European Geosciences Union
General Assembly 2022
Vienna | Austria | 23–27 May 2022

Session HS3.1: Hydroinformatics: data analytics, machine learning, systems analysis, optimization

Hydroclimatic time series analysis and clustering at multiple time scales

Georgia Papacharalampous[1], Hristos Tyralis[2], Yannis Markonis[1], and Martin Hanel[1]

[1] Department of Water Resources and Environmental Modeling, Faculty of Environmental Sciences, Czech University of Life Sciences, Kamýcá 129, Praha-Suchdol 16500, Prague, Czech Republic

[2] Air Force Projects Authority, Hellenic Air Force, Mesogion Avenue 227–231, 15561 Cholargos, Greece

papacharalampous.georgia@gmail.com @GeorgiaPapachar

Introduction

- The essential role of **time series models** (see, e.g., Hipel and McLeod 1994; Wei 2006) and **multi-scale time series analyses** (see, e.g., Markonis et al. 2018) towards a comprehensive understanding of hydroclimatic behaviours has been attentively discussed by Moss and Tasker (1987) and McKitrick and Christy (2019), respectively.
- Overviews and discussions on **large-sample hydroclimatic time series analyses** can be found in Papacharalampous et al. (2021, 2022b).
- We have proposed and extensively applied a **new methodological framework** for supporting hydroclimatic time series analyses that:
  - ✓ Are based on **large time series datasets**;
  - ✓ Focus on **multiple time series characteristics**;
  - ✓ Allow comparisons across **multiple time series types**;
  - ✓ Allow comparisons across **multiple time scales**;
  - ✓ Allow **cluster analyses** for multiple time series types and at multiple time scales.
- Such multi-scale time series analyses are missing from the previous literature.

Proposed methods and experimental design

Feature compilations for data science

(e.g., by Wang et al. 2006; Fulcher et al. 2013; Hyndman et al. 2015; Kang et al. 2017, 2020; Hyndman et al. 2020)

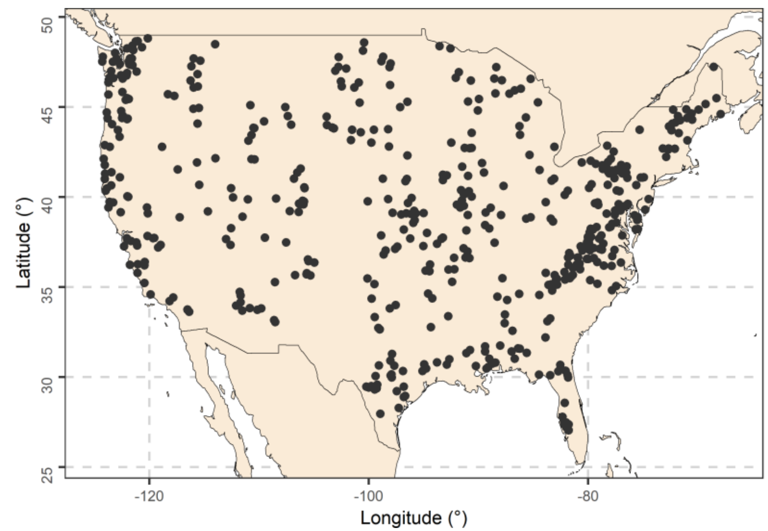
Technical and conceptual requirements

Proposed feature compilation

`x_acf1, x_acf10, diff1_acf1, diff1_acf10, diff2_acf1, diff2_acf10, seas_acf1, x_pacf5, diff1x_pacf5, diff2x_pacf5, seas_pacf, std1st_der, entropy, lumpiness, stability, nonlinearity, trend, spike, linearity, curvature, e_acf1, e_acf10, seasonal_strength`

CAMELS dataset

(Newman et al. 2015; Addor et al. 2017)



Summary of the conducted large-sample analyses

- ✓ Study of **511 geographical locations**;
- ✓ Study of **3 time series types**: temperature, precipitation, streamflow;
- ✓ Study of **9 temporal resolutions**: 1-day, 2-day, 3-day, 7-day, 0.5-month, 1-month, 2-month, 3-month, 6-month;
- ✓ Time series analysis: computation of **23 time series features**;
- ✓ Clustering using **random forests** (Breiman 2001; Tyralis et al. 2019);
- ✓ Feature importance in clustering via **explainable machine learning**.

