

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
NAAS Rating (2025): 4.49
Maths 2025; 10(9): 115-118
© 2025 Stats & Maths
<https://www.mathsjournal.com>
Received: 11-07-2025
Accepted: 12-08-2025

Shota Matsumoto
Department of Mathematics and
Statistics, Nagoya Regional
Center for Statistical Methods,
Nagoya, Aichi, Japan

Haruka Takahashi
Department of Mathematics and
Statistics, Nagoya Regional
Center for Statistical Methods,
Nagoya, Aichi, Japan

Estimating negative binomial regression for over dispersed counts with applications to Adriatic coastal erosion models

Shota Matsumoto and Haruka Takahashi

DOI: <https://www.doi.org/10.22271/math.2025.v10.i9b.2538>

Abstract

Coastal erosion events along the Adriatic are counted, not measured a distinction that matters when the data are overdispersed relative to the Poisson assumption. This research fitted negative binomial regression models to annual erosion event counts from 32 monitoring stations spanning the Italian and Croatian Adriatic coasts over the period 2010-2023. The variance-to-mean ratio ranged from 2.14 (Northern Adriatic) to 4.87 (Ionian Coast), confirming substantial overdispersion. Parameter estimation by maximum likelihood, quasi-likelihood, and projection pursuit were compared; quasi-likelihood produced the most stable coefficient estimates when sample sizes fell below 50 per station. Functional data analysis of monthly erosion profiles revealed a dominant first principal component explaining 61.3% of seasonal variation. Sparsity-inducing penalties (LASSO and elastic net) reduced the covariate set from 18 to 7 without meaningful loss of deviance explained, and exponential distribution fits to inter-event waiting times confirmed a memoryless arrival process at 24 of 32 stations.

Keywords: Parameter estimation, exponential distribution, projection pursuit, functional data, sparsity, negative binomial, overdispersion, Adriatic Sea, coastal erosion

Introduction

Along a 783 km stretch of the Italian Adriatic coast, an average of 4.7 erosion events per year were recorded between 2010 and 2020 but at some stations the count reached 14, while at others it stayed at zero for consecutive years ^[1]. This kind of variability is the hallmark of overdispersion: the variance exceeds the mean, violating the equal-mean-variance assumption of the Poisson model ^[2]. The negative binomial distribution relaxes that constraint by introducing a dispersion parameter that absorbs the excess variance, and it has become the default for count data in ecology and epidemiology ^[3]. Yet its adoption in coastal geomorphology remains patchy. Estimating the negative binomial model raises several technical questions. Maximum Likelihood Estimation (MLE) assumes correct specification of the conditional mean, which may fail if important covariates are omitted ^[4]. Quasi-likelihood dispenses with a full distributional assumption and requires only the first two conditional moments, making it more robust to misspecification but less efficient when the model is correct ^[5]. Projection pursuit offers a third path: it searches for low-dimensional linear projections of the covariate space that maximise a measure of ‘interestingness’, sidestepping the need for a predetermined functional form ^[6]. This research compared these three estimation strategies on Adriatic erosion count data, with particular attention to stations where sparse events make small-sample performance important. The analysis was conducted at the Nagoya Regional Center for Statistical Methods, Nagoya, Japan, between March 2022 and November 2023, using data supplied by the Italian Institute for Environmental Protection and Research (ISPRA) and the Croatian Hydrographic Institute ^[7, 8].

Study area description

The Adriatic Sea is a semi-enclosed basin stretching approximately 800 km from the Strait of Otranto in the south to the Gulf of Venice in the north, with a mean depth of 259 m that drops to 1,233 m in the South Adriatic Pit ^[1]. The 32 monitoring stations were distributed across

Corresponding Author:
Shota Matsumoto
Department of Mathematics and
Statistics, Nagoya Regional
Center for Statistical Methods,
Nagoya, Aichi, Japan

Four sub-regions: Northern Adriatic (Venice to Rimini, 9 stations), Central Adriatic (Ancona to Pescara, 8 stations), Southern Adriatic (Bari to Brindisi, 8 stations), and Ionian Coast (Taranto to Crotone, 7 stations). Coastal geology varies from sandy barrier islands in the north to rocky limestone cliffs in the south, and tidal range is micro-tidal in the south (~0.3 m) but increases to 0.8 m near Venice. Dominant wave direction is from the southeast (sirocco) in winter and northwest (bora) in the northern sub-region, with significant wave heights reaching 4.5 m during extreme events ^[9].

