** **77.5**	69.4 86.1	**59.7** **75.9**	70.6 89.3	77.8 101.4	70.6 89.3
Meteorological, Ala-Archa, P_a, mm	MAE RMSE	39.2 49.7	31.2 41.3	**25.6** **33.1**	38.1 50.4	42.5 53.8	33.3 44.7	38.1 50.4
Meteorological, Ala-Archa, $Tmax_a$, °C	MAE RMSE	1.3 1.6	**0.9** **1.2**	1.1 1.4	1.1 1.3	1.3 1.7	1.2 1.4	1.2 1.4
Meteorological, Ala-Archa, $Tmin_a$, °C	MAE RMSE	1.6 2.0	**1.2** **1.5**	1.3 1.7	1.2 1.6	1.6 2.0	1.4 1.8	1.5 1.9
Meteorological, Ala-Archa, $Tmax_{ma}$, °C	MAE RMSE	0.7 0.8	**0.6** **0.7**	**0.6** **0.7**	**0.5** **0.6**	0.7 0.8	0.5 0.7	0.7 0.8
Meteorological, Ala-Archa, $Tmin_{ma}$, °C	MAE RMSE	0.6 0.8	**0.5** **0.6**	**0.4** **0.5**	0.7 0.8	0.6 0.8	0.7 0.8	0.6 0.8
Hydrological Gauging, Baytik, $Dmean_a$, $m^3 s^{-1}$	MAE RMSE	0.5 0.6	**0.2** **0.1**	0.4 0.5	0.5 0.6	0.5 0.6	0.5 0.6	0.5 0.6



(**a**) (**b**)

Figure 3. Observed and reconstructed hydrometeorological data: (**a**) annual precipitation (P_a) for Baytik meteorological station; (**b**) mean annual discharge ($Dmean_a$) for Baytik hydrological-gauging station and mean annual air temperature ($Tmean_a$) for Baytik meteorological station.

Description of CMIP5, CMIP6 GCMs, and CORDEX RCMs used in our study and their spatial resolution.

Institution / Country	Models-CMIP6	Resolution	Models-CMIP5	Resolution	CORDEX	Resolution
Australian Research CouncilCentre of Excellence forClimate System Science,Australia	ACCESS-CM2	1.87 x 1.25	ACCESS13	1.90 x 1.20		
Beijing Climate Center, Beijing,China	BCC-CSM2-MR	1.12 x 1.12	BCC-CSM1.1-M	2.80 x 2.80		
Geophysical Fluid DynamicsLaboratory, NJ, USA	GFDL-ESM4	1.00 x 1.25	GFDL-ESM2G	2.50 x 2.00		
Institute for NumericalMathematics, Russia	INM-CM5-0	2.00 x 1.50	INMCM4.0	2.00 x 1.50		
Institute Pierre Simon Laplace(IPSL), Paris, France	IPSL-CM6A-IR	2.50 x 1.27	IPSL-CM5A-lr	3.70 x 1.90		
Japan Agency for Marine-Earth Science andTechnology (JAMSTEC),Kanagawa, Japa	MIROC6	1.40 x 1.40	MIROC5	1.40 x 1.40		
Max Planck Institute for Meteorology, Germany	MPI-ESM1-2-HR	0.94*0.94	MPI-ESM-mR	1.9*1.9	**RegCM4-3.v5**	**0.44 x 0.44**
	MPI-ESM1-2-LR	1.87*1.86	MPI-ESM-LR	1.86*1.87	**REMO2015.v1**	**0.22 x 0.22**
Meteorological ResearchInstitute, Ibaraki, Japan	MRI-ESM2-0	1.12 x 1.12	MRI-ESM1	1.10 x 1.10		
	NorESM2-MM		NorESM1-M	1.89*2.5	**REMO2015.v1**	**0.22 x 0.22**
Met Office Hadley Centre, UK	HadGEM3-G	1 x1	HadGEM2-ES	1.25*1.87	**REMO2015.v1**	**0.22 x 0.22**
			HadGEM2-ES	1.25*1.87	**RegCM4-3.v5.**	**0.44 x 0.44**
National Centre for Meteorological Research, France	CNRM-CM6-1	1 x 1	CNRM-CM5	1.4*1.4	**ALARO-0.v1**	**0.22 x 0.22**

