Project Title:	**Clustering of nonstationary and nonlinear time series**	
Project Number	**IMURA1030**	
Monash Main Supervisor (Name, Email Id, Phone)	Michael Burke michael.g.burke@monash.edu	*Full name, Email*
Monash Co-supervisor(s) (Name, Email Id, Phone)		
Monash Head of Dept/Centre (Name,Email)	Scott Tyo scott.tyo@monash.edu	*Full name, email*
Monash Department:	Electrical and computer system engineering	
Monash ADGR (Name,Email)	Timothy Scott	*Full name, email*
IITB Main Supervisor (Name, Email Id, Phone)	Radhendushka Srivastava rsrivastava@iitb.ac.in	*Full name, Email*
IITB Co-supervisor(s) (Name, Email Id, Phone)		*Full name, Email*
IITB Head of Dept (Name, Email, Phone)	U. K. Anandavardhanan anand@math.iitb.ac.in	*Full name, email*
IITB Department:	Mathematics	

Research Clusters:

Highlight which of the Academy's CLUSTERS this project will address? *(Please nominate JUST **one.** For more information, see www.iitbmonash.org)*	
1	*Material Science/Engineering (including Nano, Metallurgy)*
2	*Energy, Green Chem, Chemistry, Catalysis, Reaction Eng*
3	*Math, CFD, Modelling, Manufacturing*
4	***CSE, IT, Optimisation, Data, Sensors, Systems, Signal Processing, Control***
5	*Earth Sciences and Civil Engineering (Geo, Water, Climate)*
6	*Bio, Stem Cells, Bio Chem, Pharma, Food*
7	*Semi-Conductors, Optics, Photonics, Networks, Telecomm, Power Eng*
8	*HSS, Design, Management*

Research Themes:

Highlight which of the Academy's Theme(s) this project will address? *(Feel free to nominate more than one. For more information, see www.iitbmonash.org)*	
1	***Artificial Intelligence and Advanced Computational Modelling***
2	*Circular Economy*
3	*Clean Energy*
4	*Health Sciences*
5	*Smart Materials*
6	*Sustainable Societies*

The research problem

Define the problem

Nonlinear time series (parametric/nonparametric) appear in several areas including finance, geomagnetic indices, biological process etc. These time series data are mostly nonstationary and one of the sources of the non-stationarity is the mean functional. We estimate the mean functional by a nonparametric/parametric smooth estimator and adjust the time series. The residual series is believed to be a nonlinear/linear stationary time series. This type of time series data are often observed at high frequency with irregularly (randomly) spaced time instances. The time instances can also be viewed as a point process. An important question of clustering of such time series arises naturally.

There are several methods for clustering the nonlinear time series at regularly sampled data. A possible approach of clustering is based on computing the similarity (distance) measure between the spectrum estimator (feasible kernels) of the time series data.

Huang et al. (2004) proposed a SLEX (smooth local complex exponential) spectra for clustering and discrimination of nonlinear time series data. Shao and Wu (2007) have worked on the asymptotic distribution of smooth periodogram of nonlinear time series under the condition of geometric moment contraction (GMC). Bhattacharjee et. al. (2021) showed that GMC is a strong condition and restricts the parameter space of even ARCH models. Li and Merlin (2015) have worked on classification of sparse and irregularly sampled time series with mixtures of expected Gaussian kernels and random features.

We aim to develop a new clustering/classification algorithm for irregularly spaced non-linear time series in presence of non-trivial mean functional. The distance between the time series will be computed based on a proposed distance between the feasible kernels from the irregularly spaced time series data. The aim is to extract the main features of the spectra and the proposed distance function between spectra exploits this main feature well. The theoretical properties of spectra estimation will be established. Further, the results of the univariate time series will be extended to multivariate time series data.

Huang H.Y., Ombao H., Stoffer D.S. (2004) Discrimination and classification of nonstatioanry tiem series using SLEX model. *J. American Statistical Association.*

Shao X. and Wu W.B. (2007) Asymptotic Theory for nonlinear time series. *Annals of Statistics*

Bhattacharjee B, Bose A., Srivastava R. (2021) A white noise test under weak conditions. *J. Statistical Planning and Inference.*

Li S.C.X. and Merlin B. (2015) have worked on classification of sparse and irregularly sampled time series with mixtures of expected Gaussian kernels and random features. *Proceedings of Thirty-first conference on Uncertainty in Artificial Intelligence (UAI)*

Srivastava R. and Sengupta D. (2013) Nonparametric spectrum estimation under the constraint of a minimum inter-sample spacing. *Statistica Sinica.*

Project aims

Define the aims of the project

We aim to develop a new clustering/classification algorithm for irregularly spaced non-linear time series in presence of non-trivial mean functional. The distance between the time series will be computed based on a proposed distance between the feasible kernels from the irregularly spaced time series data. The aim is to extract the main features of the spectra and the proposed distance function between spectra exploits this main feature well. The theoretical properties of spectra estimation will be established. Further, the results of the univariate time series will be extended to

multivariate time series data.

How skills/experience of the IITB and the Monash supervisor(s) support the proposed project

Highlight the purpose of the collaboration and/or the complementary skills/experience that you bring to the project. Do you have any joint or independent publications in the area of the proposed project?

Dr. Radhendushka Srivastava works in time series analysis and interested in developing machine learning techniques for time series data. This project will lead to some new technique in this direction.

Dr Michael Burke works on sequential Monte carlo and Bayesian non-parametric approaches to time series analysis, with applications in robotics, human skill assessment and computer vision.

What is expected of the student when at IITB and when at Monash?

Highlight how the project will gain from the students stay at IITB and at Monash

The student will focus research on high frequency time series data along with fulfilling the requirement of the Department of Mathematics at IITB.
The student will work under the supervision of Monash and IITB supervisor in the direction of project described.

Expected outcomes

Highlight the expected outcomes of the project

Some new technique of clustering/classification based on feasible kernels of irregularly spaced high frequency time series data.
Some research articles in peer reviewed journal and conference.

How will the project address the Goals of the above Themes?

Describe how the project will address the goals of one or more of the 6 Themes listed above.

Project will develop new technique of clustering/classification of time series data.

Capabilities and Degrees Required

List the ideal set of capabilities that a student should have for this project. Feel free to be as specific or as general as you like. These capabilities will be input into the online application form and students who opt for this project will be required to show that they can demonstrate these capabilities.

Degree Requirement: Any of the following master's degree is desired: MSc. Statistics, MSc Mathematical Statistics, MSc. Applied Statistics and Informatics, MTech. CSE, MSc. Mathematics.

Desired Capability: Programming in R, Python, Machine learning methods of clustering and classification, Some experience of data analysis in academics/industry.