Kapil Gupta

PhD Candidate, Decision Sciences Area, IIM Bangalore, India

✉ kapil.gupta20@iimb.ac.in • 🌐 kapil21gupta.github.io/ • in Kapil Gupta

Research Interests

Spatial Statistics, House Price Dynamics, Spatio-Temporal Modelling, Clustering and Classification, Sports Analytics, Applications of Variable Selection Methods.

Education

Indian Institute of Management Bangalore — **Bangalore, India**
Doctoral Candidate in Decision Sciences Area — *(Expected) Nov 2024*
Recipient of the Director's Merit List Award in both the years during the coursework

Thesis: Analysing House Price Dynamics using Novel Spatio-Temporal Methods
Advisor: Dr. Soudeep Deb
Other committee members: Dr. Venkatesh Panchapagesan, Dr. Kunal Dasgupta, Dr. Anand Deo

Indian Institute of Technology Delhi — **Delhi, India**
Master of Science in Mathematics — *Jul 2016–May 2018*

Thesis: Numerical Solutions of Singularly Perturbed Linear Problems in One Dimension
Advisor: Dr. S. Chandra Sekhara Rao

Indian Institute of Information Technology, Design and Manufacturing, Kancheepuram — **Chennai, India**
Bachelor of Technology in Computer Engineering — *Aug 2012–May 2016*

Thesis: The cd-coloring of bipartite graphs
Advisor: Dr. Shalu M A

Experience

Smart City Mission, Ministry of Housing & Urban Affairs, India — **New Delhi, India**
Research Consultant — *Since Feb 2024*
- Analysing the improvement in quality of education through smart classrooms.

Real Estate Research Initiative(IIMB-RERI) — **Bangalore, India**
Research Consultant — *Since Mar 2023*
- Working on the development of a commercial rental index through spatio-temporal modeling.

Indxx, LLC — **Gurgaon, India**
Senior Data Analyst, Engineering Division — *Jun 2018–Jul 2020*
- Responsible for data analyzing and cleansing.
- Developed and calculated In-house and client based indices using SQL.
- One of two project team members chosen to visit the client's office (Qontigo) in London.

Indian Institute of Science Education & Research — **Bhopal, India**
Summer Research Intern, Department of Mathematics — *May 2014–Jun 2014*
- Reading project in ordinary differential equations under the guidance of Dr. Ashish Gupta.

Awards and Honors

First Best Paper Award- PhD Scholar Category
JOINT EVENT: 2023-ORSI & 2023-ICBAI, IISC Bangalore, India — *Dec 2023*

Director's Merit List Award
Indian Institute of Management Bangalore — *Jun 2021 & 2022*

Employee of the Year - Entrepreneurship
Indxx, LLC — *Nov 2019*

All India Rank 235
GATE Mathematics — *Mar 2018*

All India Rank 30
CSIR-NET Mathematics — *Jun 2017*

- **All India Rank 60**
  IIT JAM Mathematics *Mar 2016*

Graduate Coursework

- Probability Theory, Statistical Inference, Multivariate Statistics, Advanced statistical Methods and Computing, Multilevel Analysis, Advance Econometrics, Financial Econometrics, Marketing Models and Estimation, Mathematical Methods for Management Research, Linear Programming and Networks, Stochastic Models, Dynamic Programming

Publications

- **Gupta, K.**, Krishnamurthy, V., Deb, S. (2024). What elements of the opening set influence the outcome of a tennis match? An in-depth analysis of Wimbledon data. To appear in IIMB Management Review. Pre-print available on request.

  Abstract: This study explores the importance of first-set game elements in Wimbledon matches. Initial analysis highlights the pivotal role of winning the first set in determining overall match outcomes, with game element behavior varying across rounds. Utilizing a LASSO-induced logistic regression model on first-set data, we identify service points and average player distance as the most significant factors influencing match outcomes. Additionally, ATP rating points prove to have a consistent impact throughout the tournament. Our proposed model, incorporating a suitable random effect structure, demonstrates superior within-match forecasting accuracy during the first set compared to other statistical and machine learning approaches.

