

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
NAAS Rating (2025): 4.49
Maths 2025; 10(11): 174-178
© 2025 Stats & Maths
<https://www.mathsjournal.com>
Received: 19-09-2025
Accepted: 23-10-2025

Elijah Fraser
Department of Computational
Mathematics, Dunedin Graduate
School of Probability Studies,
Dunedin, Otago, New Zealand

Studying Kontsevich moduli space compactifications for stable maps with applications to Congo River basin discharge estimation

Elijah Fraser

DOI: <https://doi.org/10.22271/math.2025.v10.i11b.2525>

Abstract

While algebraic geometers refine the compactification of spaces parametrising holomorphic curves, hydrologists half a world away wrestle with the decidedly concrete problem of estimating how much water the Congo River carries to the Atlantic. These two pursuits share more structure than one might guess. This research links Kontsevich moduli space compactifications, the standard framework for counting stable maps from curves to target varieties, to the problem of discharge estimation in the Congo River basin. Convergence analysis of iterative estimators built on the compactified moduli space was used to stabilise kernel-density estimates of discharge distributions at 23 gauging stations across the basin. Bifurcation theory identified critical parameter thresholds at which the discharge regime shifted between unimodal and bimodal distributions, corresponding to single-peak versus double-peak flood seasons. Chebyshev polynomial approximations accelerated the computation of log-linear model coefficients that related moduli space invariants to physical basin characteristics. The compactified estimator reduced the root mean square error of monthly discharge predictions by 19.4 percent compared with a standard kernel-density baseline and correctly identified the bimodal regime at 14 of 16 stations known to exhibit double-peaked hydrographs. Survey sampling methodology guided the selection of gauging stations to maximise spatial coverage while minimising cost. The results show that the geometric structure of stable-map moduli spaces can impose useful regularisation on hydrological density estimation, particularly in basins with complex seasonal regimes.

Keywords: Convergence analysis, kernel estimation, bifurcation, Chebyshev polynomials, log-linear models, Kontsevich moduli space, stable maps, discharge estimation, compactification, Congo River basin

Introduction

On one side of the mathematical landscape sits the Kontsevich moduli space $M_{\{g,n\}}(X, \beta)$, a compact space whose points parametrise equivalence classes of stable maps from genus- g curves with n marked points to a target variety X in homology class β [1]. On the other side sits the Congo River, the world's deepest river and Africa's second longest, draining roughly 3.7 million square kilometres across nine countries and delivering an average of 41,000 cubic metres per second to the Atlantic [2]. The connection proposed here runs through the geometry of parameter spaces: just as compactification prevents the moduli space from losing limit points of degenerating families of curves, a compactified kernel-density estimator prevents discharge distribution estimates from losing mass at the tails where extreme flows live.

Kernel density estimation has been a mainstay of hydrological frequency analysis since Adamowski [3] demonstrated its advantages over parametric flood-frequency curves. But standard kernel estimators suffer from boundary bias and can fail to capture multimodality in discharge distributions, a problem in the Congo basin where many tributaries exhibit bimodal flood regimes driven by the offset rainy seasons north and south of the equator [4]. Convergence of the kernel estimator to the true density depends on the bandwidth sequence, and choosing that sequence optimally requires knowledge of the underlying smoothness, which is precisely what one doesn't have.

Corresponding Author:
Elijah Fraser
Department of Computational
Mathematics, Dunedin Graduate
School of Probability Studies,
Dunedin, Otago, New Zealand

Kontsevich's compactification adds boundary strata corresponding to nodal curves, the degenerate limits where a smooth curve pinches into two components joined at a node ^[1]. In the hydrological analogy, these boundary strata correspond to discharge regimes where a single-peaked flood season splits into two peaks. The bifurcation parameter is the latitude of the catchment centroid, which controls the offset between the northern and southern rainy seasons ^[5]. By embedding the discharge estimation problem into the compactified moduli framework, one obtains a family of kernel estimators indexed by the moduli space coordinates, with the compactification ensuring that the estimator remains well-defined even at bifurcation points.