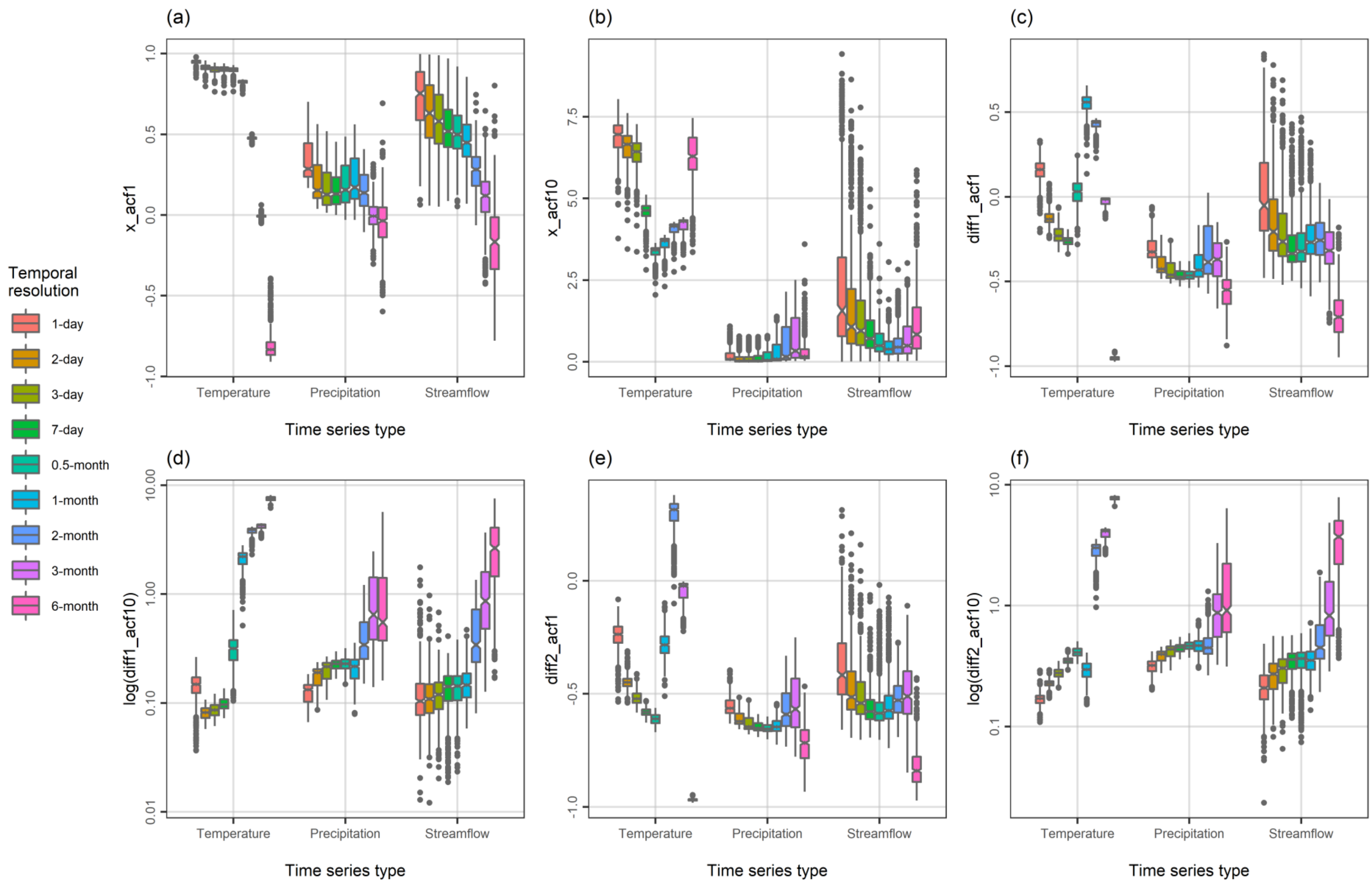
Analysis inputs and outputs

13 797 time series with varying lengths, types and temporal resolutions

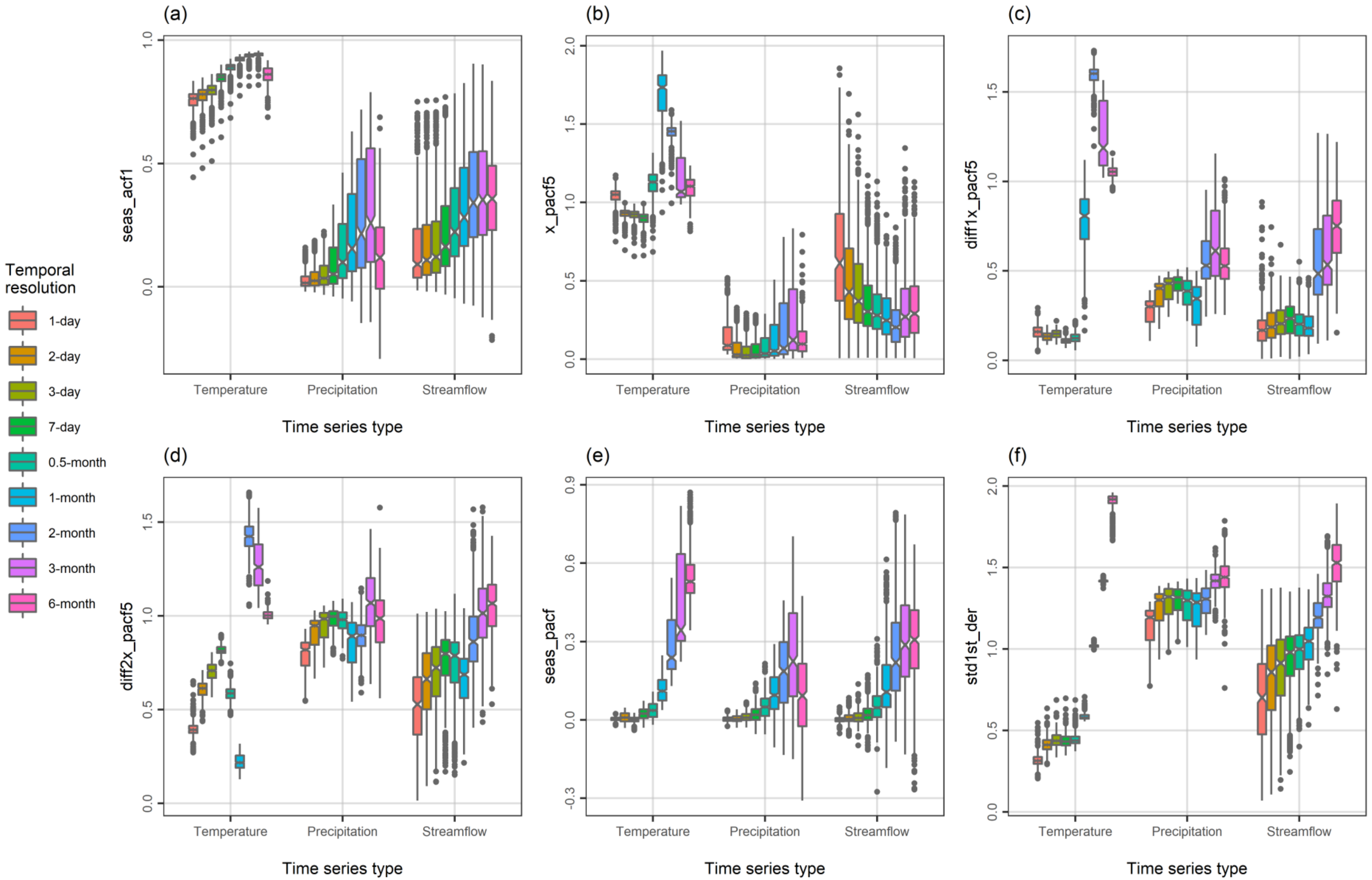