- **Gupta, K.**(2022). An integrated batting performance analytics model for women's cricket using Principal Component Analysis and Gini scores, Decision Analytics Journal. [publication]

  Abstract: This study addresses the historical lack of attention to women cricketers by quantifying their batting performances in one-day internationals (ODIs). Traditional ranking methods using simple averages overlook crucial factors like consistency and strike rates in ODIs. To address this, we propose a novel methodology using Gini-based average scores and incorporating batting strike rates in a Principal Component Analysis (PCA) framework. This approach, combining PCA with Gini scores, enhances the understanding of a player's performance for cricket fans, coaches, and managers.

Submitted Articles & Work in Progress

- **Gupta, K.**, Deb, S. (2023+) A divide-and-conquer approach for spatio-temporal analysis of large house price data from Greater London. Under review. Pre-print available on request.

  Abstract: Real estate statistical research, especially in spatio-temporal house price dynamics, grapples with slow standard Markov chain Monte Carlo (MCMC) for large datasets. We propose a divide-and-conquer approach, partitioning data into subsets, utilizing parallel Gaussian process models, and aggregating results via Wasserstein barycenter. This method accommodates multiple observations per spatial and time unit, offering enhanced benefits. Applied to London house price data from 905 areas over eight years, our approach unveils insights into amenities, trend patterns, and price relations to carbon emissions. Demonstrating good predictive accuracy with computational efficiency in cross-validation, it outperforms traditional Bayesian methods.

- Bag, S., **Gupta, K.**, Deb, S. (2022+). A review and recommendations on variable selection methods in regression models for binary data. Under resubmission at International Statistical Review. [Pre-print]

  Abstract: This paper investigates vital variable selection in logistic regression, crucial across medical, financial, and economic studies. Exploring four frequentist typologies (test-based, penalty-based, screening-based, and tree-based), we provide practitioners a comprehensive overview with underlying assumptions and theory. A simulation study assesses sixteen methods in variable selection, coefficient estimation, prediction accuracy, and time complexity across diverse setups. Real-life application with high-dimensional gene expression data enhances insights. Our findings offer practitioners practical recommendations for selecting variable methods based on simulated and real data outcomes in varied contexts.

- **Gupta, K.**, Deb, S., Panchapagesan, V. A novel spatio-temporal statistical model to analyze the real estate market in Bangalore. Manuscript in preparation.

  Abstract: Our study adds to recent statistical research in real estate by revealing spatial and temporal dependence patterns in real estate prices. Introducing a novel statistical model, we efficiently capture these dependencies using a separable Gaussian spatio-temporal process with an additive mean structure and a random error process. Implemented through a Bayesian setup for flexibility and computational advantages, our model is applied to Bengaluru house price

data from January 2015 to March 2020. Residual diagnostics confirm its effectiveness, and the model outperforms other candidates in predictive capabilities.

- **Gupta, K.**, Deb, S. Unveiling dynamics: A mixture model approach for addressing missingness in spatio-temporal data. Manuscript in preparation.

Abstract: This study addresses the challenge of missing spatio-temporal data in housing market analysis, where property transactions occur infrequently. Common assumptions in spatio-temporal analysis do not hold in this context, risking the oversight of valuable insights. We propose a novel spatio-temporal mixture model that distinguishes between observed and unobserved data, incorporating a latent variable to account for missing data points. Employing a Bayesian framework, we simultaneously model both types of data, providing a unified approach to comprehend house price dynamics by extracting insights from the inherent missingness in the spatio-temporal data.

Teaching Experience at IIM Bangalore

Course Instructor

- *Probability and Statistics (Pre-doctoral course)* — *Sep–Nov 2023*
  - Overall Rating: 4.25/5 (No. of respondents: 12)
- *Quantitative Techniques (MBA preparatory course)* — *Jun 2023*
  - Overall Rating: 4.5/5 (No. of respondents: 26)
- *R for Data Science (MBA preparatory course)* — *Jun 2023*
  - Overall Rating: 4.3/5 (No. of respondents: 65)
- *Calculus (PhD preparatory course)* — *Jun 2023*
  - Overall Rating: 3.6/5 (No. of respondents: 11)

Teaching Assistant

- *Sports Analytics (MBA)* — *Jun–Aug 2023*
  - Grade: Excellent
- *Multivariate Data Analysis (MBA)* — *Jun–Aug 2023*
  - Grade: Good
- *Data Science Doctrines: Prediction, Inference, and Causality (MBA)* — *Oct–Dec 2022*
  - Grade: Excellent
- *Decision Sciences II (MBA)* — *Oct–Dec 2022*
  - Grade: Good
- *Decision Sciences I (MBA)* — *Jul–Sep 2022*
  - Grade: Excellent

