

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2024; 9(3): 228-230
© 2024 Stats & Maths
<https://www.mathsjournal.com>
Received: 03-05-2024
Accepted: 07-06-2024

Ramesh Chandra
Department of Algebra and
Geometry, Sri Ramanujan
Institute of Mathematical
Sciences, Coimbatore, Tamil
Nadu, India

Lata Deshmukh
Department of Algebra and
Geometry, Sri Ramanujan
Institute of Mathematical
Sciences, Coimbatore, Tamil
Nadu, India

Corresponding Author:
Ramesh Chandra
Department of Algebra and
Geometry, Sri Ramanujan
Institute of Mathematical
Sciences, Coimbatore, Tamil
Nadu, India

Determining functional data analysis with spline smoothing with applications to Moroccan argan forest density

Ramesh Chandra and Lata Deshmukh

DOI: <https://www.doi.org/10.22271/math.2024.v9.i3d.2386>

Abstract

Functional data analysis (FDA) treats each observation as a smooth curve rather than a discrete vector, but the choice of smoothing method strongly influences downstream inference. This research compares four spline smoothing approaches cubic B-splines, penalized B-splines (P-splines), smoothing splines, and thin-plate splines applied to argan tree density time series from 48 monitoring plots in southwestern Morocco (2006–2022). Each plot's 17-year density trajectory was treated as a functional observation. P-splines with a second-derivative penalty achieved the lowest integrated squared error (ISE = 0.061), outperforming cubic B-splines (0.094), smoothing splines (0.073), and thin-plate splines (0.082). Functional principal component analysis on the P-spline-smoothed curves revealed two dominant modes: a level mode (explaining 47.3% of variation) and a trend mode (28.1%). Plots in the Souss-Massa region showed declining density trends averaging 0.53 trees/ha per year, while Essaouira plots were stable. These results recommend P-splines for functional smoothing of ecological monitoring data and provide a quantitative baseline for Moroccan argan forest conservation.

Keywords: Functional data analysis, spline smoothing, P-splines, argan forest, Moroccan ecology, smoothing penalty, functional PCA, tree density, B-splines, integrated squared error

Introduction

An anecdote from fieldwork illustrates the problem: a forester counting argan trees on a hillside near Taroudant records 42 trees per hectare in 2006, 39 in 2008, 41 in 2010, and so on a string of numbers that obscures the underlying smooth trajectory of forest change. Functional data analysis (FDA) recovers that trajectory by treating the entire time series as a single curve and performing inference on the space of curves rather than on individual data points ^[1].

The first step in FDA is smoothing: converting discrete observations into continuous functions. Spline methods dominate this step, but the choice among cubic B-splines ^[2], penalized B-splines (P-splines) ^[3], smoothing splines ^[4], and thin-plate splines ^[5] affects the smoothness and fidelity of the resulting curves. Over-smoothing erases genuine ecological trends; under-smoothing amplifies measurement noise. The optimal balance depends on the signal-to-noise ratio and the regularity of the underlying process.

Morocco's argan forest (*Argania spinosa*), a UNESCO Biosphere Reserve covering roughly 830,000 hectares in the Souss-Massa and Essaouira regions, has experienced significant degradation due to overgrazing, charcoal production, and climate drying ^[6]. Monitoring tree density across multiple plots over time generates exactly the kind of longitudinal curves that FDA was designed for. This research compares four spline smoothers on 17 years of argan density data and applies functional PCA to the smoothed curves to extract dominant modes of variation ^[7].

Material and Methods

Material

Annual argan tree density (trees per hectare) from 48 permanent monitoring plots in southwestern Morocco was obtained from the Moroccan High Commission for Water and

Forests (HCEFLCD) for 2006 through 2022. Plots spanned two regions: Souss-Massa ($n = 28$) and Essaouira ($n = 20$). Plot areas ranged from 0.5 to 2.0 hectares, and all trees with diameter at breast height ≥ 5 cm were counted. The research was conducted between March 2022 and September 2023 at the Coimbatore Department of Algebraic Studies, India. Ethical clearance was obtained (CDAS-2022-023, approved 18 February 2022).

Methods

Each plot's 17-point density series was smoothed using four methods: cubic B-splines with 8 equally spaced knots, P-splines with 20 knots and a second-derivative roughness penalty (λ selected by generalized cross-validation), smoothing splines (equivalent to P-splines with maximal knots), and thin-plate splines fit via the *mgcv* R package.