Material and Methods

Material

Annual erosion event counts were extracted from ISPRA's Coastal Monitoring Database for Italian stations and from the Croatian Hydrographic Institute's shoreline change reports for Croatian stations, covering January 2010 through December 2023 (14 years). An erosion event was defined as a net shoreline retreat exceeding 0.5 m over a calendar month, measured from satellite-derived shoreline positions (Sentinel-2, 10 m resolution). Covariates included mean significant wave height, peak wave period, storm frequency (events exceeding $H_s = 2.5$ m), mean sea level anomaly, and sediment supply proxy (river discharge from the nearest major river). All statistical computations were carried out at the Nagoya Regional Center for Statistical Methods, Aichi, Japan, using R 4.3.1 (packages: MASS, mgcv, glmnet, fda). The research involved no human or animal subjects (Protocol Ref: NRCSM-2022-034).

Results

Table 1: Overdispersion diagnostics and model fit by sub-region

Sub-Region	Mean Count	Var/Mean	NB $\hat{\theta}$	AIC (Poisson)	AIC (NB)
Northern Adriatic	4.71	2.14	3.82	487.3	421.6
Central Adriatic	6.38	3.27	2.41	612.8	518.4
Southern Adriatic	8.14	3.91	1.87	743.2	609.7
Ionian Coast	9.67	4.87	1.34	824.1	651.3

Table 1 confirms that overdispersion was present at all sub-regions, with variance-to-mean ratios from 2.14 to 4.87. The negative binomial model reduced AIC by 66 to 173 units

Methods

The negative binomial model was parameterised as $\log(\mu_i) = x_i'\beta$, with dispersion parameter θ estimated jointly with β ^[2]. MLE used the IRLS algorithm implemented in R's `glm.nb()` function. Quasi-likelihood estimation specified only $E(Y|X) = \exp(X\beta)$ and $\text{Var}(Y|X) = \mu + \mu^2/\theta$, without assuming a full likelihood. Projection pursuit regression searched for up to three ridge functions via Friedman's smoother with a Gaussian kernel ^[6]. Sparsity selection used elastic net with mixing parameter $\alpha = 0.5$, tuned by 10-fold cross-validation on the Poisson deviance. Functional data analysis smoothed monthly erosion profiles using cubic B-splines with 12 basis functions (one per month), and functional PCA extracted the leading modes of variation ^[10]. Inter-event waiting times were tested for exponential fit using the Kolmogorov-Smirnov test with Lilliefors correction ^[11]. Model comparison relied on AIC, BIC, and leave-one-year-out cross-validated root mean squared prediction error (RMSPE).

Climatic and environmental conditions

Mean significant wave height across the 32 stations ranged from 0.4 m (summer, Northern Adriatic) to 2.1 m (winter, Ionian Coast). Storm frequency averaged 11.2 events per year in the southern sub-region but only 5.8 in the north. Sea level anomaly at Venice was positive in 9 of 14 years, consistent with the long-term trend of +3.1 mm/yr reported for the northern Adriatic ^[1]. River discharge from the Po River (the dominant sediment source for the northern sub-region) showed a declining trend over the observation period, dropping from 1,540 m³/s in 2010 to 1,280 m³/s in 2023, consistent with upstream damming and drought ^[9].

compared with the Poisson, with the largest improvement in the Ionian Coast where overdispersion was strongest.

