BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024,
Tamil Nadu, India

Programme: M.Sc., Environmental Science and Sustainable Management

Course Title : **GEOSPATIAL DATA MODELLING AND INFORMATICS**
Course Code : 21PGCC06

Module-III
Environmental Informatics

Dr. M.Prashanthi Devi
Professor
Department of Environmental Science and Management

Environmental informatics

Introduction to Environmental Informatics

Environmental informatics is a part of applied Informatics and supports methods and procedures of information technologies which contribute to environmental data analysis and environmental protection.

It is a connection link between nature and engineering.

Applying data sampling and data analysis methods statistics simulation models and decision support systems on environmental problems and tasks one gets simulation results and prognoses to explain and to solve environmental problems.

Methods of environmental informatics form a basis of decision making processes for environmental problems using state-of-the-art computer technology.

Topics of environmental informatics

- data capture and data storage
- methods of environmental sampling
- environmental data analysis
- environmental statistics
- environmental time series
- environmental simulation models
- decision support systems

Information technology

- Data base systems (DBS),
- Laboratory information and management systems(LIMS)(documentation)
- Environmental information systems (EIS),
- Geographical information systems (GIS),
- Satellite imagery systems (SIS),
- General positioning systems (GPS).

Decision making

- Data bases (create, maintain, and control)
- simulation models (SM) (Monte Carlo model)
- knowledge systems (KS)
- decision support systems (DSS).(ex. Book seller and bank loan)

Environnemental information Systems (EIS)

- Management of data about soil, air and atmosphere, water and waste water, species and individuals in the environment.

EIS are tools to

- collect,
- analyse,
- condense,
- administrate,
- visualize
- environmental data, to use model building and simulation techniques, and to make environmental data available for decision makers.

Ecosystems

- Mostly, tools of environmental informatics are applied to ecosystems.
- An ecosystem is defined in principle as a biotic and functional system or unit, which is able to sustain life and includes all biological and non-biological variables in that unit.
- Spatial and temporal scales are not specified a priori, but are entirely based upon the objectives of the ecosystem study.

Approaches to the study of ecosystems

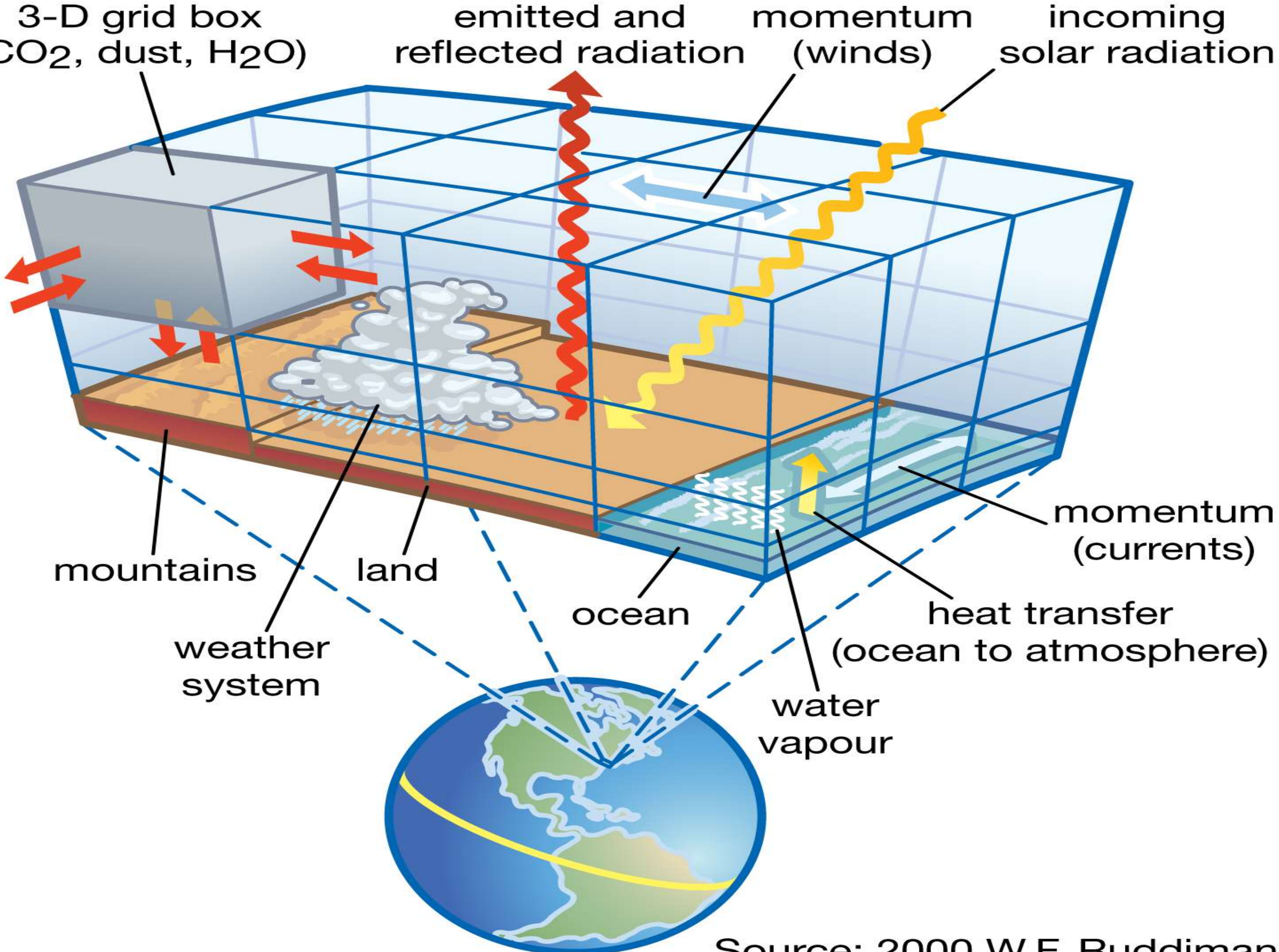
- Empirical studies ➡ where bits of information are collected. An attempt is made to integrate and assemble these into a complete picture (Restoration ecology)