317 331 feature values

Interpretable clusters

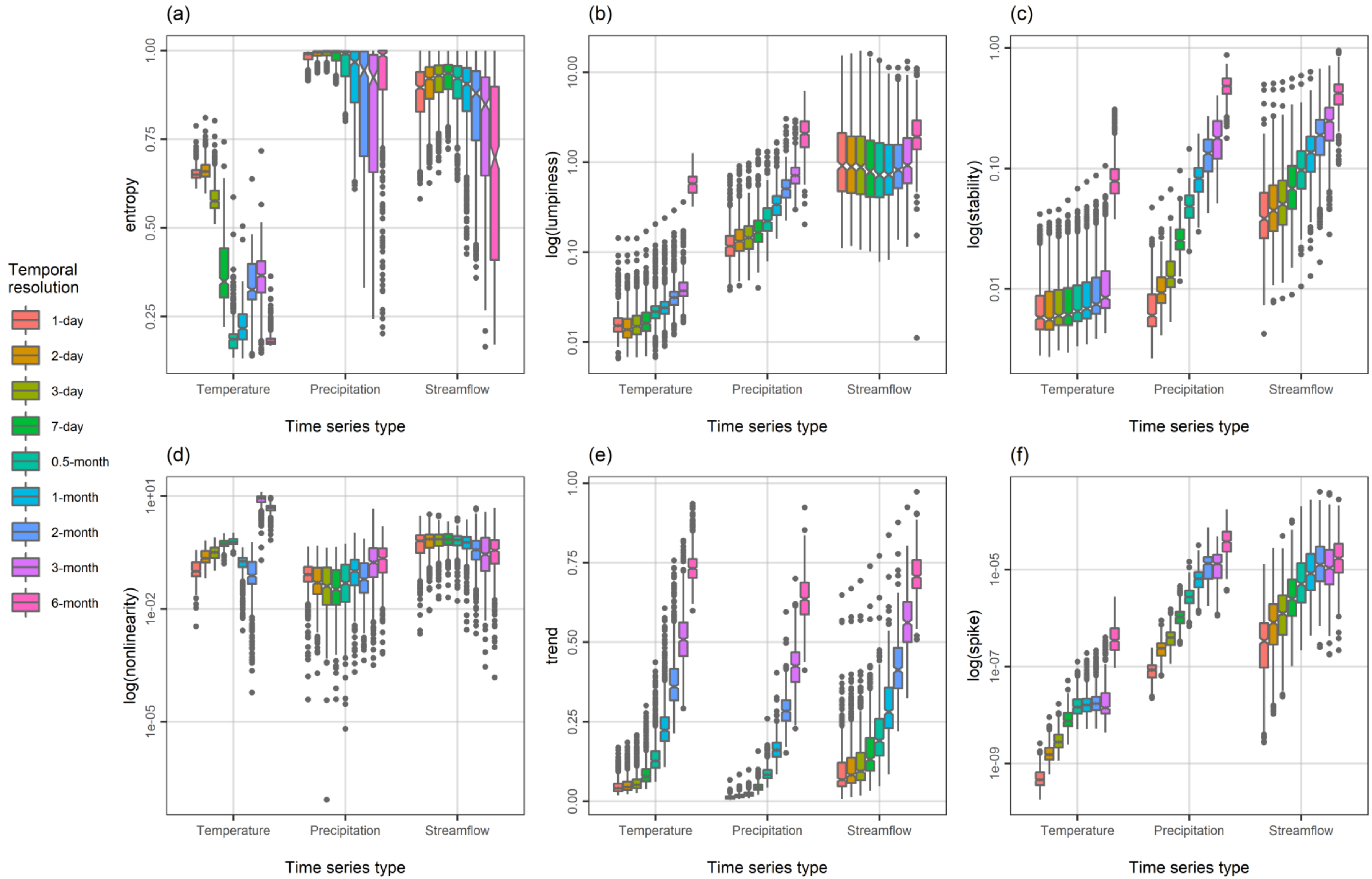
Boxplots of the feature values at multiple time scales