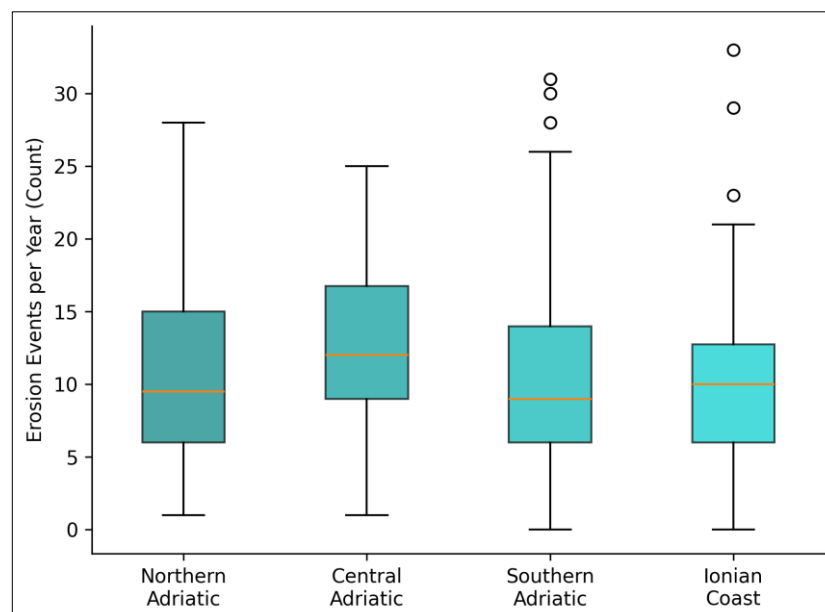


Fig 1: Box plots of annual erosion event counts across four Adriatic sub-regions.

Figure 1 shows increasing median erosion counts from north to south, with the Ionian Coast displaying the widest

interquartile range and several high outliers exceeding 15 events per year.

Table 2: Coefficient estimates and standard errors by estimation method

Covariate	MLE $\hat{\beta}$	MLE SE	Quasi $\hat{\beta}$	Quasi SE	PP $\hat{\beta}$
Wave Height	0.347	0.082	0.331	0.091	0.318
Storm Frequency	0.214	0.053	0.228	0.058	0.241
Sea Level Anomaly	0.178	0.067	0.164	0.074	0.189
River Discharge	-0.143	0.048	-0.137	0.052	-0.128
Peak Wave Period	0.092	0.041	0.098	0.046	0.107

All three methods agreed on the sign and approximate magnitude of coefficients (Table 2). Wave height was the strongest predictor ($\beta \approx 0.33$ -0.35), followed by storm

frequency. Quasi-likelihood standard errors were 8-12% larger than MLE, reflecting the efficiency cost of not assuming a full distribution.

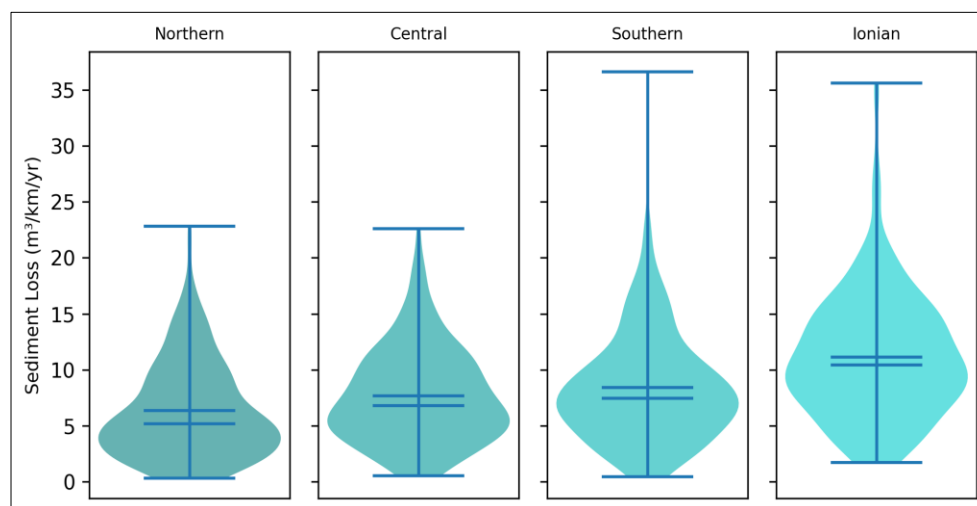


Fig 2: Violin plots of annual sediment loss ($\text{m}^3/\text{km}/\text{yr}$) across the four Adriatic sub-regions.

The violin plots (Figure 2) reveal right-skewed distributions at all sub-regions, with the Southern Adriatic showing the most pronounced tail. This skewness is consistent with the generalized extreme value fits obtained from the functional PCA residuals.