- Comparative studies ➡ where a few structural and a few functional components are compared for a range of ecosystem types (Same ecosystem but different geographic regions)

- Experimental studies ➡ where manipulations of an ecosystem are used to identify and elucidate ecological mechanisms (Human activities -altered continental ecosystems).

- Modelling and computer simulation studies ➡ to work out ecosystem management plans and to derive and eco-technological tools for goal oriented control actions.

Concept diagram of climate modeling



Source: 2000 W.F. Ruddiman

Complexity of ecosystems

- Environmental systems are complex dynamic systems. They are characterised by high dimension (expressed by the number of subsystems and interrelationships). Solving complex environmental problems the system has to be decomposed.

- Complex systems may be also characterised by the degree of computability and the uncertainty of statements.

- Problems of observability of system characteristics, controllability of system states, perturbability of system states (disturb or disquiet greatly) reachability of system states (FINAL), and robustness of mathematical model description have to be solved and questions on the reality of the mathematical model have to be answered.

- Mostly the information structure of complex systems is restricted because of aggregated information levels within the system.

Environmental objects (EO)

- The first step of data processing concerns the mapping of environmental objects(real world objects) to entities that are more abstract and can be handled by computers or directly by decision makers.

- Environmental objects can be natural entities (e.g. animals, plants, lakes, hills, landscapes) or man-made structures (e.g. houses, towns, factories). Each environmental object is mapped to a collection of environmental data objects.

- For instance, a typical environmental data object would be a time series of water quality measurements which describe the time varying behaviour of a certain chemical substance in a lake or river (the corresponding environmental objects).

.

Characterisation of an environmental system (ES)

- An environmental system (ES) is a set of interrelated objects (e.g. animals, species, elements, parts, compartments) that have certain general properties
  - It fulfils a certain function
    - It can be defined by a system purpose recognisable by an observer

- It has a characteristic constellation of (essential) system elements and an (essential) system structure which determines its function, purpose, and identity
  - It loses its identity if it is destroyed.
    - A system is not divisible, that means the system purpose can no longer be fulfilled, if one or several (essential) elements are removed.

Environmental metadata

- Environmental metadata refer to a particular environmental data object.

