# Data analysis tools

| | |
|---|---|
| **MODULE TITLE** | Data analysis tools |
| **LECTURER(S)** | D. Koutsoyiannis and C. Makropoulos |
| **ECTS VALUE** | 6 |
| **PREREQUISITES** | Mathematics |
| **COREQUISITES** | |
| **DURATION OF MODULE** | 13 weeks |

**TOTAL STUDENT STUDY TIME**

Overall, the module is expected to involve students in approximately 150 hours of learning, including 12 3-hour lectures; 44 hours assignments; 70 hours private study.

**WEB LINK** http://www.water-msc.org/en/wrem103.htm

**AIMS**

This module aims to provide a basic knowledge of tools for data analysis. It includes two sub-modules, geographical information systems, and probability and statistics. The first sub-module introduces geographic information systems, their main concepts, including inter alia, data models, geographic projections, spatial analysis and quality assurance. It further provides an introduction and hands-on tutorials for a leading, commercial, GIS software (ArcGIS by ESRI). The second sub-module provides essentials of probabilistic modelling of natural processes, introduces characteristic concepts such as return period and risk, as well as characteristics natural behaviours such as seasonality, intermittency and persistence, and provides tools to analyze data series and build probabilistic models useful for the design and management of hydraulic structures.

**INTENDED LEARNING OUTCOMES**

**1. Subject Specific Knowledge, Understanding and Skills**

By the end of this module, the students should:

a) understand the basics of geographic information systems, incl. projection, data models and data structures, spatial analysis and interpolation, catchment delineation as well as error and quality control for geographic information.

b) be able to perform a range of data visualisation, querying, manipulation and analysis tasks within ArcGIS 9*, using ArcCatalog, ArcToolbox and ArcMap and the Spatial Analyst Extension.

c) understand basics of uncertainty and apply probabilistic and statistical concepts to quantify it;

d) extract probabilistic predictions and design quantities of engineering structures.

**2. Core Academic Skills**

By the end of this module, the students should:

a) improve their general scientific background and their view of spatial information and its analysis;

b) acquire practical experience in a state-of-art GIS software;

c) improve their general scientific background and their view of the natural processes;

d) be able to critically assess research results based on data analysis;

e) have acquired some practical experience of using hydrological data and modelling tools;

f) have acquired an understanding of the impact of solutions for civil engineering works in a global and societal context.

**3. Personal and Key Skills**

By the end of this module, the students should have:

a) improved further the necessary skills for independent learning;

b) enhanced report and presentation skills;

c) enhanced using of spreadsheets and geographical information systems (GIS);

d) acquired an ability to function in multi-national teams.

**LEARNING/TEACHING METHODS**

Lectures, self-evaluation quizzes.

**ASSIGNMENTS**

Two assessed coursework assignments (2,500 equivalent words each including graphs and tables).

**ASSESSMENT**

Examination paper (60%), Course work (40%)
3-hour examination - use of notes and book allowed
2 assignments on practical application of modelling tools (2*20% = 40%, 5,000 equivalent words, including graphs and tables)

**SYLLABUS PLAN (Lectures: 3h each)**

1. **Introduction to GIS** and **the ArcGIS software:** What is a GIS? Who uses GIS and why? Key components of GIS and approaches to conceptualising it. Introduction to the GIS software to be used for hands-on tutorials: key aspects and components.
2. **Data models** and **ArcGIS interfaces:** Raster, Vector and TIN data models. Differences, uses and advantages/disadvantages for different data models. Overview of the main interfaces of ArcGIS; Introduction to ArcCatalog, ArcMap and ArcToolbox.
3. **Map Projections:** Ellipsoids, Spheroids and map projections. Coordinate systems; the GPS
4. **Spatial Analysis:** DEM/DTM/DSM, Spatial Interpolation in GIS (IDW, Spline and Kriging). Introduction to the Spatial Analyst Extension of ArcGIS.
5. **Catchment Delineation** and **ArcHydro:** The 8-pour point model; cell size and the thousand million rule; watershed and stream network delineation; Introduction to Water-specific tools within ArcGIS.
6. **Advanced data structures and data types** and **Error and Quality Control in GIS:** Geodatabases and spatio-temporal data models; sources of error in GIS, accuracy, measurement errors, locational errors; scale, resolution and accuracy**.**
7. **The utility of probability:** Historical evolution of the notion of probability; deterministic vs. indeterministic views of the world and approaches in modelling natural processes; the emergence of uncertainty; the probability as a tool to quantify uncertainty, rationalize decisions under uncertainty, and make predictions of future events under uncertainty.
8. **Basic concepts of probability:** The Kolmogorov probability system; random variables; probability distribution function; probability density function; moments; joint, marginal and conditional distribution functions; the concept of a stochastic process; the central limit theorem.

9. **Basic concepts of statistics:** Samples; estimation; estimators; confidence limits; statistical tests.
10. **Special concepts of probability theory in hydrology**: Seasonality; persistence; intermittency; skewness and long distribution tails; return period; risk. **Typical statistical analysis of a single hydrological variable:** Sample statistics; empirical distribution function; probability plots; statistical prediction; typical applications.
11. **Typical distribution functions in hydrology and water resources:** The normal and log normal distributions; gamma and beta distributions; the Pareto distribution; the extreme value distributions.
12. **Multiple time scale analysis of hydrological extremes - Rainfall intensity-duration-frequency relationships:** General mathematical framework and typical expressions; consistent parameter estimation methods; case study.

**INDICATIVE BASIC READING LIST**


1. Burrough, P.A. and McDonnell, R.A.,1998: Principles of Geographical Information Systems, 2nd Edition, Oxford University Press, Oxford.
2. Papoulis, A., 1990: *Probability and Statistics*, Prentice-Hall, New Jersey.
3. Kottegoda, N. T., and Rosso, 1997: R. *Statistics, Probability, and Reliability for Civil and Environmental Engineers*, McGraw-Hill, New York.


**EXTENDED READING LIST**


1. ESRI, 2004, ArcGIS 9 - Using ArcMap, ESRI Press, USA.
2. Bras, R. L. and Rodriguez-Iturbe, I., 1985: *Random functions and hydrology*, Addison-Wesley, USA.
3. Papoulis, A., 1991: *Probability, Random Variables, and Stochastic Processes*, 3rd ed., McGraw-Hill, New York.


**AUTHORS**


1. Demetris Koutsoyiannis, Assoc. Professor NTUA
2. Christos Makropoulos, Lecturer, NTUA