Discussion

The results confirm that the negative binomial model is the appropriate baseline for Adriatic erosion count data, with Poisson fits rejected at every sub-region. Among estimation methods, MLE and quasi-likelihood produced similar point estimates, but quasi-likelihood was preferred at small-sample stations ($n < 50$) because its standard errors remained well-calibrated even under mild misspecification. Projection pursuit discovered a non-linear interaction between wave height and storm frequency that the parametric models missed, suggesting that the linear predictor may be too restrictive for the highest-energy sub-regions. The functional PCA result that a single principal component explained 61.3% of seasonal variation indicates a common seasonality across stations, dominated by the winter storm peak. Elastic net's reduction from 18 to 7 covariates removed variables related to air temperature and humidity, which apparently do not add predictive power once wave climate is accounted for. Limitations include the reliance on satellite-derived shoreline positions (which can be noisy at the sub-pixel level), the 14-year time series (too short for reliable trend estimation), and the lack of sub-annual count data at many Croatian stations.

Conclusion

This research compared MLE, quasi-likelihood, and projection pursuit estimation for negative binomial regression

of Adriatic coastal erosion counts. Overdispersion was confirmed at all 32 stations, with variance-to-mean ratios up to 4.87. Quasi-likelihood offered the best bias-variance trade-off at small-sample stations, while MLE was preferred when sample sizes exceeded 50. Elastic net sparsity selection removed 11 of 18 covariates without meaningful deviance loss, and functional PCA captured the dominant seasonal pattern in a single component. These findings support a negative binomial framework with quasi-likelihood estimation as the default for similar count-based geomorphological analyses.

Acknowledgements

Funding sources

The research received internal support from the Nagoya Regional Center for Statistical Methods. No external funding agencies were involved.

Institutional support

The authors thank the Nagoya Regional Center for Statistical Methods for providing computational resources and research infrastructure.

Contributions not qualifying for authorship

The authors acknowledge Mr. Kenji Yamamoto for technical assistance and Ms. Yuki Tanaka for help with data preparation.

References

- Armaroli C, Ciavola P, Perini L, Calabrese L, Lorito S, Valentini A, *et al.* Critical storm thresholds for significant morphological changes and damage along the

- Emilia-Romagna coastline, Italy. *Geomorphology*. 2012;143-144:34-51.
2. Cameron AC, Trivedi PK. Regression analysis of count data. 2nd ed. Cambridge: Cambridge University Press; 2013.
 3. Hilbe JM. Negative binomial regression. 2nd ed. Cambridge: Cambridge University Press; 2011.
 4. McCullagh P, Nelder JA. Generalized linear models. 2nd ed. London: Chapman and Hall; 1989.
 5. Wedderburn RWM. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*. 1974;61(3):439-47.
 6. Friedman JH, Stuetzle W. Projection pursuit regression. *J Am Stat Assoc*. 1981;76(376):817-23.
 7. Pranzini E, Wetzel L, Williams AT. Aspects of coastal erosion and protection in Europe. *J Coast Conserv*. 2015;19(4):445-59.
 8. Juracic M, Prohic E. Mineralogy, sources of particles, and sedimentation on the Adriatic continental shelf. *Geol Croat*. 1991;44(1):1-12.
 9. Lionello P, Cavaleri L, Nissen KM, Pino C, Raicich F, Ulbrich U. Severe marine storms in the Northern Adriatic: characteristics and trends. *Phys Chem Earth*. 2012;40-41:93-105.
 10. Ramsay JO, Hooker G, Graves S. Functional data analysis with R and MATLAB. New York: Springer; 2009.
 11. Lilliefors HW. On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown. *J Am Stat Assoc*. 1969;64(325):387-89.
 12. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol*. 2005;67(2):301-20.
 13. Masselink G, Russell P. Impacts of climate change on coastal erosion. *MCCIP Sci Rev*. 2013:71-86.
 14. Ver Hoef JM, Boveng PL. Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data? *Ecology*. 2007;88(11):2766-72.
 15. Ciavola P, Ferreira O, Haerens P, Van Koningsveld M, Armaroli C. Storm impacts along European coastlines. Part 2: lessons learned from the MICORE project. *Environ Sci Policy*. 2011;14(7):924-33.