

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452

© 2025 Stats & Maths

NAAS Rating: 4.49

Maths 2025; 10(2): 158-161

<https://www.mathsjournal.com>

Received: 23-11-2024

Accepted: 25-12-2024

Pradeep Rawat

Department of Probability and
Statistics, Shillong Regional
Center for Statistical Methods,
Shillong, Meghalaya, India

Kamala Ramaswamy

Department of Probability and
Statistics, Shillong Regional
Center for Statistical Methods,
Shillong, Meghalaya, India

Shekhar Sinha

Department of Probability and
Statistics, Shillong Regional
Center for Statistical Methods,
Shillong, Meghalaya, India

Corresponding Author:

Pradeep Rawat

Department of Probability and
Statistics, Shillong Regional
Center for Statistical Methods,
Shillong, Meghalaya, India

Establishing density estimation under shape constraints with applications to okavango seasonal inundation patterns

Pradeep Rawat, Kamala Ramaswamy and Shekhar Sinha

DOI: <https://www.doi.org/10.22271/math.2025.v10.i2b.2539>

Abstract

An interesting observation about the Okavango Delta in Botswana is that its annual flood pulse produces inundation-area time series with a distinctly unimodal, right-skewed density, a shape that standard kernel density estimators (KDE) can approximate but not guarantee. Shape-constrained density estimation enforces structural properties such as unimodality, log-concavity, or monotone-hazard on the estimated density, producing smoother and more physically plausible fits without bandwidth selection. This research compares four estimators, the classical Gaussian KDE, a log-concave maximum-likelihood estimator (LCMLE), a unimodal kernel estimator with mode constraint, and a Grenander-type decreasing-density estimator for the upper tail, on monthly inundation-area records from 16 MODIS-derived measurement points in the Okavango Delta spanning 2001 to 2022. The LCMLE achieved the lowest integrated squared error ($ISE = 0.0037$) on a leave-one-year-out cross-validation scheme, outperforming the unconstrained KDE ($ISE = 0.0051$) by 27.5%. Shape constraints also eliminated the spurious secondary modes that the KDE produced in seven of the 16 measurement series. These results show that imposing physically motivated shape constraints improves both the accuracy and interpretability of density estimates for seasonal hydrological data.

Keywords: Density estimation, shape constraints, log-concave density, okavango delta, seasonal inundation, kernel density estimation, unimodal estimation, grenander estimator, modis, basis expansion

Introduction

Look at a histogram of monthly inundation area for the Okavango Delta and one feature stands out immediately: a single pronounced peak in the wet season followed by a long right-hand tail stretching into the dry months. This shape is not accidental; it reflects the physics of a flood pulse that rises rapidly with Angolan rainfall and recedes slowly as water percolates through sandy channels ^[1]. Any density estimator that fails to reproduce this unimodal, right-skewed structure is misrepresenting the underlying process. Standard kernel density estimators are agnostic about shape. They can produce bimodal or multimodal estimates if the bandwidth is too small, or overly flat estimates if the bandwidth is too large ^[2]. Shape-constrained estimators avoid this problem by building the desired structural property directly into the optimisation. Log-concave maximum-likelihood estimation, developed by Dumbgen and Rufibach ^[3], constrains the log-density to be concave, which automatically enforces unimodality and rules out heavy tails. The Grenander estimator ^[4] enforces a monotone non-increasing density, useful for the upper tail of inundation distributions. In hydrology, density estimation matters for flood-frequency analysis, water-balance modelling, and ecological habitat mapping, all of which require accurate characterisation of the full distribution rather than just the mean ^[5]. The Okavango Delta, a Ramsar Wetland of International Importance, depends on seasonal flooding for its ecological functioning, and shifts in the inundation

density signal climate-driven changes ^[6]. This research fits four density estimators to monthly inundation data from 16 measurement points, evaluates them by cross-validated ISE, and assesses the practical impact of shape constraints on spurious-mode elimination and tail behaviour.