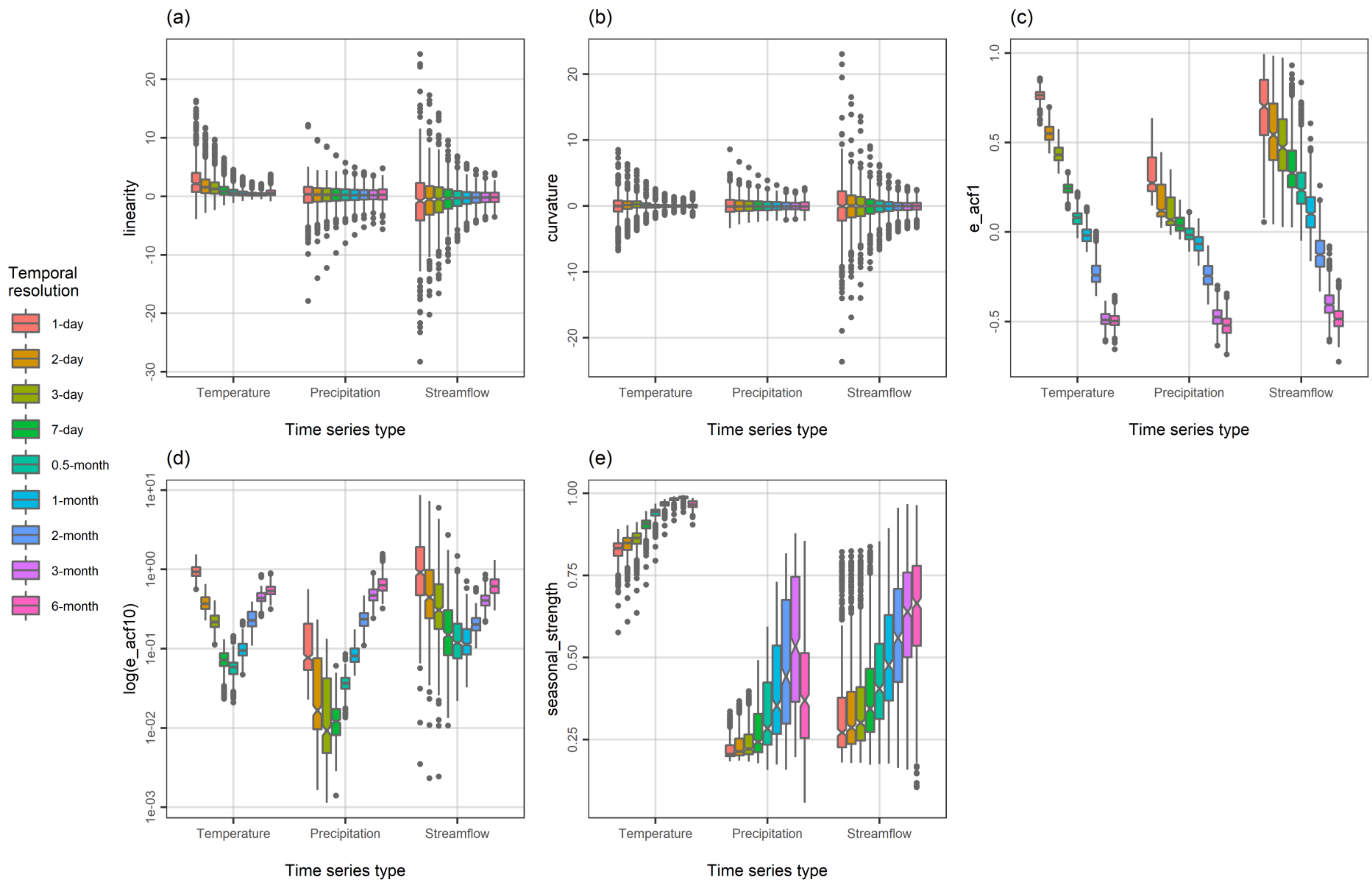
Boxplots of the feature values at multiple time scales



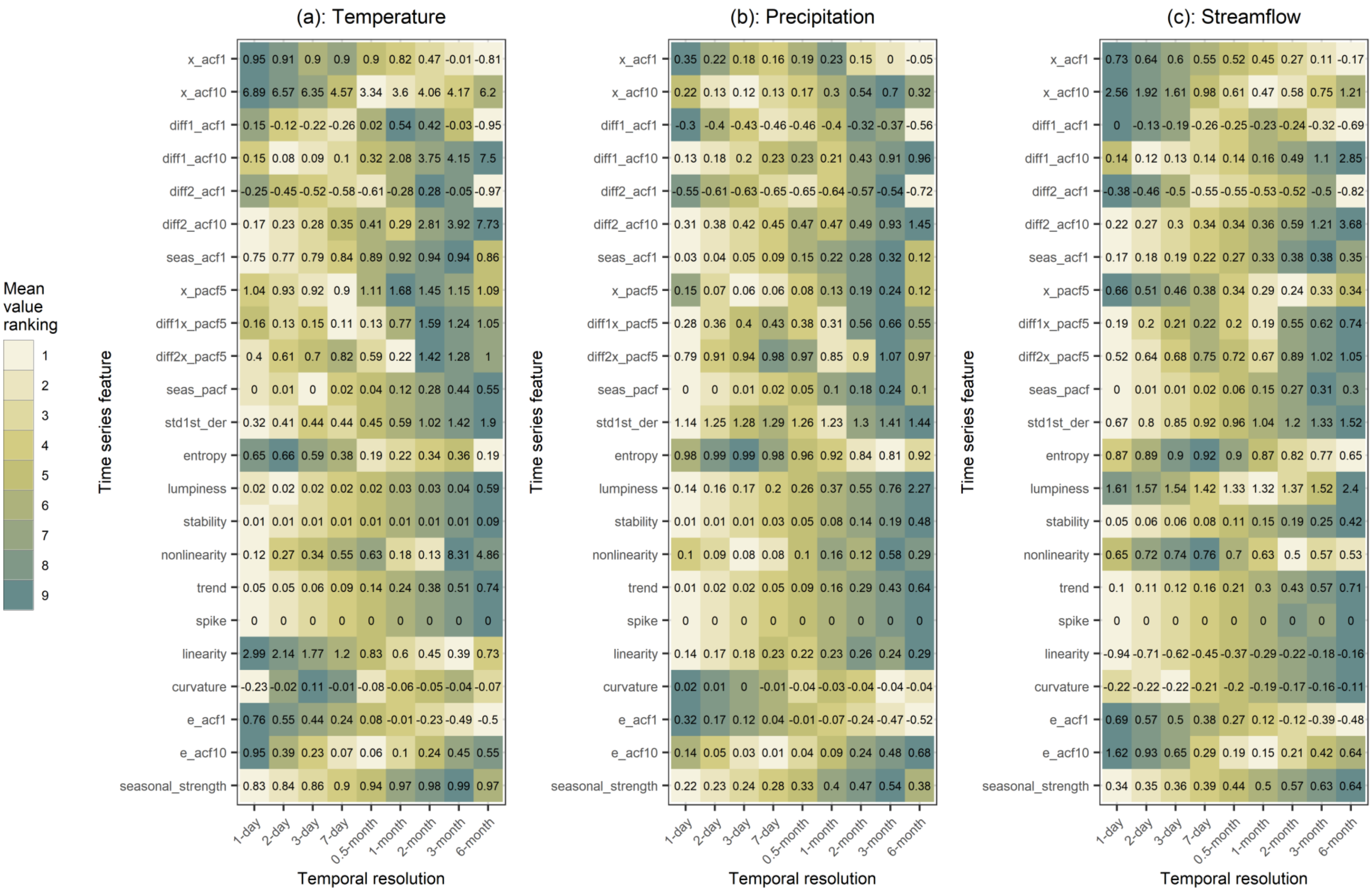
Boxplots of the feature values at multiple time scales



