Research Concepts & Skills
Volume 1: Concepts

Text with self-test questions

D G Rossiter

Version 3.21
May 23, 2013

Contents

This first volume of the text *Research Concepts & Skills* presents the conceptual skills needed for successful MSc research: the scientific method, induction and deduction, inference, statistical thinking, and scientific ethics.

1 The scientific method

In this chapter we examine what it means to "do science", what is the "scientific method", in what sense science can be said to "explain", and the logic behind scientific reasoning. We also look briefly at some social aspects of the scientific enterprise.

Key points

1. The **scientific method** is a manner of thinking and working towards more complete knowledge of the world.
2. To be scientific, a statement must, in principle, be **falsifiable**.
3. Sciences may be classified as **experimental**, **observational**, or **historical** (§1.2).
4. There are many forms of scientific **inference** (§1.3), with different logical foundations and degrees of rigour.
5. Scientific **explanation** is linked to **causality** (§1.4). A **parsimonious** explanation is preferred (§1.4.2).
6. A scientific statement may be a **fact**, **hypothesis**, **theory**, or **law**, each with a level of **certainty** (§1.5).
7. An important type of scientific reasoning is **deductive-inductive** (§1.6).
8. Scientific explanation requires sound **logical thinking** (§1.8)

1.1 What is science?

To "do science" is to follow a prescribed **method** to arrive at knowledge. The "scientific method" is not a belief system or religious dogma, but rather a manner of thinking and working towards more complete knowledge of the world. It has been proven to be extremely successful in:

- **explaining** the world as we observe it;
- **predicting** what can be further observed, e.g. new observations, new locations, repeat observations, the effect of interventions;
- **engineering**, i.e. building things that work.

Science is not **prescriptive** – it can not say what "ought" to be done. It can, however, point out the probable consequences of certain actions, as objectively as possible.

1.1.1 Characteristics of scientific knowledge

Self-criticism

Scientific knowledge is inherently **self-critical** and is never complete or final. A particularly important aspect of the scientific method is that it has a built-in mechanism to **check and revise** itself. That is, any statement in science is subject to revision or even **falsification** using the same methodology that was used to establish it in the first place. Thus it is **self-consistent** and does not allow for any **super-natural** reasoning (see §1.1.2, below).

Evidence-based

Scientific knowledge must be built up step-by-step from **experience**, including experiments and systematic observations. It can not be deduced from abstract ideas of how the world "should" work or on folk "wisdom" – although these can provide hypotheses to be investigated.

So, one way scientific knowledge advances is by accumulating more **evidence**.

Theory-based

However, science is not a disorganized collection of facts, it is a way of explaining the world. Scientists do this by constructing **theories** (also called **models**) that explain the available evidence (§1.4, §1.6).

So, the other way scientific knowledge advances is by constructing better **theories** (models) from the available evidence.

Transparency

All **methods** used in a scientific investigation and all **results** of applying the methods must be unambiguously specified and **communicated**, both within the scientific community and outside it.

This implies that:

- Science is **reproducible**: another worker can perform the same experiment or observation, and expect to obtain the same result, within the limits of experimental error;
- There is no **occult** (hidden) knowledge in science – in principle any person can acquire all the knowledge needed to do and understand science;

- Science has a built-in **self-correction** mechanism – other scientists can verify, modify, contradict or extend 'surprising' or controversial results.

In the social sciences and with historical approaches it may be impossible to exactly reproduce an observation; however the methods used are **traceable**. That is, it is clear how observations were made (transparency); so, another worker could follow the same procedure in a different setting and expect to obtain similar results, with differences due to the differences in the two situations.

No appeal to authority

Scientific explanations can not depend on any authority (religion, political ideology, established authority, even scientific "consensus" ...). Certainly with time there develops a scientific orthodoxy, based on evidence and (so far) proven theories, but this by itself does not establish the truth.

Science has suffered greatly from appeals to authority and especially ideology. Well-known examples are the so-called "Aryan physics" under the Nazis [9, Ch. 10] and Lysenko's theories of plant genetics under Stalin [9, Ch. 9].

Science has suffered also from the orthodoxy of the scientific establishment, which suppresses new ideas – but not for ever.

1.1.2 Naturalism

There is one assumption that the scientist must hold. This was well-expressed by prominent biologist Lynn Margulis, in an editorial in American Scientist (93:482); she even calls this a "faith":

> "All of us who participate in science must share one common faith. We believe that the material-energetic world is knowable, at least in large part, by the concerted activity of research: exploration, reconnaissance, observation, logic, detailed study that includes careful measurement against standards."

This assumption or 'faith' is called **naturalism** [19]. As part of the scientific method, it is not important whether it is true or not, only that science is powerless if it is not true. If Greek gods are really interfering in the natural world (for example, if Poseiden is angry and causes a tsunami, or Artemis is stirring up jealousy among people in a town because she's dissatisfied with her worshipers), there is no hope for investigating natural or social phenomena by systematic scientific research.

Naturalism is an example of what Gauch Jr. [7] calls **presuppositions** for science: things that must be true for science to be successful, but which can not be tested by science itself.

1.1.3 Science as a human activity

Science is of course carried out by humans, and the pursuit of 'truth' is subject to all their virtues and vices; fascinating discussions of how science really works may be found in books by, among others, Gower [8], Bauer [1], and Derry [5]. Gauch Jr. [7] gives an more philosophical view of the scientific method, while Okasha [19] is an especially accessible introduction to the philosophy of science underlying the scientific method. Gratzer [9] gives an entertaining, cautionary, and even chilling collection of "delusion, self-deception, and human frailty" in the pursuit of scientific knowledge.

Science is often tightly-linked to **commercial** or **political** interests. For example, there is big money at stake in the 'global warming' and 'alternate energy' fields, but also in medicine, biodiversity, etc. And wherever there is money involved, there is the possibility of corruption, fraud, or self-delusion (wishful thinking, jumping to conclusions).

However, science does include a **self-correcting mechanism** (open publication, reproducibility of methods) which at least provides some check on factual or methodological errors; it is more difficult to correct orthodox opinion as to correct explanations.

Science is **not prescriptive** – it can not say what is a "right" course of action. That is the province of human value systems, including secular humanism, ethical systems, tradition, and supernatural religions.

1.1.4 How do scientists really work?

Many years ago, Medawar [17] gave a radio talk with the provocative title "Is the scientific paper a fraud?" His point was not, of course, that scientists were frauds or cheats, but that the cool, rational, logical structure of the typical scientific paper gave a completely misleading view on how science is actually carried out. His argument was paraphrased by Webster [22]:

> "Composing a scientific paper has become a ritual, to be followed almost regardless of the subject. An 'Introduction' reviews the field; it tells how that particular branch of the subject has developed in recent years as other workers have groped towards enlightenment; it states the current situation, and then it pretends that by some process of deduction the authors (there are almost always more than one nowadays)

> arrived logically and almost inevitably at the next step, which they are about to reveal. Then comes a section entitled 'Materials and methods' and another, the 'Results'. The authors go on to discuss the results in as seamless a way as possible to show that they flow from their rational understanding of the previous research and their carefully planned experiments."

Webster asserts that this neat package misrepresents the thought processes of many scientists; in particular, that there is no room for creativity or inspiration, just rational problem-solving in designing and executing the research.

Medawar put great emphasis on the creative aspect of science:

> "Scientists should not be ashamed to admit ... that hypotheses appear in their minds along uncharted by-ways of thought; that they are imaginative and inspirational in character; that they are indeed adventures of the mind". [17]

He attributes to the English polymath William Whewell (1794-1866) the formulation of a "hypothetic-deductive" interpretation of how science is done:

1. Formulate hypotheses from experience, intuition and inspiration;
2. Design experiments to test these;
3. Deduce from the results the (partial?) truth of the hypotheses;
4. Be open to inspiration along the way.

We return to this theme in §1.3.

1.2 Types of sciences

Scientific activity can be classified as experimental, observational, or historical. All three require a separate step of model building.

1. **Experimental**: controlled conditions under which measurements are made (e.g. laboratory experiments in physics or chemistry); variable level of control of the context, but always quantifiable (e.g. temperature in a growth chamber can be controlled with a known precision);
2. **Observational**: No experiment is possible, but observations are made in uncontrolled or semi–controlled conditions:
   - e.g. we can't order up an earthquake or extreme rainfall event;

- e.g. we can't manufacture survey respondents with certain characteristics[1];
- Requires a sound **sampling design** to ensure the observations are **representative** of the process to be modelled.

3. **Historical**: There is evidence from the past, which can never be re-created experimentally (e.g. geology, archaeology):
   - these can be related to current processes, assuming that the laws of physics etc. have not changed in the meantime;
   - some of the supposed processes can perhaps be reproduced in the lab. (e.g. subject rock samples to temperatures and pressures suspected to have caused a certain mineral assemblage);
   - but this is not possible for e.g. archaeology;
   - explanation relies heavily on inference and argument.

1.2.1 Science vs. engineering

- **Scientific research** is a method to discover facts about nature and to put these in a theoretical context ('why' the observed facts are so);
- **Engineering** is the design and manufacture of objects (which may be virtual, e.g. a computer program).

They both use **logical thinking**, and during the course of an engineering project many small experiments may be carried out to improve the design. The fundamental difference is that science **investigates** the world as it is and tries to explain it, whereas engineering **changes** the world by human activity.

1.3 Scientific inference

Science attempts to reach conclusions from premises and observations, using some form of rational argument. The general term for making new statements on the basis of previous statements is **inference**. The philosophical basis is **epistemology**, i.e. the foundations of knowledge, how we know things about the world. This is one step below **explanation** (§1.4), which is concerned with 'why' something may be true (related to the philosophical concept of **ontology**), not just 'whether' it is true.

The most common form of inference is the deductive-inductive scientific method, explained in §1.6. However, there are other methods, here presented in roughly their order of rigour[2], with a few notes on their

[1] at least not with current technology ...
[2] This list was developed by Mike McCall

applicability.

1. Purely logical, e.g. mathematical theorems from postulates;
   - rigorous, by definition correct given the assumptions; unable to deal with 'messy' observations
2. **Deductive-inductive** (also called 'hypothetic-deductive');
   - see §1.6 for details
3. Cause & effect;
   - often used if there is a direct time or action sequence (e.g. method of preparing land for a crop, leading to changed soil properties)
4. Contributors & impacts;
   - a weaker form of cause & effect; a number of factors affect the outcomes, but the causal links are not explicit
   - 'co-relation', causes can't be determined with any certainty
5. Inductive patterns (classification);
   - organizes observations, then it may become clear why the particular grouping occurs
   - can be a first step to a more complete explanation
6. Case studies;
   - difficult to generalize, must identify idiosyncratic and universal factors
7. Analogy;
   - conclusions from one system are used to predict in another system, without experiments or observations of the second system
   - must argue that the two systems are analogous
   - example: 'learning a language is like learning to play music [is this true?], therefore the model of musical education can be used as a model of language education'.
8. Probabilistic;
   - organizes knowledge in a predictive model, with each outcome given a computed probability of occurrence

- for example, weights-of-evidence modelling of the probability of finding a mineral deposit; geostatistical modelling of the probability of a contaminated soil

9. Functional;
   - sometimes called 'ecological' explanations, meaning the 'environment' is sufficient to predict the outcome
   - sufficient for prediction, but do not explain anything
   - e.g. an empirical-statistical model

10. Systems explanations; 'black boxes';
    - input-output (stimulus-response) relations
    - sufficient for prediction, but do not explain anything

11. (Expert) Judgement / Wisdom / Intuition ...
    - holistic, can not be reduced to discrete steps of reasoning
    - by definition not reproducible

12. Teleological, 'higher' purpose, external cause.
    - things occur because they 'want' to ('plants want to find sunlight, so they grow out of the soil and keep growing upwards') or because some 'higher power' wants them to occur ('male and female He made them ...')
    - impossible to verify by scientific methods

1.4 Scientific explanation

To "explain" is to say **"why"** something happens or is observed – but this is very difficult, if not impossible, to establish. "Why" is ultimately an existential question (philosophically related to **ontology**, questions of existence or being), interesting to the philosopher. The nature of explanation in various scientific disciplines has been extensively discussed, e.g. for earth sciences by Kleinhans *et al.* [15].

In applied science and engineering we are mostly content with a more limited view of "why": a coherent statement that allows **prediction** of the phenomenon in the future, in other situations or at other locations besides the ones already observed.

This agrees with Harvey [10] (see Figure 1.2) who defines the process of explanation as "making an unexpected outcome an expected outcome, of making a curious event seem natural or normal"; it becomes 'natural' once the processes which gives rise to the outcome (given similar conditions) are clear.

Here is an example of a simple explanation:

- Q: Why is soil erosion faster on steep slopes?
- A: When raindrop splash detaches soil particles, they are suspended in water. The kinetic energy of flowing water carries the soil downhill before they can settle out. Gravity ensures that water flows downhill. Kinetic energy of flowing water is converted from the potential energy of gravity. On steep slopes there is more potential energy difference (represented by height difference) per horizontal unit over which the water must flow, so the flow is faster[3].

This explanation could be quantified to predict:

- Q: At what slope angle does does soil erosion become severe (defined as a soil loss limit)?

Note that this explanation allows the design of **interventions** to change the situation. For example, to reduce soil erosion:

- The slope angle could be reduced, e.g. by terracing;
- The kinetic energy of the water could be reduced, e.g. by diverting runoff from upslope;
- Detachment could be reduced by covering the soil surface, or even reducing the kinetic energy of the raindrops with an intercepting screen.

Note that all these interventions require knowledge of **causes** and **mechanisms**, at least at some level of understanding.

Explanation is generally tightly-linked to **causality**: an explanation is not very useful if it only summarizes, it must also give some idea of the causes. There is some debate among philosophers of science about this; for an interesting discussion see Okasha [see 19, Ch. 3],

Continuing the soil erosion example, we could model the dependence of erosion on slope with no attempt at causal explanation. This might be enough for prediction but not for designing interventions.

Here are some other questions, from several ITC-related fields. You are encouraged to invent plausible "explanations" and consider their relation to causality.

- Q: Why does holding separate meetings with government officials and villagers, rather than joint meetings, provide different opinions on the correct placement of a national park boundary?

[3] Of course, it's not so simple! See for example [12] or [18]

- Q: Why does a slow animation of a flood event lead to better comprehension by relief agencies of the flood effects than repeated rapid animations?
- Q: Why does a bootstrapped validation measure of a soil depth model give similar results to an independent validation set?

1.4.1 Proximate and ultimate causes

The concept of "causality" is also tricky. What appears at first to be the cause must itself have a cause, and so forth. The **proximate** (immediate) cause may be fairly easy to establish, but the deeper causes requires either more evidence or more speculation. It is probably meaningless to speak of an **"ultimate"** (last, final) cause.

Some causal links are quite clear, especially those where direct human agency is involved. For example Wu *et al.* [23] used a time-series of satellite images to argue that increased turbidity in a section of Poyang Lake, China coincided with increased sand dredging activity. Since dredging stirs up the lake bottom, and no other change in the environment besides the dredging has been observed, the statement "dredging is the cause of increased turbidity in Poyang Lake" seems sound, and no further cause is needed. The explanation of increased turbidity is causal, the cause is direct human action.

Of course, one could ask "Why is there increased dredging?", which would be answered by a separate study of the changing Chinese political and economic scene. The causal link from dredging to turbidity is fine as far as it goes, but does not allow interventions. How could dredging be limited? This depends on knowing the causes of the increased dredging.

1.4.2 Parsimony of explanation

Any evidence can have many explanations. As long as they all obey the rules of logic, how can we decide among them? This has been argued since the beginnings of philosophy, and for the purposes of science has been settled in favour of the concept of **parsimony**. Its most famous formulation is due to William of Ockham (c. 1285-1347), known as Ockham's Razor:

> "*Numquam ponenda est pluralitas sine necessitate*"

which can roughly be translated as:

> "Plurality ought never be posited without necessity"
>
> or, more idiomatically,

> "Complexity should never be added to an explanation unless necessary"

This means that if several theories equally explain the observed facts, the simplest should be used. More complexity is only justified by more evidence, whereby the simpler theory no longer explains the evidence.

Note that Ockham did not claim that a parsimonious explanation is necessarily "true", whatever that may mean. He advocated parsimony for **epistemological** (foundations of knowledge) reasons: it is the most efficient way to organize our knowledge of the world, and is most likely to lead to correct inferences.

By the way...

> This is where conspiracy theorists and scientists have an un-bridgeable conceptual (and communication) gap: the conspiracy theorist is only happier as the theory gets more complex (e.g. if many hundreds of people would have had to take part in the assassination of JF Kennedy) whereas the scientist prefers the simpler explanation *unless* there is sound evidence for more complexity.

The idea of parsimony can also be used in **hypothesis formulation**, i.e. the inductive step of the deductive-inductive method (see §1.6). A reasonable hypothesis to be tested explains the prior evidence, but no more. The experiment (deductive step) is then designed to falsify the hypothesis, i.e. to find where it can not explain. Then either:

- the simple hypothesis is made more complex to cover the new situation, or, rarely
- another simple hypothesis is invented (**paradigm shift**).

Q1 : *Recall the example of Wu* et al. *[23] (§1.4.1), who used a time-series of satellite images to argue that increased turbidity in a section of Poyang Lake, China coincided with increased sand dredging activity. Since dredging stirs up the lake bottom, and no other change in the environment besides the dredging has been observed, their explanation is "dredging is the cause of increased turbidity in Poyang Lake"*

Here is an alternative explanation: "the increased turbidity is due to meteors striking the water and stirring up the sediments just before the imagery was made. The quantity of meteors penetrating the atmosphere above Poyang Lake has steadily increased in recent years, hence the observed increase in turbidity."

Is one of the two explanations more ***parsimonious*** *than the other? Why or why not?* *Jump to A1* •

Q2 : *Given the evidence so far (known increase in dredging), which explanation is preferred?* *Jump to A2* •

Q3 : *Here is another explanation for the observed increase in turbidity: Environmental activists are trying to get sand dredging stopped in Poyang Lake. They find out about the on-going study of turbidity and the imagery that will be used. Over a period of five years, just before the overpass of the satellites that will make the images, they hire boats, go out into the lake, and use explosives to stir up the bottom sediment, so that the imagery will show more turbidity. They slowly increase this activity, so over the five years, each image will show more turbidity. They carefully calibrate the amount of sediment they stir up with the increasing number of dredging platforms. So, although the dredging isn't causing any turbidity, the activists trick the researchers, who are using satellite imagery to assess this, into making a false correlation between number of dredging platforms and increased turbidity.*

Is this explanation more or less parsimonious than the original explanation from the researchers? Explain. *Jump to A3*
•

Q4 : *What evidence would support this second explanation?* *Jump to A4* •

Parsimony in statistical inference

"Parsimony" is used as a technical term used in statistical inference to express the idea that the simplest relation that explains the data is the best [6]. In statistics, parsimony is expressed as "Fit the relation, not the noise"; this leads to maximum information, in a formal sense. The aim is not to fit just one dataset, but to determine true underlying relations. There are various statistical measures of parsimony or information: adjustments to naïve measures of model success like regression R^2 (e.g. adjusted R^2, Akaike Information Criterion).