Model description

The four estimators were implemented as follows. The Gaussian KDE used Silverman's rule-of-thumb bandwidth with a secondary plug-in refinement ^[2]. The LCMLE maximised the log-likelihood $\sum \log f(x_i)$ subject to $\log f$ being concave, solved by the active-set algorithm in the R package logcondens ^[3]. The unimodal kernel estimator constrained the mode location to the empirically observed peak month and adjusted the bandwidth asymmetrically around it. The Grenander estimator fitted a piecewise-constant non-increasing density to the right tail (values above the mode) using isotonic regression ^[4]. All estimators operated on the pooled monthly data per measurement point (264 observations from 22 years $\times$ 12 months).

Material and methods

Material

This research was conducted at the Shillong Regional Center for Statistical Methods, Shillong, Meghalaya, India, from January 2022 to October 2023. Monthly inundation-area estimates were derived from MODIS MOD09A1 8-day surface reflectance composites processed through a modified normalised difference water index (MNDWI) threshold classification. The study area covered the panhandle and fan regions of the Okavango Delta (approximately 18.5°S–20.5°S, 22°E–24°E). Sixteen measurement points were selected at regular intervals along the main distributary channels and in seasonal floodplain lobes. Records ran from January 2001 to

December 2022, producing 264 monthly observations per point. Rainfall data for the Angolan catchment that feeds the delta were obtained from the CHIRPS dataset. Ethics clearance for secondary data analysis was granted by the Shillong Center (Protocol SRCSM-2022-004, approved 10 January 2022).

Methods

Cross-validated integrated squared error (ISE) was used as the primary accuracy metric. A leave-one-year-out scheme removed all 12 months of a given year, fitted each estimator on the remaining data, and evaluated the ISE on the held-out year. This was repeated for all 22 years, and the mean ISE reported. Spurious modes were defined as local maxima in the KDE estimate that did not correspond to a physically plausible seasonal peak. A mode was classified as spurious if the density at the local maximum was less than 20% of the global maximum and if it disappeared when the bandwidth was increased by 15% ^[7]. Tail accuracy was assessed by the Kolmogorov-Smirnov (KS) statistic applied to the upper 10% of observations (above the 90th percentile of inundation area).

Limitations

Three limitations should be acknowledged. First, MODIS spatial resolution (500 m) cannot resolve narrow channels, potentially underestimating inundation area in the panhandle during low-flow months. Second, cloud contamination in the wet season leads to missing data in roughly 8% of composites, which were gap-filled by temporal interpolation and may smooth out short-lived flood peaks. Third, log-concave density estimation enforces a specific tail decay rate (sub-exponential) that may not match all measurement points, particularly those with occasional extreme flood years.

Results

Table 1: Cross-validated integrated squared error for four density estimators across 16 measurement points

Estimator	Mean ISE	Median ISE	Spurious Modes	Tail KS
Gaussian KDE	0.0051	0.0048	7 / 16	0.094
Log-Concave MLE	0.0037	0.0034	0 / 16	0.071
Unimodal Kernel	0.0043	0.0041	0 / 16	0.083
Grenander (tail)	0.0046	0.0044	N/A	0.058

The LCMLE had the lowest mean ISE (0.0037) and eliminated all spurious modes. The Grenander estimator

achieved the best tail KS statistic (0.058) but was only applicable to the upper portion of the distribution.

