UNIVERSITÀ DEGLI STUDI DI MILANO

Corso di Dottorato in Scienze della Terra

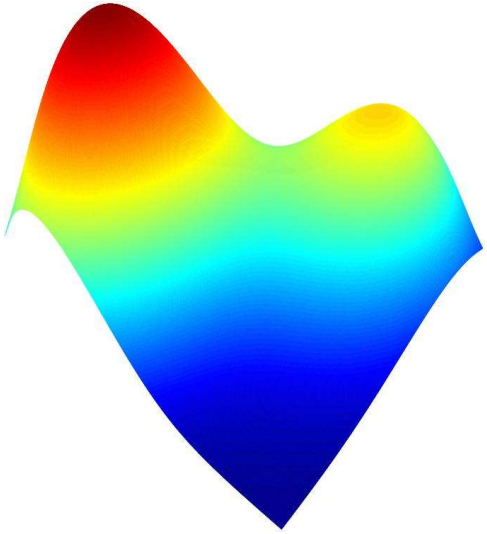
Doctoral Program in Earth Sciences
University of Milan

9-12 October 2023 - Short course (4 cfu, 20 hours)

Dipartimento di Scienze della Terra "A. Desio", via Mangiagalli 34, Milano

Advanced statistical modelling of space-time environmental data using GAMs

by Theo Economou



PROGRAM

- **Introduction to semi-parametric statistical modelling and Generalized Additive Models (GAMs)**
- **Inference in GAMs (model checking, prediction, Bayesian interpretation**
- **Spatial and temporal structures in GAMs (including hierarchical structures)**
- **Case studies and applications**

Dr Theo Economou
t.economou@cyi.ac.c
y
The Cyprus Institite

Per informazioni e iscrizione contattare:
Dr. Corrado Camera (corrado.camera@unimi.it)

Course description

The course aims to provide an introduction to data modelling via the framework of Generalised Additive Models (GAMs), combined with Bayesian inference for full uncertainty quantification. The focus is on environmental data, although the scope of the methodology is much wider. This course is designed to provide an exposition to data analysis and modelling using the learning framework of Bayesian inference in conjunction with GAMs. The course will introduce the building blocks behind GAMs, and placing the framework within the world of AI and data science. Besides GAMs, the course will introduce basic concepts such as (aleatoric and epistemic) uncertainty, statistical inference and prediction, Bayesian analysis and decision theory. The course is focused on exposing the general concepts rather than going over theoretical justifications and is thus accessible to a wide range of participants. All concepts are demonstrated using a series of examples, from simple to more complicated ones, using the statistical programming language R. Familiarity with R is not a requirement; it is merely used here as a tool.

Course structure

The course is divided into 4 topics:

Topic 1: Introduction probabilistic data modelling

Topic 2: Introduction Generalized Additive Models

Topic 3: GAMs as a general modelling tool

Topic 4: Modelling space and time data with GAMs

Course aims:

- Introduce a unifying framework with which one can model environmental data (not provide a “cookbook” of recipes for tackling various data challenges)
- Show how the framework can be used to learn about processes that generated the data and use the learning to solve problems (answer scientific questions).
- Introduce the idea of using probability as a way to quantify uncertainty.
- Introduce Generalised Additive Models as a unifying data modelling framework via Bayesian inference (learning).
- Introduce decision theory as a prescriptive tool for making decisions under uncertainty.
- Demonstrate how GAMs can be used as a general framework to model space-time environmental data.

Learning outcomes:

- Distinguish between buzzwords such as Machine Learning, Artificial Intelligence, Data Science etc. mean exactly and what the can/can't do.
- Understand how particular data modelling methods can be viewed as specific cases of a general framework.
- Appreciate uncertainty, its sources, and why/how it is quantified.
- Understand Bayesian inference as a general approach to learning from data.
- Understand decision theory (the idea of loss/utility functions) as a method with which predictions from data models can be interpreted and thus actual use.
- Experience the R language as a data modelling tool by going over basic worked examples.
- Understand what semi-parametric statistical modelling is.
- Understand the basic mathematical concepts underlying GAMs
- Understand GAMs and how many regression methods can be viewed as specific cases.
- Understand Bayesian predictive inference and model checking.
- Appreciate the use of the mgcv R package as a flexible data modelling tool for environmental data.

Programme for “Advanced statistical modelling of space-time environmental data using GAMs”

09 Oct: Topic 1: Introduction probabilistic data modelling

9.00-12.00 – Lectures - **Aula M02** (Via Mangiagalli N° 31)

13:30-15:30 – Practical session - **Aula M02** (Via Mangiagalli N° 31)

10 Oct: Topic 2: Introduction Generalized Additive Models

9.00-12.00 – Lectures - **Aula M02** (Via Mangiagalli N° 31)

13:30-15:30 – Practical session - **Aula M02** (Via Mangiagalli N° 31)

11 Oct: Topic 3: GAMs as a general modelling tool

9.00-12.00 – Lectures - **Aula C10** (Via Mangiagalli N° 25)

13:30-15:30 – Practical session **Aula C10** (Via Mangiagalli N° 25)

12 Oct: Topic 4: Modelling space and time data with GAMs

9.00-12.00 – Lectures **Aula C06** (Via Mangiagalli N° 25)

13:30-15:30 – Practical session **Aula C10** (Via Mangiagalli N° 25)

Course level and prerequisites:

The course is appropriate for those involved in any type of data analysis, where data can be observational or in the form of physical model output. It does not require a high level of theoretical mathematical/statistical knowledge but it does assume some familiarity with the notion of a probability distribution.

For the practical element of the course, you will need to bring your own laptop with the R software (or ideally RStudio) installed.