This is covered in detail in §7.

1.5 Levels of certainty

We use the words “fact”, “hypothesis”, “theory” and “law” in common speech with a variety of meanings which often overlap. When discussing scientific certainty we must be more precise.

1.5.1 Facts

A **fact** is something directly observable and measurable. A “fact” always has some uncertainty, since no instrument is perfect. Note that the uncertainty is not from definition or interpretation, only from measurement imprecision. The uncertainty of a “fact” can be quantified, e.g. from instrument characteristics or sampling theory.

1.5.2 Hypotheses

A **hypothesis** is a tentative theory, not yet tested; it is what we believe to be the true explanation or true state of nature, based on previous work or first principles.

A dictionary definition is [11]:

> “[An] idea or a suggestion that is based on known facts and is used as a basis for reasoning or further investigation.”

Note the emphasis on “starting point”.

This agrees with Harvey [10], who calls a hypothesis ’logically consistent controlled speculation’ – note that a hypothesis must at least be internally-consistent (’logical’).

1.5.3 Theories

A **theory** is a **conceptual framework** which:

- **explains** existing facts;
- allows **predictions**; and
- is in principle **falsifiable** (some experiment or observation could contradict it or force its modification).

A dictionary definition is [11]:

> “[A] reasoned supposition put forward to explain facts or events.”

Note the emphasis on “reasoned”, meaning that a theory must be supported by **evidence** and **logical argument** from this.

Harvey [10] proposes a stronger definition of theory, to differentiate it more from a hypothesis. A theory is a “highly articulate systems of statements of enormous explanatory power”; that is, there is enough evidence behind the theory, and it is expressed in enough detail, to allow many complex predictions. This detail is expressed in a **model**.

So, a theory differs from a hypothesis mainly in its level of abstraction (conceptual framework), detail (model) and amount of supporting evidence,. In this definition, a "theory" only has to be a "reasoned supposition". So some theories are tentative, based on scanty evidence and easily-falsifiable, while others have much evidence behind them and are approaching "laws" (see next).

1.5.4 Laws

A **law** is theory with overwhelming evidence, including the conditions under which it is true. A similar definition is a **theory whose falsification, within its context, is almost inconceivable.**

> The classic example is Newton's laws of motion (1687); these are precise mathematical statements, consistently applied to bodies of all sizes and at all distances. These laws are not always true; there is a limiting condition: velocities must be low compared to the speed of light, so that relativistic effects are not significant.

In fields such as geography, it is quite unlikely we can formulate laws in the same sense as in physics; it is perhaps better to speak of 'law-like statements', for example, von Thünen's "law" of land use related to distance to markets.

The boundaries in the sequence [hypothesis ⇒ theory ⇒ law] are of course fuzzy.

Q5 : *The following statement must be placed on the cover of secondary school science textbooks in some states of the USA:*

> *"Evolution of species by variation and natural selection is a theory, not a fact."*

According to the definitions given above, is this a technically-correct statement? *Jump to A5* •

Q6 : *What impression do you think it is intended to give to young students?* *Jump to A6* •

Q7 : *Given the evidence since Darwin and Wallace [e.g. 4], where does* *this "theory" of evolution by variation and natural selection fit in the sequence [fact, hypothesis, theory, law]?* *Jump to A7* •

1.6 The deductive-inductive scientific method

The best-known scientific method is known as the "deductive-inductive" approach. It has the following structure:

1. **Observe** and synthesize general knowledge of the world;
2. **Invent a theory** to explain the observations $\Longrightarrow$ *abduction*;
3. Use the theory to **make predictions** $\Longrightarrow$ *deduction*;
4. **Design experiments** or **Make more observations** to test these predictions;
5. **Modify the theory** in the light of results $\Longrightarrow$ *induction*;

Repeat from step 3 **until** you can't think of any new predictions that might falsify or modify the theory.

Step 4 is the crucial stage of **experimental design**: make new observations where they are **most likely to contradict what is expected** or where **an unexpected result would make maximum damage** to the theory. That is, the maximum information from a new experiment or observation comes either when the outcome is least predictable, or when it so predictable that an unusual result would be devastating.

Since we don't start from the beginning, the "Observe" and "Theory" steps are based on others' previous work and our general knowledge. This is nicely-shown in a famous diagram by Box *et al.* [2] (Fig. 1.1):

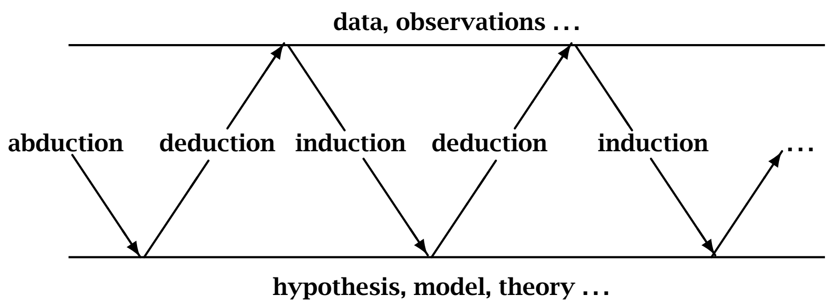

Figure 1.1: The deductive-inductive iterative approach to scientific knowledge (after Box *et al.* [2])

- The **abductive** step is the formation of the first hypothesis. It differs from induction in that no systematic experiments or observations have been made, so that the hypothesis is based on common knowledge, general principles, and any observations that are available.

  > "Based on what I know from previous experience, and what I can observe, I formulate the following hypothesis: ...",

- The **deductive** step goes from existing theory or hypothesis to design a new experiment or set of observations, with expected results:

  > "If my theory is true, and if I do this experiment (or make these observations), I should obtain these results."

- This is then compared to the actual results, leading to an **inductive** step where the existing theory is modified to account for the new results:

  > "My experiment did not give all the expected results. (My observations are not all as I expected.) However, if I modify my theory this way, then the experiment (observations), as well as my previous knowledge, would fit this new theory".

  This continues until the researcher is satisfied (and can satisfy others) that the theory is complete within its assumptions.

The logical basis of abduction, induction and deduction is more fully explored below (§1.8). A natural resources example is given in §1.10.

An expanded view of the inductive-deductive method is due to Harvey [10] (Figure 1.2). This view emphasizes the **asymmetric** roles of "successful" and "unsuccessful" experiments, i.e. those that confirmed or contradicted the hypotheses:

- After an **unsuccessful** outcome, we have to re-consider our mental and formal models of the world; they may be really wrong; and we're more or less starting over in trying to explain the world; the **hypotheses** for the following experiments must be adjusted.

- After an **successful** outcome, we were able to upgrade our **hypothesis** into a **theory** or even a **law**; this confirms our mental and formal models so far, and allows them to be expanded (generalized, or applied to more phenomena).

After many iterations we may be confident enough to claim to have **explained** some aspect of the world.

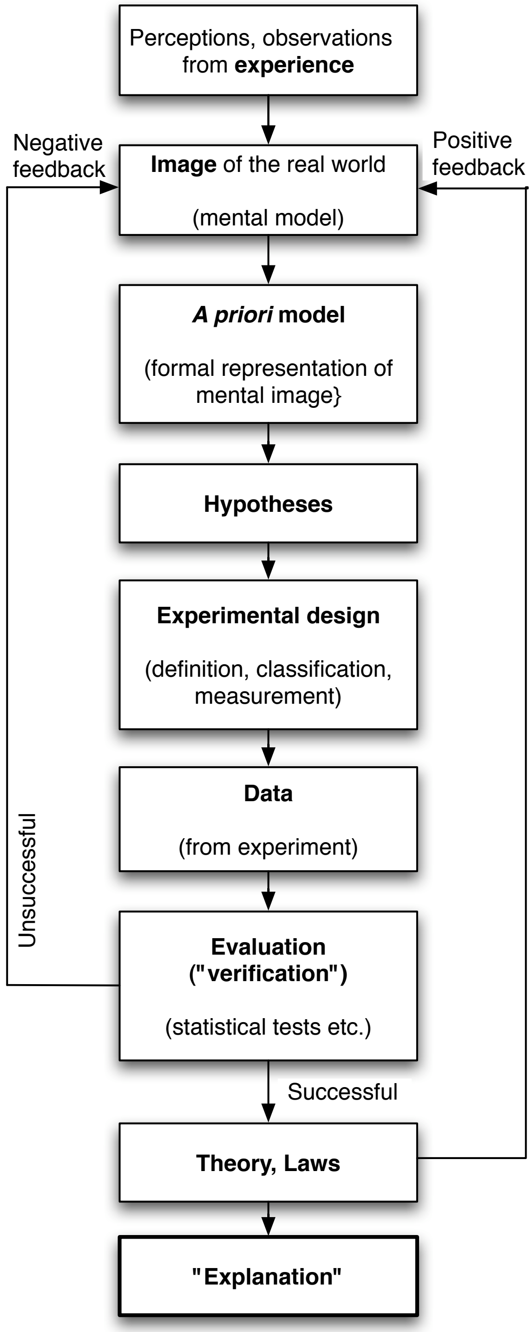

Figure 1.2: An approach to scientific knowledge (after Harvey [10])

In reality, data from experiments is messy – some confirms the hypothesis while other contradicts it. So often the positive and negative feedback go together to **modify** our mental, and then formal, models.

1.7 Is a hypothesis necessary for science?

A **hypothesis** as defined above is a reasonable first explanation of the true state of nature based on previous work or first principles; the research must be designed to test or challenge this hypothesis. The research will either:

1. confirm;
2. contradict; or
3. cause a modification of ...

... the hypothesis.

Here's a simple example:

Research question: Do students preferentially associate with others of their own nationality in academic activities at ITC?

Hypothesis: No, ITC students mix freely; any preference of association is due only to personal preference.

Experiment: Observe groups formed by free association (not by instructors) and compare their national composition to one that would be expected by random association.[4]

Revised hypothesis: Confirm, reject or update the original from the experiment.

Iterate until results are consistent.

Some philosophies of science advocate **hypothesis-free** research, since just by stating a hypothesis we are constructing a context for the research and limiting its outcomes. This is often advocated in social sciences where researchers immerse themselves in communities with "no preconceptions" and "allow the theory to follow the observations".

This appears impossible in principle. No person can escape their life experiences, which form an implicit hypothesis (even theory) of how things (including societies) work. It is better to make these hypotheses explicit and then design the research to test them. This is well-expressed by Medawar [17]:

> "Innocent observation is a mere philosophic fiction. There is no such thing as unprejudiced observation. Every act of

[4] This would have to be expanded into a detailed experimental design.

> observation we make is biased. What we see or otherwise is a function of what we have seen or sensed in the past." [17]

One could use the same argument for a natural resources survey. If the soils of a region have never been studied, how can the surveyor have a hypothesis of what soils are there, and how they are distributed on the landscape? Should observations be made without any theory? That is, should the sampling design be based on total ignorance?

As is often the case, Charles Darwin made a strong argument, which has become a well-known quotation:

> "About thirty years ago there was much talk that geologists ought only to observe and not theorise; and I well remember some one saying that at this rate a man might as well go into a gravel-pit and count the pebbles and describe the colours. How odd it is that anyone should not see that all observation must be for or against some view if it is to be of any service!" [3, Letter 3257, Darwin, C. R. to Fawcett, Henry, 18 Sept 1861]

This attitude of total ignorance is:

- **inefficient**: because sampling can not be directed to extract the maximum information (i.e. to confirm or disprove the hypothesis); and
- **wasteful**: because it ignores previous work on soils and soil geography in other regions; the surveyor can reason from first principles of soil behaviour and from analogous regions elsewhere in the world, so is not truly in a state of ignorance.

So in fact a soil survey in an un-mapped area must begin with a set of hypotheses based on previous knowledge (in this case, theories of soil formation and reasoning from similar areas). Then the survey can be designed to confirm or, more likely, modify that hypothesis.

Another famous example of hypothesis-driven research is the discovery by Otto Frisch that splitting a U atom with a single low-energy neutron could release a large amount of energy due to the mass-energy equivalence. The physicist Max Perutz commented:

> "The violence of the reaction had remained unnoticed [by other research teams] without a hypothesis predicting it; and Frisch detected it by an experiment designed to falsify the hypothesis [but which instead provided evidence for it]." [20]

Perutz's point is that without having stated the hypothesis, and designed an experiment to test it, the phenomenon was not observed, even though several groups by this time (February 1939) had carried out experiments with the same reaction, but without looking at the energetics.

1.8 Logic in scientific explanation

In science we use a combination of strict **deductive logic**, generalization by **inductive reasoning**, and holistic **abduction**. How we actually think 'logically' in science is a fascinating topic [21]; here we give only a simplified view.

1.8.1 Explanation

Following Kleinhans *et al.* [15] we distinguish between **ab**duction, **in**duction and **de**duction. These three kinds of explanation are shown diagrammatically in Fig. 1.3:

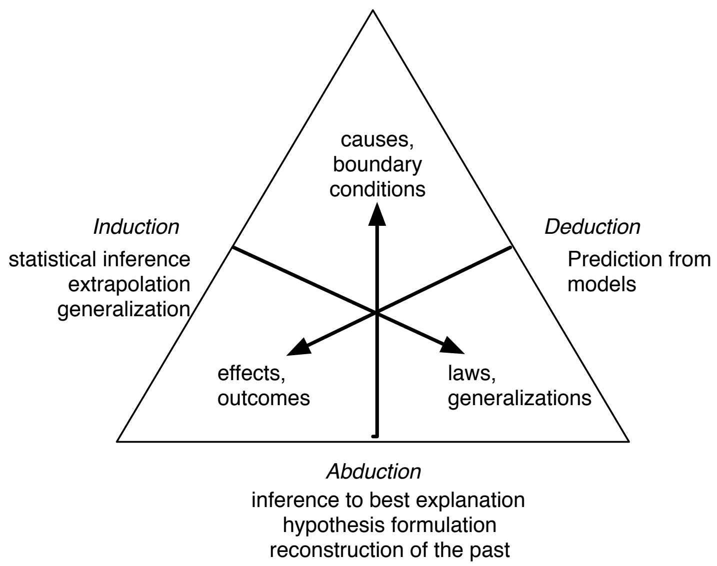

Figure 1.3: Three types of explanations (after [14])

- **Abduction**: Propose a **first hypothesis** from common knowledge, first principles, and available observations (perhaps the results of other people's experiments).

  The term "abduction" in this sense was introduced by the American philosopher C. S. Peirce around 1900. It is distinguished from induction (below) because we have only inferred, not designed any observations or experiments which might falsify the hypothesis.

  > **Example**: If we go to the market and see some red apples displayed for sale, and behind them a large crate of red

apples, we can infer (by abduction) that the loose apples were taken from that crate.

Note that there are many other possible explanations for the presence of those loose apples; there is no logical necessity that they came from the crate. Note also that we've done nothing to prove or disprove the hypothesis.

- **Deduction**: **specialise** from a general law to a specific case
  - provides ideas for experiments or observations
  - "If this theory is true, then the following should occur or be observed"

  Example: At the market we see a crate of red apples, and buy a bag of apples taken from that crate. When we get home, without looking in the bag, we can deduce that any apple that we take out of our bag will be red.

  Note that if a selected apple is not red, our premise must be false: the apples in our bag were not in fact taken from the crate. We have **falsified** our theory.

- **Induction**: **generalise** from observations to theories:
  - Logical process of **inference**;
  - From a **particular** set of observations to a **universal** statement;
  - This is how we make hypotheses, theories and laws.

  Note that an inductive argument does not assert that its conclusion is necessarily true. In technical terms, the premises of the argument do not entail the conclusion, they only provide a proximate explanation. If the premises are true, it strongly suggests the conclusion is true, but not necessarily. This is not a logical proof; i.e. there could be other explanations.

  Example: We have a closed bag of apples which may be of any colour. We pick one without looking, it is red. We pick another, it is also red. We continue in this way, and after a number of apples have been picked, we infer by induction that the apples in this bag all came from the crate of red apples.

  At this point we have a theory, and can then make a deduction (the next apple we pick will be red), which can be falsified by experiment.

1.8.2 Assumptions

- Taken as true in the context of this research;
- Can not be tested within the time, budget or experimental design;
- If they are not true, the research is not valid;
- Often difficult to express, "taken for granted" at many levels;
- Established laws are often taken as assumptions, without explicit mention (e.g. we don't repeat the laws of universal gravitation each time we model landslide hazard);
- The more problematical should be made explicit;
- Could an **assumption** be a good **research question**? I.e. maybe the "assumption" should be tested!

Assumption vs. Hypothesis

Note the main difference between an **assumption** and a **hypothesis**: the latter is tested as part of the research, the former not.

1.8.3 Proof

In science very little is actually 'proven' in the strict sense of the word. Nature is very complex and subtle; simple answers are almost never satisfactory. Instead of 'proof' in the strict sense, **accumulate evidence**; additional evidence should support a good theory;

However, additional evidence may cause the theory to be **modified**: either simplified or made more complex.

Sometimes the theory falls under its own weight (too complex, a 'house of cards'); leading to a **paradigm shift** (completely new way of conceptualising a set of observations), as famously explained by Kuhn [16].

> "My results make no sense in view of my current hypothesis.
> I have to abandon it and formulate a new one".

In practice, we are looking for proof within some **context**. Experiments are often designed to find the **limits of applicability** of a theory.

1.8.4 Statistical inference

Statistical inference is a formal way to accumulate evidence in support of a **model** which is a mathematical expression of the corresponding **hypothesis**. A statistical model can not by itself prove anything; there must also be a **meta-statistical** argument about **causes** and, if possible, **mechanisms**.

This is covered in detail in §7.