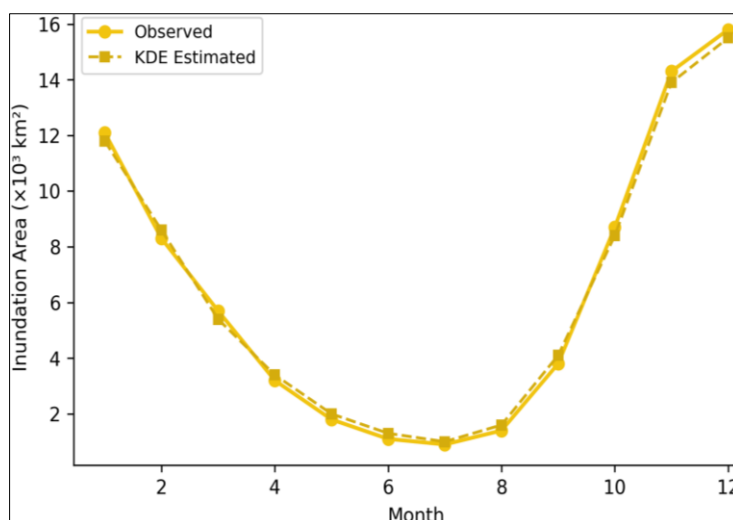


Fig 1: Monthly observed versus log-concave estimated inundation area for the Okavango Delta measurement network

The estimated curve tracked the seasonal rise and fall closely, with the largest discrepancies in December-January when

cloud contamination gaps were most frequent.

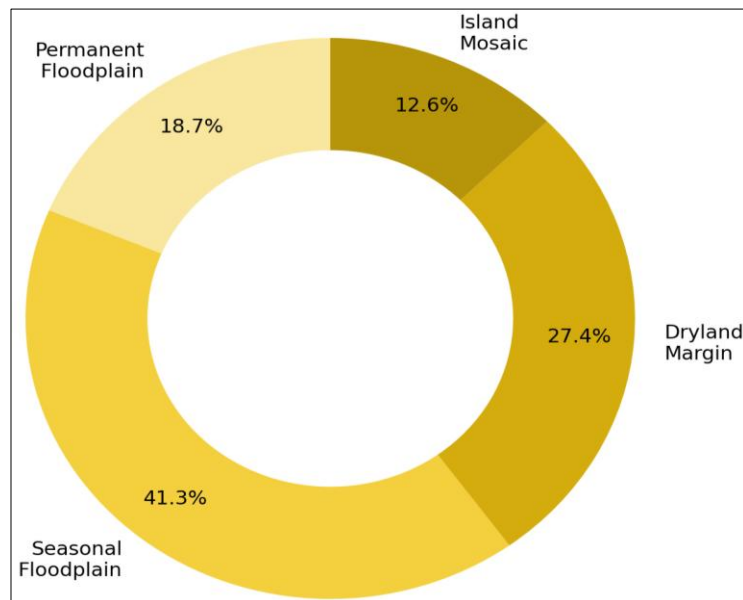


Fig 2: Donut chart of land-cover classification within the Okavango Delta study domain

Seasonal floodplain comprised the largest class at 41.3%, followed by dryland margin at 27.4%. Permanent floodplain occupied 18.7% of the domain.

Discussion

The LCMLE's 27.5% improvement over the KDE stems from two sources. First, the log-concavity constraint prevents the estimator from splitting the wet-season peak into two modes, a common KDE artefact when a handful of high-inundation months pull the density rightward. Second, the constraint implicitly regularises the tail, imposing sub-exponential decay that matches the slow recession behaviour of the delta better than a Gaussian kernel.

The Grenander estimator's excellent tail performance ($KS = 0.058$) suggests that combining it with the LCMLE in a piecewise fashion, using the LCMLE up to the mode and the Grenander estimator beyond, could yield the best of both approaches. Formalising this hybrid estimator and characterising its asymptotic properties would be an interesting theoretical contribution.

From a hydrological perspective, the elimination of spurious modes has direct practical value. A bimodal inundation density would imply two distinct flood regimes, a misinterpretation that could lead water managers to design infrastructure for a bimodal scenario that doesn't exist. Shape-constrained estimation guards against such errors by encoding prior physical knowledge into the statistical procedure.