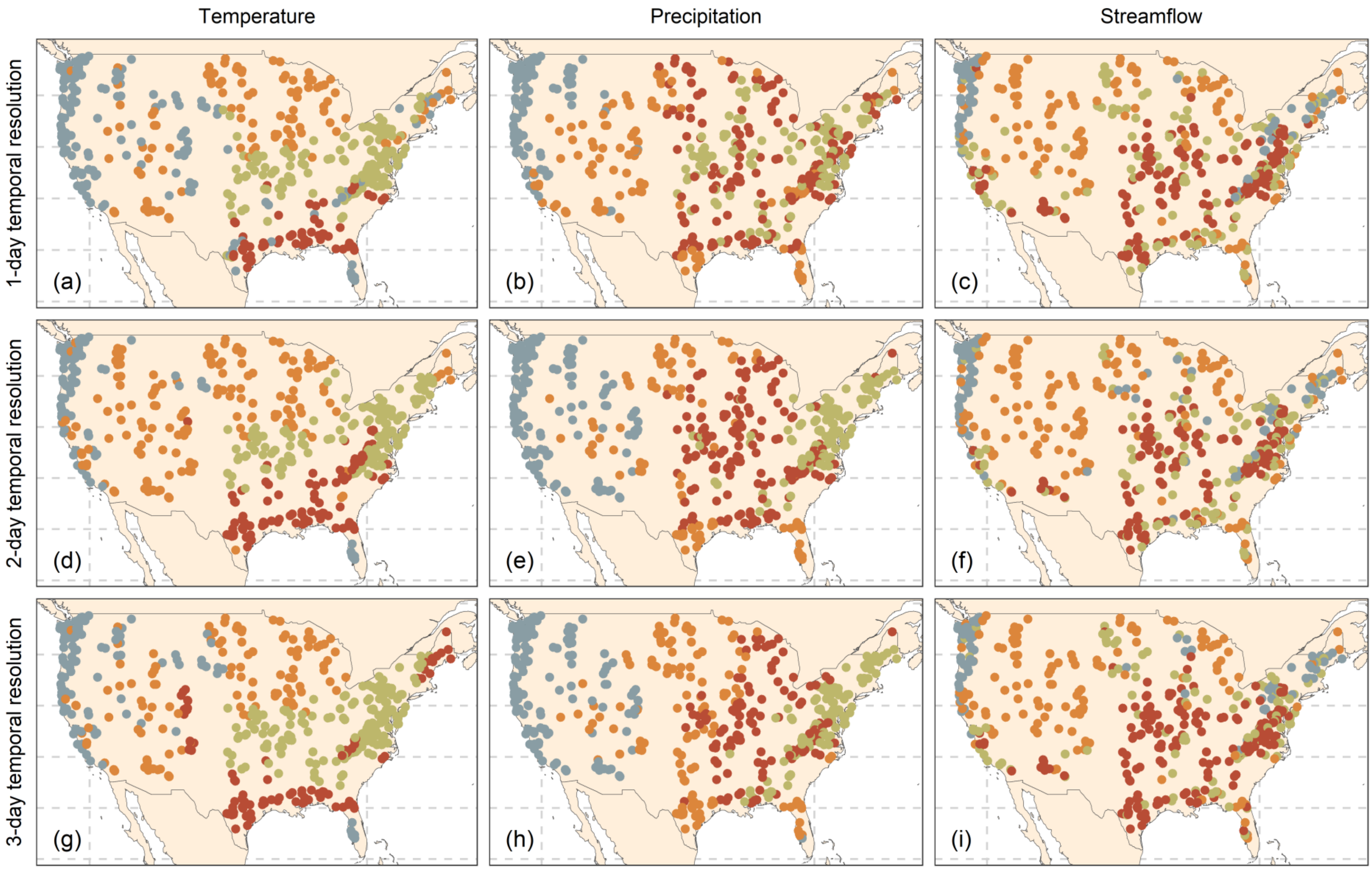
Boxplots of the feature values at multiple time scales



