TEMPERATURE DATA GENERATION USING QUANTILE REGRESSION FORESTS FOR DIFFERENT CLIMATE CHARACTERISTICS IN TÜRKIYE

CENK SEZEN

1. Faculty of Engineering, Ondokuz Mayıs University, Türkiye

1. Background and Goals

- Performing the synthetic temperature data generation using the quantile regression forests (QRFs)
- Comparing the synthetic simulation results using different evaluation criteria, statistics, and density plots
- Evaluating synthetic daily mean temperature generation results under different climate characteristics.

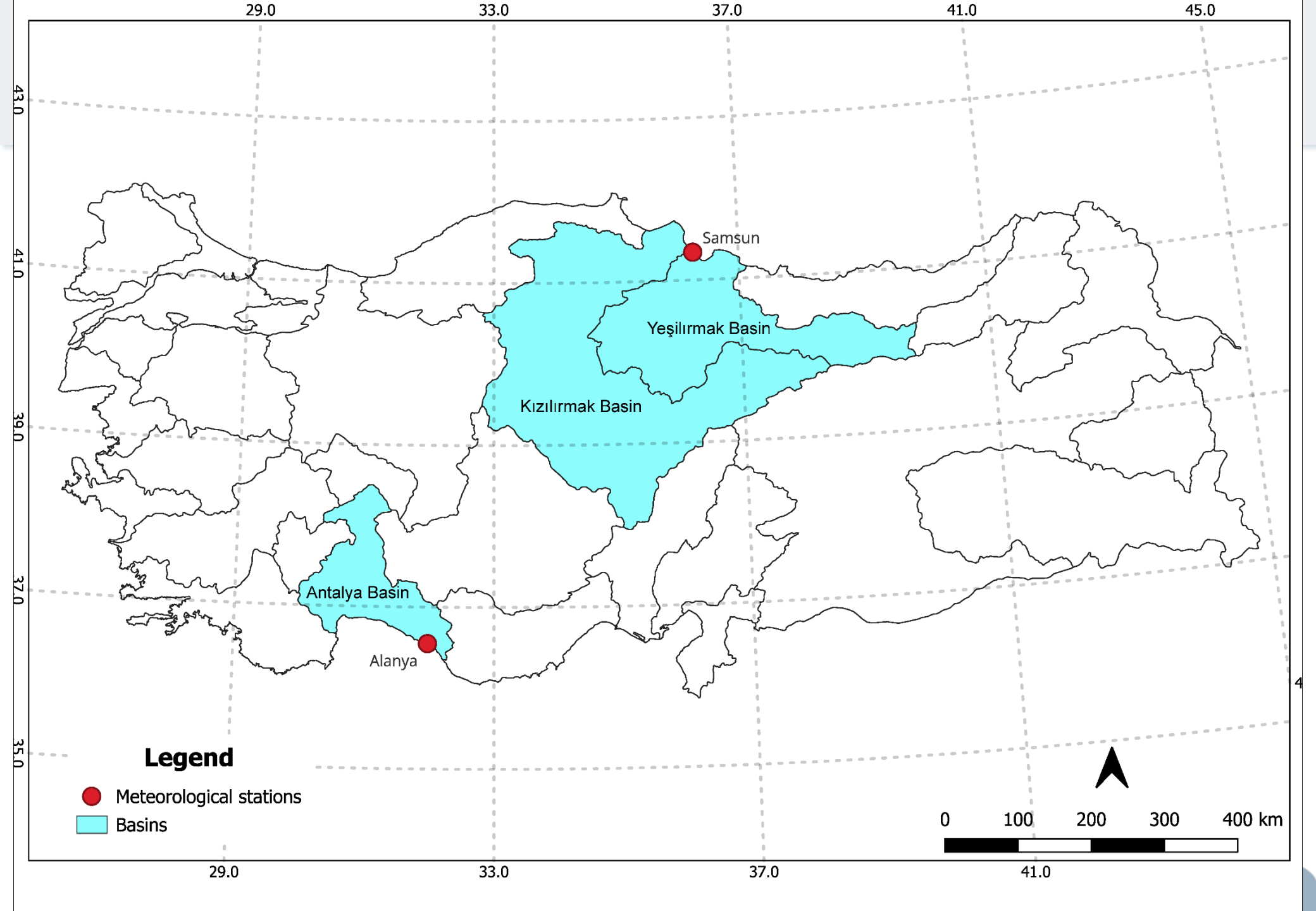
2. Outline of the Presentation

1. Introducing the study area and data characteristics

2. Giving brief information related to the adopted methods

3. Presenting and discussing the synthetic temperature generation results

4. Summarizing the findings


29.0
33.0
37.0
41.0
45.0
43.0
41.0
39.0
37.0
35.0
Samsun
Yeşilırmak Basin
Kızılırmak Basin
Antalya Basin
Alanya
Legend
Meteorological stations
Basins
0
100
200
300
400 km

3. Study area and data characteristics

Daily mean, minimum, and maximum temperature statistics for Samsun for the periods 01.10.1999-30.09.2015 and 01.10.2015-30.09.2020

Station	Period	Climatic variables	Minimum (°C)	Maximum (°C)	Mean (°C)	Standard deviation(°C)	Skewness	Kurtosis
Samsun	01.10.1999-30.09.2015	Tmin	-6.60	26.80	11.60	6.60	-0.01	-1.01
		Tmax	-1.80	38.40	18.95	7.21	-0.20	-1.01
		Tmean	-3.00	29.50	15.12	6.90	-0.08	-1.06
	01.10.2015-30.09.2020	Tmin	-4.10	26.70	12.68	6.52	0.00	-1.10
		Tmax	0.30	35.10	19.56	7.06	-0.18	-1.00
		Tmean	-2.00	28.80	15.90	6.77	-0.07	-1.09

3. Study area and data characteristics

Daily mean, minimum, and maximum temperature statistics for Alanya for the periods 01.10.1996-30.09.2012 and 01.10.2012-30.09.2017