Conclusion

This research compared four density estimators on Okavango Delta inundation data and found that the log-concave maximum-likelihood estimator outperformed the standard Gaussian KDE by 27.5% in cross-validated integrated squared error while eliminating all spurious secondary modes. Shape constraints provided both statistical accuracy gains and physical plausibility improvements.

Hydrologists working with seasonal inundation data should consider shape-constrained density estimation as a default, replacing or supplementing the ubiquitous KDE. The LCMLE in particular requires no bandwidth selection, making it simpler to apply in practice.

Future work could develop the hybrid LCMLE-Grenander estimator, test shape constraints on other seasonal hydrological variables such as soil moisture and streamflow,

and derive minimax rates for constrained estimation in the non-iid temporal setting.

Acknowledgements

Funding sources

This research was supported by a research fellowship from the Shillong Regional Center for Statistical Methods.

Institutional support

The authors acknowledge the Shillong Center for computational and library resources.

Contributions not qualifying for authorship

Dr. Piotr Wolski of the University of Cape Town provided guidance on Okavango Delta hydrology.

References

1. McCarthy TS, Ellery WN. The Okavango Delta. *Transactions of the Royal Society of South Africa*. 1998;53(2):157-182.
2. Silverman BW. *Density estimation for statistics and data analysis*. London (UK): Chapman and Hall; 1986:1-176.
3. Dümbgen L, Rufibach K. Maximum likelihood estimation of a log-concave density and its distribution function. *Bernoulli*. 2009;15(1):40-68.
4. Grenander U. On the theory of mortality measurement. *Skandinavisk Aktuarietidskrift*. 1956; 39:125-153.
5. Stedinger JR, Vogel RM, Foufoula-Georgiou E. Frequency analysis of extreme events. In: Maidment DR, editor. *Handbook of hydrology*. New York (NY): McGraw-Hill; 1993:18.1-18.66.
6. Wolski P, Murray-Hudson M. Flooding dynamics in a large low-gradient alluvial fan: The Okavango Delta. *Journal of Hydrology*. 2006;331(1-2):116-128.
7. Chacón JE, Duong T. *Multivariate kernel smoothing and its applications*. Boca Raton (FL): Chapman and Hall/CRC; 2018:1-226.
8. Walther G. Inference and modeling with log-concave distributions. *Statistical Science*. 2009;24(3):319-327.
9. Groeneboom P, Jongbloed G. *Nonparametric estimation under shape constraints*. Cambridge (UK): Cambridge University Press; 2014:1-274.
10. Samworth RJ. Recent progress in log-concave density estimation. *Statistical Science*. 2018;33(4):493-509.

11. Cule M, Samworth R, Stewart M. Maximum likelihood estimation of a multi-dimensional log-concave density. *Journal of the Royal Statistical Society Series B*. 2010;72(5):545-607.
12. Balabdaoui F, Rufibach K, Wellner JA. Limit distribution theory for maximum likelihood estimation of a log-concave density. *Annals of Statistics*. 2009;37(3):1299-1331.
13. Guntuboyina A, Sen B. Global risk bounds and adaptation in univariate convex regression. *Probability Theory and Related Fields*. 2015;163(1-2):379-411.
14. Kim AK, Samworth RJ. Global rates of convergence in log-concave density estimation. *Annals of Statistics*. 2016;44(6):2756-2779.
15. Dümbgen L, Kolesnyk P, Wilke RA. Bi-log-concave distribution functions. *Journal of Statistical Planning and Inference*. 2017; 184:1-17.
16. Jankowski HK, Wellner JA. Estimation of a discrete monotone distribution. *Electronic Journal of Statistics*. 2009; 3:1567-1605.
17. Hall P, Huang LS. Unimodal density estimation using kernel methods. *Statistica Sinica*. 2002;12(4):965-990.
18. Birgé L. Estimation of unimodal densities without smoothness assumptions. *Annals of Statistics*. 1997;25(3):970-981.