1.9 Answers to self-test questions

A1 : *They are equally parsimonious: both explain the cause of turbidity (the physical effect of dredging or of a meteor strike) and the cause of its increase (more dredging, more meteors striking the lake).* *Return to Q1* •

A2 : *The first explanation (dredging) is more persuasive, because increased dredging activity has actually been observed. The second explanation (meteorites) needs evidence, so far non-existent, of increased (or any) meteor strikes on the lake.* *Return to Q2* •

A3 : *This explanation is much more complex, hence less parsimonious. It requires many more steps to explain what is observed than the original explanation.* *Return to Q3*
•

A4 : *Such an environmental activist group must exist, they must have somehow known about the research from its inception (this link must be demonstrated – perhaps there was a spy in the research group?), they must have owned or hired boats, they must have had access to explosives. If all this is true, this alternate explanation is at least possible. But then, it must be demonstrated that dredging does not in fact stir up sediment.* *Return to Q4*
•

A5 : *Yes, this is a "reasoned supposition put forward to explain facts or events".* *Return to Q5* •

A6 : *It is clearly intended to undermine the young student's confidence in the evidence for organic evolution, and increase their susceptibility to the argument: "It's just a theory (implied: among many possibilities)".* *Return to Q6* •

A7 : *Organic evolution by mutation and natural selection is fairly described as a law, i.e. "a theory whose falsification, within its context, is almost inconceivable". Details of the process are less certain; these are theories which are the object of active research.* *Return to Q7* •

1.10 A natural resources example of abduction, induction and deduction

A soil mapping project[5] is a good example of a deductive-inductive approach with many iterations.

1. The **abductive** step: From background knowledge of soil forming process, the geological and tectonic environment, and surveys of presumably similar areas, the soil geographer forms a hypothesis of how soils have formed in this landscape, reasoning as follows:.

   > "There is evidence of recent (Pleistocene) glaciation in this region. The bedrock is known to be hard, massive limestone. The surface topography is an oriented field of rounded hills oriented in the presumed direction of the glaciation. My hypothesis is that the glacier flowed over existing small hills and formed them into a streamlined shape. As it did this, it removed the pre-glacial soil and scraped the bedrock almost bare. Since de-glaciation (about 12,000 years ago), soils have formed in the humid cool continental climate with typical northern hardwood vegetation."

   Furthermore, there is evidence from a somewhat similar region:

   > "In New England the same landform is observed on hard granite bedrock; these soils are indeed thin organic layers directly on the scraped rock."

2. The first **deductive** step: From the hypothesis, the mapper can predict the soil types expected in each location; these are typically characterised by their landscape position. **Observations** are planned where they will best test this theory, for example, where the surveyor considers the most typical of each landscape position. In this case, the surveyor may reason as follows:

   > "If my theory is right, the shallowest soils should be on the steep side hillslope where the ice stream had the most pressure; there should be hard limestone close to the surface and even out-cropping where erosion has been most active. There should be a dark topsoil with substantial organic matter, neutral to slightly alkaline, in equilibrium with the underlying limestone; there should be fragments of weathering limestone in the topsoil and increasing into the subsoil; total soil thickness over the hard limestone bedrock should not exceed 50 cm."

[5] based on the well-known drumlin field of the Lake Ontario plain between Syracuse and Rochester, NY (USA) [13]

3. The soil is examined in the field and turns out to be a very thick, slightly acid, clay with no bedrock within tens of meters of the surface and no rock fragments. Obviously, something is wrong with the hypothesis!

4. The first **inductive** step: The theory must be re-formulated. Some possibilities:

   (a) There never was a glacier here;

   (b) The glacial period was much longer ago, so soil formation had longer in which to operate;

   (c) There were no pre-existing hard rock hills for the glacier to mold.

   The first two have much regional evidence against them, so we hesitate to propose them, when there is a simpler alternative. Then we have to account for the oriented streamlined shape of the hills and their composition; i.e. we have to invent a new theory. One possibility is that the glacier encountered a flat lake plain with clayey soils, and molded these into a regular pattern of streamlined hills.

5. The second **deductive** step: The surveyor may now reason as follows:

   > "If the hills were formed by the glacier molding local clayey material, this same soil material should be found in the valleys between the small hills. Furthermore, the clay must have been formed in the inter-glacial period, so it should consist of medium- to low–activity clay minerals."

   Again, **observations** are planned where they will best test this theory, in this case in the valleys directly between two small hills. Furthermore, a laboratory determination must be made of the type of clay mineral in both landscape positions.

6. These observations show indeed the same type of clay in the hills and small valleys; furthermore the composition of the clay minerals is as predicted; the theory so far is **not contradicted** (is **supported**); note it is never fully **confirmed**.

This is only the beginning of the story; the soil geographer must build up a coherent theory of all the soils in the region and their inter-relation, in order to make a complete map. In addition, the reasoning (both induction and deduction) is much more complicated in reality.

1.11 References

Bibliography

[1] Bauer, H. H. 1992. *Scientific literacy and the myth of the scientific method*. Urbana, IL: University of Illinois Press 5

[2] Box, G. E. P.; Hunter, W. G.; & Hunter, J. S. 1978. *Statistics for experimenters: an introduction to design, data analysis, and model building*. New York: Wiley 16

[3] Burkhardt, F.; Browne, J.; Porter, D. M.; & Richmond, M. (eds.). 1994. *The correspondence of Charles Darwin*, volume 9. Cambridge, England: Cambridge University Press 20

[4] Coyne, J. A. 2009. *Why evolution is true*. New York: Viking 15

[5] Derry, G. N. 1999. *What science is and how it works*. Princeton, NJ: Princeton University Press 5

[6] Gauch Jr., H. G. 1993. *Prediction, parsimony & noise*. *American Scientist* **81**:468–478 13

[7] Gauch Jr., H. G. 2002. *Scientific method in practice*. Cambridge: Cambridge University Press 5

[8] Gower, B. 1996. *Scientific method : an historical and philosophical introduction*. London ; New York: Routledge 5

[9] Gratzer, W. 2000. *The undergrowth of science: delusion, self-deception and human frailty*. Oxford: Oxford University Press 4, 5

[10] Harvey, D. 1969. *Explanation in geography*. London: Edward Arnold 9, 14, 17, 18

[11] Hornby, A. S. (ed.). 1995. *Oxford advanced learner's dictionary*. Oxford: Oxford University Press 14

[12] Hudson, N. 1995. *Soil conservation*. Ames, IA: Iowa State University Press, 3rd edition 10

[13] Hutton, F. Z. 1971. *Soil survey of Cayuga County, New York*. Washington, DC: US Government Printing Office 25

[14] Kleinhans, M.; Buskes, C.; & de Regt, H. 2010. *Philosophy of earth science*. In Allhoff, F. (ed.), *Philosophy of the sciences: a guide*, pp. 213–236. Wiley-Blackwell 21

[15] Kleinhans, M. G.; Buskes, C. J. J.; & de Regt, H. W. 2005. Terra Incognita*: Explanation and reduction in earth science*. *International Studies in the Philosophy of Science* **19**(3):289 – 317 9, 21

[16] Kuhn, T. S. 1970. *The structure of scientific revolutions*. Chicago: University of Chicago Press, 2nd edition 23

[17] Medawar, P. B. 1963. *Is the scientific paper a fraud? The Listener (BBC)* **70**:377–8 5, 6, 19, 20

[18] Morgan, R. 1995. *Soil erosion and conservation*. Harlow, England: Longman, 2nd edition 10

[19] Okasha, S. 2002. *Philosophy of science: a very short introduction*. Oxford: Oxford University Press 4, 5, 10

[20] Perutz, M. F. 1997. *A passion for science: Review of "Lise Meitner: A Life in Physics" by Ruth Lewin Sime. New York Review of Books* **44**(3) 20

[21] Toulmin, S. E. 1958. *The uses of argument*. Cambridge: University Press 21

[22] Webster, R. 2003. *Let's rewrite the scientific paper. European Journal of Soil Science* **54**(2):215–218. URL: `http://dx.doi.org/10.1046/j.1365-2389.2003.00515.x` 5

[23] Wu, G.; de Leeuw, J.; Skidmore, A. K.; Prins, H. H. T.; & Liu, Y. 2007. *Concurrent monitoring of vessels and water turbidity enhances the strength of evidence in remotely sensed dredging impact assessment. Water Research* **41**(15):3271–3280. URL: `http://dx.doi.org/10.1016/j.watres.2007.05.018` 11, 12

2 Research

In the previous chapter science, the scientific worldview, and the scientific method were presented. Here we specify what it means to do scientific **research**.

Key points

1. **Research** is discovering something new about the natural world, the built world, or society;
2. Research may also include the development of new **methods**, **systems** or **models**;
3. Research has a **general structure**: posing questions, gathering evidence, making claims, discussing claims (§2.1);
4. **Discourse** in research can be divided into description, review and argumentation (§2.3).
5. **Types of research** include (1) designed experiments; (2) systematic observations; (3) review and synthesis; (4) system design; (5) social sciences (§2.4);

2.1 General structure of research

The term “research” is from the French *rechercher*, “to look for (again)”, and so by extension “to investigate”, “to [attempt to] find out”. This general term implies that to do research is to **discover** something that was previously completely or partially unknown or not understood.

The “something new” that is discovered by research may be:

- new **facts** about the **natural** world, the **built** (engineered) world, or human **society**;
- new understanding of the **processes** in these;
- new or improved **methods** to investigate the above;
- new or improved **systems**;
- new or improved **models**; or
- a new **synthesis** (conceptual framework) of existing facts.

In a following section (§3.1) different specific thesis structures are discussed, for example the stereotypical ‘Introduction, Methods, Results, Discussion, Conclusion’. There is however a more **general** or **abstract** structure of thesis research which covers this and other specific structures. This general structure is:

1. raising (or, posing) **questions**;
2. providing **evidence** to answer these questions; this requires some appropriate **methods** to gather the evidence;
3. making **claims**: a statement of what has been achieved, based on this evidence;
4. a **discussion** of the **reliability** and **relevance** of the claims.

In short:

Questions $\Longrightarrow$ Evidence $\Longrightarrow$ Claims $\Longrightarrow$ Context

---

Q8 : *Could we have evidence before questions?* *Jump to A8* •

Q9 : *Should we make claims before having any evidence?* *Jump to A9* •

Q10 : *The context is placed last; why not first?* *Jump to A10* •

2.2 Research stages

What is "research" also depends on how much is known about the subject. For subjects where little is known, a **three-stage** approach may be appropriate:

1. A **reconnaissance** stage of unstructured observation;
2. A **reflective** stage, during which **hypotheses** are generated;
3. A **testing** stage, where experiments or structured observations are designed to verify these hypotheses.

Reconnaissance ⇒ Reflection ⇒ Testing ⇒ Conclusions

The reconnaissance stage itself can be considered research, but only if it leads to productive and testable hypotheses. This is a common approach in anthropology or descriptive linguistics: a mass of data is recorded and then "mined" for hypotheses. At the very least, this stage must conclude with a **conceptual framework** relating the observations.

For a typical MSc project, the **literature review** substitutes for the reconnaissance stage.

The reflective stage is not by itself research; rather, it produces research questions, and so is part of the research process. A research that ended with untested questions could hardly be presented.

Finally, the testing stage is where the hypotheses are confronted with evidence, giving results and answers (perhaps partial) to the research questions. The conclusion of the testing stage is the acceptance, rejection, or modification of the hypotheses formulated in the reflective stage.

Even in the reconnaissance stage, observers must have some idea what to look for, but too much prejudice may lead them to ignore important observations. This leads to what McKnight [3] calls the balance of "inductive inquiry" (I2) and "hypothesis-driven" (HD) approaches:

Inductive inquiry
: Unguided and unlimited exploration, attempting to collect facts.

  This is speculative and with no guarantee of success – certainly facts will be collected but can they be put into a meaningful framework?

Hypothesis-driven
: Built on previous scholarship (published hypotheses with evidence for their validity), and fundamentally driven by theory.

> If the hypothesis is well-formulated and reasonable in light of previous results, and the methodology is well-designed to address it, a valid scientific result (positive or negative) is almost guaranteed.

In practice, pure I2 is impossible – even reconnaissance must have some simple working hypotheses. Similarly, pure HD is also impossible – during an experiment or observation, the scientist must be open to unexpected and unexplained observations.

2.3 Description, review, and argumentation

Another abstract view of research is by the type of writing. At various points in the document the author will be using one of three approaches:

1. **Description**;
2. **Review**; or
3. **Argumentation**.

2.3.1 Description

Accurate and complete description is of course a pre-requisite for good science. A large part of a thesis must be descriptive, with no review or argumentation: the presentation of the research questions, the methods actually applied, and the results obtained from applying the methods.

Here are some examples of descriptive text, in three contexts:

- Question: "Can metal roofs be distinguished from roofs made from plant material by their multi-spectral signature?"
- Method: "Normalized principal components analysis (PCA) was applied to the six bands, and the components that cumulatively accounted for more than 80% of the total variance were used in subsequent classification steps."
- Results: "The first two principal components accounted for 83.5% of the total variance."

These are non-controversial descriptive texts. They just have to be correct, clear and complete. The reader is being informed about the questions, methods and results.

2.3.2 Review

The thesis can not be purely descriptive. Another important aspect is the link to the scientific and societal **context**. That is, the thesis must not only **review** what has been done before in the relevant science, but also how that work **fits into** the larger societal context.

A review is not original work of the thesis, although the **synthesis** from multiple sources into a coherent review certainly requires a creative effort.

The obvious place of reviewing in the thesis is the literature review (covered in a separate topic). But, reviews also are useful in the introduction (motivation, societal importance) and conclusions (links to past and proposed work, links to societal problems).

For example:

- “Coping with soil variation has never been an easy task for soil surveyors. Soil variables vary not only horizontally but also with depth, not only continuously but also abruptly. In comparison with vegetation or land use mapping, soil mapping requires much denser field inspections. Moreover, soil horizons and soil types are fuzzy entities, often hard to distinguish or measure.”
- “The main aim of image segmentation is to distinguish homogeneous regions within an image and then to split the image into these regions. There are three general classes of segmentation techniques: thresholding, edge-based segmentation, and region-based segmentation.” (adapted from [1])
- “Competing claims on natural resources have been identified as one of the main drivers of ethnic conflict in *CountryX*. The present research showed, however, that at least in the studied region, conflicts over natural resources do not follow ethnic lines.”.

Note that all these would have supporting citations if presented in a thesis document or research paper; they are presented here without this backing for simplicity.

Q11 : *What could this author of the first example be aiming at with this review?* *Jump to A11* •

Q12 : *What will most likely follow the review in the second example?* *Jump to A12* •

Q13 : *What will most likely follow the review in the third example?* *Jump to A13* •

2.3.3 Argumentation

Description and review present information: description on what the present research intended, carried out and observed; review on the scientific and societal context. But scientific discourse requires more than that: the description and review must be used as the basis for **argumentation**.

Among the points to be argued in a thesis are:

- Why certain **questions** are **interesting** to address;
  - Are they important?
  - Are they unsolved?
  - Is there a promising approach to address them?
- Why particular **methods** have been **chosen**;
  - Why are they appropriate to the question?
  - Are there others, and if so, why were they not chosen?
- Why and how the **results** contribute to the **advancement of science**;
  - How significant are the results?
  - How much do they confirm or contradict other work?
  - If they seem to contradict or extend other work, why? What is different about this study?
  - What are possible followups to this work?
- Why and how the **results** contribute to **societal goals**.
  - Who cares about the results?
  - What should be done with the results in order to have a positive effect on society? Note that this can be an indirect argument: progress in knowledge or methodology is *prima facie* advancing the human project.

Types of argumentation, including sound logic, are discussed in topic "Argumentation" (§5).

2.4 Types of research

Under the general term "research" we can identify several kinds of research projects:

1. **Designed experiments**, e.g. laboratory or field research where the researcher imposes the **treatments** in a (semi-)controlled situation and measures the system response;
2. **Systematic observations**, e.g. resource survey or community meetings, where the researcher makes **measurements or observations** according to a plan but without complete control of the process;
3. **Data mining**, where the researcher looks for unexpected patterns in large datasets (§2.4.3).
4. **Synthesis**, where the researcher imposes a new **conceptual framework** on previous data and establishes that this is a better or more unifying explanation;
5. **System design**, where the researcher **designs** a system (database, visualization, modelling ...) and shows that it is somehow "better" than previous designs; this includes design of algorithms and methods;
6. **Modelling**, where the researcher builds a conceptual or (more commonly) computational model of a process; the model is evaluated by its success in **reproducing the behaviour** of the natural or social system.
7. **Comparative studies**, where the researcher compares existing situations in order to determine the reasons for the observed differences. The researcher must argue that all relevant factors have been considered; thus only close analogues should be used.

Q14 : *Why is casual observation not included in this list? Should it be added?* *Jump to A14* •

2.4.1 Natural vs. social sciences

There is in general a distinction between the natural and the social sciences:

Natural sciences — The principal object of study is "nature", i.e. physical reality; there is a clear separation between observer and observed; argumentation is as logical and objective as possible;

Social sciences — The principal object of study are humans and human society (including **organizations** and **governments**); so we can not impose treatments at will; we are studying ourselves or our social constructs, so it is very difficult to avoid subjectivity; argumentation grades into humanities; see also §3.1.13.

2.4.2 Object vs. methodology

The main focus of research can be on the **object**, also called the **thing in itself** (e.g. natural world, the built world, society), or on the **methodology**[1] used in the study, i.e. how the "thing in itself" is best studied.

For example:

Object — Changes in land use in a study area; commerce patterns in a district; audit of a reconstruction project after a natural disaster;

Methodology — How to assess land-use changes with multiple satellite sensors of different resolution; how to visualise spatio-temporal commerce patterns; how to map reconstructed buildings from high-resolution imagery using image segmentation techniques.

Often an ITC thesis includes both aspects; we are interested in the thing in itself (ITC development relevance) but also the methodology (ITC technology focus).

2.4.3 Data mining

The advent of very large datasets and powerful computers has opened a new kind of research, similar in philosophy to the pure reconnaissance stage or "inductive inquiry" explained in §2.2, called **data mining**. This has been defined by Foody [2] as "the extraction of patterns from data". He points out that "a vast amount of geographical data is now acquired routinely and *often without prior hypothesis*" (emphasis added). So the task of data mining is to find hidden patterns that then become hypotheses.

An example of this research approach is the study of Rogan *et al.* [4] to compare the performance of three machine learning algorithms for mapping land-cover modifications. These authors had no prior theory on which algorithms would give the best results, nor why.

The innovation in a data mining study is in the method by which patterns are found.