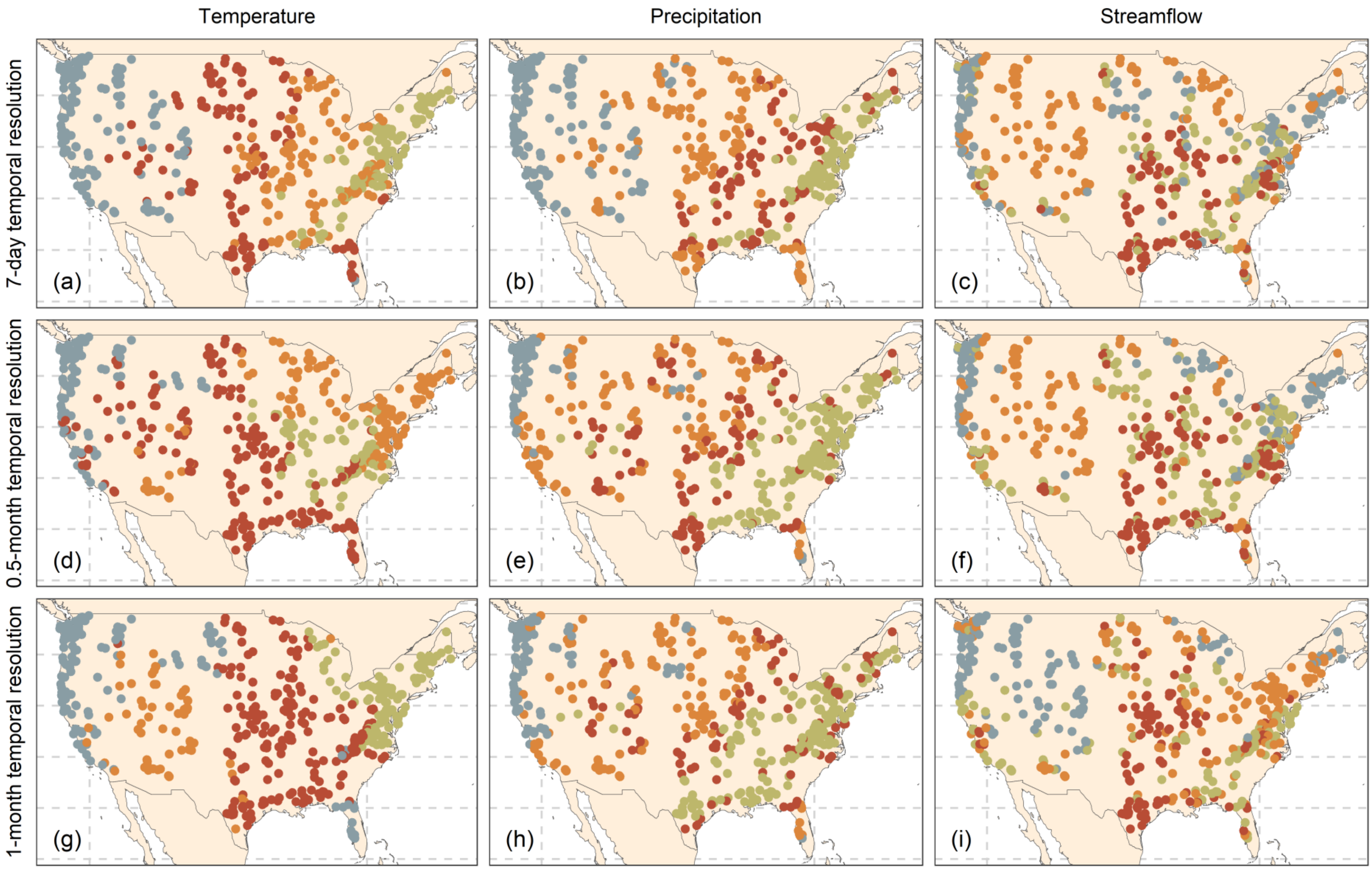
Means of the feature values at multiple time scales



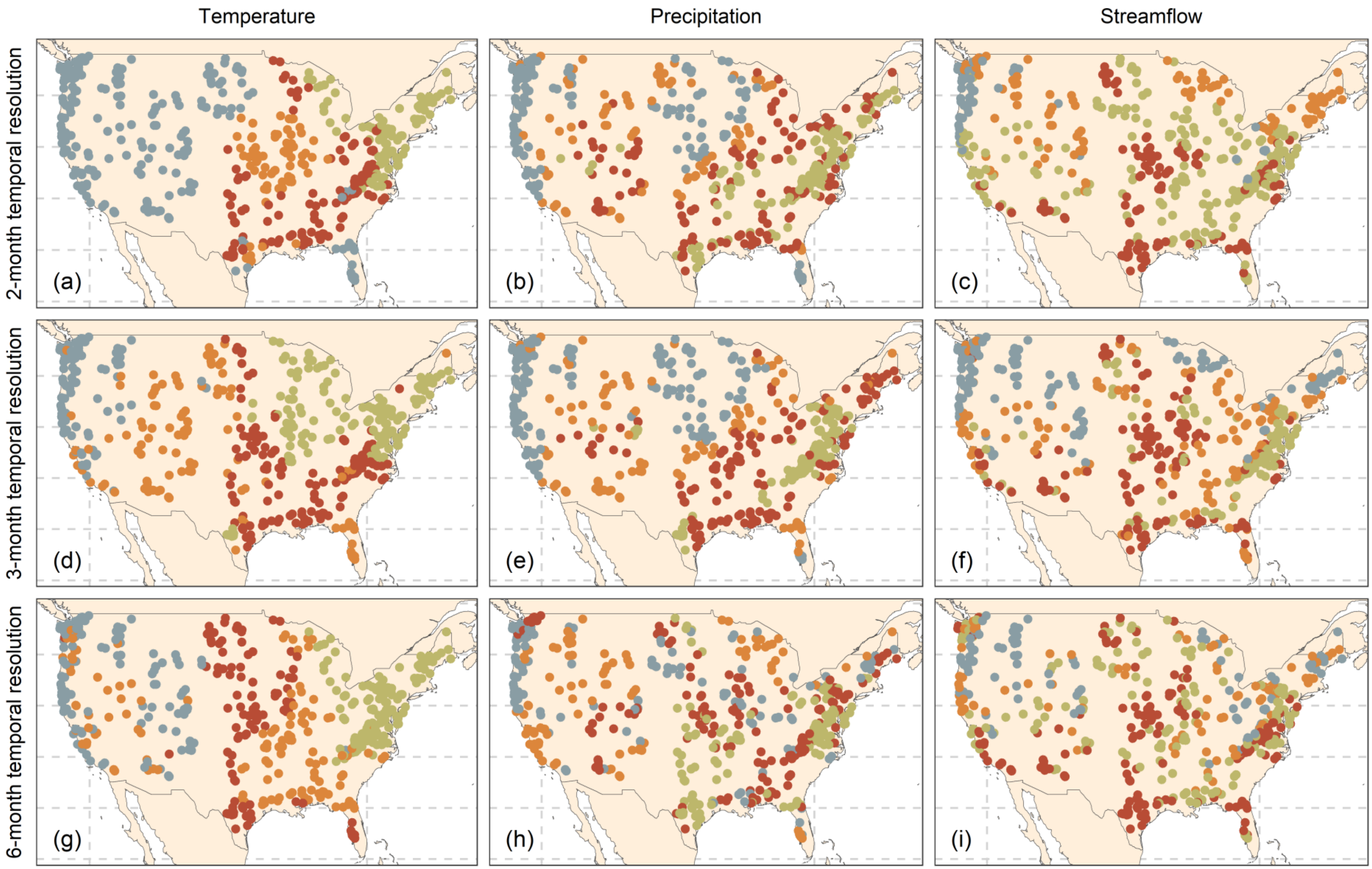
Hydroclimatic clusters at multiple time scales