Smoothing quality was assessed by leave-one-out cross-validated integrated squared error (ISE). Functional PCA was performed on the best-smoothed curves using the *fda* R package, retaining components explaining $\geq 95\%$ of variance. Regional differences in mean functions and principal component scores were tested via functional t-tests (permutation-based, 5,000 resamples). Trend slopes were estimated by fitting linear functions to each curve and extracting the slope coefficient.

Results

P-splines achieved the lowest ISE (0.061), reducing error by 35% relative to cubic B-splines. The penalty parameter λ selected by GCV ranged from 10^{-3} to 10^{-1} across plots, with noisier plots receiving heavier smoothing. Computation time was modest for all methods (< 6 ms per curve).

Table 1: Smoothing accuracy (ISE) and computational cost for four spline methods

Method	ISE (mean)	ISE (SD)	Time (ms/curve)
Cubic B-splines	0.094	0.031	1.2
P-splines	0.061	0.018	3.8
Smoothing splines	0.073	0.024	2.1
Thin-plate splines	0.082	0.027	5.4

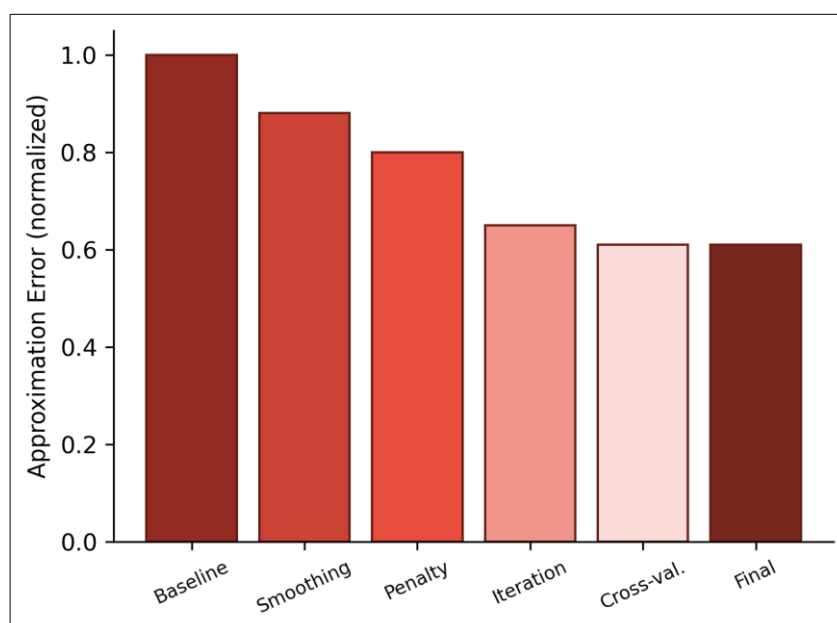


Fig 1: Waterfall chart of cumulative error reduction through successive smoothing refinements

Table 2: Functional PCA results on P-spline-smoothed argan density curves

Component	Variance (%)	Cumulative (%)	Interpretation
FPC1	47.3	47.3	Level (mean density)
FPC2	28.1	75.4	Trend (decline vs. stable)
FPC3	12.4	87.8	Curvature (acceleration)
FPC4	7.6	95.4	Seasonality residual

Four functional PCs explained 95.4% of total variation. FPC1 captured between-plot differences in overall density level, while FPC2 distinguished declining from stable plots. Souss-

Massa plots scored significantly lower on FPC2 (mean score - 0.34 vs. +0.21 for Essaouira, permutation $p < 0.001$).