2.5 Answers to self-test questions

A*8* : *Certainly we could have evidence before questions; in fact there is evidence all around us in the form of previous studies as well as our own unstructured observations. This prior evidence is used to formulate questions.*

[1] "study of methods"

The 'evidence' in this sequence is the evidence we collect in our own work. *Return to Q8* •

A*9* : *Certainly we can make claims before having any evidence; in fact we should make* ***tentative*** *claims before gathering evidence, which the evidence should confirm, deny or modify.*

The tentative claims help us select methods. *Return to Q9* •

A*10* : *In fact there is an implicit feedback here: the context at the end of one research becomes the context at the start of the next one.* *Return to Q10* •

A*11* : *The author is reviewing the difficulties to date with soil survey. Most likely he will propose a new method to overcome these difficulties. Thus the review is motivating the research.* *Return to Q11* •

A*12* : *The author will most likely pick one of the reviewed techniques, and justify this choice with argument.* *Return to Q12* •

A*13* : *The author will try to explain why the results of this study seem to contradict other studies; this will then be argumentation.* *Return to Q13* •

A*14* : *Casual (opportunistic) observation may lead to hypotheses but it can not be considered research, because it is not driven by questions, hypotheses, and appropriate systematic methods.*

Some researchers disagree, and would include casual observation as research, arguing that, after all, something new is discovered and reported. The counter-argument is that without a systematic method it is impossible to put these observations in any framework and draw any reliable conclusions. *Return to Q14* •

2.6 References

Bibliography

[1] Du, Y. 2008. *Verification of tsunami reconstruction projects by object - oriented building extraction from high resolution satellite imagery*. MSc, ITC. http://www.itc.nl/library/papers_2008/msc/gfm/duye.pdf 33

[2] Foody, G. M. 2003. *Uncertainty, knowledge discovery and data mining in GIS*. *Progress in Physical Geography* **27**(1):113–121 36

[3] McKnight, S. L. 2009. *Unconventional wisdom*. *Cell* **138**(5):817–819 31

[4] Rogan, J.; Franklin, J.; Stow, D.; Miller, J.; Woodcock, C.; & Roberts, D. 2008. *Mapping land-cover modifications over large areas: A comparison of machine learning algorithms*. *Remote Sensing of Environment* **112**(5):2272–2283 36

3 The MSc proposal and thesis

After completing this Research Skills module, the student should be able to **conceptualize** and **structure** their research proposal and MSc thesis according to accepted standards of the ITC course and the scientific discipline of the thesis topic.

This implies the following end objectives:

1. To define research **problems**;
2. To define research **objectives** to address the problems;
3. To define research **questions** to meet the objectives;
4. To select appropriate **methods** to address the questions;
5. To define **anticipated results** ("hypotheses") expected from the application of the methods;
6. To **structure** these into a thesis proposal and, eventually, a thesis.

We first discuss how to structure research in a research proposal (problems, questions, claims, methods). We then discuss how the research proposal becomes transformed into a research thesis.

3.1 The MSc research proposal

Key points

1. The **research proposal** establishes the relevance, novelty, methodological soundness, and feasibility of the project; it **convinces** the reviewer that the research should be undertaken.
2. The research proposal often has a conventional structure: **Problem** ⇒ **Objectives** ⇒ **Questions** ⇒ **Hypotheses** ⇒ **Methods** (§3.1.1).
3. The **research problem** is a general statement of what is not known and should be discovered by research (§3.1.2); the **research objectives** are statements of what is expected as the results of the research (§3.1.3); the **research questions** are specific questions that the research can answer (§3.1.5); the **hypotheses** are expected answers to each question (§3.1.6).
4. **Research methods** are chosen in order to answer the research questions (§3.1.8).
5. If the research is carried out in a specific geographic area, the **study area** must be described (§3.1.10).
6. A **design thesis** must present a new design that is demonstrably "better" in some sense than existing systems (§3.1.11).
7. A **social** or **organizational thesis** emphasizes concepts and definitions and argues from diverse evidence (§3.1.13).

The research proposal establishes the relevance, novelty, methodological soundness, and feasibility of a thesis project. It should convince the reviewer that the research should be undertaken.

In a previous section (§2.4) a classification of research was presented. The ITC thesis will typically fall in one of these categories, but may have elements of several. We first discuss aspects of the proposal that are common to all thesis types, and then aspects specific to each type.

3.1.1 Common elements of a research proposal

A research proposal has a conventional structure:

Problem ⇒ **Objectives** ⇒ **Questions** ⇒ **Hypotheses** ⇒ **Methods**

The problem, objectives, questions and hypotheses are usually in one section called **Introduction**. This is usually followed by a **Literature review** and then the **Methods**.

The thesis will then have several more sections, covering the results, discussion, conclusions and recommendations; this is explained in §3.1. But at the time the proposal is written, there are no results, so these sections are not included in the proposal.

Q15 : *Why might it be useful, in the proposal writing stage, to outline the sections on results, discussion, conclusions and recommendations even before any results are available?* *Jump to A15* •

Following these concepts in order is a systematic way to approach research. It must first fit a known **problem** (so that it is important), then it must have a defined **objective** (so that it is clear what it should accomplish), which is then specified as a list of **questions** that the research should answer. For each question, the researcher must have a **hypothesis**, i.e. what answer is expected.

Examples of research proposals

These concepts are illustrated here with examples modified from three ITC MSc theses, one in Naivasha, Kenya [22], one in Lake Cuitzeo, Mexico [12], and one from the Netherlands [38]. The Naivasha example deals with the applicability of Small-format Aerial Photography (SFAP) to monitor wind erosion; we will also examine a different thesis topic (not actually carried out) that deals with the same area and general problem, but more conceptually. The Lake Cuitzeo study is an example of a survey-oriented thesis, where the principal problem that there is no map of something of interest, so mapping methods must be developed and applied. The Dutch study is an example of method design, here using animation to represent uncertainty in regional planning maps, with a case study in North Brabant province.

3.1.2 Research problem

The research problem is a general statement of **why** the research should be done. This is something that is not well-understood or solved and can be addressed by research.

There are many social problems (poverty, environmental destruction, war, ...), but an MSc thesis rarely addresses these directly. Rather, social problems can motivate research and prove its societal relevance. The social problem leads to research problems.

The fundamental questions:

1. Why should anyone **care about the outcome** of this research?
2. Who would use the **results** of this research? and **for what?**
3. Why should anyone **sponsor** this research?

A reasonable answer to the first question might be "because it's intrinsically interesting to know … " or "because it's not known"[1]. This undoubtably advances overall knowledge; however for most ITC research there should be a more concrete reason to undertake it.

> And of course, one important outcome of MSc research is that the successful candidate will receive an important professional qualification (the MSc degree) and further career … so the candidate surely cares about the outcome of the research … but that's not sufficient for the sponsor or supervisor!

Research problems can be categorized as follows:

- **Social**: something wrong with human society;
- **Environmental**: something wrong with the natural world;
- **Management**: a deficiency in managing a social or environmental problem;
- **Technical**: a deficiency in methods to solve problems;
- **Information**: a lack of information, facts that are not known;
- **Knowledge**: a lack of understanding: why things happen.

These categories are not mutually exclusive; a research project can address combinations of these. For example, social, environmental, management and technical problems often reveal an additional information or knowledge problem. And there can be technical aspects of the other problem categories.

The novelty of the research must be supported by a **literature review**. If someone else has already solved the problem, why re-do the work? This is explored further in the topic "Literature Review".

Example (Naivasha SFAP): This is an example of a technology-oriented thesis, where new methods must be developed:

> "Wind erosion is causing widespread destruction of crop land and pastures in the rift valley of Kenya."

[1] As in Mallory's answer to the question as to why he wanted to climb Mount Everest: "Because it's there."

> "We do not know the priority areas for intervention."

> "It is impractical to monitor wind erosion over large areas by ground survey or conventional aerial photography."

Q16 : *Classify these three statements according to the list of problem types just above.* *Jump to A16* •

This sequence of problems leads naturally to an **objective**, namely to find a cost-effective way to monitor wind erosion over large areas and from these surveys to determine priority areas for intervention.

There are many problems implicit in this example, and they could lead to useful research problems:

- It is not known how to monitor wind erosion over large areas in a cost-effective manner;
- Priority areas for intervention have not been identified;
- There are no established methods for identifying priority areas;
- It is not known what land-use practices are most **associated** with wind erosion;
- The physical and social **causes** of wind erosion in this area are not known;
- Interventions to minimize erosion are not known.

These problems are inter-related: we must be able to monitor before we can determine the priority areas; and the monitoring is the basis for associating land-use practices with erosion. This association is then used as evidence when arguing about the physical and social causes; and once we know these causes we can design interventions.

The question for the MSc student is: **What can I realistically accomplish in the thesis research?** and what part do I leave for others? This is partly determined by the status of the problem. For example, if no good monitoring method is available, we should work on this, because the other problems depend on it.

Example (Naivasha causes): This is an example of a cause-oriented thesis, where the emphasis is on determining **why** something is occurring.

> "Wind erosion is causing widespread destruction of crop land and pastures in the rift valley of Kenya."

> "Previous studies have established the extent of the problem and its historical development."

> "We do not know the proximate or ultimate causes."

The emphasis here is on determining **why** some known phenomenon is occurring.

Q17 : *Classify these three statements according to the list of problem types just above.* *Jump to A17* •

Example (Lake Cuitzeo): This is an example of a survey-oriented thesis, where the principal problem that there is no map of something of interest.

> "The water of Lake Cuitzeo is used for multiple purposes, including irrigation and human consumption." (context)

> "Almost nothing is known about its quality, but large areas are suspected to be sub-standard for both purposes." (an **information** problem)

> "There is no map of the different water quality parameters in the lake." (also an **information** problem)

> "It is not known whether there is any trend in the quality." (also an **information** problem)

> "Nothing is known about the causes of poor water quality, although it is suspected that high-input irrigated farming is a major contributor."

In the limited time available for an MSc thesis, only some of this lack of knowledge can be addressed. In particular the time series necessary to determine a trend can not be collected in one field visit. The final problem depends on the knowledge that is lacking as expressed by the second and third problems.

Example (animation): This is a design-oriented thesis, where the principal problem is that current designs (in this case, of representations of spatial uncertainty) are not adequate.

> "The province of North Brabant (NL) is digitizing land-use plans and making them available in this form to the public and professional planners."

> "Not all planning objects are comparable, because some of them are uncertain or fuzzy."

> "Uncertainty and fuzziness are hard to perceive in traditionally-mapped data."

> "Planners are not able to correctly judge how some planning objects that are continuous in reality influence land-use options ... because these continuous features are represented by crisp boundaries on the map."

> "Planning objects, of which the location, boundaries, orientation, size and/or shape are not well-defined, can not be judged exactly."

> "Static graphic variables have been used to represent uncertainty and fuzziness, but dynamic visualization methods have not yet been integrated with these."

Q18 : *What could be a research problem that the author could address?*
Jump to A18 •

Difficulties defining research problems There is often confusion between a **social** and a **research** problem. For example, consider the following social problem:

> "In recent decades, urban sprawl of City X has dramatically increased, especially new-style gated residential areas. The combination of private roads, high buildings, poor sidewalk design, and few gates in these residential areas results in difficult access to public transit."

This leads to many other social problems: inequity, lost productivity, increased automobile usage leading to congestion, pollution etc. But, where is the **research** problem, that is, what is not known, related to this social problem, that can be determined by research? Some possibilities are:

1. It is not known what proportion of people use public transit, how they reach the stops, how much time is required from various locations (3D) – an **information** problem;
2. It is not known what are people's motivations for using public transit vs. other forms (automobile, walking, bicycle) – a **knowledge** problem;
3. The optimal placement of public transit stops and routes is not known – a **management** problem;
4. Current methods for route optimization don't take into account the time people spend within tall buildings – a **technical** problem;

Note that all of these would be justified by literature review.

3.1.3 Research objectives

These are statements of what is expected as the **output** of the research. Each of the objectives must be at least partially met at the end of the project. They are not operational – they say what the author wants to accomplish but not in enough detail (yet) to plan the research,

> **Note:** Operationalization is the task of the **research questions**, as discussed in the next section (§3.1.5).

There is usually a single **general** objective which is not operational, which is then broken down into a list of **specific** objectives which can be addressed by operational research methods.

Example of a general objective (Naivasha SFAP):

> "To determine the applicability of Small-format Aerial Photography (SFAP) to wind erosion mapping and monitoring in the rift valley of Kenya, and the main factors which affect its success."

Example of a general objective (Naivasha causes):

> "To determine the causes of wind erosion in the rift valley of Kenya."

Example of a general objective (Lake Cuitzeo):

> "To map the water quality of Lake Cuitzeo on one sampling date and suggest possible causes for any spatial variation in water quality."

Example of general objectives (animation):

> "To develop methods to effectively visualize uncertainty and fuzziness in animated representations by various combinations of graphic and dynamic visualization variables."

> "To select or develop a method by which the usability of uncertainty and fuzziness display in spatial planning maps can be evaluated."

Q19 : *How are the two objectives of the animation proposal related?* *Jump to A19* •

The **specific objectives** should be built up from simple (easy to formulate and investigate) to complex. If there is an inventory to be done, the objective is simply to do it; this may be followed by objectives that require more inference.

The thesis should at least partially meet all the objectives.

Example of specific objectives (Naivasha SFAP):

> "To determine which wind erosion features, and of what dimensions, can be visually interpreted on SFAP"

> "To determine the accuracy with which SFAP can be georeferenced with single-receiver GPS and mosaicked into a seamless image"

> "To determine the costs of a SFAP mission in local conditions"

Example of specific objectives (Naivasha causes)

In this case, the word "causes" is very broad, and it is customary to distinguish between **proximate** (immediate) and **ultimate** causes, also between **factors** and **processes**.

> "To determine factors related to wind erosion in the study area"

> > "To determine which land-use practices are most associated with wind erosion"

> > "To determine which soil properties are most associated with wind erosion"

> "To relate these factors with presumed processes"

> "To identify and quantify the proximate and ultimate causes of wind erosion in the study area"

Questions about **factors** may be answered by investigating the **association** (roughly speaking, "correlation") between them and the erosion; this then is information to be analyzed in terms of the **processes** by which wind erosion occurs, to finally discuss **causes**.

Example of a specific objective (Lake Cuitzeo):

> "To determine the water quality status of the central and Eastern parts of the lake"

> "To map the spatial distribution of the water quality components measured at one sampling time"

> "To map the spatial distribution of aquatic vegetation density"

> "To determine if there is a relationship between the reflectance values of optical multi-spectral sensors and measured water quality parameters including vegetation density"

> "To determine whether land use affects water quality, and if so, which constituents are affected by which land uses"

It's clearly easier to simply sample and map the water quality (first two objectives) than to determine why the water is of higher quality in some areas than in others (last objective). The third and fourth objectives (vegetation density and whether multi-spectral sensors can detect this) are technology objectives in support of the other research objectives. These objectives could be the objectives of another (different) thesis if a difficult enough problem, since this would be a different research focus.

3.1.4 Expected outputs

In some thesis projects there may also be a **list of outputs**, i.e. what is expected to be produced, e.g.:

- maps
- databases
- computer programs

all of which are specified in detail.

These are logically part of the objectives ("an objective is to produce a map of ... ") and are most common in "design" theses.

3.1.5 Research questions

These specify what the research will actually address. **Each research question must be answered** by the thesis, therefore it must be a specific question to which an answer can be given. Questions follow objectives and may be simple re-statements in **operational form**, i.e. where an experiment or sample can answer it.

Questions are of two main types:

observational — 'What', 'where' or 'which' questions;

analytical — 'Why' or 'how' questions.

Answers to analytical questions are the real objective of a research thesis; answers to observational questions provide the evidence.

"Question" words

Here are some words that can be used to introduce research questions; first for those that do not require much analysis:

- **"Where?"** (mapping), e.g. "Where (in the study area) is the most severe accelerated erosion"
- "Is there" or "Does" (presence, existence), e.g. "Is there a water quality gradient with depth?"; this could be re-formulated "Does water quality vary with depth?"
- **"Which?"** (identification), e.g. "Which land areas are currently used for smallholder cassava production?"; "Which aspects of current land use plans are most controversial?"

- **"Can?"** (technique), in the sense of "Is it possible?", e.g. "Can a light aircraft with GPS carry out a photo mission to specified accuracy standards?"; "Is it possible to see blow-outs on an air photo?"
- **"What?"** (results of a technique), e.g. "What is the accuracy of georeferencing?"
- **"What?"** (is encountered in the field), e.g. "What are the most common species of trees planted in domestic gardens?"
- **"How?"** (observational), e.g. "How has water quality changed since the establishment of the irrigation project?"; this could be re-formulated "What, if any, are the change in water quality ... ".

Another type of question requires deeper analysis:

- **"What is?"** (effects), e.g. "What is the effect of increased grazing on vegetation density?"
- **"What is?"** (relation), e.g. "What is the relation between increased grazing and vegetation density?"; this must be answered with a statistical model.
- **"Why?"** (causes), e.g. "Why does increased grazing affect vegetation density?"; this must be answered with some proposed mechanism.
- **"How?"** (function), e.g. "How does increasing pesticide use in surrounding farmland affect reproductive success of migratory bird species in the lake?"

Example of research questions (Naivasha SFAP):

> "What are the photo-interpretation elements for different wind erosion features?" (e.g. in this case the blowouts may be darker because of the different ash in subsoil; elongated form in wind direction etc.)
>
> "Can blow-outs and dunes caused by wind erosion be seen on SFAP, and if so, of what dimensions?"
>
> "What is the smallest wind erosion feature than can be recognised, measuring both vertically and horizontally?"
>
> "Can sufficient group control points be established to convert the set of SFAP photos to orthophoto mosaic?"
>
> "What is the accuracy of such a conversion, using a single GPS receiver for ground control?"
>
> "What is the cost of a SFAP mission and how does this compare with conventional survey?"

> "What is the time required to organise a SFAP mission and produce an wind erosion assessment, and how does this compare with conventional survey?"

Note that although the general objective speaks of "monitoring", there is no research question directly related to this, because the research is only done in one time period. However, several questions relate to monitoring: what can be detected, and how much a mission costs. So in the conclusion the author can use the answers to these questions to discuss the applicability of the method for monitoring.

Example of research questions (Naivasha causes):

> "What are the land-use practices in the study area?"

> "Which of these are most associated with wind erosion features?"

> "What is the quantitative relation between the intensity of specific land uses and wind erosion?"

> "What is the physical process which relates the intensity of a specific land use to wind erosion?"

> "What are the synergistic or antagonistic effects of specific land uses and other causative factors?"

> "What is the principal cause of wind erosion in the study area?"

The above list only mentions land-use intensity; other causative factors should be added. Note that the questions go from easiest to answer to hardest. The last question can not really be answered as such; instead we can **argue** from the results of the previous questions to a more-or-less convincing story about causes.

Example of research questions (Lake Cuitzeo):

> "What is the water quality (turbidity, salinity) and depth at representative sample points in the Central and Eastern parts of Lake Cuitzeo?" (sampling)

> "What is the spatial structure of the lake depth as modelled by (i) geographic trend surface, (ii) distance from shore, and (iii) ordinary variograms?