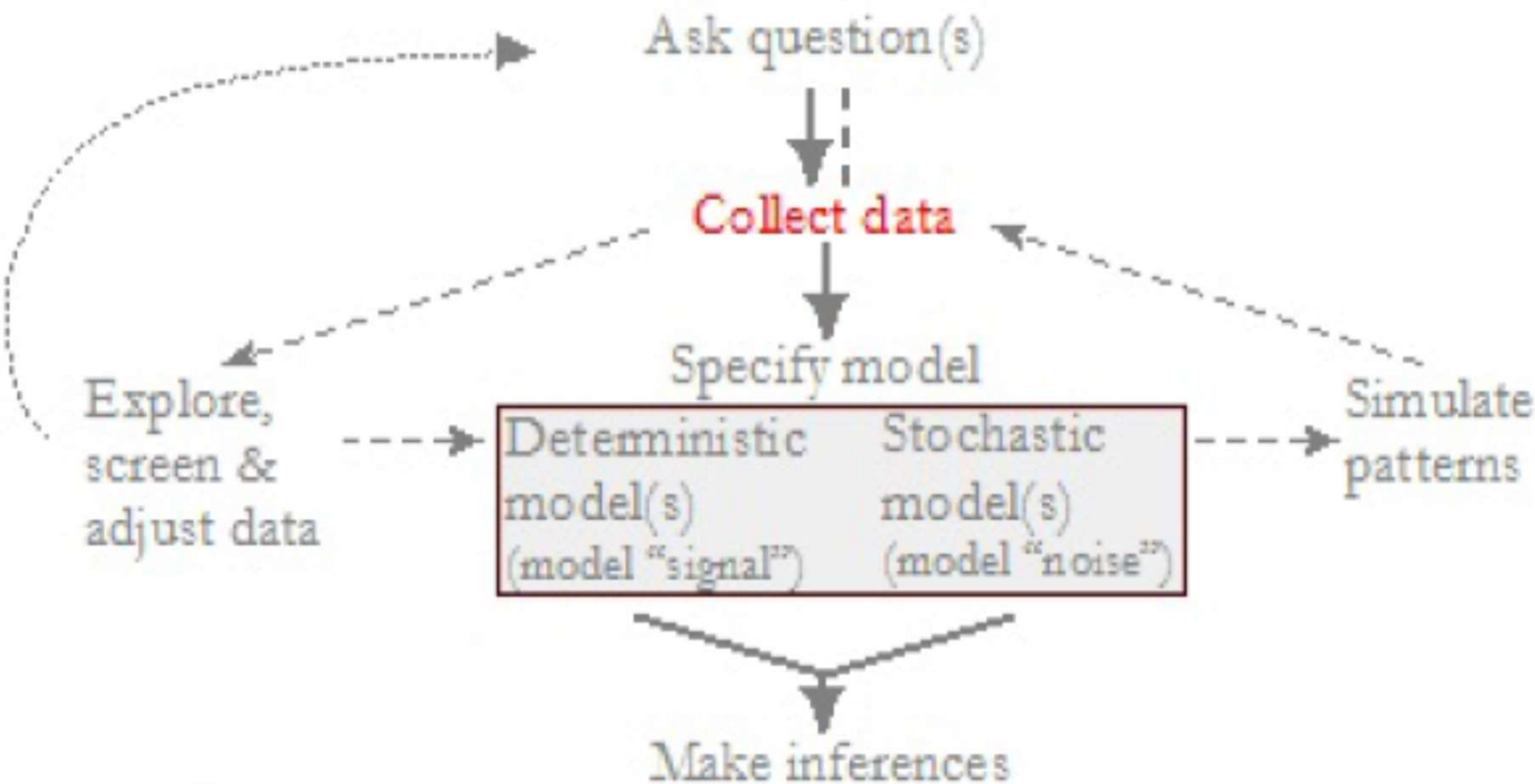
- Each environmental data object is associated with one or more metadata objects that specify its format and contents. (ex. author, date created, date modified, and file size)

- Data on spatial and/or temporal scales are included.

Environmental data analysis

- Applications of data analysis methods are helpful to form a base for environmental modelling and simulation as well as for decision making and environmental impact assessment.
- Disturbances of data analysis are caused by small sets of representative regular sampled data which are available. The power of external and internal driving forces on environmental indicators influences the quality of data to be obtained. The a-priori process information on environmental indicators is low.
- Re-sampling of data can be done by interpolation or approximation methods to place data on a regular time and space grid. Interpolation methods deliver equidistant data while functional relationships result from approximations.

