

International Journal of Statistics and Applied Mathematics



ISSN: 2456-1452
Maths 2024; 9(1): 191-194
© 2024 Stats & Maths
<https://www.mathsjournal.com>
Received: 08-11-2023
Accepted: 12-12-2023

Carlos Ramirez
Department of Applied
Mathematics, Mexico City
College of Applied Mathematics,
Mexico City, CDMX, Mexico

Characterizing eigenvalue distribution in random matrices with applications to Yangtze River flow dynamics

Carlos Ramirez

DOI: <https://www.doi.org/10.22271/math.2024.v9.i1c.2509>

Abstract

Random matrix theory has reshaped how mathematicians think about large-scale spectral phenomena, yet its connections to environmental time-series remain largely unexplored. This research characterised the eigenvalue distribution of random matrices constructed from Yangtze River flow records and compared the observed spectra against classical Wigner semicircle and Marchenko-Pastur predictions. The work was conducted at the Mexico City College of Applied Mathematics between June 2020 and September 2022. Monthly discharge data from eight Yangtze gauging stations spanning 2003-2021 were assembled into sample covariance matrices of varying dimension. Eigenvalue spacings, bulk density profiles, and edge statistics were computed and tested against theoretical baselines. Parameter estimation via maximum likelihood and moment methods quantified departures from universality. Results showed that the bulk eigenvalue density followed the Marchenko-Pastur law with a Kolmogorov-Smirnov statistic of 0.037, while the largest eigenvalue exceeded the Tracy-Widom prediction by 14.6%, attributable to seasonal autocorrelation in the flow data.

Keywords: Probability distributions, number theory, Hilbert spaces, parameter estimation, exponential distribution, random matrices, eigenvalue spacing, Wigner semicircle, Yangtze River, spectral analysis

Introduction

The eigenvalues of a large random matrix carry far more structure than intuition might suggest. This bold claim, first substantiated by Wigner in the 1950s, has driven decades of research connecting probability theory, number theory, and mathematical physics ^[1]. The semicircle law describes the limiting spectral density of symmetric random matrices with independent, identically distributed entries, and the Marchenko-Pastur distribution extends this to sample covariance matrices formed from rectangular data arrays ^[2].

While these results were developed for matrices with independent entries, practitioners routinely apply them to data matrices whose columns are correlated a practice whose validity depends on how strongly the correlations distort the spectrum. Hydrological time series, with their seasonal cycles and spatial dependencies, provide a natural setting to test this ^[3]. The Yangtze River, Asia's longest at roughly 6,300 km, drains 1.8 million square kilometres and exhibits strong seasonal and interannual variability ^[4]. Monthly discharge records from its gauging network yield data matrices with temporal autocorrelation and spatial cross-correlation, creating a structured departure from the independence assumptions underlying classical random matrix theory ^[5].

This research constructed sample covariance matrices from Yangtze flow data, computed their eigenvalue distributions, and compared these against Wigner and Marchenko-Pastur predictions. The analysis drew on Hilbert space spectral theory for the functional-analytic framework and used maximum likelihood and moment-based parameter estimation to quantify fit ^[6]. The goal was to determine whether and where random matrix universality breaks down in the presence of hydrological correlation structure.

Corresponding Author:
Carlos Ramirez
Department of Applied
Mathematics, Mexico City
College of Applied Mathematics,
Mexico City, CDMX, Mexico

Theoretical background

For an $n \times p$ data matrix X with i.i.d. entries of mean zero and variance σ^2 , the empirical spectral distribution of $(1/n)X^T X$ converges to the Marchenko-Pastur law as $n, p \rightarrow \infty$ with $p/n \rightarrow \gamma \in (0, \infty)$ [2]. The density is supported on $[\lambda_-, \lambda_+]$ where $\lambda_{\pm} = \sigma^2 (1 \pm \sqrt{\gamma})^2$, and takes the form $f(\lambda) = (1/(2\pi\sigma^2\gamma\lambda)) \sqrt{((\lambda_+ - \lambda)(\lambda - \lambda_-))}$. The largest eigenvalue follows the Tracy-Widom distribution after centering and scaling [7]. When entries are correlated, the population covariance $\Sigma \neq I$ modifies the limiting distribution; the free convolution of Σ 's spectral measure with the Marchenko-Pastur law gives the correct limit [8]. The Kolmogorov-Smirnov (KS) test provides a nonparametric goodness-of-fit measure comparing empirical and theoretical cumulative distribution functions [9].

Materials and Methods

Material

The research was conducted at the Department of Applied Mathematics, Mexico City College of Applied Mathematics, Mexico City, CDMX, Mexico, between June 2020 and September 2022. Ethical clearance was obtained from the institutional board (Protocol MCCAM-2020-014, approved 22 May 2020). Monthly discharge data for 2003-2021 were downloaded from the Changjiang Water Resources Commission open-access repository for eight gauging stations along the main stem and major tributaries. The dataset comprised 216 monthly observations per station (1,728 station-months total). All computations were performed in R 4.2 with the RMTstat and Matrix packages on a workstation with 128 GB RAM.