Table 4. Statistical summary of the comparisons between the CMIP5, CMIP6 GCMs, and CORDEX RCMs simulations and observations at Baytik and Ala-Archa meteorological stations. KGE: Kling–Gupta efficiency, RMSE: root-mean-square error, P_a: annual precipitation, and $Tmean_a$: mean annual air temperature; Baytik station metrics: standard deviation (SD) of the P_a −104 mm/year, SD of the $Tmean_a$ −0.7 °C/year; Ala-Archa station metrics: SD of the P_a −99 mm/year, SD of the $Tmean_a$ −0.5 °C/year.

	Models	Meteorological Station Baytik				Meteorological Station Ala-Archa			
		P_a		$Tmean_a$		P_a		$Tmean_a$	
		KGE	RMSE, mm/year	KGE	RMSE, °C/year	KGE	RMSE, mm/year	KGE	RMSE, °C/year
CMIP5	ACCESS1-3	0.15	113	0.12	1.7	0.03	113	−0.12	2.3
	BCC-CSM1-1m	0.00	160	0.19	1.4	−0.05	149	−0.98	5.2
	CNRM-CM5	−0.13	160	−0.73	1.6	−0.20	149	−0.62	2.1
	HadGEM2-ES	0.08	116	0.08	0.6	0.07	105	0.02	1.0
	INMCM4.0	−0.12	197	−0.05	4.2	−0.05	182	−0.26	1.0
	IPSL-CM5A-LR	−0.19	158	−0.25	5.5	−0.12	143	−0.37	1.6
	MIROC5	−0.01	136	−0.01	2.5	0.07	117	−1.41	6.3
	MPI-ESM-MR	−0.05	143	0.10	3.0	−0.02	127	0.11	1.4
	MRI_ESM1	−0.08	156	0.17	0.9	0.03	145	−0.69	4.2
	NorESM1-M	−0.11	184	0.17	1.0	−0.01	168	−0.82	4.1
CMIP6	ACCESS-CM2	0.13	112	0.06	1.1	0.03	128	−0.85	4.5
	BCC-CSM2-MR	−0.13	162	0.18	1.1	−0.06	147	−0.85	4.6
	CNRM-CM6	−0.16	306	−0.15	5.0	−0.17	302	−0.04	1.2
	HadGEM3-GC31-MM	−0.24	282	−0.16	5.0	−0.25	274	0.03	1.3
	INM-CM5-0	0.02	127	0.05	1.1	0.05	113	−0.76	4.5
	IPSL-CM6A-LR	0.11	112	−0.10	3.3	0.04	103	−0.20	1.3
	MIROC6	−0.19	177	−0.21	4.5	−0.22	165	−1.85	8.1
	MPI-ESM1-2-LR	−0.17	170	0.06	3.2	−0.17	153	0.08	1.2
	MRI-ESM2-0	0.10	223	−0.05	2.1	0.05	222	−1.13	5.7
	MPI-ESM1-2-HR	−0.11	147	0.19	1.1	−0.16	135	−0.39	3.3
	NorESM2-MM	−0.02	251	0.18	1.3	−0.05	242	−0.90	4.9
CORDEX	ALARO-0.v1_CNRM	−0.15	268	−0.38	4.6	−0.43	326	−1.09	4.8
	RegCm4-3v5_HadCEM2-ES	−0.11	298	−0.01	3.9	−0.22	308	0.01	0.9
	RegCm4-3v5_MPI-ESM-MR	−0.35	315	0.24	2.6	−0.39	321	−0.07	1.7
	REMO2015_HadGEM2-ES	0.06	131	−0.07	4.2	0.07	128	0.11	0.8
	REMO2015_MPI-M-ESM-LR	0.17	137	−0.33	3.9	0.24	131	−0.37	0.8
	REMO2015_NCC-NorESM1-M	−0.08	164	−0.03	3.6	−0.11	160	0.02	1.0