Environmental Data



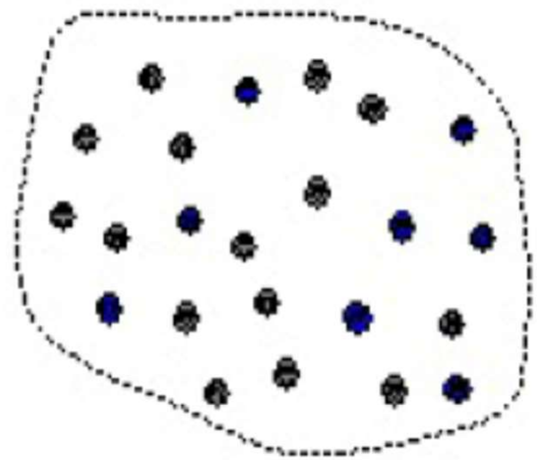
- There are many important aspects to the collection of environmental data relating to study design and sampling method that influence the type and strength of statistical inferences that can made

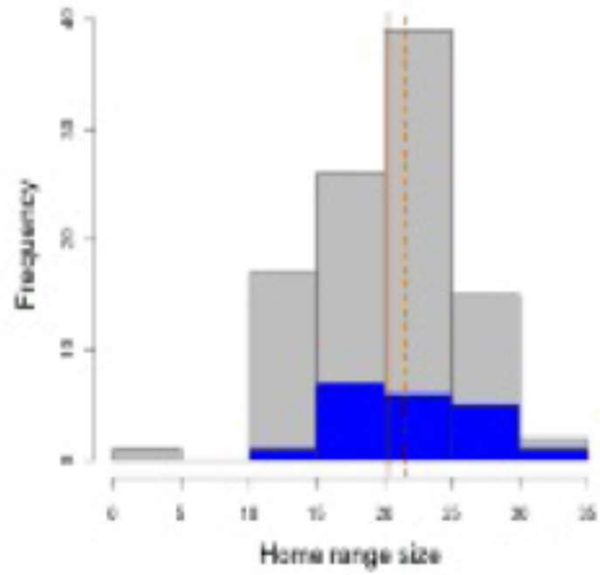
Purpose of data collection

- strong inferences
- Influence the type and strength of statistical inferences
- identifying the desired scope of inference
- choosing appropriate experimental (sampling) units
- choosing the types of data to collect.

Environmental Data... samples and populations

- We usually wish to make inferences about a *population* (statistical, not biological), which is defined as the collection of all the possible observations of interest (size=N)
- We usually collect only a subset of observations from the population, called a *sample* (size=n)
- We *infer* characteristics of the population from the sample; e.g., estimate parameters, test hypotheses, compare models, and make predictions





Environmental Data... sampling units

Sampling units are the units in space and time that we either experimentally manipulate or observe, and the following:

- *Scale dependence*... all sampling units have an explicit spatial and temporal scale that defines the boundaries of a single unit
- *Source of variability of interest*... the sampling units exhibit variability that is the focus of our analysis
- *Subsampling*... sampling units may be subsampled for reasons of study design (e.g., multiple trials), but these subsamples are combined in some appropriate way (e.g., mean) to represent a single observation for each sampling unit
- *Statistical population*... the full collection of all sampling units within the spatial and temporal scope of desired inference represents the statistical "population"

Environmental Data... what to measure

- Given our question, we first need to determine *what* data to collect for each of the sampling units:
  - *Relationships among variables...* independent vs dependent, interdependent
  - *"Type" of data...* continuous, count, proportions, binary, time at death, time series, etc.