Station	Period	Climatic variables	Minimum (°C)	Maximum (°C)	Mean (°C)	Standard deviation (°C)	Skewness	Kurtosis
Alanya	01.10.1996-30.09.2012	Tmin	-0.50	33.00	17.30	6.33	0.00	-1.13
		Tmax	6.50	40.80	24.66	6.45	0.01	-1.13
		Tmean	3.30	35.40	20.58	6.43	0.02	-1.20
	01.10.2012-30.09.2017	Tmin	0.40	29.00	17.72	6.22	-0.02	-1.07
		Tmax	6.50	43.70	24.90	6.48	-0.04	-1.02
		Tmean	3.20	35.70	20.99	6.35	-0.03	-1.09

3. Study area and data characteristics

- The **Black Sea climate** in Turkey is characterized by **high precipitation** year-round, resulting in lush forests.
- The Black Sea climate features **mild temperatures**, with cool, wet winters and warm, humid summers.
- The **Mediterranean climate** in Turkey is characterized by **hot, dry summers** and **mild, rainy winters**.
- This pattern supports distinctive vegetation like maquis (chaparral) and is common in the coastal regions of the Aegean and Mediterranean Seas.

4. Methods

Quantile Regression Forests

- Quantile Regression Forests is a tree-based ensemble method for estimating conditional quantiles.
- The conditional distribution function of the response value, Y , given X = x, can be given as follows:

$$F(y|X = x) = P(Y \leq y|X = x) = E(1_{\{Y \leq y\}}|X = x) \quad (1)$$

- The last expression is appropriate to draw analogies with the random forest approximation of the conditional mean E(Y|X = x).
- Just as E(Y|X = x) is approximated by a weighted mean over the observations of Y, an approximation to $E(1_{\{Y \leq y\}}|X = x)$ can be defined by the weighted mean over the observations of $1_{\{Y \leq y\}}$:

$$\hat{F}(y|X = x) = \sum_{i=1}^{n} w_i(x) 1_{\{Y_i \leq y\}} \quad (2)$$

where, $w_i(x)$ refers to the weights (Meinshausen, 2006)