Table 5. Statistical summary of comparisons between the CMIP6 GCMs with and without bias correction simulations and observations at Baytik meteorological station. KGE: Kling–Gupta efficiency, RMSE: root mean square error, P_a: annual precipitation, and $Tmean_a$: mean annual air temperature; Baytik station metrics: standard deviation (SD) of the P_a −104 mm/year, SD of the $Tmean_a$ −0.7 °C/year.

	Models	Baytik Meteorological Station							
		P_a		Bias Corrected P_a		$Tmean_a$		Bias Corrected $Tmean_a$	
		KGE	RMSE, mm/year	KGE	RMSE, mm/year	KGE	RMSE, °C/year	KGE	RMSE, °C/year
CMIP6	ACCESS-CM2	0.13	112	0.19	112	0.06	1.1	0.06	1.0
	BCC-CSM2-MR	−0.13	162	−0.10	125	0.18	1.1	0.19	1.0
	CNRM-CM6	−0.16	306	−0.06	153	−0.15	5.0	0.13	1.1
	HadGEM3-GC31-MM	−0.24	282	−0.16	173	−0.16	5.0	0.12	0.9
	INM-CM5-0	0.02	127	0.04	118	0.05	1.1	0.06	0.9
	IPSL-CM6A-LR	0.11	112	0.11	110	−0.10	3.3	0.01	1.1
	MIROC6	−0.19	177	−0.10	165	−0.21	4.5	0.00	1.1
	MPI-ESM1-2-LR	−0.17	170	−0.10	168	0.06	3.2	0.19	0.9
	MRI-ESM2-0	0.10	223	0.15	146	−0.05	2.1	−0.01	1.0
	MPI-ESM1-2-HR	−0.11	147	−0.05	143	0.19	1.1	0.20	0.9
	NorESM2-MM	−0.02	251	0.06	117	0.18	1.3	0.19	1.0
Mean	MMEs	−0.06	188	0.01	139	0.01	2.6	0.10	1.0