Environmental Data... types of variables

Independent versus Dependent

In most cases, we are interested in relating one or more independent variables to one or more dependent variables

- *Independent variable*... typically the variable being manipulated or changed; controlled or selected by the experimenter to determine its relationship to an observed phenomenon (i.e., the dependent variable); also known as "X ", "predictor," "regressor," "controlled," "manipulated," "explanatory," "exposure," and/or "input" variable
- *Dependent variable*... the observed result of the independent variable being manipulated; usually cannot be directly controlled; also known as "Y", "response," "regressand," "measured," "observed," "responding," "explained," "outcome," "experimental," and/or "output" variable

Environmental Data… scales of variables

"Scale" of variable

- Refers to the scale of measurement or observation (although scale has other dimensions) of any variable, dependent or independent
- Function of the intrinsic nature of the variable and the researcher's choice of how to quantify the variable
- Affects the form of the statistical model (if dependent variable) and details of the model (if independent)

"Scales" of data:

Qualitative:

- Nominal categorical (A, B, C)

Unordered ex Male/Female

Quantitative:

- Ordinal (rank ordered, e.g., counts)
- Interval (arbitrary zero, e.g., temperature)
- Ratio (true zero, e.g., mass)

Environmental Data… types of data

"Type" of data

- Refers to the form of the <u>dependent</u> or response variable (not the independent or predictor variables)
- Function of both the scale of the dependent (e.g., continuous versus categorical) and the way in which the data was collected owing to the study design
- Largely determines the overall form of the statistical model (especially regarding the stochastic component)

"Types" of data:

- Continuous
- Count
  - Simple count
  - Categorical
- Proportion — ↑quantity of 1 variable other fixed
- Binary
- Time to death/failure
- Time series
- Circular — Aspect

Environmental Data… types of data

Continuous Data

- Data in which the observations of the sampling units for the dependent variable can be measured on a *continuum* or scale; can have almost any numeric value; can be meaningfully subdivided into finer and finer increments, depending upon the precision of the measurement system.

Examples:

- Temperature
- Mass
- Distance
- Etc.

Some methods:

- Regression
- Analysis of variance

Environmental Data... types of data

Count Data

- Data in which the observations of the sampling units for the dependent variable can take only the *non-negative integer values* {0, 1, 2, ...} and have no upper bound, and where these integers arise from counting rather than ranking.

Examples:

- #territories
- #detections in each habitat type
- Etc.

1) Simple counts

Plot	#Infected
1	2
2	11
3	7
...	...

2) Categorical data

	Town	
Species	A	B
Sugar maple	4	9
Red maple	2	3
Norway maple	21	43

Some methods:

- Log-linear models
- Contingency tables

Environmental Data... types of data

Proportion Data

- Data in which we have multiple observations (trials) for each sampling unit (#trials/unit = trial size) and we know in how many of the trials the event of interest (dependent variable) occurred and how many times the event did *not* occur, so that we can express the result as a *proportion*.

Trial size	#Infected	#Not infected
10	8	2
15	11	4
12	9	3
...	...	...

Examples:

- Percent mortality
- Percent infected
- Sex ratio
- Etc.

Some methods:

- Logistic regression

Environmental Data... types of data

Binary Data

- Data in which we have a single observation (trial) for each sampling unit (trial size=1) and the <u>dependent</u> variable can take only one of two values; i.e., when the response is *binary* (yes/no) and you can have unique values of one or more explanatory variables for each observational unit. Special case of proportion data when trial size equals 1.

Individual	Infected
1	0
2	1
3	1
...	...

Examples:

- Present or absent
- Dead or alive
- Male or female
- Etc.

Some methods:

- Logistic regression

Environmental Data... types of data

Time to Death/Failure

- Data in which the observations of the sampling units for the <u>dependent</u> variable take the form of measurements of the *time to death* (or *time to failure* or *time to success*); each entity is followed until it dies (or fails or succeeds), then the time of death (or failure or success) is recorded.

Individual	Time to death
1	7
2	10
3	1
...	...

Examples:

- Animal/plant longevity
- Snag fall
- Roof failure (leakage)
- Etc.

Some methods:

- Survival analysis

Environmental Data... types of data

Time Series

- Sequence (vector) of data points, in which a variable (generally considered a <u>dependent</u> variable) is measured typically at *successive* times (or locations), spaced at (often uniform) time (or space) intervals; i.e., serially correlated data.

Time	Measurement
1	0.07
2	1.20
3	0.61
...	...

Examples:

- Population size
- Annual temperature
- Etc.