4. Methods

Evaluation of the simulation results

- Twenty synthetic temperature datasets were generated, and the success of the generation process was evaluated using statistics (i.e., mean, standard deviation, minimum, maximum, kurtosis, skewness).
- Density plots, and evaluation metrics, such as the Kling-Gupta Efficiency (KGE), Nash-Sutcliffe Efficiency (NSE), and root mean square error (RMSE).
- Box plots were also employed to assess the performance of the generated temperature datasets.

5. Results and Discussion

- The 21-year datasets were used for the synthetic mean temperature at the Samsun and Alanya stations.
- The 16-year and 5-year datasets were used as training and testing periods.
- The previous lags (t-1, t-2 and t-3) of minimum, maximum, and mean temperatures were used as input data for generating the mean temperature.
- Seasonality was also considered for the temperature data generation process.
- 20 synthetic temperature datasets were generated for each temperature dataset.

5. Results and Discussion

Statistics for the generated daily mean temperature values for Samsun

	Minimum (°C)	Maximum (°C)	Mean (°C)	Standard deviation (°C)	Skewness	Kurtosis
Generation 1	-1.90	28.00	14.48	6.81	-0.04	-1.07
Generation 2	-1.90	27.82	14.54	6.97	-0.06	-1.09
Generation 3	-1.10	28.20	14.69	6.80	-0.05	-1.06
Generation 4	-3.00	28.70	14.70	6.80	-0.06	-1.00
Generation 5	-1.27	28.00	14.64	6.80	-0.11	-1.04
Generation 6	-1.90	28.70	14.68	6.84	-0.07	-1.01
Generation 7	-0.40	28.20	14.56	6.63	-0.06	-1.06
Generation 8	-1.60	28.20	14.67	6.86	-0.07	-1.04
Generation 9	-0.40	28.20	14.65	6.87	-0.10	-1.03
Generation 10	-2.40	28.20	14.57	6.77	-0.10	-1.02
Generation 11	-1.20	28.40	14.57	6.92	-0.06	-1.11
Generation 12	-1.20	28.00	14.72	6.88	-0.09	-1.08
Generation 13	-1.90	28.00	14.50	6.92	-0.05	-1.10
Generation 14	-2.40	27.80	14.78	6.97	-0.15	-1.01
Generation 15	-1.90	28.20	14.65	6.79	-0.05	-1.04
Generation 16	-1.20	28.40	14.68	6.81	-0.10	-0.96
Generation 17	-1.90	28.40	14.47	6.84	-0.09	-0.99
Generation 18	-2.30	28.00	14.44	6.88	-0.08	-1.03
Generation 19	-1.90	28.40	14.71	6.75	-0.05	-0.98
Generation 20	-2.40	28.20	14.65	6.93	-0.10	-1.07
Mean_Generation	**-1.71**	**28.20**	**14.62**	**6.84**	**-0.08**	**-1.04**
Observed	**-2.00**	**28.80**	**15.90**	**6.77**	**-0.07**	**-1.09**