Hydroclimatic clusters at multiple time scales



Hydroclimatic clusters at multiple time scales



Feature importance in clustering at multiple time scales

Ranking (1–23)

(a): Temperature

Time series feature	1-day	2-day	3-day	7-day	0.5-month	1-month	2-month	3-month	6-month
x_acf1	8	10	14	2	2	1	7	14	3
x_acf10	1	2	7	10	7	7	5	5	1
diff1_acf1	10	11	1	6	6	8	3	1	5
diff1_acf10	5	18	3	8	3	3	2	2	2
diff2_acf1	6	17	6	12	13	14	4	3	12
diff2_acf10	20	14	4	13	14	11	1	4	9
seas_acf1	4	3	10	15	8	9	11	16	8
x_pacf5	12	12	11	3	4	5	20	9	11
diff1x_pacf5	11	8	8	5	10	4	13	6	19
diff2x_pacf5	14	4	2	7	11	20	10	7	20
seas_pacf	23	19	23	22	20	12	8	8	13
std1st_der	7	9	13	1	1	2	6	13	6
entropy	2	6	15	9	12	22	21	21	22
lumpiness	18	20	19	21	17	15	18	17	14
stability	15	15	17	18	16	13	12	11	4
nonlinearity	22	22	20	16	22	18	15	12	16
trend	16	16	18	17	19	16	16	18	15
spike	9	13	16	19	9	10	14	15	10
linearity	19	21	21	20	21	19	19	19	17
curvature	21	23	22	23	23	23	23	22	23
e_acf1	13	1	5	4	15	17	17	20	18
e_acf10	17	7	12	11	18	21	22	23	21
seasonal_strength	3	5	9	14	5	6	9	10	7

Temporal resolution

(b): Precipitation

Time series feature	1-day	2-day	3-day	7-day	0.5-month	1-month	2-month	3-month	6-month
x_acf1	2	2	2	2	2	1	7	16	9
x_acf10	7	11	10	5	4	9	2	2	5
diff1_acf1	11	7	6	4	9	10	4	12	6
diff1_acf10	10	8	8	8	10	11	13	6	1
diff2_acf1	13	12	12	10	14	14	9	9	3
diff2_acf10	12	13	13	13	15	15	15	14	2
seas_acf1	17	15	15	14	8	7	3	1	7
x_pacf5	4	3	3	3	3	4	5	4	8
diff1x_pacf5	6	6	5	6	5	3	12	5	12
diff2x_pacf5	9	9	7	11	11	5	16	13	17
seas_pacf	23	22	22	19	16	13	11	7	11
std1st_der	1	1	1	1	1	2	8	15	10
entropy	8	10	9	7	6	8	6	8	16
lumpiness	21	21	21	22	21	22	21	22	19
stability	15	18	18	21	18	16	14	11	14
nonlinearity	20	17	16	15	12	17	22	17	23
trend	14	16	17	17	20	20	18	18	15
spike	18	19	19	18	13	12	10	10	13
linearity	19	20	20	20	19	19	17	21	21
curvature	22	23	23	23	23	23	23	23	22
e_acf1	3	4	4	9	17	18	20	19	18
e_acf10	5	5	11	16	22	21	19	20	20
seasonal_strength	16	14	14	12	7	6	1	3	4