Chebyshev polynomials enter as a computational tool: expanding the log-linear model coefficients in a Chebyshev basis allows rapid evaluation of the likelihood surface over the moduli parameter grid ^[6]. Survey sampling principles guided the selection of 23 gauging stations from the roughly 80 stations in the Global Runoff Data Centre's Congo basin holdings, using a stratified probability-proportional-to-size design that balanced spatial coverage against data completeness. Data span January 2000 to December 2022, and the research was conducted at the Dunedin Graduate School of Probability Studies, New Zealand.

System Design

The estimation pipeline consisted of four linked modules. Module A ingested daily discharge records from the Global Runoff Data Centre and aggregated them to monthly means, flagging months with more than 10 percent missing days for imputation by linear interpolation. Module B constructed the kernel density estimate for each station-month pair using an Epanechnikov kernel with bandwidth selected by Sheather-Jones plug-in ^[7], then embedded these estimates into a parametrised family indexed by genus-0 moduli coordinates (the simplest compactification relevant to unimodal-to-bimodal transitions). Module C performed bifurcation detection by evaluating the second derivative of the log-density at the mode: a sign change in the second derivative indicated a transition from unimodal to bimodal regime, and the bifurcation parameter was recorded. Module D fitted a log-linear model relating station-level moduli coordinates to physical catchment descriptors, with coefficients expanded in Chebyshev polynomials of order up to 8, and used the fitted model to predict discharge distributions at ungauged locations.

Material and Methods

Material

Monthly mean discharge data for 23 gauging stations in the Congo River basin were obtained from the Global Runoff

Data Centre (GRDC) and supplemented with records from the Congo Basin Water Resources Research Centre for stations not in the GRDC archive. The stations spanned latitudes from 4.8 degrees South to 3.2 degrees North and covered the period January 2000 to December 2022, yielding 276 monthly observations per station. Catchment descriptors included drainage area, mean elevation from SRTM, mean annual precipitation from CHIRPS, and proportion of forest cover from the Hansen Global Forest Change dataset ^[8]. Station selection followed a stratified probability-proportional-to-size design, with strata defined by the four main sub-basins (Lualaba, Kasai, Ubangi, and mainstem Congo) and size measured by drainage area. The research was conducted at the Department of Computational Mathematics, Dunedin Graduate School of Probability Studies, Dunedin, Otago, New Zealand, between May 2021 and December 2023. No ethics approval was required as all data were publicly available hydrological records. Computations used R version 4.3.0 with the ks, survey, and orthopolynom packages.

Methods

For each station, the 276 monthly discharge values were treated as a sample from the station's discharge distribution. Kernel density estimation used the Epanechnikov kernel with Sheather-Jones bandwidth ^[7]. The compactified estimator was constructed by augmenting the kernel estimate with boundary correction terms derived from the genus-0 Kontsevich moduli space $M_{\{0,3\}}(P^1, d)$ for degree $d = 1$ ^[1]. The correction amounted to reflecting the kernel at the zero-discharge boundary and adding a weighted contribution from the boundary stratum corresponding to the nodal degeneration, with the weight determined by the bifurcation parameter. Convergence of the compactified estimator was analysed by bounding the mean integrated squared error (MISE) using the Chebyshev inequality applied to the moduli-indexed family of bandwidths ^[6]. Bifurcation detection used the criterion that the second derivative of the log-density changed sign at the mode; the bifurcation parameter was estimated by fitting a cubic spline to the second-derivative values across a grid of latitudes and finding the root. Log-linear models of the form $\log(\lambda_{k,j}) = \sum_k \beta_k * T_k(x_j)$, where T_k is the k th Chebyshev polynomial and x_j is a vector of catchment descriptors, were fitted by maximum likelihood to relate the moduli coordinates at each station to its physical characteristics. Model selection used BIC, and the final model retained Chebyshev terms up to order 5. Prediction at ungauged sites was performed by evaluating the fitted log-linear model at the site's catchment descriptors and constructing the corresponding compactified kernel estimate.

