IITB-Monash Research Academy

An Indian-Australian research partnership

Indian Institute of Technology Bombay | MONASH University

CSIRO

Project Title:	Statistical analysis and visualisation of spatially distributed big time series electricity usage data
Project Number	**IMURA0722**
Monash Main Supervisor (Name, Email Id, Phone)	Rob Hyndman Rob.Hyndman@monash.edu
Monash Co-supervisor(s) (Name, Email Id, Phone)	Dianne Cook DiCook@monash.edu
Monash Head of Dept/Centre (Name,Email)	Prof Heather Anderson Heather.Anderson@monash.edu
Monash Department:	Business Econometrics
Monash ADRT (Name,Email)	Fang Cooke Fang.Cooke@monash.edu
IITB Main Supervisor (Name, Email Id, Phone)	Radhendushka Srivastava radhe@math.iitb.ac.in
IITB Co-supervisor(s) (Name, Email Id, Phone)	
IITB Head of Dept (Name, Email, Phone)	Neela Nataraj
IITB Department:	Department of Mathematics
Data61 Supervisors	Peter Toscas Peter.Toscas@data61.csiro.au

Full name, Email (Monash Main Supervisor)

Full name, Email (Monash Co-supervisor(s))

Full name, email (Monash Head of Dept/Centre)

Full name, email (Monash ADRT)

Full name, Email (IITB Main Supervisor)

Full name, Email (IITB Co-supervisor(s))

Full name, email (IITB Head of Dept)

Research Clusters:

Highlight which of the Academy's CLUSTERS this project will address? *(Please nominate JUST **one.** For more information, see www.iitbmonash.org)*

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	*Material Science/Engineering (including Nano, Metallurgy)*	***1***	***Advanced computational engineering, simulation and manufacture***
2	*Energy, Green Chem, Chemistry, Catalysis, Reaction Eng*	**2**	***Infrastructure Engineering***
3	*Math, CFD, Modelling, Manufacturing*	3	*Clean Energy*
4	***CSE, IT, Optimisation, Data, Sensors, Systems, Signal Processing, Control***	4	*Water*
5	*Earth Sciences and Civil Engineering (Geo, Water, Climate)*	5	*Nanotechnology*
6	*Bio, Stem Cells, Bio Chem, Pharma, Food*	6	*Biotechnology and Stem Cell Research*
7	*Semi-Conductors, Optics, Photonics, Networks, Telecomm, Power Eng*		
8	*HSS, Design, Management*		

The research problem

Define the problem

Electricity smart meter technology is increasingly being deployed in residential and commercial buildings. The technology facilitates the collection of energy usage data at much finer temporal scales than was possible previously, such as at hourly or less intervals. The introduction of smart meters affords the opportunity to better model and understand residential and business energy usage patterns between months, between days and within days, something that is not possible using only quarterly energy usage information. Using the geographical data of the households also makes it possible to identify spatial and spatio-temporal patterns. This work will concentrate on undertaking research into spatial and spatio-temporal visualisation and inferential methods for cognostics of spatially distributed big time series electricity usage data.

Project aims

Define the aims of the project

The research activity has three main aims:

1. **Cognostics methods for electricity usage time series data with spatial and spatio-temporal structure.** The aim is to test existing metrics for time series data (e.g. Fulcher, Little, and Jones, 2013) to ascertain if they provide spatially useful diagnostic statistics when the spatial dimension of the spatially distributed big time series data is accounted for. If not, to propose new metrics that do provide spatially useful diagnostic statistics, and to then extend the methods to provide useful spatio-temporal diagnostic statistics. The methods will need to be able to account for spatially distributed explanatory variables such as demographics, building size and material, and household behavioural patterns. The methods developed will be reliant on parallel processing using multiple multi-core computers and platforms such as Hadoop, Spark or Tessera.