Presentation and Talks

- **JOINT EVENT: 2023-ORSI & 2023-ICBAI**, **IISC Bangalore, India**
  "A Novel Spatio-Temporal Statistical Model to Analyze Real Estate Market in Bengaluru, India". — *Dec 2023*

- **Annual International Research Conference (AIRC)**, **IIM Lucknow, India**
  "A Novel Spatio-Temporal Statistical Model to Analyze Real Estate Market in Bengaluru, India". — *Dec 2023*

- **Indo-German workshop on Data Mathematics and Scientific Computing** , **IIT Tirupati, India**
  "Efficient Divide-and-Conquer Approach for Spatio-Temporal Modeling of Real Estate Data". — *Sep 2023*

- **Invited Talk: 6th International Conference on Econometrics and Statistics**, **Tokyo, Japan**
  "Efficient Divide-and-Conquer Approach for Spatio-Temporal Modeling of Real Estate Data". — *Aug 2023*

- **IMR Doctoral Conference, 2023**, **IIM Bangalore, India**
  "What elements of the opening set influence the outcome of a tennis match? An in-depth analysis of Wimbledon data". — *Feb 2023*
  - One of the 10 papers accepted out of 98 submissions.

- **Management Doctoral Colloquium Shodh Samagam**, **IIM Visakhapatnam, India**
  "Measuring Batting Performance in Women's Cricket - An In-Depth Analysis of One-Day International Matches". — *Dec 2021*

- **8th MathSport International Conference**, **University of Reading, UK**
  "Does the outcome of a tennis match hinge on the opening set? An in-depth analysis of the Wimbledon data". — *Jun 2021*

Professional Memberships

- *The International Society for Bayesian Analysis* — *Jan 2024-Present*
- *International Indian Statistical Association* — *Dec 2022-Present*
- *Institute of Mathematical Statistics* — *Dec 2021-Present*

Positions of Responsibility

- **Reviewer**
  Research in Transportation Business & Management Journal — *2024*
- **Session Moderator (Young Researchers from India)** — **Online**
  International Day of Women in Statistics and Data Science — *Oct 2023*
- **Session Chair (Spatial Statistics Session)** — **Tokyo, Japan**
  6th International Conference on Econometrics and Statistics — *Aug 2023*
- **Indian Institute of Management Bangalore** — **Bangalore, India**
  PhD Students Academic Representative — *Nov.2020–Oct.2021*
  - Represented the whole PhD student community at IIM Bangalore.
- **Indian Institute of Technology Delhi** — **Delhi, India**
  MSc Mathematics 2016 batch class representative — *Jul 2016–May 2018*
  - Represented the class of 60 MSc students at IIT Delhi.
- **Indian Institute of Information Technology Kancheepuram** — **Chennai, India**
  Quality Management Service core in annual techno-cultural fest Samgatha — *Jul 2014–Jun 2015*
  - Worked as an organiser for the Samgatha, and led the team of 40 volunteers.

Summer School & Workshop

- **Workshop on Trustworthy AI** — **Bangalore, India**
  Microsoft Research, University of Pennsylvania, and Wadhwani AI — *Jan 2023*
  - One of the 35 participants selected from a pool of of 150+ student applicants from India.
- **The Summer Institutes in Computational Social Science (SICSS)** — **Delhi, India**
  Ashoka University & CSIR-Central Scientific Instruments Organisation — *Jun 2022*
  - One of the 25 participants chosen from a pool of applications from scholars all over the world.

Computer Skills

R, SQL, LaTeX, Python, MATLAB, C, C++, Microsoft Office.

References

Prof. Soudeep Deb
Assistant Professor
Decision Sciences Area
Indian Institute of Management, Bangalore
Email: soudeep@iimb.ac.in

Prof. Venkatesh Panchapagesan
Associate Professor
Finance & Accounting Area
Indian Institute of Management, Bangalore
Email: Venky@iimb.ac.in

Prof. Rishideep Roy
Assistant Professor
School of Mathematics, Statistics and Actuarial Science
University of Essex, Colchester Campus
Email: rishideep.roy@essex.ac.uk

Prof. Anand Deo
Assistant Professor
Decision Sciences Area
Indian Institute of Management, Bangalore
Email: anand.deo@iimb.ac.in