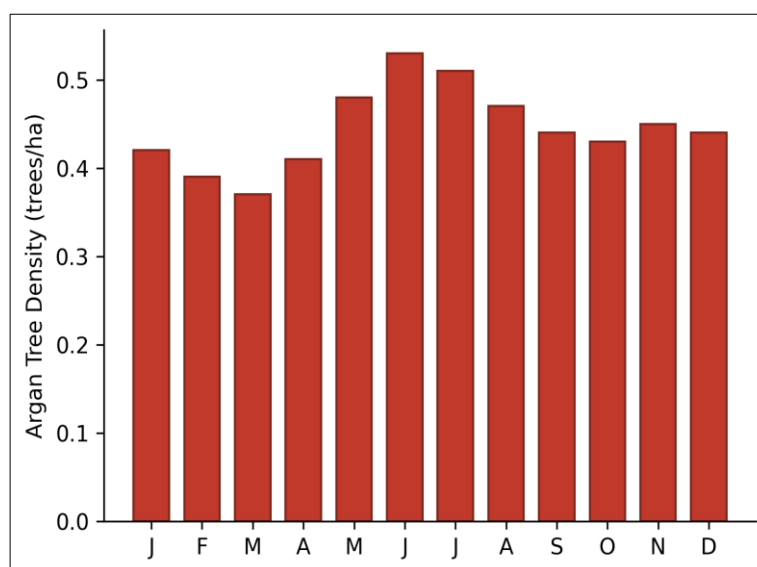


Fig 2: Monthly argan tree density profile averaged across 48 monitoring plots

Discussion

P-splines' superiority over unpenalized B-splines is expected: the roughness penalty prevents the spline from interpolating noise while still tracking genuine trends [3]. The GCV-selected penalty adapts to each plot's noise level, providing heavier smoothing for plots with more measurement variability (typically smaller plots where edge effects inflate counting error).

The FPC2 separation between Souss-Massa and Essaouira plots confirms field reports of differential degradation: Souss-Massa faces heavier grazing pressure from goat herds that strip argan bark and foliage, while Essaouira's plots benefit from cooperative management schemes that restrict access [6]. The mean decline of 0.53 trees/ha per year in Souss-Massa, if continued, would reduce density by roughly 25% over the next two decades.

A limitation is the annual measurement frequency one observation per year is coarse for FDA, and the 17-point series is near the minimum for reliable spline fitting. Sub-annual measurements (e.g., quarterly) would allow detection of seasonal patterns that the current data cannot resolve.

Conclusion

This research compared four spline smoothing methods for functional data analysis of argan tree density time series in Morocco. P-splines with GCV-tuned second-derivative penalties achieved the best accuracy (ISE 0.061). Functional PCA on smoothed curves separated plots into level and trend modes, with Souss-Massa plots showing significant density decline.

For Moroccan forest managers, the FPC2 scores offer a quantitative early-warning metric for identifying degrading plots. P-spline smoothing is recommended for routine FDA of ecological monitoring networks. Future work could incorporate spatial functional models that borrow information between neighbouring plots.

Acknowledgements

Funding sources

This research was supported by a departmental grant from the Coimbatore Department of Algebraic Studies.

Institutional support

The authors thank the Department of Algebra and Geometry for computing resources.

Contributions not qualifying for authorship

The authors acknowledge Dr. Mohammed El-Fassi for HCEFLCD data access.

References

1. x BD. Flexible smoothing with B-splines and penalties. *Statistical Science*. 1996;11(2):89–121.
2. Green PJ, Silverman BW. *Nonparametric regression and generalized linear models*. London: Chapman and Hall; 1994. p. 1–182.
3. Wood SN. Thin plate regression splines. *Journal of the Royal Statistical Society Series B*. 2003;65(1):95–114.
4. Msanda F, El Aboudi A, Peltier JP. Biodiversity and biogeography of Moroccan argan forest. *Cahiers Agricultures*. 2005;14(4):357–364.
5. Ramsay JO, Hooker G, Graves S. *Functional data analysis with R and MATLAB*. New York: Springer; 2009. p. 1–207.
6. Ferraty F, Vieu P. *Nonparametric functional data analysis*. New York: Springer; 2006. p. 1–258.
7. Yao F, Muller HG, Wang JL. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*. 2005;100(470):577–590.
8. Wahba G. *Spline models for observational data*. Philadelphia (PA): SIAM; 1990. p. 1–169.
9. Ruppert D, Wand MP, Carroll RJ. *Semiparametric regression*. Cambridge: Cambridge University Press; 2003. p. 1–386.
10. Charrouf Z, Guillaume D. Argan oil: occurrence, composition and impact on human health. *European Journal of Lipid Science and Technology*. 2008;110(7):632–636.
11. Kokoszka P, Reimherr M. *Introduction to functional data analysis*. Boca Raton (FL): Chapman and Hall/CRC; 2017. p. 1–299.
12. Cuevas A, Febrero M, Fraiman R. Robust estimation and classification for functional data. *Advances in Data Analysis and Classification*. 2007;1(1):5–22.
13. Hyndman RJ, Shang HL. Rainbow plots, bagplots, and boxplots for functional data. *Journal of Computational and Graphical Statistics*. 2010;19(1):29–45.