> "What is the spatial structure of the water quality parameters as modelled by (i) geographic trend surface; (ii) distance from shore; (iii) depth; (iv) ordinary variograms; (v) residual variograms from the trend and feature space models?" (modelling)

> "How much of the spatial structure can be explained by these models and how much remains unexplained?" (success of modelling)

> “What is the spatial distribution of water quality parameters and depth?” (mapping, using the models)

> “What is the relationship between reflectance values of optical multispectral sensors and water quality parameters including vegetation density?” (modelling, depends on the previous map)

> “What is the spatial distribution of water quality parameters including vegetation density as mapped from optical multi-spectral sensors?” (mapping, using the models)

> “What land uses are associated with areas of poorer water quality?”

Note how the water quality parameters are now specified. The analytic methods (trend surfaces, variograms) are also specified. Some questions depend on the results of others. For example, if there is no relation between aquatic vegetation and MSS, it is impossible to make a map.

This is a long list of questions and may be too much for a single study. Not all questions may be answered to the same depth.

Example of research questions (animation):

> “Which planning objects are uncertain and fuzzy in spatial planning maps?”

> “What characteristics of these objects play a role in the plan preparation phase of spatial planning?”

> “How can these objects be represented in an interactive animated way by combination of graphic and dynamic visualization variables?”

> “How can the annoyance of some users by some animated effects, e.g. moving or blinking objects, be eliminated, while still communicating the uncertainty?”

> “Which combinations of variables can best aid spatial planners in making better decisions?”

Research Questions related to Research Objectives

One way to organize research **questions** is to list them as a sub-list under each research **objective**. This shows which questions, if answered, will meet the objective. Here is one of the examples from the previous section, re-organized in this way.

Example of specific objectives and related questions (Naivasha SFAP):

1. “To determine which wind erosion features, and of what dimensions, can be visually interpreted on SFAP”

(a) "What are the photo-interpretation elements for different wind erosion features?"

(b) "Can blow-outs and dunes caused by wind erosion be seen on SFAP, and if so, of what dimensions?"

(c) "What is the smallest wind erosion feature than can be recognised, measuring both vertically and horizontally?"

2. "To determine the accuracy with which SFAP can be georeferenced with single-receiver GPS and mosaicked into a seamless image"

   (a) "Can sufficient group control points be established to convert the set of SFAP photos to orthophoto mosaic?"

   (b) "What is the accuracy of such a conversion, using a single GPS receiver for ground control?"

3. "To determine the costs of a SFAP mission in local conditions"

   (a) "What is the cost of a SFAP mission and how does this compare with conventional survey?"

   (b) "What is the time required to organise a SFAP mission and produce an wind erosion assessment, and how does this compare with conventional survey?"

3.1.6 Hypotheses

Hypothesis: "[An] idea or suggestion that is based on known facts and is used as a basis for reasoning or further investigation" [15]

In the context of research, these are the researcher's ideas on what the research will show, before it is carried out. They are statements that can be **proved**, **dis-proved**, or (most likely) **modified** by the research. They are based on previous work, usually discovered in the literature review. They should match the research questions one-to-one.

Another definition of hypothesis in this sense is **anticipated results**.

The hypothesis must be **specific**, not a general statement. For example, given the research question "What is the effect of grazing intensity on vegetation density?" we can formulate the corresponding hypotheses:

- **Wrong**: "Grazing affects vegetation density"
- **Right**: "Above a threshold (to be determined), vegetation density is reduced linearly (coefficient to be determined) with grazing intensity, measured as animal-months.

The “to be determined” could be filled in with reference to results reported in the **literature review**, or from **first principles**. The first hypothesis is too general, “affects” could be anything.

Example of hypotheses (Naivasha SFAP): The following statements refer to SFAP at a nominal photo scale of 1:5 000:

- “Blow-outs and dunes caused by wind erosion can consistently be seen on SFAP”
- “Both blow-outs and dunes with a vertical relief difference of as little as 1 m, and an minimum horizontal dimension of 5 m can be seen.”
- “It is always possible to find sufficient points for direct linear transformation within a single SFAP.”
- “SFAP can be converted to an orthophoto mosaic with a horizontal accuracy of 5 m using GPS ground control.”
- “The cost of a SFAP mission is an order of magnitude less than a conventional air photo mission.”
- “The time required to organise a SFAP mission and produce an wind erosion assessment is less than two weeks.”

Example of hypotheses (Naivasha causes):

- “The principal land uses are small-scale subsistence farming, paddock grazing of cattle, and extensive grazing.”
- “Wind erosion is found only in paddock grazing.”
- “No erosion is observed until grazing intensity reaches a threshold, after which the extent increases exponentially with grazing intensity until the whole area is destroyed.”
- “Overgrazing leads to removal of the surface cover (grasses), exposing the soil to the full kinetic energy of the wind.”
- “Fine-grained volcanic ash soils are more susceptible to wind erosion, when exposed by overgrazing, than coarse-textured ash and lacustrine soils.”

These hypotheses came from previous land use and soil studies in the study area, wind erosion studies in similar areas, and general physical principles. They look like conclusions but they are not! They are hypotheses to be verified, modified, or refuted.

Note especially the third hypothesis, giving the **form** of the presumed quantitative relation.

Example of hypotheses (Lake Cuitzeo):

In this case there is very little known about the study area, so the hypotheses are not very specific. An examination of an optical image gives some clues.

> "The central part of the lake is shallower and more turbid than the eastern part of the lake; salinity is absent to moderate."

> "There is no geographic trend to depth; depth increases quadratically (bowl-like) with distance from shore; there is strong spatial dependence at ranges to 1 km."

> "There is an east-to-west geographic trend in salinity; turbidity increases quadratically (bowl-like) with distance from shore; ... "

> "Models explain about 80% of the spatial variability." (this based on studies in "similar" areas)

> (No hypothesis for the corresponding question, output is the map)

> "Turbidity is linearly related to blue reflectance."

> (No hypothesis for the corresponding question, output is the map)

> "Areas of the lake receiving discharge water from high-input irrigated agriculture have the poorest water quality."

Example of hypotheses (animation): This thesis does not state any hypotheses.

Some of the questions can not be re-formulated into hypotheses. For example, "Which combinations of variables can be recommended to aid spatial planners in making better decisions, based on user tests?" depends on the results of the tests. But, to design the tests, the researcher must have some idea (hypothesis) about which combinations might be best, so that these can be included in the test.

Another example is "What characteristics of these objects play a role in the plan preparation phase of spatial planning?" This will be answered by direct observation of the planners at work; it is impossible to prejudge.

But, from literature review and the researcher's own ideas, there should be design decisions which are proposed as superior:

> "Replacing blinking graphical objects with subtle low-frequency changes in colour enhances comprehension and reduces user fatigue".

Statistical hypotheses Another use of the word "hypothesis" is in frequentist statistical inference; see Chapter §7. Here the so-called **null**

hypothesis (abbreviated H_0) is a numerical statement about some population that is to be tested on the basis of some sample; the so-called **alternate hypothesis** (abbreviated H_1 or H_a) is its complement.

Example: "There is no difference in mean height between third-grade boys and girls in school district X"; or "The difference in mean height between third-grade boys and girls in school district X is 5 cm."

	Null hypothesis H_0 is really ...	
Action taken	True	False
Reject	Type I error committed	success
Don't reject	success	Type II error committed

This definition is too narrow for use in a research proposal, especially because frequentist testing is not the only approach to statistical inference.

3.1.7 Assumptions

Assumptions are preconditions for research, things that are taken as true and which are not questioned during the proposed research. They are two kinds: conceptual and logistical.

Q20 : *What is the essential difference between an assumption and a hypothesis?* *Jump to A20* •

If an assumption is false, the research is (at least partly) invalid (for concepts) or infeasible (for logistics). It is often difficult to specify assumptions, precisely because we "assume" them, and they are thus difficult to make explicit.

Questioning an "assumption" is one way to think of challenging research questions. Maybe what "everyone knows" to be true isn't.

Assumptions are of different levels of abstraction and certainty. Some are about nature; often these are so deeply internalized that we do not consider them explicitly, e.g., we *assume* that gravity and light continue to operate as usual (water flows downhill, the energy of light depends on the wavelength, and similar for physical laws. These are not mentioned in the thesis.

Other "laws" are not so universal as gravitation, yet they form the (often unspoken) basis of work in a discipline. Roughly speaking these are the well-established **textbook results** that have reached the status of common knowledge.

Other assumptions are about factors we will not confirm but which are necessary for our results:

- "Soils are fairly homogeneous in the study area, so any differences in biodiversity are due to other factors (the ones we will study)."
- "Social structure in the study area is based on strong kinship ties."

This forms part of the argument for the research approach (problem, objectives, questions). They should be backed up by literature (in this case, a soil survey and a previous social survey); the validity of this justification can be evaluated with the thesis proposal.

Q21 : *In this example, what should the researcher do if it can not be assumed that the soils are homogeneous?* *Jump to A21* •

This kind of assumption is sometimes called *ceteris paribus*, Latin for "all other things being equal". Each research question leads to one or more hypotheses (i.e., proposed or possible answers to the question). So, the methods chosen to test the hypothesis only consider those factors mentioned in the hypothesis. But, if there are other factors (untested) that in fact cause experimental or observational differences, conclusions based on the "answer" to the hypothesis (true, false, modified) are wrong or incomplete. This is especially important in causal hypotheses: "The reason that X occurs is Y"; when only $Y_1, Y_2 \ldots$ are considered, but the real reason is Z.

Other assumptions are preconditions for research **logistics**, and should be made explicit in the proposal, e.g.:

- "The study area is accessible;"
- "Permission to access the study area will be granted by local authorities"
- "A translator will be assigned to the research team;"
- "Samples will be processed by a laboratory correctly and within a given time;"
- "A model will be updated by its author prior to the time it is needed in this research;"
- "ITC will acquire a license for a specialized computer program."

These form part of the argument for the research **methods**. Their validity can be judged along with the proposal.

Verifying assumptions

In some research projects research methods are designed to verify "assumptions" (e.g. that the soils are relatively homogeneous in an area).

But properly speaking, these are then not assumptions, but rather hypotheses, and should be included in the list of research questions ("Are the soils homogeneous?") and hypotheses ("yes"), with appropriate methods to test them.

Thus the research has several stages, some of which are pre-conditions for others.

3.1.8 Research methods

At this point the research has been structured as:

1. Social, contextual problems
2. Research **problems**
3. Research **objectives**
4. Research **questions**, several per objective
5. Research **hypotheses** for each question
6. Research **assumptions**, not to be tested

The next step is to select **research methods** to answer the questions.

Methods are chosen in order to answer the research questions. This is why specific questions are so important.

Note that some methods may not be applicable to your situation. For example, laboratory tests for the cation exchange capacity soils developed for young soils in temperate climates give very misleading results for most tropical soils, for which other methods and measures are appropriate. As another example, some image processing methods may only be feasible for small images. You must argue that the selected methods fit the research context.

Finding methods

There are many books describing methods; these are specific to given disciplines. For example:

- Miles & Huberman [21]: methods for qualitative data analysis, e.g. in social sciences research
- Ryerson & American Society for Photogrammetry and Remote Sensing [33]: methods in remote sensing

- Maidment [20]: methods in hydrology
- de Gruijter *et al.* [13], Cochran [5]: sampling designs
- Knuth [18]: algorithms
- FAO [9]: methods for describing soil profiles
- van Reeuwijk [27]: laboratory methods for soil analysis

You should know the methods references for your field.

There are also review articles or book chapters that describe and compare methods; these are excellent resources to help you choose among methods. For example:

- Foody [11]: review of methods for accuracy assessment of land cover maps.

Advanced **textbooks** often explain and compare methods. This is common in statistics, for example:

- Bishop *et al.* [2] on discrete multivariate statistics.
- Legendre & Legendre [19] on statistical methods in ecology
- Davis [6] on statistics and data analysis in geology

Searching the **internet** can be a useful starting point, but very rarely provides a definitive method. Use it to find reliable references in other forms. Some handbooks may have been placed on-line as a convenience.

There are some complete handbooks on-line; if from a reputable source they can be used and cited, for example:

- National Institute of Standards and Technology [23] on statistical methods for quality control.

Example methods

Example of methods (Naivasha SFAP):

For example, to answer the question "Can blow-outs and dunes caused by wind erosion be seen on SFAP, and if so, of what dimensions?", we must:

1. Make a legend of wind erosion features and their characteristics to be measured in the field;
2. Identify test features in the field and geo-reference them;
3. Produce the SFAP;
4. Geo-reference the SFAP;

5. Interpret the SFAP at the locations of test features according to the legend;
6. Compare the interpreted features with the known features;
7. Quantify the degree of agreement.

All of these require definite methods. In this case we also have to protect against photo-interpreter bias: knowing the features in the field will the interpreter imagine them on the image? Perhaps the photo-interpretation should be before the field visit? Or should a block be photo-interpreted, not just specific features? It requires careful thought to make the methods able to answer the questions.

Example of methods (animation):

This author drew a flow chart, classifying the methods as:

1. Task analysis
2. Development of conceptual framework
3. Creation of animated representations
4. Evaluation
5. Synthesis and recommendations

Q22 : *The "task analysis" is a study of the sources of uncertainty and fuzziness in spatial planning maps. Why is this placed before the development of a conceptual framework?* *Jump to A22*

•

Description of methods

For each research method selected, the Methods section of the proposal (and thesis) should state:

1. Either:
   (a) the **name** of the method that was chosen, with a **reference** to the literature that describes it; *or*
   (b) a **detailed description** of the method, if it is being developed as part of this project;

   In either case, the method must be described in sufficient detail (either here or in the references) for someone else to be able to apply it.
2. The **materials** necessary to apply the method;

3. **Why** this method was chosen:
   (a) Why is it applicable in this study?
   (b) Why is it preferred to other methods that could have been applied?
      · For example: cheaper, faster, more precise, adapted to the specific environment ...
4. What are the **assumptions** for applying this method, and how are they met in this study?
   · For example, a 1-dimensional water flow model (vertical flux only) assumes that there are no lateral fluxes (in the other two dimensions); this assumption is met in horizontally-homogeneous soils on level landscapes, so if such a model is applied the modeller must prove that these conditions are met.

Note that it is not sufficient to describe a method; the justification for selecting it is also needed.

Consider the following description of a sampling scheme, adapted from the thesis of Fekerte Arega Yitagesu [10], "Spectroscopy to derive engineering parameters of expansive soils":

> "Sampling sites were selected by stratified random sampling. The geographic strata were based on information from previous studies by the Geological Survey of Ethiopia on the expansion potential of the soils. The first stratum is the CMC and Bole area, where frequent problems due to expanding soils have been reported. These were also observed in the fieldwork: e.g. deformation of road pavements, cracking of foundation slabs and walls of houses. The second stratum is the Kotebe and Ferensay area where such problems are rarely reported. The third stratum is a small area near the international airport where extreme problems have been reported."

Q23 : *Why is there no citation for the term "stratified random sampling"? Should a citation be given?* *Jump to A23*
•

Q24 : *Why does the author choose a stratified sampling scheme?* *Jump to A24* •

Q25 : *Is the justification for this scheme thorough and convincing?* *Jump to A25* •

Q26 : *Are alternate schemes discussed?* *Jump to A26* •

Sequence of methods

The order in which methods are described should be logical; often they follow the time sequence of the research. This makes it easy for the reader to understand how the research was carried out.

In the case of a project with fieldwork, a typical breakdown is:

1. pre-fieldwork;
2. fieldwork;
3. post-fieldwork.

These are broken down further by activity. For example Fekerte [10] classifies methods as follows[2]:

1. Field data collection
   (a) Sampling scheme
   (b) Site description procedure
   (c) Soil sampling procedure
2. Laboratory analysis
   (a) Atterberg limits
   (b) Free swell tests
   (c) Cation exchange capacity determination
   (d) Spectral measurements
      i. The ASD field spectrometer
      ii. The PIMA field spectrometer
      iii. Measurement of soil reflectance

Q27 : *What is the reason that field data collection methods are presented before laboratory analysis methods?* *Jump to A27*
•

[2] Slightly adapted

Q28 : *In the field methods, is the sequence (1) sampling scheme; (2) site description procedure; (3) soil sampling procedure appropriate?* *Jump to A28* •

In the case of a project without fieldwork, the sequence is typically one of **dependence**: which steps must be performed before others. For system design (see §3.1.11), this might be:

1. System specification methods;
2. System design methods;
3. System implementation methods;
4. System evaluation methods.

3.1.9 The "research" thesis

The research thesis structure is applicable to both designed experiments and systematic observations. It has a conventional scientific structure following the deductive-inductive approach. In addition to the common proposal elements listed above (§3.1.1), a research thesis often has the additional element of a study area.

3.1.10 Study area

If the research is carried out in a specific geographic area, the **study area** must be described. This is true for fieldwork, but also for desk studies with secondary data from a specific area (e.g. imagery).

The choice of study area must be justified. Even if the area is assigned and not selected by the student, it still should be appropriate.

- Where is the study area located? Almost always a location map is presented.
- What are its geographic limits?
- Is the entire geographic area included or are only some sub-areas investigated? If sub-areas, a map of these should be presented.
- Why was this area selected? What makes it appropriate for the research problem?
- If sub-areas ("windows") were selected, why were these sub-areas chosen? Are they representative of the whole area? If not, what are their special characteristics?

- What are the characteristics of the study area that are relevant to the problem? For example: demographics, land-use pattern, geology, geomorphology, soils. Note this is related to the previous question: the characteristics make an area suitable.

The justification for selecting a study area has three aspects:

1. **scientific**: the area should be suitable to answer the research questions;
2. **practical**: there should be sufficient secondary data; primary data collection should be feasible; among the issues to consider are access, permissions, transport, language, and security.
3. **social/contextual**: the area should be important to the social problem; e.g. transport planning in a city with known acute mobility problems.

Q29 : *A study area is proposed in a zone of active armed conflict, where the researcher does not speak the local language, where permission to travel in the zone depends on warlords, and where it is illegal for civilians to carry detailed topographic maps or GPS receivers. Should the research proposal be approved?* *Jump to A29* •

Q30 : *A study area is proposed for which there is a large amount of secondary data, but not yet in the possession of ITC or the student. Should the research proposal be approved?* *Jump to A30* •

3.1.11 The "design" thesis

Another type of research is a **design**, of for example a computer program, a user interface, a database structure, or an algorithm. Here the key question is why a design should be considered research and not just a project. This is essentially the difference between engineering "research" and "development". The MSc thesis of Zhang [38] is an example.