5. Results and Discussion

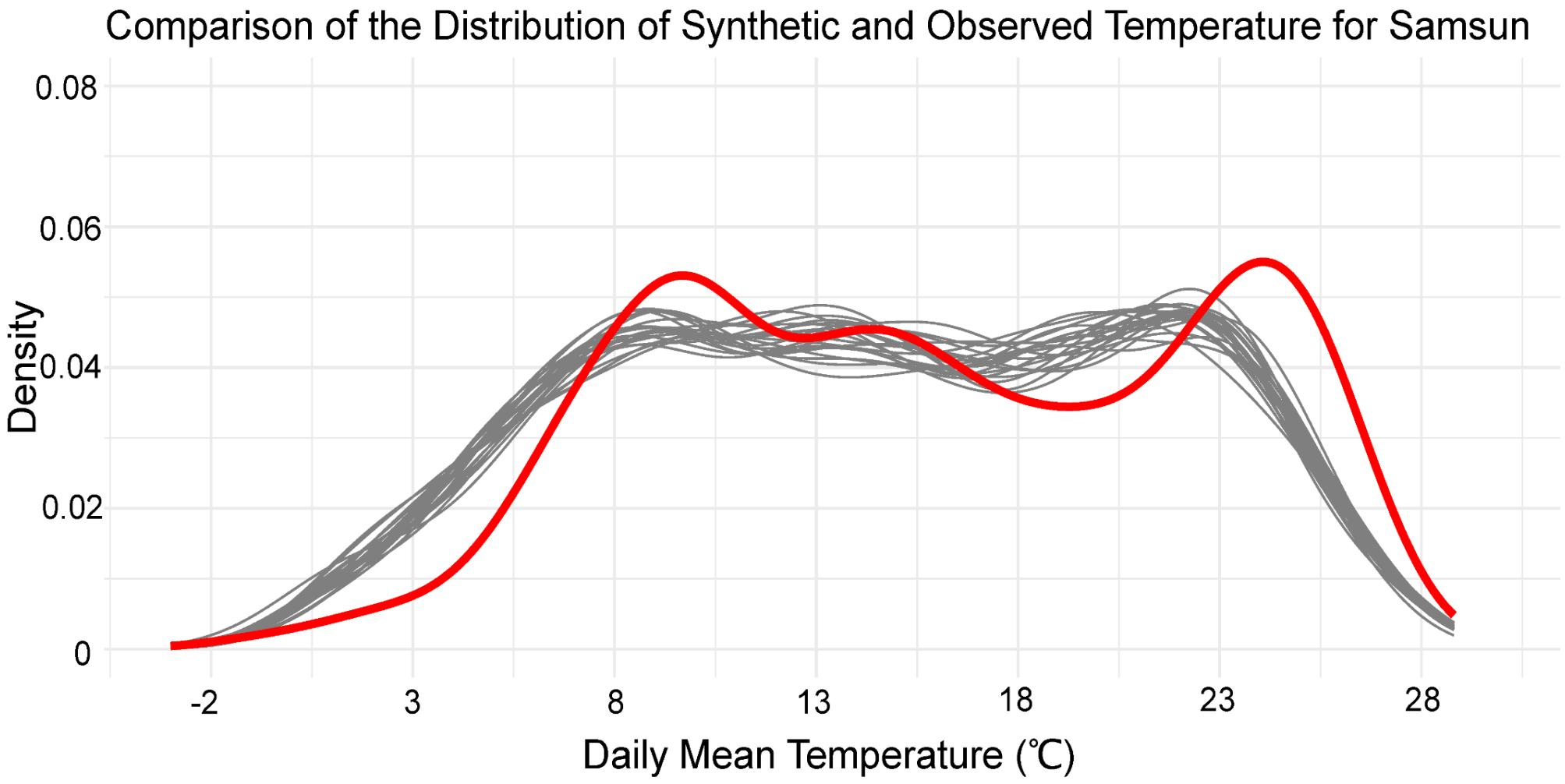
Statistics for the generated daily mean temperature values for Alanya

