INDIAN INSTITUTE OF TECHNOLOGY MADRAS
Department of Chemical Engineering
**CH 5350 Applied Time-Series Analysis**

# Project Statement

Due: Friday, November 29, 2019

---

**[Fuzzy time-series modelling]**

The course helped us master modelling of linear stationary processes that are possibly intertwined with trends, seasonality and random walk effects. Real-life processes, however, exhibit a variety of characteristics that are outside the purview of additive / multiplicative seasonal (plus trend) auto-regressive, integrated, moving average models. Non-linear processes are prime examples of such excursions.

There exist several modelling paradigms for non-linear random processes as new ones continue to emerge. A popular paradigm is the *fuzzy time-series (FTS)* model originally proposed by Song and Chissom (1993). The FTS paradigm rests on the notion of fuzzy logic introduced by Zadeh (1965). FTS method consists of four steps, namely, (i) partitioning the given data into intervals (ii) defining fuzzy sets and establishing a logical relationship between the fuzzy sets (iii) forecasting for each set and (iv) defuzzifying the forecasts. A paper by Duru and Yoshida (2012) details the principles in fuzzy time-series forecasting A step-wise illustration of FTS on Shipping and Transportation Index is provided in Chou (2018). Finally, an excellent introduction to FTS is available at `https://towardsdatascience.com/a-short-tutorial-on-fuzzy-time-series-dcc6d4eb1b15`

The objective of your project is to predict **relative humidity** using FTS modelling and compare the forecasts with a standard optimally tuned SARIMA model that takes temperature as exogenous regressors. The specific instructions for the project follow after a description of the data below.

Data is obtained from an Automated Weather Station (AWS) located at Sriharikota during the period May 15 - July 07, 2009. The data set comprises measurements of different meteorological variables namely, air temperature, wind speed, wind direction, atmospheric pressure, relative humidity (RH), rainfall and sunshine. In addition, other meta details such as the station code, latitude / longitude, date and time of measurements are included.

1. Read the data into R using the `read.csv` routine (you may use the "Import Dataset" option in RStudio). Ignore the first four columns for the remainder of analysis.

2. Observe that the data is missing at different date / time locations. Write an R script that identifies these missing locations based on the TIME column and inserts "NA" for corresponding values in the variables.

3. Use your favourite missing data replacement algorithm to impute ("fill in") the missing observations. Partition the data into training and test data for model fit and validation, respectively. For this purpose, you may reserve the last

4. Create TS objects for RH and temperature variables, respectively.

5. Build the best fuzzy time-series model by either writing your own R script or using the `AnalyzeTS` package. Call the resulting model $\mathcal{M}_1$. Report all your findings, especially the forecast metrics. Discuss the goodness of your model.

6. Develop a linear SARIMA model with temperature as the exogenous input to RH. Determine the appropriate past values of temperature to be included using a mix of cross-correlation analysis and trial-and-error approach. Call this model $\mathcal{M}_2$. The procedure (for developing the model) should be systematic and include all checks. Report the model and all other relevant details.

7. Compare the forecasts of $\mathcal{M}_1$ and $\mathcal{M}_2$ on the basis of training and test performance to select the better model among the two. Provide suitable justification for your choice.

8. Using the model that you select in Step 7, replace the imputed values in Step 3 with their predictions. Re-build the model with this interpolated data. Does the resulting model and / or the forecasts appear any significantly different from its predecessor? If yes / no, why?

---

**Instructions for submitting the report:**

- All reports should be submitted in the **form of `R` Markdown and the resulting PDF files only.** No other format is acceptable.
- A maximum page size of 15 and minimum font size of 11 should be observed.
- All `R` scripts (and functions) should be embedded in the markdown file. Individual submission of script files is not permitted.
- Each finding should be aptly supported or backed by statistical and logical arguments.
- Figures should be neatly annotated and be well-explained in the report.
- All equations and mathematical symbols (if any) should be typeset in LaTeX only.

---

## References

Q. Song, and B.S. Chissom (1993). Fuzzy Time Series and Its Models, *Fuzzy Sets and Systems*, vol. 54, pp. 269-277.

L.A. Zadeh (1965). Fuzzy Sets, Inform and Control, vol. 8, pp. 338-353.

O. Duru and S. Yoshida (2012). IEEE Conference on Computational Intelligence for Financial Engineering & Economics, pp. 1-7, March 29-30, New York, USA.

Ming-Tao Chou (2018). Fuzzy Forecast Based on Fuzzy Time Series, Time Series Analysis - Data, Methods, and Applications, Chun-Kit Ngan, IntechOpen, DOI: 10.5772/intechopen.82843. Available from: `https://www.intechopen.com/books/time-series-analysis-data-methods-and-applications/fuzzy-forecast-based-on-fuzzy-time-series`