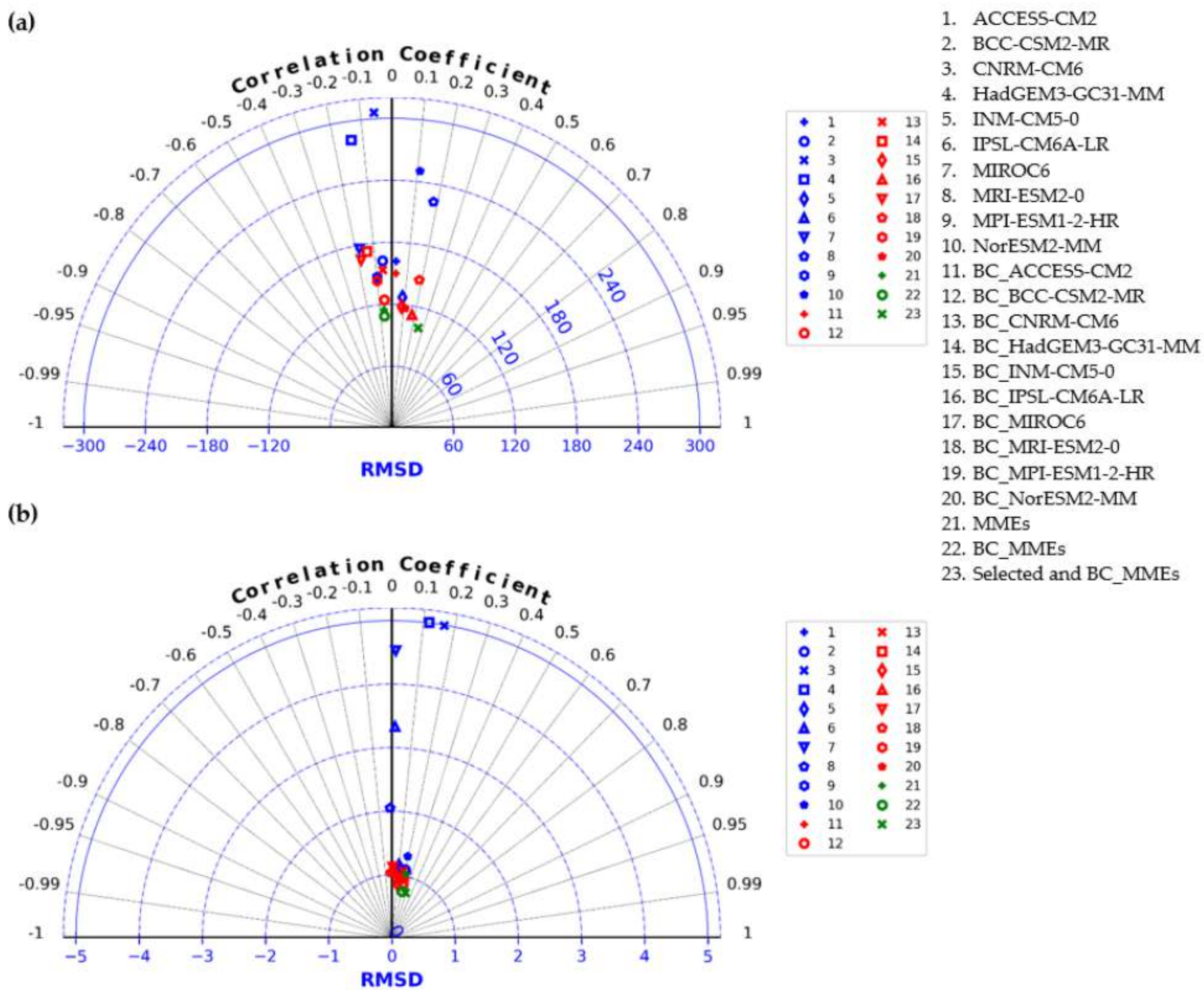
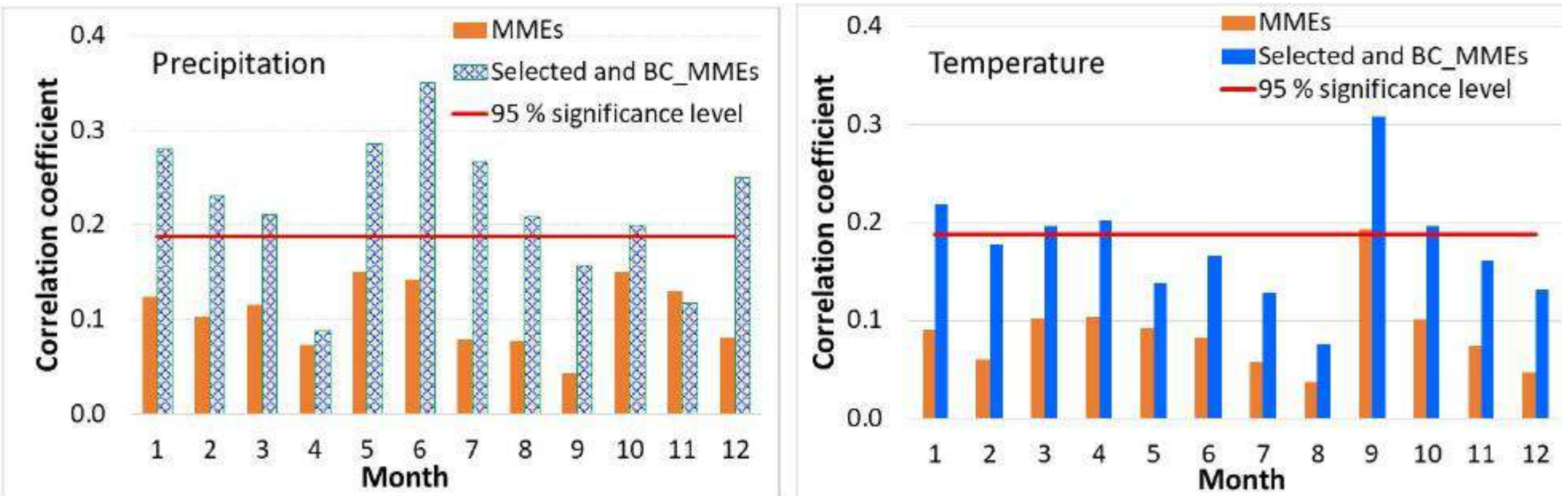