	Minimum (°C)	Maximum (°C)	Mean (°C)	Standard deviation (°C)	Skewness	Kurtosis
Generation 1	6.20	33.30	20.31	6.38	0.04	-1.26
Generation 2	3.30	31.90	20.38	6.40	0.01	-1.23
Generation 3	3.80	34.20	20.35	6.42	0.02	-1.24
Generation 4	5.60	34.30	20.38	6.42	0.09	-1.24
Generation 5	3.30	34.20	20.38	6.42	0.03	-1.22
Generation 6	3.30	32.80	20.33	6.39	0.04	-1.21
Generation 7	3.80	32.20	20.28	6.48	0.02	-1.19
Generation 8	3.75	32.30	20.38	6.42	0.02	-1.20
Generation 9	3.80	33.30	20.29	6.46	0.02	-1.26
Generation 10	6.20	35.40	20.26	6.43	0.06	-1.24
Generation 11	3.80	34.30	20.39	6.36	0.01	-1.24
Generation 12	6.20	34.21	20.36	6.31	0.04	-1.23
Generation 13	4.90	34.30	20.28	6.40	0.02	-1.15
Generation 14	4.00	34.20	20.35	6.42	0.01	-1.20
Generation 15	4.90	35.00	20.31	6.47	-0.01	-1.26
Generation 16	3.80	32.50	20.33	6.42	0.04	-1.21
Generation 17	3.80	32.30	20.39	6.42	0.01	-1.22
Generation 18	3.30	31.30	20.32	6.42	0.01	-1.23
Generation 19	3.30	34.30	20.28	6.37	0.00	-1.19
Generation 20	5.50	33.20	20.28	6.46	0.01	-1.21
Statistical mean values of generations	**4.33**	**33.48**	**20.33**	**6.41**	**0.02**	**-1.22**
Observed	**3.20**	**35.70**	**20.99**	**6.35**	**-0.03**	**-1.09**