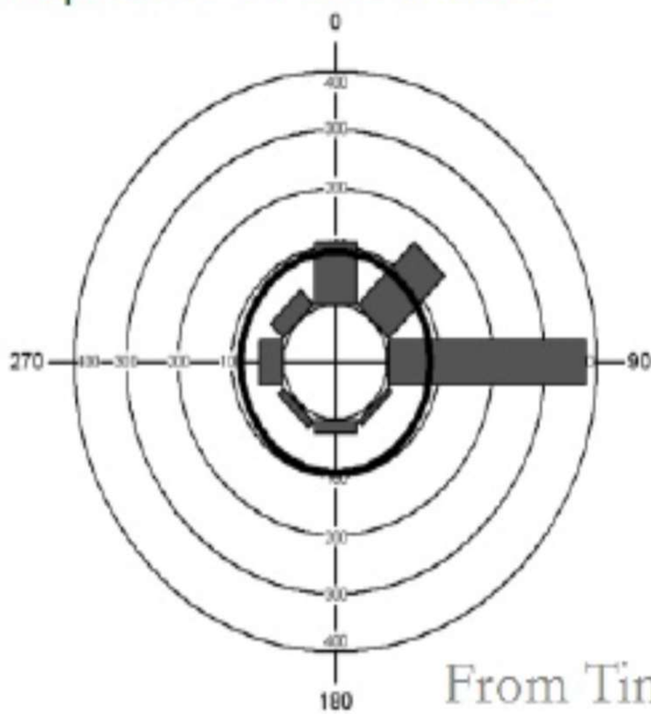
Some methods:

- Autocorrelation
- Spectral analysis
- Wavelet analysis

Environmental Data... types of data

Circular Data

- Data in which the observations of the dependent variable are *circular* in nature; where the beginning and end of the sequence is the same.



From Timm et al (2007)

Examples:

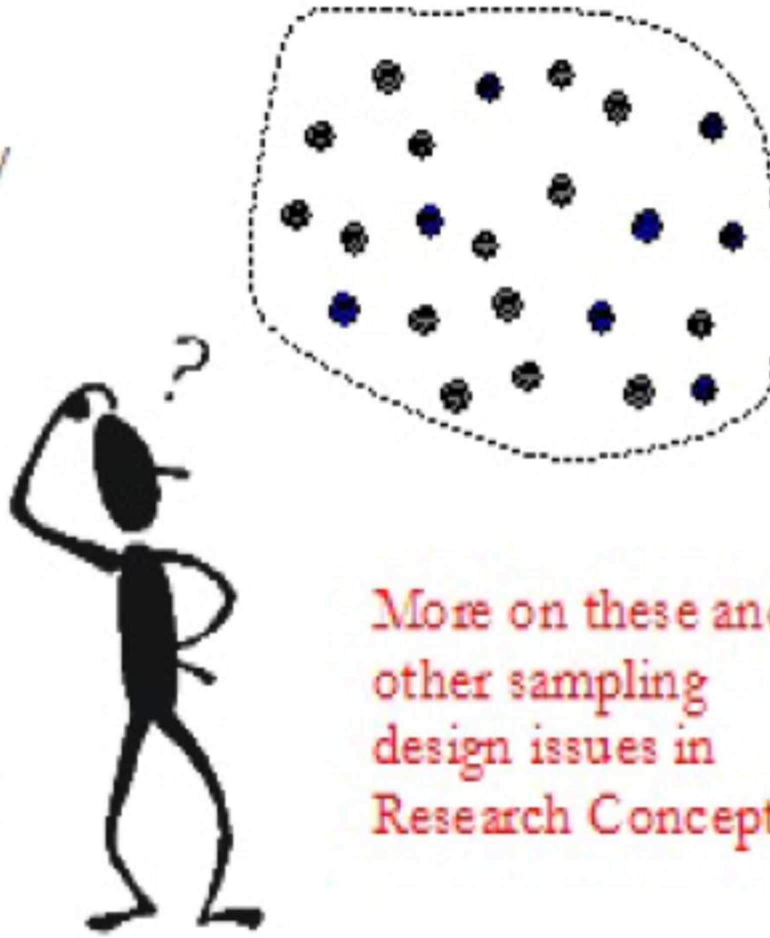
- Aspect
- Turning angle
- Day of year*
- Etc.