Figure 4. Taylor diagram comparing the observed data with the CMIP6 GCMs: (**a**) annual precipitation (P_a) for Baytik meteorological station; (**b**) mean annual air temperature ($Tmean_a$) for Baytik meteorological station. Blue marks are for simulations without bias correction, red marks are for simulations with bias correction (BC), and green marks are as follows: 21—mean for multi-model ensembles (MMEs), 22—MMEs with BC, and 23—selected the best MMEs with BC. RMSD: root-mean-square deviation.



(a) (b)

Figure 5. Correlation coefficients between observed monthly data at Baytik meteorological station and monthly climate-model data: (**a**) monthly precipitation; (**b**) monthly mean air temperature; MMEs—multi-model ensembles, selected and BC_MMEs—selected best models and bias-corrected MMEs, $n = 99$.

Discussion

- **The reconstructed data show increases in mean annual temperature and mean annual discharge . However, the precipitation trends are not significant , likely because increasing temperatures are melting glaciers, which contribute to a higher average annual discharge in the Kashka-Suu River.**

- **Improvement in model resolution from CMIP5 to CMIP6 improved the model's performance. This better performance of some of the CMIP6 GCMs may be due to enhanced parameterization.**

- **the highest performance in simulating precipitation and temperature was obtained using the CORDEX models, which replicated historical $Tmean_a$ with KGE = 0.24 and P_a with KGE = 0.24. This is due to the high spatial resolution of CORDEX–REMO models (0.22°). Thus, spatial resolution plays a key role for regions with complex orography, both for modeling atmospheric processes and for model validation.**

- **The selected GCMs with BC showed best results**

Conclusions

(1) Instrumental observations in Kyrgyzstan are insufficient for assessing long-term climate and hydrological changes. Due to their annual resolution and sensitivity to climate, tree rings provide reliable proxies that can be used to extend instrumental records, as shown in our findings between climate and discharge changes in drylands of Kyrgyzstan. We also provide qualitative information on long-term hydrologic variability in the region that can inform water managers, stakeholders, and decision-makers.

(2) ML algorithms that combined RFR, KNN with HPT, and XgbR with HPT performed best, and these were used to reconstruct hydrometeorological data in Kyrgyzstan for the first time.

(3) Increases in the average annual temperature and mean annual discharge of the Kashka-Suu River were associated with more rapid glacier melting; however, precipitation did not significantly change with time.

(4) The CORDEX models best simulated precipitation and temperature over northern Tien Shan. These successfully replicated historical $Tmean_a$ (KGE = 0.24) and Pa (KGE = 0.24), due to their high spatial resolution (0.22°), indicating that spatial resolution plays a key role in complex mountain regions both for modeling atmospheric processes and model validation.

(5) Multi-model ensembles with selected GCMs and bias correction significantly increased performance of climate models.

Since climate models are important tools for assessing future climate change effects, their accuracy must be demonstrated before they can be employed to estimate climate change scenarios. Accurate climate models can then provide the basis for climate change adaptation as well as disaster reduction and prevention in Kyrgyzstan. In the future, similar research in other climatic zones of Kyrgyzstan is planned for more effective implementation of climate change adaptation strategies.

Спасибо за внимание!
Конул бурганызга чон рахмат!
Thank you for your Attention!

Link to publication:
https://www.mdpi.com/2073-4441/14/15/2297