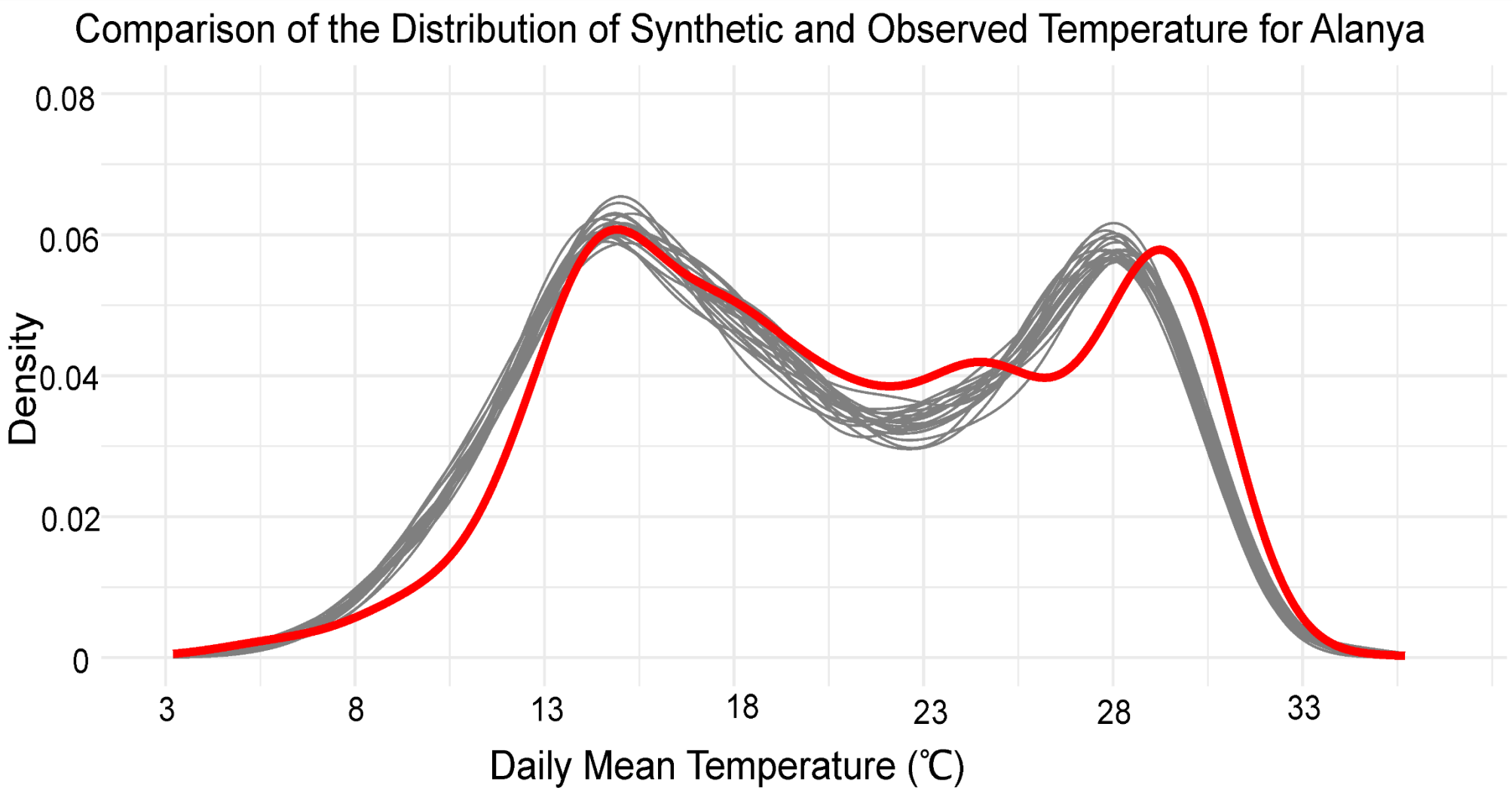
5. Results and Discussion

	Samsun			Alanya		
	NSE	RMSE (°C)	KGE	NSE	RMSE (°C)	KGE
Generation 1	0.524	4.670	0.767	0.800	2.838	0.901
Generation 2	0.512	4.728	0.765	0.817	2.720	0.909
Generation 3	0.534	4.621	0.771	0.798	2.854	0.900
Generation 4	0.541	4.586	0.774	0.792	2.899	0.897
Generation 5	0.512	4.727	0.761	0.799	2.845	0.900
Generation 6	0.528	4.650	0.769	0.802	2.826	0.902
Generation 7	0.490	4.833	0.745	0.796	2.872	0.898
Generation 8	0.551	4.534	0.780	0.806	2.799	0.903
Generation 9	0.521	4.684	0.767	0.804	2.813	0.902
Generation 10	0.536	4.610	0.772	0.797	2.858	0.900
Generation 11	0.525	4.662	0.770	0.811	2.761	0.906
Generation 12	0.535	4.617	0.773	0.808	2.784	0.903
Generation 13	0.516	4.709	0.766	0.793	2.893	0.897
Generation 14	0.531	4.634	0.773	0.806	2.796	0.904
Generation 15	0.508	4.746	0.759	0.792	2.899	0.896
Generation 16	0.531	4.635	0.770	0.798	2.855	0.900
Generation 17	0.521	4.682	0.767	0.808	2.783	0.904
Generation 18	0.490	4.833	0.754	0.802	2.829	0.901
Generation 19	0.505	4.761	0.756	0.800	2.839	0.901
Generation 20	0.540	4.592	0.776	0.809	2.773	0.905
Mean	**0.523**	**4.676**	**0.767**	**0.802**	**2.827**	**0.901**