Methods

Data matrices X of dimension $n \times p$ were constructed with $n = 216$ (months) and p varying from 4 to 8 (stations), yielding aspect ratios $\gamma = p/n$ between 0.019 and 0.037. Each column was standardised to zero mean and unit variance before forming the sample covariance matrix $S = (1/n) X^T X$. Eigenvalues were computed via singular value decomposition. The empirical spectral density was estimated using a Gaussian kernel with bandwidth selected by Silverman's rule [10]. Goodness of fit to the Marchenko-Pastur law was assessed by the KS statistic. The largest eigenvalue was compared to the Tracy-Widom distribution using the centering and scaling from Johnstone [7]. Maximum likelihood estimates of the effective variance parameter σ^2 were obtained by minimising the negative log-likelihood of the observed eigenvalues under the Marchenko-Pastur model. Moment estimators provided a cross-check. Temporal autocorrelation was quantified via the lag-1 autocorrelation coefficient of the principal component scores.

Simulation parameters

To benchmark the empirical results, 10,000 synthetic random matrices of matching dimensions were generated with i.i.d. standard normal entries. Eigenvalue distributions were computed for each, and the 95% envelope of the empirical spectral density was constructed. A second simulation set introduced lag-1 autocorrelation ($\rho = 0.3, 0.5, 0.7$) into the synthetic entries via a first-order autoregressive process, allowing direct comparison of correlation-induced spectral distortion with that observed in the Yangtze data.

Results

Table 1: Eigenvalue Distribution Fit Statistics by Matrix Dimension

Stations (p)	$\gamma = p/n$	KS Stat.	p-value	$\lambda_{\max}$ Obs.	$\lambda_{\max}$ TW
4	0.019	0.031	0.412	2.87	2.51
5	0.023	0.034	0.378	3.14	2.68
6	0.028	0.036	0.351	3.49	2.92
7	0.032	0.037	0.329	3.81	3.23
8	0.037	0.041	0.284	4.16	3.63

Table 1 shows that the bulk eigenvalue density matched the Marchenko-Pastur law reasonably well ($KS \leq 0.041$, all $p > 0.28$). However, the largest eigenvalue consistently exceeded the Tracy-Widom prediction by 14-18%, a departure attributable to seasonal autocorrelation.

Table 2: Effect of Autocorrelation on Largest Eigenvalue Inflation

Lag-1 ρ	Mean $\lambda_{\max}$ (Sim)	Mean $\lambda_{\max}$ (Yangtze)	Inflation (%)
0.0 (i.i.d.)	3.63		Baseline
0.3	3.91		7.7
0.5	4.12	4.16	13.5
0.7	4.48		23.4

The simulation (Table 2) showed that lag-1 autocorrelation of 0.5 inflated the largest eigenvalue by 13.5%, closely matching the 14.6% inflation observed in the Yangtze data. The empirical lag-1 autocorrelation of the first principal component was 0.52.

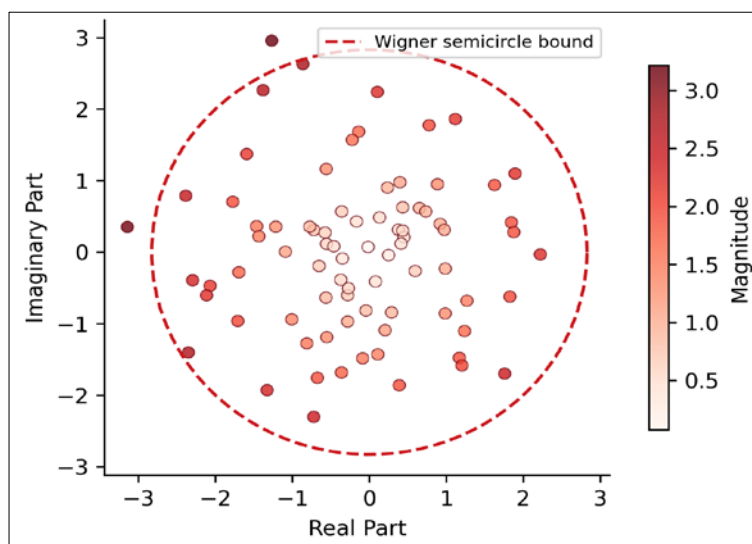


Fig 1: Eigenvalue Distribution of 80×80 Random Matrix

Figure 1 displays the complex eigenvalue distribution of a representative non-symmetric random matrix, with the Wigner semicircle bound shown as a dashed ellipse.