A "research"-level design must have:

- A clear **research objective**: defined as results from the proposed project that others can use, including the audience that should be interested in the results.
- **Research questions** that make the objective explicit: if these questions can be answered, the objective has been reached.
- A high level of **innovation**; in particular it must create something really new, or at least a new synthesis;

- It must result in a design that is demonstrably **better** in some sense than the alternatives;
- The thesis must both **define** and **demonstrate** this superiority.

The **hypothesis** of the "research" thesis is then replaced with a statement of the proposed **innovation** and evaluation criteria to assess this: in what sense is the new design better than previous designs? This has several aspects:

1. Where is the innovation?
2. What defines "better"?
3. How can this superiority be established?

In a "design" thesis, superiority is often established by a demonstration that certain **design criteria** have been met, which were not met in other products.

Similarly, in the **Results** section of the thesis, the discussion of how well the results support the research hypothesis is replaced by an argument that the design is "better".

An example of a design thesis project is a proposal for a new structure of a soil geographic database. Here "better" could be defined as "allows the representation of real-world objects that can not be represented in any existing design" or "supports a class of queries that can not be carried out in any existing design". The thesis would have to:

1. Establish that there is a demand for a design;
2. Review existing designs and identify their shortcomings;
3. Show the proposed design and its **innovations**;
4. Show how it is used on some sample data, i.e. a **proof-of-concept**, the design really can represent what was promised;
5. Show that it can represent concepts that are impossible with existing designs.
6. Show that this improved design is useful for answering a richer class of questions; for example, the database user can easily extract parameters for a defined class of models.

The "demand" for a design replaces the "research problem" of the research thesis. The question is still why anyone should bother to undertake this work.

Another example of a "design" thesis project is the design and implementation of an improved user interface for statistical modelling. Here, "improved" could be tested by a series of designed experiments with target users, with measurable outcomes, e.g. how quickly or correctly users could accomplish a particular task.

3.1.12 Example of a design thesis

The concepts of a design thesis are illustrated by a PhD thesis [29] which included the design and implementation of the *ALES* microcomputer program to assist in land evaluation [30, 32].

This thesis has sections on:

1. System demand;
2. System objectives;
3. System requirements;
4. System design;
5. System implementation;
6. System application to the problem field.

These are argued as follows:

Demand: Here the **demand for a system** is established. In this example we find sentences such as "There is today a high demand worldwide for information on the suitability of land for a wide range of land uses" (**social** problem) and "There are no comprehensive computer programs that allow the land evaluator to organise knowledge from diverse sources . . . " (**demand** in the narrow sense; identifies potential users).

Objectives: Here the **general and specific objectives of the system** (i.e. why it is being built) are presented. In this example the stated objective is ". . . to allow land evaluators to collate, systematise, and interpret this diverse information using the basic principles of the FAO's *Framework for Land Evaluation* [8], and to present the interpreted information in a form that is directly useful to land-use planners"; this is then expanded to discuss the target group and the type of models that should be automated.

Requirements: Here the specifics of **what the system must be able to do** are listed. In this example, the system should allow expert judgement on the interactions between land characteristics to be captured in the system.

Design: Here the **innovative aspects of the system design** are explained. In this example, it is proposed to represent the interactions by decision trees; the question is then how to represent these in the system.

Implementation — This section contains an **explanation of how the system is implemented**. By itself it is only a project, not research; however as part of the wider discussion, and if justified, it becomes part of research. In this example, the selected computer language and database system are explained, and the program control flow is presented.

Application — This section **demonstrates that the system can meet the stated requirements**. In this example, several models were built in the system, some as duplication of existing manual methods [16, 34, 35] to show that the system could replace these, and some that were beyond the capabilities of existing methods. The proof of the system was that these models could be built, they could be applied to data, and results (in this case land evaluation tables) could be produced.

3.1.13 The "social" or "organizational" thesis

Another class of thesis project is a social analysis, i.e. the study of humans and human societies or their organizations. These are complicated and even contradictory objects of study, and it is notoriously difficult to come to firm conclusions. Also, ethical and practical considerations make designed experiments either impossible or inadvisable. Still, since this species is so powerful, it seems unavoidable that it must sometimes be studied. And, since organizations are the means by which so much is accomplished, they too must be studied.

Here the "hypothesis" takes the same form as a research thesis, but the research **method** is different; in particular the **evidence can be subjective and anecdotal**, rather than the objective result of a measurement. The **Results** section of the thesis then takes the form of a reasoned **argument** from evidence as interpreted by the researcher.

A "social" thesis usually needs a section on **Definitions** or **Concepts**, where terms such as "participatory", "sustainable", "equitable" etc. are well-defined, so that they can be consistently identified in the research. A good example is the paper by Roling [28] on concepts of sustainability.

Social

A typical ITC example is the hypothesis that "participatory" land-use planning is "more successful than" top-down or technocratic approaches. Such a thesis also must clearly address the concept of "better" (as in the "design" thesis): what defines "better", and how can this be established?

The “social” thesis project may include some structured interviews or meetings, but these are much less controlled because of the unfortunate tendencies of human beings (both researchers and subjects) to distortion, fabrication, imagination, wishful thinking, etc. Social scientists have developed a range of techniques for increased objectivity, which should be used if possible (e.g. questions that ask for what should be the same information in different ways).

Organizational

Geoinformation technology is one pillar of ITC’s mission, but this technology is almost always applied in an organizational context. Examples are so-called “e-Governance”. Controlled experiments are almost impossible in a “living” organization, so the research project often takes the form of an extended argument from diverse evidence; a good example is by Bekkers & Homburg [1].

3.1.14 The “modelling” thesis

Another type of research is where the researcher builds a conceptual or (more commonly) computational **model** of a process. Models are very important in management applications. For example, models of river basin hydrology are used to predict floods, droughts, and navigable periods, and to plan release and storage in reservoirs. Spatial models of soil erosion [e.g. 17] are used to plan soil conservation measures and design sediment controls.

These models are evaluated by their success in **reproducing the behaviour** of the natural or social system. Key issues in a modelling thesis are therefore **calibration** and **validation** [31, 36].

However, real field data with which to calibrate and validate may be sparse, and may not represent the full range of the phenomenon being modelled. For example, an erosion model should predict soil loss and runoff for small, medium and large storms; during a given field season there may be no large storms. So, **synthetic** data (“pseudo-data”) can be manufactured to represent the full range of conditions under which the model should perform. An important evaluation is then the results of a **sensitivity analysis**: variation in model output as model input is varied.

Empirical (statistical) models are also valuable tools for estimating complex or expensive-to-measure properties from simpler or cheaper-to-measure properties [e.g. 24]. These models are evaluated by their success in predicting the complex properties from the simpler ones.

Models should also be assessed by comparison to existing models: it should be possible to show that the new or modified model is an **im-**

provement, in some sense to defined by the author, compared to existing models. The model could give more accurate or precise predictions, require less or less-expensive input data, be easier to parameterize, be applicable in a wider range of scenarios, etc. Standard datasets from well-studied sites are often used to compare models.

Models should also be assessed by their contribution to problem-solving; in particular, why is the proposed model better than existing solutions? This is generally argued by comparing the performance of models in several case studies.

3.1.15 Answers to self-test questions

A*15* : *Although results are not known, they should be anticipated; this is the basis of the hypotheses (§3.1.6) which are part of the proposal. Thinking about what the thesis might look like once results are obtained can help focus thinking on questions, hypotheses and methods.* *Return to Q15* •

A*16* :

1. *"Wind erosion is causing widespread destruction of crop land and pastures in the rift valley of Kenya.": this is an* ***environmental*** *problem, also with* ***social*** *implications;*
2. *"We do not know the priority areas for intervention.": this is a* ***management*** *problem;*
3. *"It is impractical to monitor wind erosion over large areas by ground survey or conventional aerial photography.": this is a* ***technical*** *problem.*

Return to Q16 •

A*17* :

1. *"Wind erosion is causing widespread destruction of crop land and pastures in the rift valley of Kenya.": same as previous answer; this is an* ***environmental*** *problem, also with* ***social*** *implications;*
2. *"Previous studies have established the extent of the problem and its historical development.": this is not a problem, but rather an* ***opportunity****, a basis for deeper investigation;*
3. *"We do not know the proximate or ultimate causes.": this is a* ***knowledge*** *problem.*

Return to Q17 •

A*18* : *Only the last statement ("... dynamic visualization methods have not* *yet been integrated with these."); this is what is new.* Return to Q18 •

A*19* : *The second objective (evaluation method) is necessary to satisfy the first (representation method), since the method developed for the first objective must be "effective". There must be some way to evaluate the effectiveness.* Return to Q19 •

A*20* : *A hypothesis must be tested; an assumption is not tested.* Return to Q20 •

A*21* : *The research questions and hypotheses must be changed to include the effect of soil heterogeneity on biodiversity, and the experimental design must be changed include a description of relevant soil factors and how they will be related to biodiversity metrics.* Return to Q21 •

A*22* : *It is not possible to place concepts in a framework until the concepts are clearly defined. This will be done in the task analysis.* Return to Q22 •

A*23* : *The author evidently feels that this term is standard and can be found in many references, further that this term will already be familiar to the intended reader.*

A reference could be added to a standard text, e.g. Cochran [5] or de Gruijter et al. *[13].* Return to Q23 •

A*24* : *Since the purpose of the study is to relate engineering properties of expansive soils to their spectral properties, the author wants to make sure to select some expansive, non-expansive and extremely expansive soils.* Return to Q24 •

A*25* : *Yes, the prior evidence is from a trustworthy source (Geological Survey of Ethiopia) and is the best information prior to fieldwork. There is also justification from the author's own observation during fieldwork that anecdotally confirms the prior evidence.* Return to Q25 •

A*26* : *No other schemes are presented and then discarded. The author evidently feels that the justification for this scheme is sufficient, and does not spend time rejecting others.* Return to Q26 •

A*27* : *It is not possible to understand the laboratory procedures without knowing how the samples for laboratory analysis were obtained.* Return to

Q27 •

A*28* : *This is an appropriate sequence, because it is in time order: first design the sampling scheme, then go to the field and describe each site, then at each site take the soil sample.* *Return to Q28* •

A*29* : *Clearly, the research has almost no chance of success; the proposal should be rejected. This is presented here as an extreme case, but less extreme cases can still present insurmountable problems, so the research must be abandoned.* *Return to Q29* •

A*30* : *This is a potentially disastrous situation. Unless the student can be completely sure that the data will be provided as promised, the proposal can not be accepted.* *Return to Q30* •

3.2 From proposal to thesis

> **Key points**
>
> 1. The thesis goes beyond the thesis proposal: after the research is complete, the author can present **results**, **discussion**, **conclusion**, and **recommendations**;
> 2. The **results** are what was actually observed when methods were applied; the **discussion** places these in scientific context (§3.2.3);
> 3. The **conclusions** present the author's view of the most important findings; the **recommendations** present the author's view on what should be done with the results of the research (§3.2.5).

The research proposal, as discussed in the previous section, has a conventional structure: **Problem** ⇒ **Objectives** ⇒ **Questions** ⇒ **Hypotheses** ⇒ **Methods**

The problem, objectives, questions and hypotheses are usually in one section called **Introduction**. This is usually followed by a **Literature review** and **Methods**.

This is sufficient for the proposal, since no original research has been done at this point. The proposal can be taken over almost completely into the thesis, with appropriate change of verb tense, and modified as necessary to fit what was actually done in the research.

The thesis must then contain four more elements:

> **Results** ⇒ **Discussion** ⇒ **Conclusions** ⇒ **Recommendations**

These elements may be organized in several ways, as now discussed.

3.2.1 Thesis structure

The thesis is the story of a research project. As with all literature, the structure of the writing depends on the most effective way to tell the story.

If the project has one main line, a simple structure suffices:

1. Introduction (problems, objectives, questions, hypotheses)
2. Literature review
3. Study area (if relevant); Data description (if relevant)
4. Methods
5. Results & Discussion

6. Conclusion & Recommendations

In this structure, the Conclusions and Recommendations are often included in a single chapter. Similarly, the Results and Discussion may be (see §3.2.3).

However, other structures are possible and even preferable. The goal is to bring the reader along, from problem through methods to results. If the thesis can be naturally divided into a sequence of **sub-topics** which follow in logical sequence, this can be the basis of an effective thesis structure.

Example 1: Geothermal exploration

This is from the 2006 MSc thesis of Hendro Wibowo: "Spatial Data Analysis and Integration for Regional-Scale Geothermal Prospectivity Mapping, West Java, Indonesia" [37][3].

1. Introduction
2. Geothermal exploration – a review
3. Study area
4. Conceptual model of geothermal prospectivity
5. Analysis of geophysical data for indications of geothermal prospectivity
6. Analysis of Landsat TM data for indications of geothermal prospectivity
7. Regional-scale predictive modelling of geothermal prospectivity
8. Conclusions and recommendations

Q31 : *From the thesis and chapter titles, outline the flow of the "story" of this thesis.* *Jump to A31* •

Example 2: Flood modelling

The 2007 MSc thesis of Saowanee Prachansri [26] is structured as follows:

1. Introduction
2. Research procedure
3. Literature review
4. Study area
5. Soil properties in relation to land use (plot scale)
6. Surface runoff modelling (hillslope scale)
7. Flood modelling (catchment scale)
8. Flood hazard assessment with land use change scenarios
9. Conclusions & Recommendations

[3] This work has since been revised and published as a peer-reviewed journal paper [4]

Note the sequence of scales: plot, hillslope, and catchment. The results of modelling at finer scales are inputs to the model at the next coarser scale. These are then followed by an integrating chapter: flood hazard (catchment) as affected by land use changes (plot), as revealed by the three-step modelling.

Note also the chapter on research procedure, which outlines the models at the three scales and the flow of information between them.

3.2.2 Revising the proposal

A well-designed proposal should, in theory, not have to change during the research, except for verb tense: "will be" ⇒ "were", etc. But of course not everything goes as planned during the research. Some of the proposal may thus have to be revised before beginning the results and discussion.

- The **problem** and **objectives** should not change; that would imply that the original research could not be carried out at all;
- Some **research questions** may have been impossible to address, and so must be removed; during the research others may have suggested themselves and now be possible to answer;
- **Hypotheses** must be added for new questions, but otherwise not modified;
- The **methods** must be as actually applied; it may not have been possible to carry them out as originally planned.

The suggested time to do this is right after fieldwork or attempting to apply the methods from the proposal.

3.2.3 Results & Discussion

Julius Caeser is reputed[4] to have reported on his style of war-making with the three words *Veni, vidi, vici*, i.e. "I came, I saw, I conquered"[5]. We can use these words to describe three steps of reporting on research:

I came — the problem, objectives, questions and **methods** applied to attack them;

I saw — the **results**, what happened when the methods were applied;

[4] Plutarch (75), *Parallel Lives*, "Caesar", translated by John Dryden

[5] "When he gave Amantius, a friend of his at Rome, an account of this action [the battle at Pontus] to express the promptness and rapidity of it, he used words: I came, saw, and conquered, which in Latin, having all the same cadence, carry with them a very suitable air of brevity"

I conquered what these results mean with relation to the research questions; the **discussion** of results.

So, the **results** are what was actually observed when methods were applied; the **discussion** places these in scientific context. These are written in two different styes:

- **Results** are presented **neutrally**: writing style is "reporting";
- The **discussion** is the reasoned **opinion** (or view) of the author: writing style is "argument".

Q32 : *Should a "negative" result be presented? For example, measurements of soil erosion which show no erosion? What then is the role of the discussion?* *Jump to A32* •

Each result should be discussed, with points such as the following:

- What is the interpretation of this result?
- If the result is presented as a figure or table, what is the reader supposed to infer?
- Is the result as expected (hypothesized)? If not, why not?
- Is the result in agreement with previous research? If not, why not? (What makes this case different?)

Figures and tables must be referred to in the text, and then interpreted. For example (see Figure 3.1 and Table 3.1):

> "Model results were compared with measured discharge at the catchment outlet. The simulated and measured hydrographs for three events are shown in Figure 6-3, and their comparative summary statistics in Table 6-8. The model closely fits the peak discharge volume and time, except for the 26-Sept-2005 event, where the predicted peak is too large and early. This is likely due imprecision in the measured hydrograph, due to the sparse recording interval (every three hours)." – [adapted from 26, §6.3].

Q33 : *How much speculation (reasoning based on opinion but not fact) should the discussion contain?* *Jump to A33* •

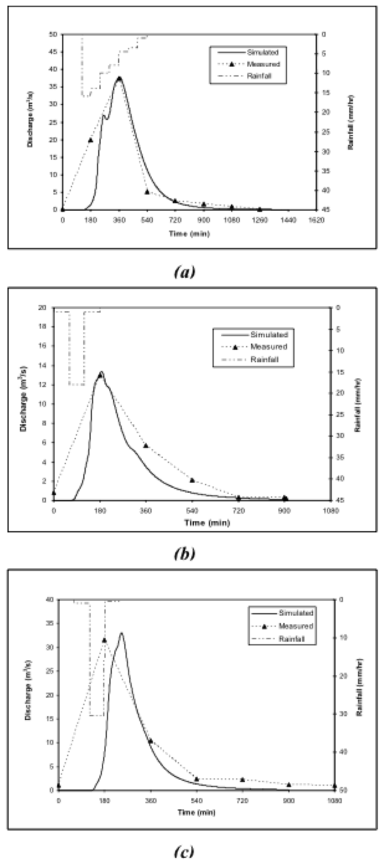

Figure 6-3: Measured and simulated discharge in Nam Chun catchment on 060905(a), 180905(b) and 260905(c) events.

Figure 3.1: [adapted from 26, §6.3]

Events	Rainfall	Peak discharge	
		observed	simulated
(ddmmyy)	(mm)	$m^3 s^{-1}$	$m^3 s^{-1}$
calibration			
060905	52.6	37.5	37.9
180905	18.4	13.0	13.4
260905	29.5	31.5	33.0
validation			
070905	16.6	04.2	04.7
090905	10.6	01.0	01.3
120905	26.7	15.4	23.8

Table 3.1: Observed and simulated peak discharges

Revisiting the literature review

The same literature review used to justify the research is used in discussion and conclusions. Because of the time lapse between the two phases (four to six months), there may well be new literature relevant to the topic.

So before writing the discussion, the literature search should be repeated (using the same search strategy as during the proposal stage[6] and the literature review should be updated with the new references.

Q34 : *What is the role of the literature review in the discussion?* *Jump to A34* •

Organization of the results & discussion

Since in general there are several questions, each with a hypothesis and methods, the results and discussion must be expanded. There are two ways, results followed by discussion, or each result with its own discussion.