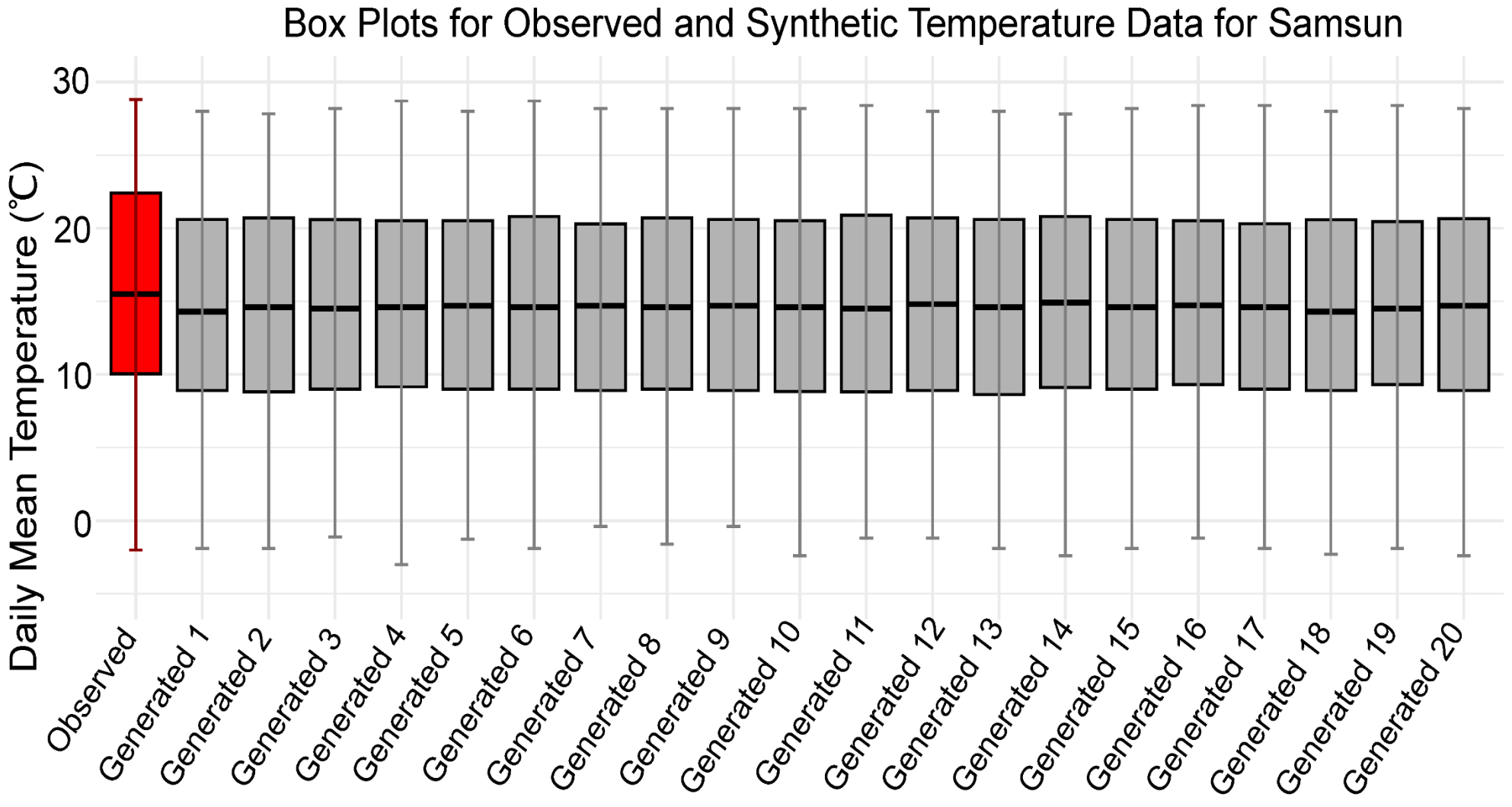
5. Results and Discussion



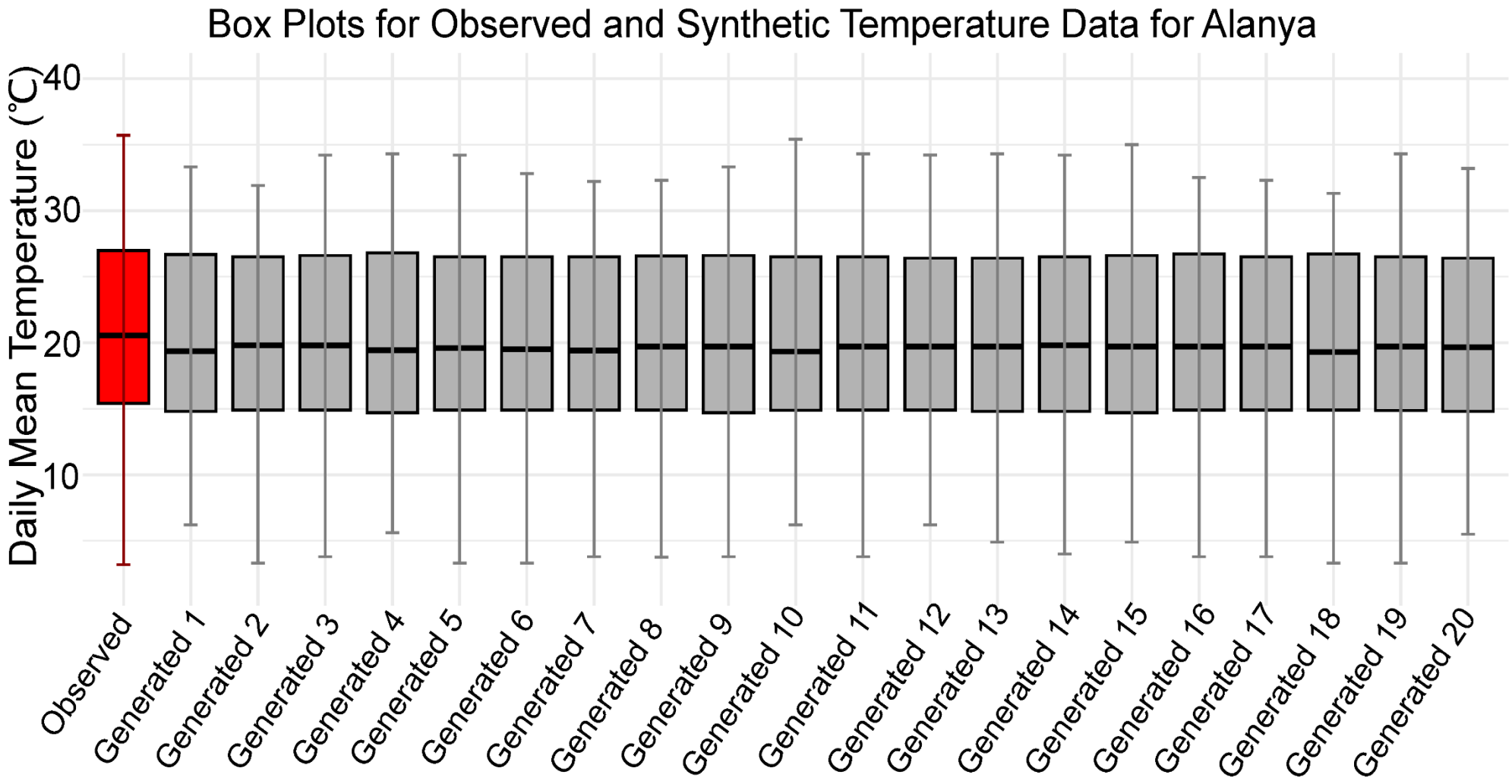
5. Results and Discussion



5. Results and Discussion



5. Results and Discussion



6. Conclusions

- In this research, the synthetic daily mean temperature data were simulated using QRF.
- For this purpose, minimum, maximum and mean temperature data of Samsun and Alanya in Türkiye, which have different climate characteristics, are used.
- The mean, minimum, maximum, standard deviation, skewness, and kurtosis statistics demonstrated that the synthetic temperature generation results are more promising in Alanya than in Samsun.
- Density and box plots also showed that temperature data generation is more satisfactory in Alanya.
- The poor performance in Samsun can be related to the more dramatic changes in mean temperature when considering the training and testing periods.

Thank you!

Cenk Sezen
cenk.sezen@omu.edu.tr

6. Reference

N. Meinshausen (2006) "Quantile Regression Forests", Journal of Machine Learning Research 7, 983-999 https://jmlr.csail.mit.edu/papers/v7/