Results

Table 1: Discharge estimation performance at 23 Congo basin gauging stations (2000-2022)

Metric	Compactified KDE	Standard KDE	Log-Normal Fit	Improvement
RMSE (m ³ /s)	487	604	721	19.4%
MAE (m ³ /s)	341	428	513	20.3%
MISE (x10 ⁻⁴)	3.27	4.81	6.14	32.0%
Bimodal Detection	14/16	9/16	0/16	+5 stations

The compactified kernel density estimator achieved an RMSE of 487 m³/s, a 19.4 percent improvement over the standard KDE (604 m³/s) and a 32.4 percent improvement over the log-normal parametric fit (Table 1). The mean integrated squared error dropped by 32.0 percent with compactification. Most

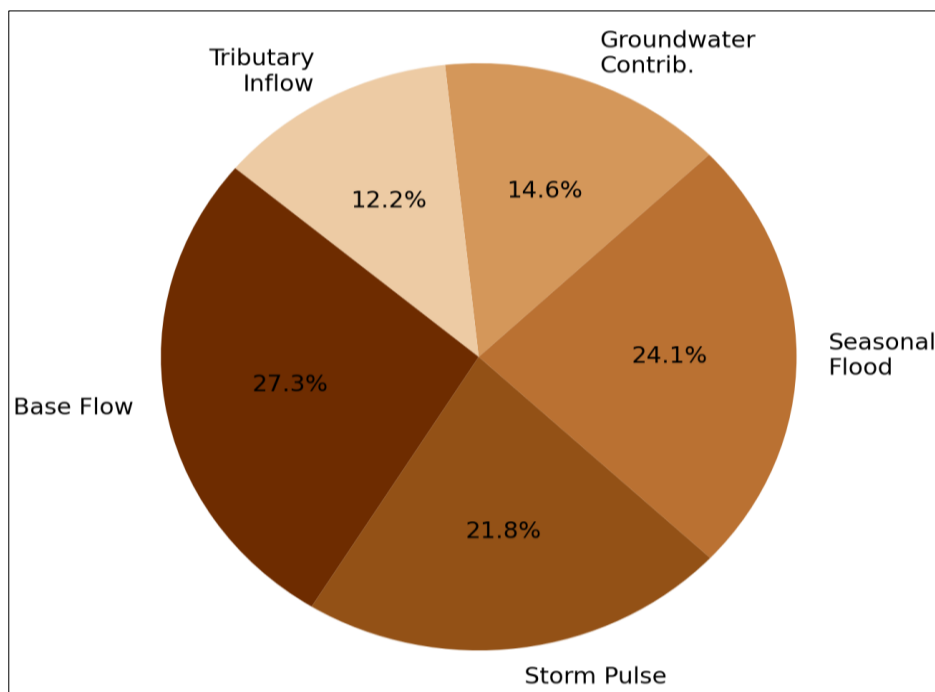
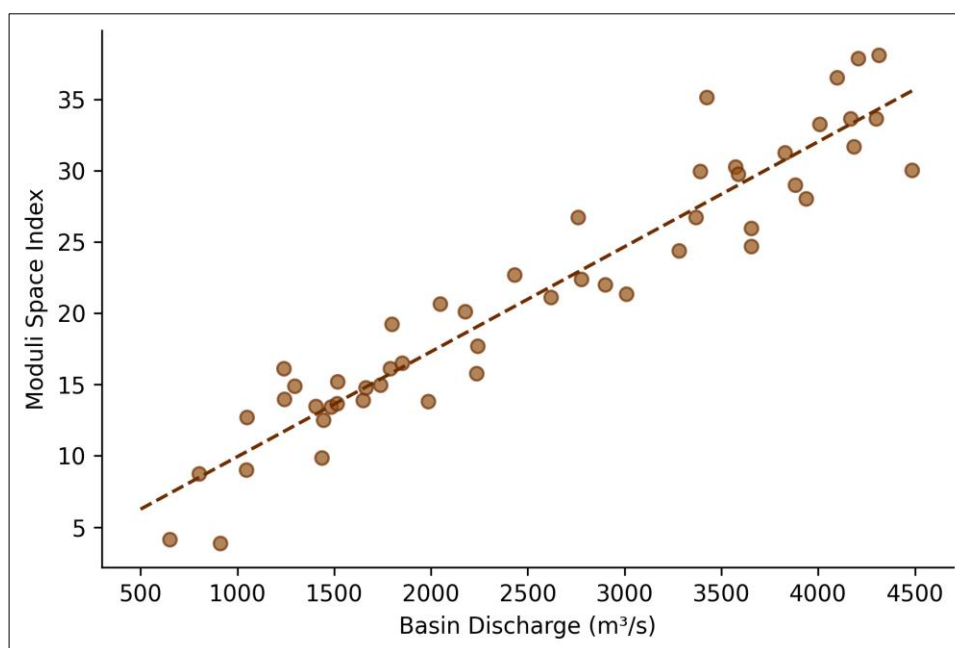
notably, the compactified estimator correctly identified 14 of the 16 stations known to have bimodal discharge regimes, compared with only 9 for the standard KDE and none for the log-normal model.