Parallel structure

The first structure is **parallel**:

1. Results
   (a) Result for question 1
   (b) Result for question 2
   (c) ...
   (d) Result for question n
2. Discussion
   (a) Discussion of result 1, with respect to question 1
   (b) Discussion of result 2, with respect to question 2
   (c) ...
   (d) Discussion of result n, with respect to question n

This structure was followed in an MSc "Verification of tsunami reconstruction projects by object-oriented building extraction from high resolution satellite imagery" [7], which has a parallel structure in the methods, results and discussion chapters:

1. Methods
   (a) Image to image registration
   (b) Building footprint extraction
   (c) Classification accuracy assessment
   (d) Detection of new buildings
2. Results
   (a) Image to image registration
   (b) Building footprint extraction
   (c) Classification accuracy assessment
   (d) Detection of new buildings

[6] You did use and document a search strategy, didn't you?

3. Discussion
    (a) Building footprint extraction
    (b) Classification accuracy assessment
    (c) Detection of new buildings

Apparently there was nothing to discuss about image registration; the results were considered sufficient without further discussion.

Sequential structure

Some authors prefer to merge the discussion with the relevant results, in a **sequential** structure. This format is then expanded:

1. Results & Discussion
    (a) Result for question 1; discussion with respect to question 1
    (b) Result for question 2; discussion with respect to question 2
    (c) ...
    (d) Result for question nl discussion with respect to question n

---

Q35 : *What is the advantage of the first structure (all results first, then discussion)? What is the advantage of the second structure (each result with its discussion)?* *Jump to A35* •

In the first structure, the discussion can refer to all results; in the second, only to those presented so far. Thus the second structure is preferred for fairly independent research questions.

The chapter headings in the second structure should refer to the subtopic.

Results in a modelling thesis

One result of a modelling thesis is a comprehensive list of the **equations** that underly the proposed model. For example (adapted from [14]):

> "The change in soil moisture storage is calculated according to mass balance as:
>
> $$\frac{dS}{dt} = P_n - ET_a - R_p - E_o(SUST) - Q \tag{3.1}$$
>
> where S is the soil water content $[L]$, defined by volumetric soil moisture content (W) and an effective rooting depth (D) such that $S = WD$. P_n is the net precipitation $[L]$, ET_a is the actual evapotranspiration rate $[LT^{-1}]$ and R_p is the flux below the root zone $[LT^{-1}]$. $E_o(SUST)$ is the evaporated fraction of surface-ponded water $[LT^{-1}]$ and Q is surface runoff $[LT^{-1}]$."

Note that all terms in the equation are defined, along with their dimensions.

3.2.4 Common mistakes in the results & discussion

The results and discussion are where you show what happened when you applied the selected research methods. This section reviews common mistakes in writing these sections. We categorize them as:

- **Under**-interpretation: not getting full value from the results; and
- **Over**-interpretation: making unsubstantiated claims.

Under-interpretation

Results must be interpreted, not just presented. The thesis author must explain what the results mean, in terms of the research questions. In particular, every table and figure must be discussed: what is the reader supposed to understand from it?. This is *not* a repetition of the table or figure contents; the reader can see this for themselves. Rather, it is drawing attention to the most important (remarkable, surprising, . . .) results and explaining them. For example, consider Table 3.2. First, the text must refer to the table and explain what it shows, e.g.,

> "Table 3.2 shows the linear regression coefficients, their standard errors, coefficients of determination, and number of observations, for the prediction of cross-sectional sapwood area from cross-sectional stem area for the nine species."

Then, the contents of the table must be discussed. The wrong way is to repeat the numbers in the table with no interpretation:

Wrong

> "The slope for *Acacia erioloba* was 0.6133 ± 0.0163, with $R^2 = 0.983$ (n=24), for *Acaia fleckii* 0.6757 ± 0.0238, with $R^2 = 0.971$ (n=24) . . . "

Summarizing the table is a bit better:

Better

> "Slopes: ranged from 0.392 (*Dichrostachyus cinerea*) to 0.7692 (*Burkea africana*); standard errors of the slopes from . . . ; R^2 from . . .

Even better is to interpret the table's numbers:

Even better

> "Slopes varied by a factor of almost two, from 0.392 (*Dichrostachyus cinerea*) to 0.7692 (*Burkea africana*); standard errors were all quite small (0.078 to 0.024) relative to the slopes. All models explained almost all the variance ($R^2 > 0.971$)."

Best is to discuss the implications of the table's numbers:

Species	Slope	SE slope	R^2	n
Acacia erioloba	0.6133	0.0163	0.983	24
Acaia fleckii	0.6757	0.0238	0.971	24
Acacia luederitzii	0.615	0.0135	0.99	21
Burkea africana	0.7692	0.0091	0.998	18
Boscia albitrunca	0.7121	0.0268	0.967	24
Dichrostachyus cinerea	0.392	0.0077	0.991	23
Lonchocarpus nelsii	0.7943	0.0132	0.995	18
Ochna pulcra	0.6581	0.011	0.994	23
Terminalia sericea	0.5317	0.009	0.995	18

Table 3.2: Per-species linear models, Sapwood area vs. Stem area

Best

"Slopes varied by a factor of almost two, from 0.392 (*Dichrostachyus cinerea*) to 0.7692 (*Burkea africana*). This large variation is due to the major differences in tree morphology. *Dichrostachyus* species have very thick trunks relative to their height . . . This result clearly shows that these relations must be species-specific."

Note how the author here explains what the table implies in terms of the research questions.

Like tables, all figures should be discussed. The main aim is to draw the reader's attention to the outstanding features shown in the graphic – what should the reader be looking at? Further the figure should be interpreted in two ways: (1) what does this imply about "nature" (the thing being studied)?; (2) what does this imply about the analysis (steps to be followed)?

Here is an example of a figure and its interpretation (emphasized by **bold text**):

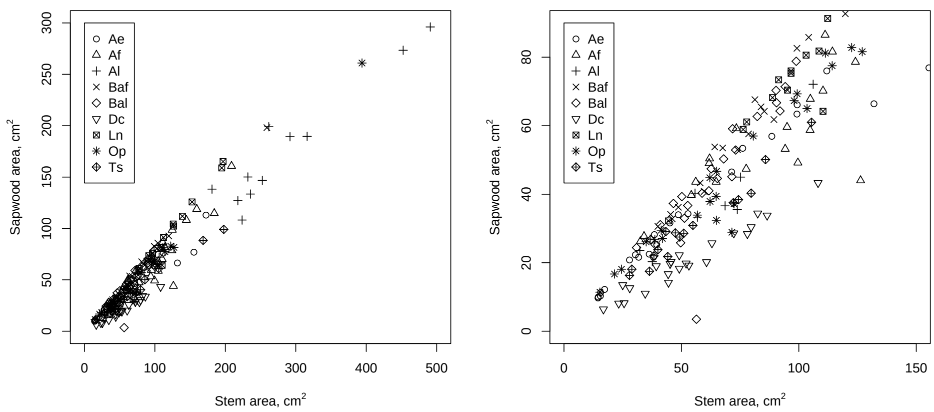

Figure 3.2: Kalahari tree species, sapwood area vs. stem area

> “Figure 3.2 shows the relation between sapwood area and stem area for the nine species. There is generally a linear relation, especially for the smaller trees; however for the largest trees there seems to be a smaller increase in sapwood for a corresponding increase in stem area. Further, the relation for *Dichrostachyus cinerea*) is clearly **anomalous** . . . These discrepancies **can be explained by** . . .
>
> “It is clear that not all species have the same relation, even if we consider only the smaller trees. Thus, **per-species statistical relations must be developed**.”

Over-interpretation

Statements must be supported by your results, possibly in conjunction with results from other studies. For example:

> “Nowadays coastal areas are affected by increasing frequency of extreme events like tsunami, storm surges and cyclones as a result of global climate change.”

In the context of a MSc thesis: (1) where is the proof of “increasing frequency”?; (2) even if this is proven, where is the proof of “as a result of global climate change”?

Note that both of these are very hard to prove, given the short time-series. Are they required by the MSc study?

A better way to introduce such strong statements is with moderation in interpretation: restrict discussion to facts and direct inferences from these. For example:

> “Major tsunamis have affected the . . . coast in 1865, 1920, 1985 and 2007 [reference]. The last-named resulted in . . . deaths and . . . Rp. damage [reference]. As the population in the coastal areas has steadily increased [reference], combined with the national policy on concentrating economic activity in these areas [reference], vulnerability to tsunamis has increased accordingly.”

Facts vs. interpretations

Do not be afraid to interpret, but do not extrapolate beyond what the evidence suggests:

- **Statement of fact**: “The usability test with planning staff in Province X was successful: 80% of the participants (18 of 20) could complete the tasks well within the allotted time”

- **Reasonable interpretation**: "Since this province's planning department was selected as representative (see Methods, §2.2), we expect that similar results would be obtained in other provinces; therefore the system seems ready for country-wide implementation."
- **Excessive interpretation**: "Planning agencies in all Southeast Asia should immediately implement the planning support system developed during this thesis project".

3.2.5 Conclusions & recommendations

The most interesting section of the thesis for many readers is the conclusion. What finally does the author conclude about their work? Further, can the author make any recommendations about how better to address the research question, or what follow-up steps should be taken?

Some authors combine these, because the recommendations flow directly from conclusions; others prefer to present conclusions about the present work, and after that recommendations for future work.

Conclusions

The conclusions are a summary of the results and discussion, without (here) any justification. Readers who want justification will look back into the body of the thesis.

The conclusions refer to the objectives and answer the questions posed in the introduction. It may not be possible to answer all the questions fully; here you can state that, but the reason for this unsatisfactory conclusion is presented in the discussion of the relevant question.

Note that this is not the place to introduce new ideas, let alone results! There should be nothing new in the conclusion, other than the synthesis and direct response to the research questions.

The conclusions should address issues such as these:

- Were the research questions proper and sufficiently specific to be addressed?
- Were the methods applied satisfactory for the purpose of answering the research questions? If not, what should have been done instead?
- Were the data collected sufficient? If not, what additional data should have been collected?
- Was the case study or study area appropriate to answer the questions? If not, what characteristics should have been changed?

- How widely are these results applicable? I.e. how generic are they? If the same methods were applied to other cases, would similar results be expected? Why or why not?
- To what degree do the results answer the question? If not fully, what further information is required to do so?

Recommendations

After spending substantial time with a research topic, the author should have developed some ideas about what should be done next.

- Should any **action** be taken based on the results of this work? For example, should a methodology developed in a research project be operationalized (put into daily practice)? If so, what modifications might be needed, who should do this, etc.
- Does this research suggest **followups**? "We have come this far, the next step is ..." What should be done to overcome any limitations in the present work?
- Or, is the work complete, and the problem **solved**? Then the recommendation is to move on to something else, there is nothing useful to be done here.
- If this or similar work should be **re-done**, what should be changed from the way you went about it? Were there mistakes in planning (e.g. sampling strategy), methods applied, logistics?

A useful way to think about the conclusion and recommendations is with part of the SWOT methodology[7]:

Strengths	What did the research accomplish well?
Weaknesses	What did the research not accomplish so well, or what were its limitations?
Opportunities	What paths does this research open up for us?
Threats	What other approaches could be better to address this problem?

The conclusion and recommendations sections are not written with these "SWOT" headings; the SWOT are listed to help you think about your research in context.

Another way to think about recommendations is to consider these popular thesis defence questions:

- If you had to start this research over again, what would you do differently?

[7] Thanks to Prof. Alfred Stein for this idea

- If you are now given the opportunity to continue working on this research for one (two, three ...) more months , what would you do with the time?

3.2.6 "Selling" your thesis

Of course, the thesis (or a journal article) must be sound science, but it should also convince the reader that:

- the work is important;
- the proper questions have been asked;
- proper methods have been applied;
- you have properly interpreted the results, leading to strong (but not exaggerated!) conclusions;
- your recommendations are supported by the research.

The following points help"sell", i.e., convince the reader of the above points:

- Use a clear **structure**, e.g., outline, paragraphs with topic sentences;
- The **text** should be short and to the point, but without sacrificing relevant detail; this makes it easy to read;
- The language should be clear and concise, using precise statements and modal qualifiers as required;
- Logic and argumentation must be sound;
- Statements should have just the right strength, depending on the facts;
- There must be a clear link to related work, in the introduction, discussion, and conclusion;
  - Why is your work important? What did the others not do?
  - How are your results related to others? Can you confirm, modify or reject their work?

3.2.7 Answers to self-test questions

A*31* : *The thesis aims to map the potential for geothermal energy in west Java (title). The literature review discusses exploration methods which might be applied. The study area is described in a separate chapter; note that any earth science thesis must be based firmly on the geologic reality of a particular study area. A conceptual model of prospectivity is then followed by two chapters of*

analysis by two contrasting methods: geophysical data and Landsat TM data. These are then integrated into an overall regional prediction, Finally, conclusions are drawn. Return to Q31
•

A*32 : Yes, if a method was applied according to the research plan, the results obtained should be reported.*

The discussion should explain why, in the author's opinion, the negative result was observed. In this example, perhaps there were no rainfall events of sufficient size during the observation period. Or, even more difficult to explain, perhaps all the factors that are supposed to lead to soil erosion were present but still there was no erosion. This can lead to a new or modified theory on how soil erosion occurs. Return to Q32 •

A*33 : The discussion can contain ample speculation, as long as it is logically argued and well-related to what is known from this and previous research.*

A good example is the discussion of how generic the applied methods or achieved results are. Clearly, the author has only done the present case study, and must speculate on its extensibility. However, this argument is from the known characteristics of the case study, in comparison with the known characteristics of other cases. Return to Q33
•

A*34 : The literature review gives the scientific context for the question, which led to the hypothesis, the method, and the results. Thus when discussing the results, the relevant literature must be referred to, often directly:*

> *"These results agree with previous studies [1, 4, 12], which found that . . . "*

> *"These results contradict several previous studies [1, 4, 12], which found that . . . The discrepancy can be explained by . . . "*

Return to Q34 •

A*35 : In the first structure the author has no distractions in reporting the results, and lets the reader see clearly what was observed as a result of applying the methods. Discussion is deferred until all results are reported, so that the discussion can merge several results.*

In the second structure the reader sees the author's interpretation of each result directly as it is reported. The result is fresh in the mind of the reader.
Return to Q35 •

3.3 References

Bibliography

[1] Bekkers, V. & Homburg, V. M. F. 2007. *Myths of e-government : looking beyond the assumptions of a new and better government. The Information Society* **23**:373–382. ITC network: http://intranet.itc.nl/papers/ext2000/bekkers_myths_2007.pdf 67

[2] Bishop, Y.; Fienberg, S.; & Holland, P. 1975. *Discrete multivariate analysis: theory and practice.* Cambridge, MA: MIT Press 58

[3] Booth, W. C.; Colomb, G. G.; & Williams, J. M. 1995. *The craft of research.* Chicago: University of Chicago Press

[4] Carranza, E. J. M.; Wibowo, H.; Barritt, S. D.; & Sumintadireja, P. 2008. *Spatial data analysis and integration for regional-scale geothermal potential mapping, West Java, Indonesia. Geothermics* **37**(3):267–299 72

[5] Cochran, W. 1977. *Sampling Techniques.* New York: John Wiley, 3rd edition 58, 69

[6] Davis, J. C. 1986. *Statistics and data analysis in geology.* New York: Wiley 58

[7] Du, Y. 2008. *Verification of tsunami reconstruction projects by object - oriented building extraction from high resolution satellite imagery.* MSc, ITC. http://www.itc.nl/library/papers_2008/msc/gfm/duye.pdf 76

[8] FAO. 1976. *A framework for land evaluation.* Soils Bulletin 32. Rome, Italy: FAO 65

[9] FAO. 2002. *Guidelines for soil profile description.* Rome: Food and Agriculture Organization of the United Nations, 4th edition 58

[10] Fekerte Arega Yitagesu. 2006. *Spectroscopy to derive engineering parameters of expansive soils.* MSc, ITC. http://www.itc.nl/library/papers_2006/msc/ereg/fekerte.pdf 60, 61

[11] Foody, G. M. 2002. *Status of land cover classification accuracy assessment. Remote Sensing of Environment* **80**(1):185–201 58

[12] Galindo De Obario, M. 2005. *Water quality and its spatial variability in Lake Cuitzeo, Mexico.* MSc, International Institute for Geo-information Science & Earth Observation (ITC), Enschede (NL). http://www.itc.nl/library/papers_2005/msc/wrem/galindo.pdf 41

[13] de Gruijter, J.; Brus, D.; Bierkens, M.; & Knotters, M. 2006. *Sampling for Natural Resource Monitoring*. Springer 58, 69

[14] Hiwot Ghiday Gebreyohannes. 2008. *Groundwater recharge modelling: a case study in the central Veluwe, The Netherlands*. MSc, ITC. http://www.itc.nl/library/papers_2008/msc/wrem/hiwot.pdf 77

[15] Hornby, A. S. (ed.). 1995. *Oxford advanced learner's dictionary*. Oxford: Oxford University Press 52

[16] Hudson, N. 1981. *Soil conservation*. Ithaca, NY: Cornell University Press, 2nd edition 66

[17] Jetten, V.; Govers, G.; & Hessel, R. 2003. *Erosion models: Quality of spatial predictions*. *Hydrological Processes* **17**(5):887–900 67

[18] Knuth, D. E. 1997. *The art of computer programming*. Reading, Mass.: Addison-Wesley, 3rd edition 58

[19] Legendre, P. & Legendre, L. 1998. *Numerical ecology*. Oxford: Elsevier Science, 2nd english edition 58

[20] Maidment, D. R. (ed.). 1993. *Handbook of hydrology*. New York: McGraw-Hill 58

[21] Miles, M. B. & Huberman, A. M. 1994. *Qualitative data analysis : an expanded sourcebook*. Thousand Oaks, CA: Sage Publications, 2nd edition 57

[22] Nagelhout, A. 2001. *Performance analysis of small format aerial photography SFAP in assessing current status and trends in wind erosion : a case study in the Longonot – Kijabe hill area, Naivasha district, Kenya*. MSc thesis, International Institute for Aerospace Survey and Earth Sciences, Enschede (NL) 41

[23] National Institute of Standards and Technology. 2003. *NIST/SEMATECH e-Handbook of Statistical Methods*. http://www.itl.nist.gov/div898/handbook/ 58

[24] Pachepsky, Y. A. & Rawls, W. J. 2005. *Development of pedotransfer functions in soil hydrology*. Developments in Soil Science, 30. Elsevier 67

[25] Parsons, T. & Knight, P. G. 1995. *How to do your dissertation in geography and related disciplines*. London: Chapman and Hall

