**Data Questions:**

*What is ‘big data’?*

* Big data generally refers to massive volumes of data not readily handled by the usual data tools and practices and present unprecedented opportunities for advancing science and informing resource management through data-intensive approaches. (Big data and the future of ecology – Hampton et al)

* The definition of big data is data that contains greater variety, arriving in increasing volumes and with more velocity. This is also known as the three Vs.

* The 5 V's:
  + Volume: i.e. global, long-term data sets
  + Veracity: the “truth” or accuracy of data and information assets - e.g. dodgy politics, co-morbidity, accuracy of numbers
  + Velocity: the speed at which data is received, stored and managed
  + Variety: the diversity and range of different data types, including unstructured data, semi-structured data and raw data
  + Value: the benefit/value of the information being collected . How insightful it is.

*What kind of big data sets might we find that are relevant to biology, conservation, ecology, restoration, and environment?*

* Ecological big-data systems include in situ and remote sensors, community data resources, biodiversity databases, citizen science, and permanent stations (Farely et al 2018)
* Long-term monitoring of contamination levels state wide
* Covid-19 data

*Why is big data such a trendy term?*

* This big data revolution, which encompasses techniques to capture, process, analyse and visualize large datasets in a rapid timeframe, has led to an explosion in data variety over the last five decades (Runting et al, 2020)

*What are the differences between big data sets and ‘traditional’ data sets derived from a controlled experiment?*

* In-situ vs. ex-situ
* While traditional data is based on a centralized database architecture, big data uses a distributed architecture. Computation is distributed among several computers in a network. This makes big data far more scalable than traditional data, in addition to delivering better performance and cost benefits.

*What are some of the concerns or cautions or issues with big data sets?*

* big data present an array of challenges related to large data volumes, high data heterogeneity (variety), varying quality and uncertainty (veracity), and a need for timely information (velocity). (Farely et al 2018)

* Variety and veracity are the defining challenges in ecoinformatics, given the heterogeneity of ecological data, practitioners, and ecosystems, but volume and velocity are growing rapidly (Farely et al 2018)

* Other challenges include complex relationships among highly varied ecological datasets that require sophisticated data models, computational costs of macroscale ecological forecasting, network bandwidth limits, and the need for flexible systems that support multiple and novel data uses (Farely et al 2018)

*How do I decide what kind of data I need to address my scientific question?*

*Where and how can we find the data?*

* development of open code- and data-sharing platforms (Farely et al 2018)
* Crowd sourcing
* Citizen science
* Public data forums

*What can I do if I can’t find the data?*

* Simulate it in R

*Once I’ve got the data, how do I address the question and obtain results?*

* Clean your data, identify trends, and model it to answer your specific research questions

**Meta-analysis Questions:**

*What is meta-analysis?*

* Meta-analysis is a statistical technique for combining the findings from independent studies.
* Good meta-analyses aim for complete coverage of all relevant studies, look for the presence of heterogeneity, and explore the robustness of the main findings using sensitivity analysis.

* i.e. results from the individual studies are combined using appropriate techniques (meta-analysis), significant benefits of treatment may be shown.

<https://www.betterevaluation.org/sites/default/files/Meta-An.pdf>

* The quantitative procedures of meta-analysis help to address some of the challenges introduced by the existence of multiple answers to a given research question.

<https://www.annualreviews.org/doi/pdf/10.1146/annurev.psych.52.1.59>

* Meta-analyses, which combine the results from many trials, have more power to detect small but clinically significant effects.

*What is a systematic review?*

* The validity of the meta-analysis depends on the quality of the systematic review on which it is based.

* Systematic review methodology is at the heart of meta-analysis.
* This stresses the need to take great care to find all the relevant studies (published and unpublished), and to assess the methodological quality of the design and execution of each study.