Some methods:

- Circular stats

Environmental Data... sampling design

- Next, we need to determine *where*, *when*, and *how often* to collect the data:
  - *Scale*... matching observational/ experimental units to the ecological question
  - *Randomization*... obtaining an unbiased sample
  - *Replication*... minimizing uncertainty
  - *Control*... accounting for important sources of variation



More on these and other sampling design issues in Research Concepts

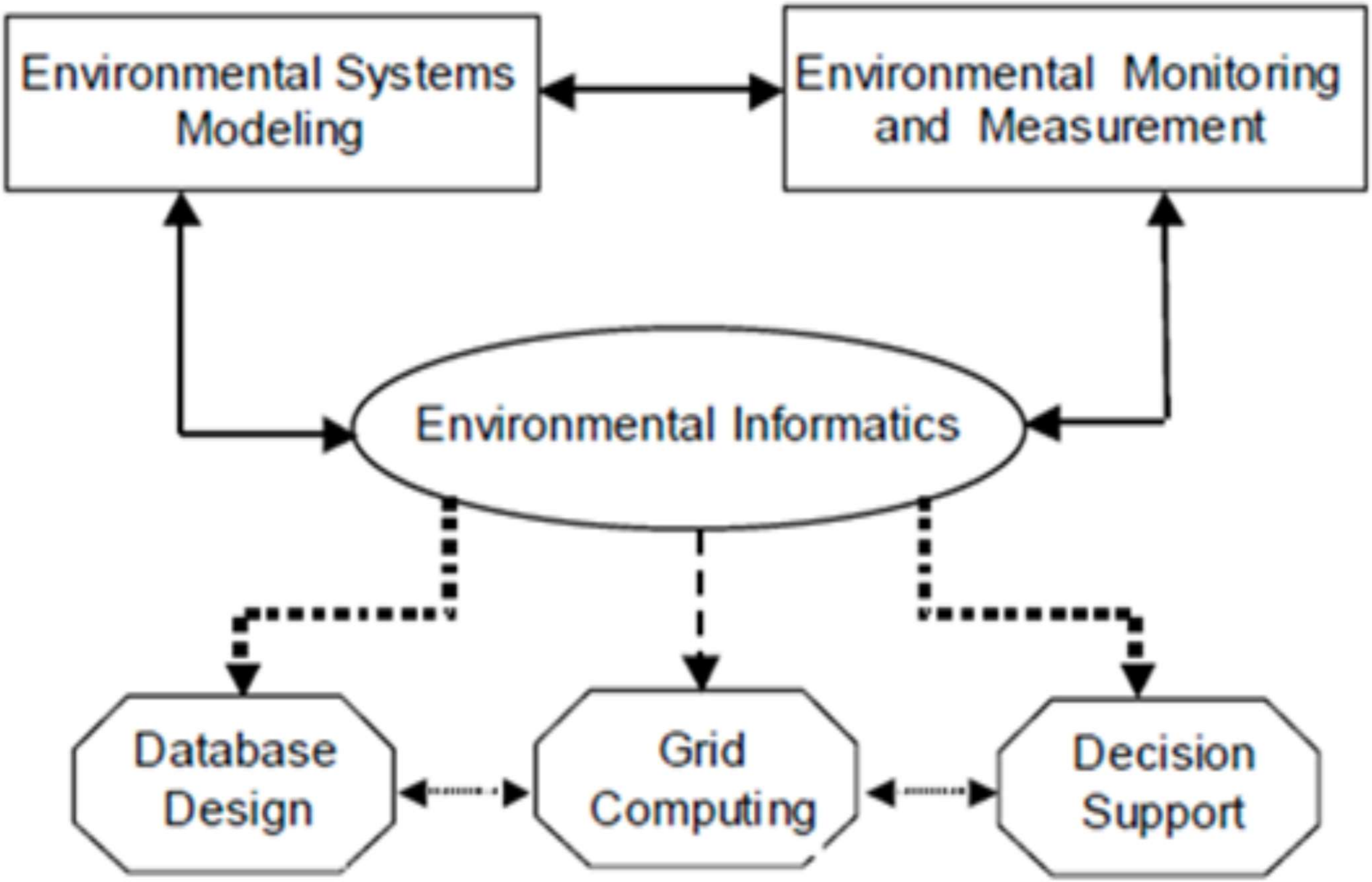


Figure 1: Components of environmental informatics and their relevant interactions.