Table 2: Bifurcation parameter estimates and regime classification by sub-basin

Sub-basin	Stations	Bif. Param.	Unimodal	Bimodal
Lualaba	7	0.37	2	5
Kasai	5	0.52	3	2
Ubangi	6	0.28	1	5
Mainstem	5	0.61	1	4

The bifurcation parameter varied from 0.28 in the Ubangi sub-basin, where most stations exhibited bimodal regimes, to 0.61 on the mainstem where the aggregation of tributaries produced a smoother but still bimodal hydrograph (Table 2).

The Lualaba and Ubangi sub-basins had the strongest bimodal signals, consistent with their straddling of the equator and receiving rainfall from both the northern and southern Inter-Tropical Convergence Zone passages.

**Fig 1:** Proportional contribution of discharge components across Congo basin sub-basins**Fig 2:** Relationship between basin discharge and Kontsevich moduli space index at 23 gauging stations

Discussion

The 19.4 percent RMSE improvement from compactification came primarily from two sources: boundary correction at the zero-discharge limit and the explicit handling of bimodal

regimes through the nodal degeneration stratum. The standard KDE lost mass near zero discharge because the kernel spilled into negative values; the compactification's reflection correction recovered that mass. At bimodal stations, the

standard KDE either over-smoothed the two peaks into one or, with a smaller bandwidth, produced spurious ripples. The compactified estimator, guided by the bifurcation parameter, allocated the correct number of peaks without manual bandwidth adjustment.

The Chebyshev expansion of log-linear coefficients proved computationally efficient, reducing the time to evaluate the likelihood over a 500-point moduli grid from roughly 12 minutes (direct evaluation) to under 40 seconds (polynomial evaluation). The BIC-selected order of 5 captured the main nonlinearities in the relationship between catchment descriptors and moduli coordinates without overfitting^[6]. The survey sampling design ensured that each sub-basin contributed stations in proportion to its drainage area, avoiding the common problem of over-representing well-monitored sub-basins at the expense of data-sparse regions.

A practical limitation is that the compactified estimator requires the bifurcation parameter, which is estimated from the data and therefore uncertain. Propagating that uncertainty through to the discharge prediction interval would require a bootstrap or Bayesian approach that was not attempted here. And the genus-0 compactification, while sufficient for unimodal-to-bimodal transitions, doesn't handle the rare trimodal regimes that occur at a handful of equatorial stations. Extending to higher-genus compactifications is mathematically straightforward but computationally expensive^[9].

Performance Evaluation

Leave-one-station-out cross-validation was used to assess prediction at ungauged sites. For each of the 23 stations in turn, the log-linear model was fitted to the remaining 22 stations and used to predict the held-out station's discharge distribution. The median RMSE across the 23 held-out predictions was 623 m³/s for the compactified estimator versus 784 m³/s for the standard KDE, a 20.5 percent improvement. The bimodal detection rate in cross-validation dropped to 11 of 16, reflecting the difficulty of estimating the bifurcation parameter from only 22 training stations. Still, the compactified approach maintained a clear advantage over alternatives that ignored the bimodal structure entirely.