2. **Develop visualisation methods for spatial and spatio-temporal cognostics for spatially distributed big time series.** The goal is to provide visualisation methods for supporting the analysis, to find anomalies, possibly errors in the data, or unusual uses, and explore patterns like clusters of behaviour, and summarise the behaviour (e.g. Javed et al, 2010). Cognostics provide efficient numerical summaries that can be understood better with plots of the time series, in the spatial context where the data arises. The new challenge for visualisation is handling the volume of data supplied by smart meters, especially to provide interactive graphics (e.g. Cheng et al, 2016). In addition, providing visual explanations is helpful for decision makers to understand patterns or changes in patterns. There is a need to develop visualisation tools that can be used to describe associations between different variables, clustering in spatial location, demography, behavioural or building type, that can change over time (e.g. Wickham et al, 2012). New methods are needed to visualise changing spatial-

temporal patterns in an intuitive and interactive way.

3. **Develop inferential methods for spatially distributed large time series electricity usage data.** It is important to determine if any identified clusters or patterns are indeed statistically meaningful. That is, what confidence do we have that the identified patterns actually exist and are not a random artefact? As the electricity energy usage data will likely exhibit complex multi-seasonality, serial and spatial correlation, making comparisons against permuted observations assumed to be independent will not be appropriate. The permutations against which comparisons will be undertaken will need to respect the multi-seasonality and build on the seasonal block bootstrap (Dudek *et al*., 2014), the serial correlation (Kreiss and Lahiri, 2012) and the spatial correlation (García-Soidán, Menezes and Rubinos, 2014).

References

Cheng, X., Cook, D. and Hofmann, H. (2016) Enabling Interactivity on Displays of Multivariate Time Series and Longitudinal Data, Journal of Computational and Graphical Statistics, To Appear. Available at http://amstat.tandfonline.com/doi/ abs/10.1080/10618600.2015.1105749.

Dudek, A., Leśkow, J., Paparoditis, E. and Politis, D. N. (2014). A generalized bootstrap for seasonal time series. *Journal of Time Series Analysis* **35**, 89 – 114.

Fulcher, B. D., Little, M. A., Jones, N. S. (2013). Highly comparative time-series analysis: the empirical structure of time series and their methods. *Journal of the Royal Society Interface* **10**: 20130048.

García-Soidán, P., Menezes, R. and Rubinos, O. (2014). Bootstrap approaches for spatial data. *Stochastic Environmental Research Risk Assessment* **28**, 1207 – 1219.

Javed, W., McDonnel, B., and Elmqvist, N. (2010). Graphical Perception of Multiple Time Series. IEEE Transactions on Visualization and Computer Graphics, **16**(6), 927–934.

Kreiss, J.-P. and Lahiri, S. N. (2012). Bootstrap methods for time series. *Time Series Analysis: Methods and Applications* **30**, doi: 10.1016/B978-0-444-53858-1.00001-6.

Wickham, H., Hofmann, H., Wickham, C. and Cook, D. (2012) Glyph-maps for Visually Exploring Temporal Patterns in Climate Data and Models, Environmetrics, invited submission, **23**(5):151–182.

Expected outcomes

Highlight the expected outcomes of the project

- Applied, theoretical and/or computational journal and conference papers
- Prototype software for analysing and visualising large spatially distributed time-series data sets.

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.

The project addresses the first goal: Advanced computation engineering, simulation and manufacture. It explores the statistical methodologies for spatially distributed big time series data of energy consumption and aims to develop soft visualisation tools for spatial and spatio-temporal for the big time series data. These tools would improve the understanding of energy consumption.

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.

Required capabilities

- Degree in statistics, econometrics or related field
- Experience with time series modelling
- Statistical modelling skills, particularly for linear and nonlinear models, statistical learning and data visualisation
- Excellent programming skills in R and C/C++

Desirable

- Spatial and spatio-temporal modelling skills
- Skills in parallel processing using multi-core computers and platforms such Hadoop, Spark or Tessera

Potential Collaborators

Please visit the IITB website www.iitb.ac.in OR Monash Website www.monash.edu to highlight some potential collaborators that would be best suited for the area of research you are intending to float.

Select up to **(4)** keywords from the Academy's approved keyword list **(available at http://www.iitbmonash.org/becoming-a-research-supervisor/))** relating to this project to make it easier for the students to apply.

6. Data Science, optimisation, algorithms
14. Signal processing
25. Computer simulation