What is a Model ?

Simplified Representation of a Real or Theoretical System at some particular point of time and space intended to provide the understanding of the system.

Whether a model is good or not depends on the extent to which it provides understanding.

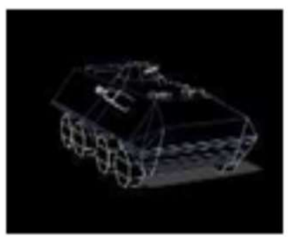
What Level of Model Detail ?

All the models are simplification of reality.
Exact copy of a reality can only be the reality itself.

There is always a trade off as to what level of detail is included in the model:

–Too little detail: risk of missing relevant interactions.

– Too much detail: Overly complicated to understand




What is a Simulation ?

- **It is an experiment in a computer where the real system is replaced by the execution of the program.**

- **It is a program that mimics (imitate) the behaviour of the real system**

- **A Simulation is the imitation of the operation of a real-world process or system over time. It can be done by hand or on a computer.**

- **The behaviour of a system as it evolves over time is studied by developing a simulation model.**

- **This model takes the form of a set of assumptions concerning the operation of the system.**

- **The assumptions are expressed in**
  - Mathematical relationships
  - Logical relationships
  - Symbolic relationships between the entities of the system.

Modelling and Simulation

Discipline of understanding and evaluating the interaction of parts of a real or theoretical system by;

Designing its representation (model)

Executing (running) the model including the time and space dimension (simulation).

Why Simulation?

Accurate Depiction of Reality

Parts of the system may not be observable (e.g., internals of a silicon chip or biological system)

Insightful System Evaluations

It may be too difficult, hazardous, or expensive to observe a real, operational system

Uses of Simulations

1. **Analyse systems before they are built**
2. **Reduce number of design mistakes**
3. **Optimize design**
4. **Analyse operational systems**
5. **Create virtual environments for training, entertainment**

When to use Simulation

Over the years tremendous developments have taken place in computing capabilities and in special purpose simulation languages, and in simulation methodologies.

The use of simulation techniques has also become widespread.

Following are some of the purposes for which simulation may be used.

Simulation is very useful for experiments with the internal interactions of a complex system, or of a subsystem within a complex system.

Simulation can be employed to experiment with new designs and policies, before implementing

Simulation is very useful in determining the influence of changes in input variables on the output of the system.

Simulation helps in suggesting modifications in the system under investigation for its optimal performance.

Simulation can be used to verify the results obtained by analytical methods and reinforce the analytical techniques.

When Simulation is Not Appropriate

Simulation should **not be used when the problem can be solved using common sense.**

Not, if the problem can be **solved analytically.**

Not, if it is easier to perform **direct experiments.**

Not, if the **costs exceeds** savings.

Not, if the **resources or time are not available.**

No data is available, not even estimate simulation is not advised.

If there is no **enough time or the people are not available**, simulation is not appropriate.

If managers have **unreasonable expectation say, too much soon** – or the power of simulation is over estimated, simulation may not be appropriate.

If system behaviour is too complex or cannot be defined, simulation is not appropriate.

Goal of modeling and simulation

- A model can be used to investigate a wide verity of “what if” questions about real-world system.
  - Potential changes to the system can be simulated and predicate their impact on the system.
  - Find adequate parameters before implementation
- So simulation can be used as
  - Analysis tool for predicating the effect of changes
  - Design tool to predicate the performance of new system
- It is better to do simulation before Implementation.

Types of models

- **Time scale**
  - Probability
  - Steady state
  - Event
  - Continuous
- **Determinism**
  - Deterministic
  - Stochastic
- **Spatial representation**
  - Lumped
  - Semi-lumped
  - Distributed
- **Mathematical solution**
  - Conceptual
  - Empirical
  - Mechanistic

Modelling Terminology

- **System**
  - a limited part of reality containing interrelated components
- **Model**
  - simplified representation of a system
- **Boundary**
  - edges of the system
- **Simulation**
  - mathematical representation of a system
- **Environment**
  - set of conditions outside the system being modelled
- **State and rate variables**
  - State of components
  - Rate of change
- **Parameters**
  - Constant values for a system
- **Driving variables**
  - External variables driving change
- **Feedback**
  - negative, positive