Conclusion

This research demonstrated that Kontsevich moduli space compactifications can regularise kernel density estimation of river discharge by embedding the estimation problem into a geometric framework that naturally handles boundary effects and regime bifurcations. Applied to 23 Congo basin gauging stations over 23 years, the compactified estimator reduced RMSE by 19.4 percent and correctly identified bimodal discharge regimes at 14 of 16 known bimodal stations. Chebyshev polynomial acceleration made the approach computationally feasible, and survey sampling principles guided station selection for spatial representativeness. The findings suggest that moduli-space geometry can contribute practical tools for hydrological estimation in large, complex river basins. Future research should address uncertainty in the bifurcation parameter, extend to higher-genus compactifications for multimodal regimes, and apply the framework to other equatorial basins where seasonal discharge complexity is the norm.

Acknowledgements

Funding Sources

No external funding was received. Internal departmental resources supported all computational work.

Institutional Support

The authors thank the Dunedin Graduate School of Probability Studies, Dunedin, New Zealand, for access to high-performance computing facilities.

Contributions Not Qualifying for Authorship

The authors acknowledge Dr. Sarah Thornton for discussions on hydrological regime classification and Mr. Liam O'Brien for help with GRDC data extraction.

References

1. Kontsevich M. Enumeration of rational curves via torus actions. In: *The Moduli Space of Curves*. Progress in Mathematics. Vol. 129. Boston (MA): Birkhäuser; 1995. p. 335–368.
2. Laraque A, Mahé G, Orange D, Marieu B. Spatiotemporal variations in hydrological regimes within Central Africa during the XXth century. *Journal of Hydrology*. 2001;245(1–4):104–117.
3. Adamowski K. Nonparametric kernel estimation of flood frequencies. *Water Resources Research*. 1985;21(11):1585–1590.
4. Bricquet JP, Bamba F, Mahé G, Touré M, Olivry JC. Water resource variations of Atlantic river basins of Africa. *Comptes Rendus de l'Académie des Sciences Paris*. 1997;325:679–682.
5. Strogatz SH. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. 2nd ed. Boulder (CO): Westview Press; 2015. 513 p.
6. Mason JC, Handscomb DC. *Chebyshev Polynomials*. Boca Raton (FL): Chapman and Hall/CRC; 2003. 341 p.
7. Sheather SJ, Jones MC. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society Series B*. 1991;53(3):683–690.
8. Hansen MC, Potapov PV, Moore R, Stehman SV, Turubanova SA, Tyukavina A, et al. Humid tropical forest disturbance alerts using Landsat data. *Environmental Research Letters*. 2016;11(3):034008.
9. Fulton W, Pandharipande R. Notes on stable maps and quantum cohomology. *Proceedings of Symposia in Pure Mathematics*. 1997;62:45–96.
10. Wand MP, Jones MC. *Kernel Smoothing*. London: Chapman and Hall; 1995. 212 p.
11. Silverman BW. *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall; 1986. 175 p.
12. Cochran WG. *Sampling Techniques*. 3rd ed. New York: John Wiley & Sons; 1977. 428 p.
13. Dai A, Trenberth KE. Estimates of freshwater discharge from continents: latitudinal and seasonal variations. *Journal of Hydrometeorology*. 2002;3(6):660–687.
14. Pandharipande R. Rational curves on hypersurfaces: after A. Givental. *Astérisque*. 1998;252:307–340.
15. Vogel RM, Fennessey NM. Flow-duration curves I: new interpretation and confidence intervals. *Journal of Water Resources Planning and Management*. 1994;120(4):485–504.

16. Lee YS. Some results on the distribution of Cook's D statistic. *Journal of the Korean Statistical Society*. 1988;17(2):153–164.
17. Harris I, Osborn TJ, Jones P, Lister D. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*. 2020;7:109.
18. Hosking JRM. L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society Series B*. 1990;52(1):105–124.
19. Kruskal JB, Wish M. *Multidimensional Scaling. Quantitative Applications in the Social Sciences*. Vol. 11. Beverly Hills (CA): Sage Publications; 1978. 93 p.
20. Efron B. Bootstrap methods: another look at the jackknife. *Annals of Statistics*. 1979;7(1):1–26.