<https://www.annualreviews.org/doi/pdf/10.1146/annurev.psych.52.1.59>

* The objective of systematic reviews is to present a balanced and impartial summary of the existing research, enabling decisions on effectiveness to be based on all relevant studies of adequate quality.
* Frequently, such systematic reviews provide a quantitative (statistical) estimate of net benefit aggregated over all the included studies. ***Such an approach is termed a meta-analysis.***

*How does a systematic review differ to a literature review or a scoping review?*

* The danger of unsystematic (or narrative) reviews, with only a portion of relevant studies included, is that they could introduce bias.
* reports may be more likely to be included in a review than those which show no significant differences; and informal synthesis may be tainted by the prior beliefs of the reviewer.
* Meta-analysis carried out on a rigorous systematic review can overcome these dangers – offering an unbiased synthesis of the empirical data

*Why would we do a meta-analysis and/or systematic review?*

* Meta-analysis allows the combining of numerical results from a few or many studies, the accurate estimate of descriptive statistics and the explanation of inconsistencies as well as the discovery of moderators and mediators in bodies of research findings.

* Meta-analysis allows researchers to arrive at conclusions that are more accurate and more credible than can be presented in any one primary study or in a nonquantitative, narrative review.

<https://www.annualreviews.org/doi/pdf/10.1146/annurev.psych.52.1.59>

*How do we get the data for a meta-analysis?*

* electronic databases

*What kind of data do we need for a meta-analysis?*

*What kind of issues should be accounted for in meta-analysis related to confidence and variation?*

* Good meta-analyses aim for complete coverage of all relevant studies, look for the presence of heterogeneity, and explore the robustness of the main findings using sensitivity analysis.

*What kind of issues should be accounted for in meta-analysis related to independence?*

* All from same lab/researcher

*What kind of issues should be accounted for in meta-analysis related to publication bias?*

* However competent the meta-analysis, if the original review was partial, flawed or otherwise unsystematic, then the meta-analysis may provide a precise quantitative estimate that is simply wrong.

<https://www.betterevaluation.org/sites/default/files/Meta-An.pdf>

**Scientific Modelling Questions:**

*What is a model? What does ‘model’ mean?*

* An abstracted or simplified representation of an object system or concept (From seminar)

* From Haefner book:
* A **model** is a description of a system
* A **system** is any collection of interrelated objects
* An **object** is some elemental unit upon which observations can be made, but whose internal structure either does not exist or is ignored
* A **description** is a signal that can be decoded or interpreted by humans

*Why do modelling? What are the goals or purposes of modelling? How do these depend on our field of research?*

* From Haefner book:
* Models facilitate the discussion of systems (a tool)
* Models are used to communicate a view of the world

* 3 primary scientific uses of models:
* **Understanding** - of either a real, physical system or of a system of logic such as another scientific theory
* **Prediction** - of the future or of some state that is currently unknown
* **Control** - to constrain or manipulate a system to produce a desirable condition

* From seminar:
* Why model? :
  + to clarify
  + to understand
  + to compare
  + to predict
  + to manage
  + to educate
  + to communicate ideas.
* What is the model aim?
  + Visualisation?
  + Communication of results and ideas?
  + Part of a larger model?
  + Control and management? (decision support)
  + Accurate prediction?
  + Theoretical explanation?
  + Theoretical frame work for experimental investigations

*What types of formal models are there? How can models be classified?*

* *Empirical, mechanistic, process-based, descriptive, ontological, phenomenological, static, dynamic, stochastic, deterministic, space & time: continuous or discrete, DEs/FDEs/PDEs, compartment/transport/particle/individual-based/agent-based, finite state automata/cellular automata/state-transition, Eulerian/ Lagrangian...)*

* From Haefner book:
* Forms of models:
  1. **Conceptual or verbal** - description in a natural language
  2. **Diagrammatic** - graphical representations of the objects and relations (e.g. ecological 'box and arrow' diagrams of energy flow, physiological diagrams of metabolic pathways)
  3. **Physical** - a real, physical mock up of a real system or object
  4. **Formal** - mathematical (usually using algebraic or differential equations