Temporal resolution

(c): Streamflow

Time series feature	1-day	2-day	3-day	7-day	0.5-month	1-month	2-month	3-month	6-month
x_acf1	1	2	3	2	1	2	11	10	11
x_acf10	7	9	9	10	12	9	6	5	2
diff1_acf1	8	7	5	4	4	4	15	15	4
diff1_acf10	15	16	16	15	16	18	4	3	1
diff2_acf1	10	10	10	11	13	10	16	16	3
diff2_acf10	11	11	12	12	11	14	13	8	6
seas_acf1	13	13	14	14	8	8	1	2	8
x_pacf5	3	3	2	3	3	1	14	6	7
diff1x_pacf5	17	14	11	7	10	13	3	1	10
diff2x_pacf5	9	8	6	5	5	5	10	13	16
seas_pacf	23	23	21	20	20	16	7	11	13
std1st_der	2	1	1	1	2	3	9	9	9
entropy	6	5	7	8	6	6	5	7	15
lumpiness	20	19	19	19	19	20	17	18	18
stability	19	18	18	18	17	15	12	12	12
nonlinearity	18	20	20	21	23	23	23	20	19
trend	16	17	17	17	18	17	18	17	17
spike	14	15	15	16	15	11	8	14	14
linearity	21	21	22	22	21	22	21	23	23
curvature	22	22	23	23	22	21	20	21	21
e_acf1	4	4	4	6	9	12	19	19	20
e_acf10	5	6	8	9	14	19	22	22	22
seasonal_strength	12	12	13	13	7	7	2	4	5

Temporal resolution

Summary, further reading and future research

- A detailed **methodological framework** for multifaceted and automatic hydroclimatic time series analysis at **multiple time scales** has been proposed.
- Various **similarities and differences** between temperature, precipitation and streamflow time series have been identified with respect to the evolution patterns of their features with increasing (or decreasing) temporal resolution.
- **Feature-based clustering** has been performed for investigating the **spatial variability** of the temperature, precipitation and streamflow features across the contiguous United States and across temporal resolutions.
- The significance of using **a variety of features** in assessing hydroclimatic similarity has been acknowledged based on results obtained through **explainable machine learning**.
- Details on the proposed methodological framework and extensive discussions on the obtained results can be found in Papacharalampous et al. (2022a).
- Future research could, among others, extent the already conducted investigations to other temporal resolutions and time series types.

References

Addor N, Newman AJ, Mizukami N, Clark MP (2017) The CAMELS data set: Catchment attributes and meteorology for large-sample studies. Hydrology and Earth System Sciences 21:5293–5313. doi:10.5194/hess-21-5293-2017

Breiman L (2001) Random forests. Machine Learning 45(1):5–32. doi:10.1023/A:1010933404324

Fulcher BD, Little MA, Jones NS (2013) Highly comparative time-series analysis: The empirical structure of time series and their methods. Journal of the Royal Society Interface 10(83):20130048. doi:10.1098/rsif.2013.0048

Hipel KW, McLeod AI (1994) Time Series Modelling of Water Resources and Environmental Systems, Elsevier. ISBN 978-0-444-89270-6

Hyndman RJ, Wang E, Laptev N (2015) Large-scale unusual time series detection. 2015 IEEE International Conference on Data Mining Workshop (ICDMW), Atlantic City, NJ, pp. 1616–1619. doi:10.1109/ICDMW.2015.104

Hyndman RJ, Kang Y, Montero-Manso P, Talagala T, Wang E, Yang Y, O'Hara-Wild M (2020) tsfeatures: Time Series Feature Extraction. R package version 1.0.2. https://CRAN.R-project.org/package=tsfeatures

Kang Y, Hyndman RJ, Smith-Miles K (2017) Visualising forecasting algorithm performance using time series instance spaces. International Journal of Forecasting 33(2):345−358. doi:10.1016/j.ijforecast.2016.09.004

Kang Y, Hyndman RJ, Li F (2020) GRATIS: GeneRAting TIme Series with diverse and controllable characteristics. Statistical Analysis and Data Mining: The ASA Data Science Journal 13(4):354–376. doi:10.1002/sam.11461

Markonis Y, Hanel M, Máca P, Kyselý J, Cook ER (2018) Persistent multi-scale fluctuations shift European hydroclimate to its millennial boundaries. Nature Communications 9(1):1–12. doi:10.1038/s41467-018-04207-7

McKitrick R, Christy J (2019) Assessing changes in US regional precipitation on multiple time scales. Journal of Hydrology 578:124074. doi:10.1016/j.jhydrol.2019.124074

Moss ME, Tasker GD (1987) The role of stochastic hydrology in dealing with climatic variability. The Influence of Climate Change and Climatic Variability on the Hydrologie Regime and Water Resources. IAHS Publications 168:201–207

Newman AJ, Clark MP, Sampson K, Wood A, Hay LE, Bock A, Viger RJ, Blodgett D, Brekke L, Arnold JR, Hopson T, Duan Q (2015) Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance. Hydrology and Earth System Sciences 19:209–223. doi:10.5194/hess-19-209-2015

Papacharalampous GA, Tyralis H, Papalexiou SM, Langousis A, Khatami S, Volpi E, Grimaldi S (2021) Global-scale massive feature extraction from monthly hydroclimatic time series: Statistical characterizations, spatial patterns and hydrological similarity. Science of the Total Environment 767:144612. doi:10.1016/j.scitotenv.2020.144612

Papacharalampous GA, Tyralis H, Markonis Y, Hanel M (2022a) Hydroclimatic time series features at multiple time scales. arXiv:2112.01447

Papacharalampous GA, Tyralis H, Pechlivanidis IG, Grimaldi S, Volpi E (2022b) Massive feature extraction for explaining and foretelling hydroclimatic time series forecastability at the global scale. Geoscience Frontiers 13(3):101349. doi:10.1016/j.gsf.2022.101349

Tyralis H, Papacharalampous GA, Langousis A (2019) A brief review of random forests for water scientists and practitioners and their recent history in water resources. Water 11(5):910. doi:10.3390/w11050910

Wang X, Smith K, Hyndman RJ (2006) Characteristic-based clustering for time series data. Data Mining and Knowledge Discovery 13:335–364. doi:10.1007/s10618-005-0039-x

Wei WWS (2006) Time Series Analysis, Univariate and Multivariate Methods, second edition. Pearson Addison Wesley