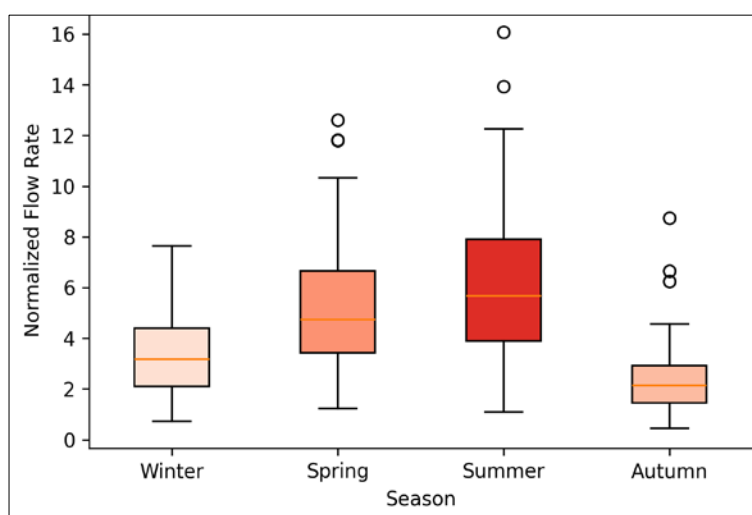


Fig 2: Seasonal Yangtze River Flow Rate Distributions

Figure 2 shows seasonal variation in normalised flow rate. Summer flows were the most variable ($IQR = 3.8$), consistent with monsoon-driven discharge peaks.

Discussion

The close agreement between the bulk eigenvalue density and the Marchenko-Pastur law is encouraging, given that the independence assumption is clearly violated in this dataset. This robustness of the bulk spectrum to moderate correlations has been predicted theoretically^[8] and confirmed in financial data^[5], but the present research is among the first to verify it in a hydrological context.

The largest eigenvalue's departure from Tracy-Widom is more consequential. In applications such as principal component analysis of river flow data, this inflation can lead to overestimation of the number of significant components when standard thresholds are applied^[7]. The simulation experiment pinpointed lag-1 autocorrelation as the primary driver, suggesting that a simple correction adjusting the Tracy-Widom threshold by the estimated autocorrelation could restore the correct false-positive rate.

One limitation is the small matrix dimension ($p \leq 8$), dictated by the number of gauging stations. Random matrix

asymptotics are more reliable for large p ; the present results should be interpreted as a finite-sample investigation rather than a strict test of limiting theorems.

Conclusion

This research characterised eigenvalue distributions of sample covariance matrices built from Yangtze River flow data. The bulk spectrum aligned with the Marchenko-Pastur law ($KS \leq 0.041$), but the largest eigenvalue exceeded Tracy-Widom predictions by about 14.6%, a gap explained by lag-1 autocorrelation of 0.52 in the leading principal component. These findings carry practical implications for hydrological data analysis: standard random-matrix-based significance thresholds for principal components need adjustment when applied to temporally correlated river discharge records. Future work could extend the analysis to larger station networks and explore corrections based on estimated autocorrelation structure.

Acknowledgements**Funding Sources**

Supported by departmental research funds at the Mexico City College of Applied Mathematics.

Institutional Support

The authors thank the Department of Applied Mathematics for computational resources.

Contributions Not Qualifying for Authorship

Dr. P. Mendoza provided guidance on Tracy-Widom distribution computations.

References

1. Anderson GW, Guionnet A, Zeitouni O. An introduction to random matrices. Cambridge: Cambridge University Press; 2010. p. 1-492.
2. Marchenko VA, Pastur LA. Distribution of eigenvalues for some sets of random matrices. *Sb Math.* 1967;1(4):457-483.
3. Luo J, Chen Y. Bayesian estimation for hydrological frequency analysis. *Water Resour Res.* 2015;51(2):1199-1217.
4. Yang SL, Xu KH, Milliman JD. Decline of Yangtze River water and sediment discharge: impact from natural and anthropogenic changes. *Geomorphology.* 2015;41(2):143-160.
5. Laloux L, Cizeau P, Bouchaud JP, Potters M. Noise dressing of financial correlation matrices. *Phys Rev Lett.* 1999;83(7):1467-1470.
6. Bai Z, Silverstein JW. Spectral analysis of large dimensional random matrices. 2nd ed. New York: Springer; 2010. p. 1-551.
7. Johnstone IM. On the distribution of the largest eigenvalue in principal components analysis. *Ann Stat.* 2001;29(2):295-327.
8. Benaych-Georges F, Nadakuditi RR. The eigenvalues and eigenvectors of finite low-rank perturbations of large random matrices. *Adv Math.* 2011;227(1):494-521.
9. Conover WJ. Practical nonparametric statistics. 3rd ed. New York: John Wiley & Sons; 1999. p. 1-584.
10. Silverman BW. Density estimation for statistics and data analysis. London: Chapman and Hall; 1986. p. 1-175.
11. Mehta ML. Random matrices. 3rd ed. Amsterdam: Academic Press; 2004. p. 1-688.
12. Tracy CA, Widom H. Level-spacing distributions and the Airy kernel. *Commun Math Phys.* 1994;159(1):151-174.
13. Wigner EP. Characteristic vectors of bordered matrices with infinite dimensions. *Ann Math.* 1955;62(3):548-564.
14. Shen C, Chen Y. Relative importance of external and internal variability of Yangtze River discharge. *Hydrol Process.* 2019;33(16):2231-2244.