* Mathematical classifications:
  1. Does the mathematics have an explicit representation of mechanistic processes?
* YES: **Process-orientated** or **mechanistic models** - e.g. populations dynamics models w/ details of reproductive physiology
* NO: **Descriptive** or **phenomenological models** - e.g. Boyle's law relating temperature, pressure and volume, or density-dependent population dynamics w/ re-production represented as a single parameter

1. Does the mathematic have an explicit representation of future system states or conclusions?

* YES: **Dynamic** models (e.g. leaky bucket model - how it changes over time)
* NO: **Static** models (e.g. linear regression equation relating variables x and y)

1. Does the mathematics represent time continuously?

* YES: **Continuous** models - time may take on any values e.g. 3.3 seconds
* NO: **Discrete** models, time is an integer only

1. Does the mathematics have an explicit representation of space?

* YES: **Spatially heterogeneous** models e.g. objects have a position in space, or occupy a finite region of space.
  1. **Discrete**: space is represented as spatially homogeneous
  2. **Continuous**: every point in space is different e.g. diffusion equation in physics
* NO: **Spatially homogenous** models e.g. simple equation of population dynamics

1. Does the model allow for random events?

* YES: **Stochastic** models - e.g. random temp values may produce random changes in the intrinsic rate of increase in population dynamic models.
* NO: **Deterministic** models i.e. all parameters are constant

* **Compartment** models - models that describe the flow of physical material between physical or biological storage compartments
* **Transport** models - those that transport material, energy or momentum from point to point in physical space. They are similar to compartment models but use special mathematical structures (partial differential equations) and conservation principles
* **Particle** models - those that follow the fate of individual particles moving in space (e.g. individual blood cells flowing through veins) or individual organisms changing their condition (e.g. body size)
* **Finite state automata** - are models that represent an object as being in only a few, fintite number of states or conditions e.g. might model weather dynamics as a system that has only 'good, bad or intermediate' weather quality. This is different from compartment models such as the leaky bucket, there the container could have any volume of fluid.

Definition. Partial differential equation (PDE) models are sets of equations describing the evolution of a physical quantity, not only with time, but also according to a structure variable such as space.

*Model as hypothesis – the classical view (Popper) versus the strong inference view (Platt).*

Platt characterizes strong inference as the repetition of three essential steps:

(1) devise alternative hypotheses;

(2) devise a crucial experiment that will exclude one or more hypotheses; and

(3) perform the experiment and obtain a clean result.

(Then, (1′) recycle the procedure to refine the possibilities that remain.)

1) Devising alternative hypotheses; 2) Devising a crucial experiment (or several of them), with alternative possible outcomes, each of which will, as nearly as possible, exclude one or more of the hypotheses; 3) Carrying out the experiment so as to get a clean result; 1') Recycling the procedure, making sub-hypotheses or sequential hypotheses to refine the possibilities that remain; and so on.

*What is involved in the modelling process?*

*What is the relationship between modelling and empirical data? How does modelling fit into empirical data collection and how does data fit into modelling?*

*One with out the other and is adaptive - can make a model and use different data*

*What trade-offs are involved in modelling? (Levins’ triangle 1966, Orzack and Sober 1993, Levins 1993)*

*Constraints on models in haefen book*

*What is ‘individual-based modelling’ or ‘agent-based’ modelling? What advantages/disadvantages do these have compared to other types of modelling?*

*How have population dynamics growth models been adapted to take into account age-structure of populations; more than one interacting species with competition, predation etc; density-dependence; stochasticity etc?*

* You can change the parameter to represent what is required within you system/model
* Like birth rate, death rate in mice experiment
* Add in species interactions with one limiting the other

* *Systems analysis, conceptual modelling, qualitative model formulation*

* *Turning the concepts into formulas and/or code: quantitative model formulation*

* *Getting values for parameters – calibration/parameterisation*

* *Simulation and computational/numerical issues*

* *How do we evaluate a model? Validation and Testing*

* *Model analysis: Sensitivity Analysis*

* *Model analysis: Analysis of Dynamics and Stability*

* *Models for Management Decision Support*

* *Optimisation – for calibration, analysis, or... Optimisation*

* *Stochasticity and Uncertainty*