[26] Prachansri, S. 2007. *Analysis of soil and land cover parameters for flood hazard assessment : a case study of the Nam Chun watershed, Phetchabun, Thailand*. Msc, ITC. http://www.itc.nl/library/papers_2007/msc/aes/prachansri.pdf 72, 74, 75

[27] van Reeuwijk, L. P. (ed.). 2002. *Procedures for soil analysis*. ISRIC Technical Paper 9. Wageningen: ISRIC, 6th edition 58

[28] Roling, N. 1997. *The soft side of land: socio-economic sustainability of land use systems*. *ITC Journal* pp. 248–262 66

[29] Rossiter, D. G. 1988. *The Automated Land Evaluation System: a microcomputer program to assist in land evaluation*. PhD, Cornell University 65

[30] Rossiter, D. G. 1990. *ALES: A framework for land evaluation using a microcomputer*. *Soil Use & Management* **6**(1):7–20 65

[31] Rossiter, D. G. 2003. *Biophysical models in land evaluation*. In Verheye, W. (ed.), *Encyclopedia of Life Support Systems (EOLSS)*, volume Article 1.5.27 in Theme 1.5. Oxford: EOLSS Publishers 67

[32] Rossiter, D. G. & Van Wambeke, A. R. 1997. *Automated Land Evaluation System: ALES Version 4.65 User's Manual*. SCAS Teaching Series No. T93-2, Revision 6. Ithaca, NY: Cornell University, Department of Soil, Crop & Atmospheric Sciences, February 1997 edition 65

[33] Ryerson, R. A. & American Society for Photogrammetry and Remote Sensing. 1998. *Manual of remote sensing*. New York: Wiley, 3rd edition 57

[34] Sánchez, P.; Couto, W.; & Buol, S. 1982. *The fertility capability soil classification system: interpretation, applicability and modification*. *Geoderma* **27**(4):283–309 66

[35] Sys, C.; Van Ranst, E.; Debaveye, J.; & Beernaert, F. 1993. *Land evaluation, Part 3 : Crop requirements*. Agricultural Publications 7. Brussels: General Administration for Development Cooperation 66

[36] Tedeschi, L. O. 2006. *Assessment of the adequacy of mathematical models*. *Agricultural Systems* **89**(2-3):225–247 67

[37] Wibowo, H. 2006. *Spatial data analysis and integration for regional scale geothermal prospectivity mapping, West Java, Indonesia*. Msc, ITC. `http://www.itc.nl/library/papers_2006/msc/ereg/wibowo.pdf` 72

[38] Zhang, Q. 2008. *Animated representatation of uncertainty and fuzziness in spatial planning maps*. MSc, ITC. `http://www.itc.nl/library/papers_2008/msc/gfm/qiuju.pdf` 41, 63

4 Frameworks

Key points

1. In the context of MSc research, a **framework** is structure to organize concepts or steps of the research process.
2. A **conceptual** framework is used to organize concepts and show their inter-relation.
3. An **analytical** or **mathematical** framework is the set of defining equations.
4. A **research** framework is used to organize steps of the research and show their inter-relation and dependence; these may contain **feedback** (adaptive steps).

The Oxford English Dictionary [6] defines a *framework* as "A structure composed of parts framed together", for example the framework of a building. By extension it can refer to a mental or conceptual "structure", so that a framework provides an organization and shows the inter-relation between parts.

Frameworks are used to:

- **Limit** what is being discussed (the "universe of discourse")
- **Organize** concepts or steps of a process
- Clarify **relations** between concepts or between steps of a process

The key point is that if some thing or some relation is *not* included in a framework, it can not be discussed or investigated, because the concept or relation is not even identified.

Warning! An **incorrect framework** necessarily leads to **incorrect research**. So, careful definition of frameworks goes a long way to ensuring success of the research.

There is a great deal of confusion and overlap between various kinds of "frameworks" in the literature.

Here we distinguish four kinds of frameworks:

- **conceptual** (§4.1);
- **analytical** (mathematical) (§4.2);

- **process** (§4.3); and
- **research** (§4.4).

In addition there are other uses of the word "framework" that may be encountered, including the European Research Framework Programme and "logical frameworks" for international development projects (§4.6).

4.1 Conceptual frameworks

The word "concept" is defined as "...an idea of a class of objects, a general notion or idea" [6]. Humans organize their world by concepts, because there are too many details to process individually.

Some concepts are fairly **concrete** and easy to define, e.g., *"animal"*, *"map"*, *"motion"*; but even here there may be problems. For example, what exactly is an *"animal"*? Still, most speakers can agree on a definition.

Other concepts are more **abstract** and more difficult to define, for example, *"poverty"*, *"sustainability"*, *"improvement"*. There is more chance of disagreement, so the importance of **precise definition** becomes greater. The definition goes a long way to defining the direction and limits of research.

A **conceptual framework** is a written description or drawing showing how concepts are **linked**: which concepts **influence** which other concepts.

- The conceptual framework **limits** the scope of the research;
- If a concept or link is not in the framework, it can't be studied; it is effectively **invisible**;
- We hope it's also **non-existent** or at least **irrelevant**!

4.1.1 Example: Human Development Index

A simple example of a conceptual framework is the definition of a so-called "Human Development Index" by the UNDP [8], shown graphically in Figure 4.1. This shows the **very abstract** concept "Human Development" is defined by three **abstract** concepts "Health", "Education", and "Living standards"; these are in turn defined by **concrete** (or, specific) concepts that can be (more or less) easily-measured, quantified, and combined into an **indicator**, i.e., the HDI.

Q36 : *What indicators are proposed to measure the concept "health"?*

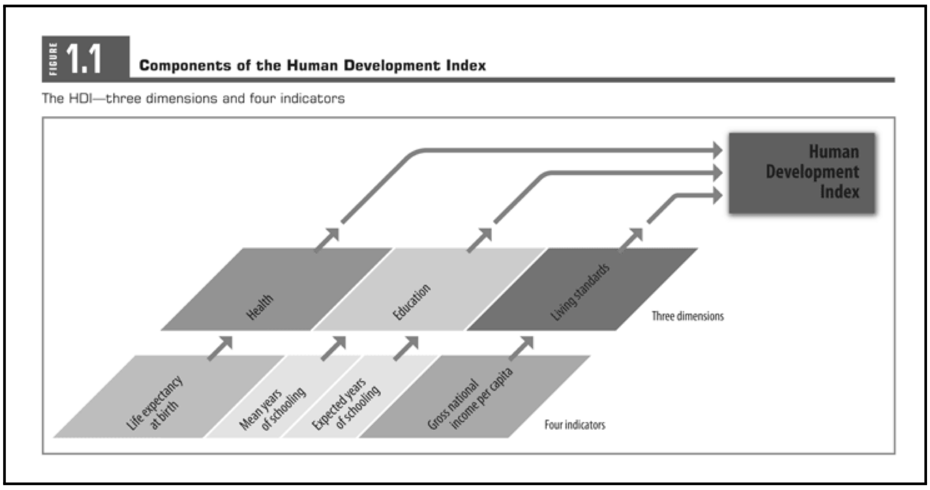

Figure 4.1: Conceptual framework "Human development"

How well does this indicator matches the general concept "human development"? *Jump to A36*
•

4.1.2 Example: Hierarchical definitions

Another example is shown in Figure 4.2. This presents the hierarchical relation between concepts of uncertainty; the various concepts such as "Error" and "Vagueness" are identified and put in relation to the very abstract (top-level) concept "Uncertainty".

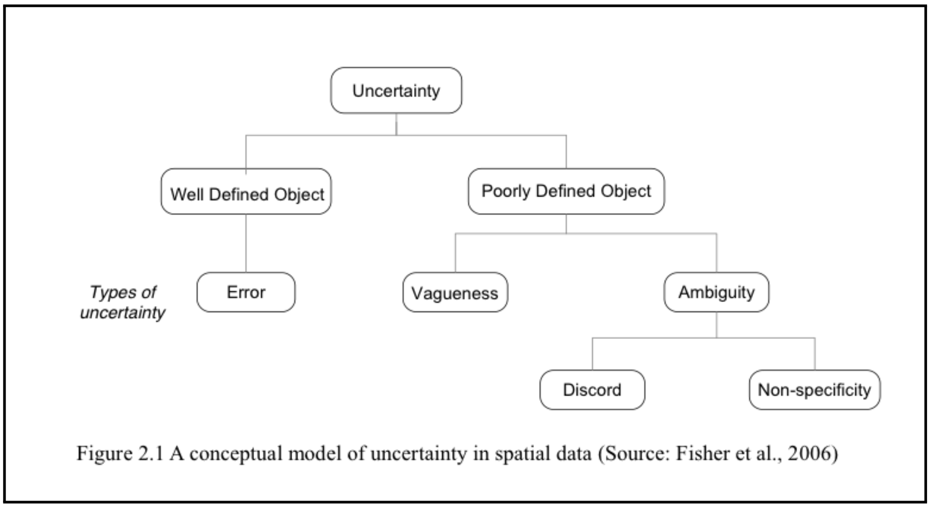

Figure 4.2: Hierarchical concepts (Source: Zhang [10])

Q37 : *According to this conceptual framework, can we speak of* error *in the position of a* poorly-defined object*?* *Jump to A37* •

4.1.3 Example: A complex conceptual framework

Figure 4.3 shows a complex framework.

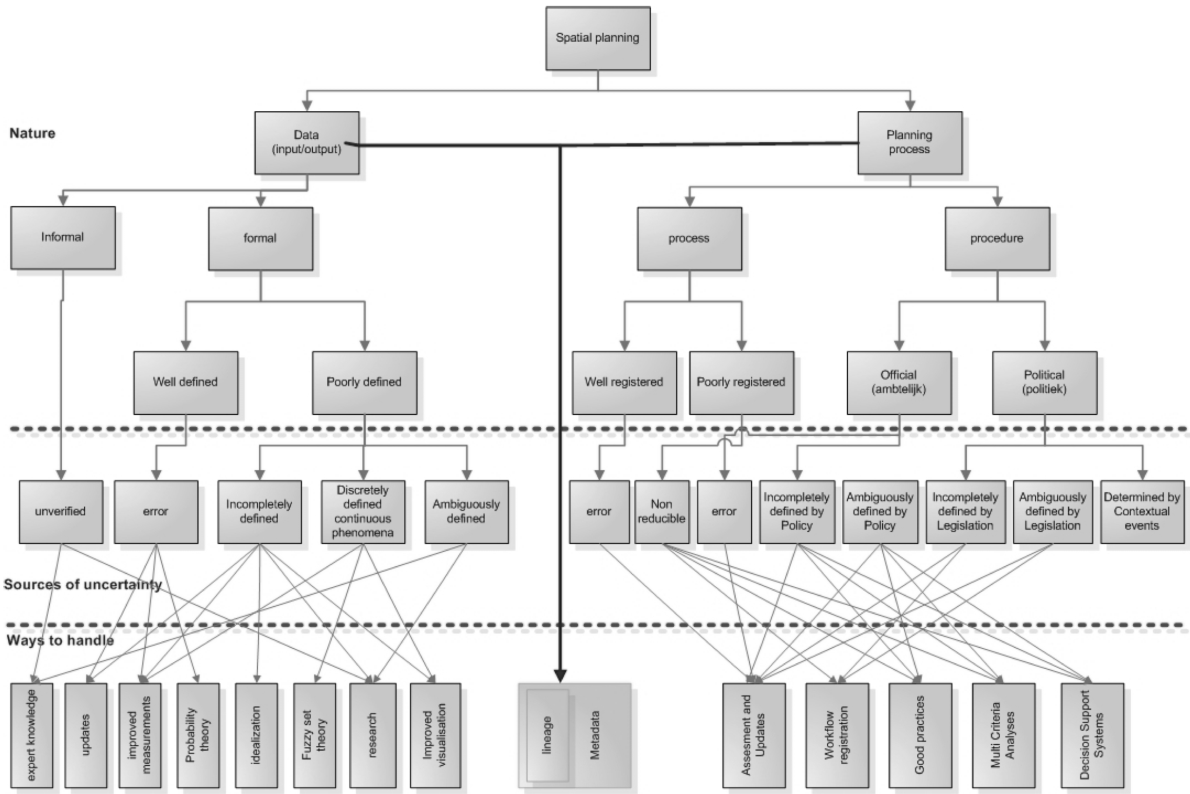

Figure 4.3: Uncertainty in spatial planning (Source: [9, Figure 2])

Things to notice here are:

- Concepts are separated into three sequential "compartments"
  - *nature*;
  - *sources of uncertainty*;
  - *ways to handle* uncertainty.
- Some sources of uncertainty can be handled in **one, many, or even no ways.**

Note: It may be best to break such a complex diagram into a **nested hierarchy** of diagrams

Q38 : *Is the sequence* nature → sources of uncertainty → ways to handle

logical, within the overall concept (root of diagram) spatial planning*?*
Jump to A38 •

4.1.4 Example: Information flow between concepts

Figure 4.4 shows four ways to conceptualize **land use change** ("change" in the figure).

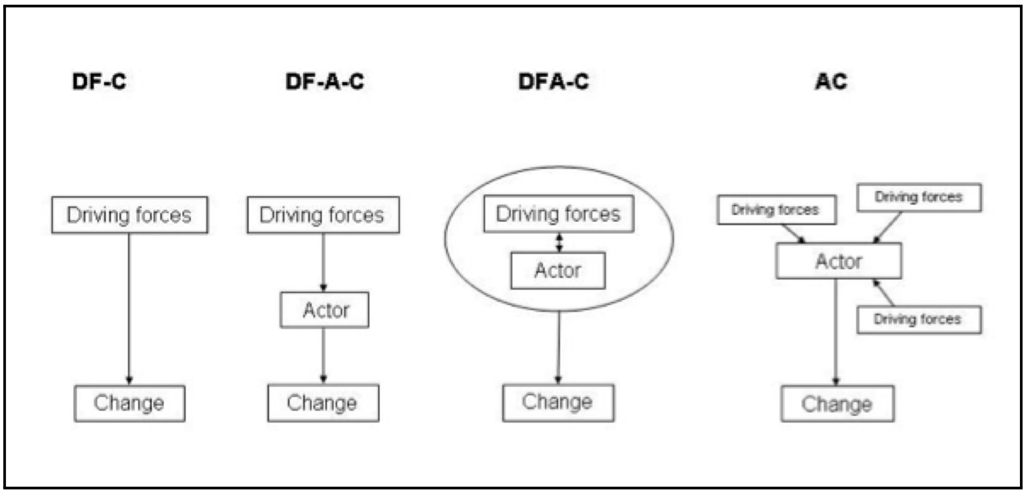

Figure 4.4: Conceptual frameworks for "land use change" (Source: [1, Fig. 1])

Here we see three concepts: *"driving force"*, *"actor"*, *(land use) "change"*; but linked in different ways. These clearly show how **different frameworks** lead to **different research**.

DF-C model: Here, the *"actor"* concept does not appear; *driving forces* operate **directly** (without people as *actors*) to effect *land use change*.

- There may be people, of course, but that concept is not modelled, so can't be studied.
- This could lead to an **empirical statistical model** ("regression") of *land use change* in response to *driving forces*, the **mechanism** is not specified.

DF-A-C model: Here, the *forces* cause the *actor* to effect *land use change*.

- So, ***actors* must be identified**
- The effect of *causes* on *actors* must be specified
- These effects on the *actors* **indirectly** cause *land use change*
- The *actor* is identified and makes **decisions**, but only responding to **one** *driving force*.

The new element here is the *actor*'s **decision making**

AC model: Here, the *actor* is central:

- The *actor* influenced by **several** *driving forces*
- Nothing happens to cause *land use change* until the *actor* does something
- The *actor* **integrates** various *driving forces* and makes decisions.

The new element here is the *actor*'s **integration** of different *driving forces*.

DFA-C model:

- Here there is give-and-take between the *actor* and the *driving forces*
- The *actor* is an **active** participant in the process in **both** directions.
- The new element here is that the *actor* is able to exert some **influence** on a *driving force* (e.g., market prices; but not the weather)
- Now we can model **negotiation** between the *actor* and *driving force*; the ellipse surrounding these two implies repeated **feedback** until the *actor* finally decides to effect some *land use change*

Q39 : *What would be the major advantage and disadvantage of choosing the* ***DF-C*** *model?* *Jump to A39* •

4.1.5 Example: Inter-relation between concepts

Figure 4.5 shows part of a conceptual framework for selecting environmental indicator sets.

Points to notice are:

- There are a large number of **concepts**, e.g., *"water transparency"*, *"fine sediment load"*, *"algae and plant populations"*;
- These concepts must be **defined**, and there must be **operational methods** to **quantify** them;
- The diagram asserts that *"fine sediment load"* and *"algae and plant populations"* **affect** *"water transparency"*;
- This influence must also be **specified** (model form, model parameters . . .)
- There is *not* a causal link back from *"water transparency"* to *"fine sediment load"* – so this can not be accounted for, should it exist.
- Notice that nutrient loads *"P concentration"*, *"N concentration"* do *not* **directly** influence *"water transparency"*; this conceptual model

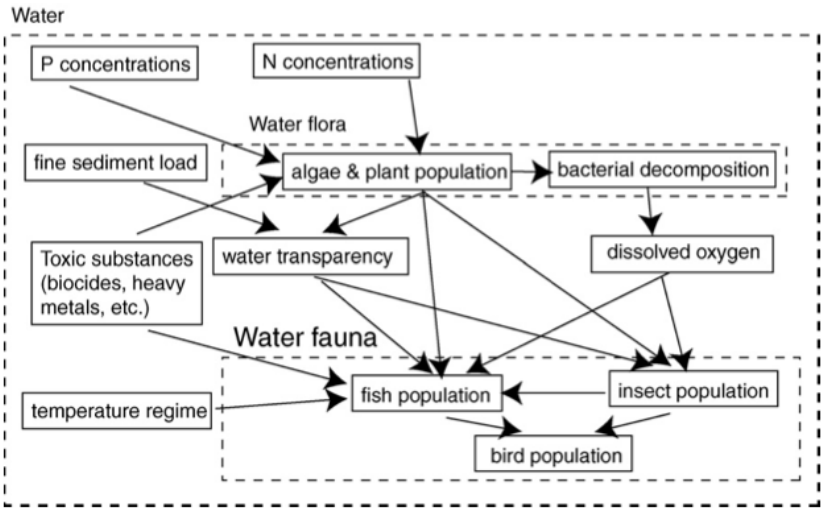

Fig. 3 – The water component of our sample causal network in more detail.

Figure 4.5: Conceptual framework for environmental indicators (Source: [4])

says they affect *"algae and plant populations"* which then affects *"water transparency"*, i.e., an **indirect** influence.

- This **adds complexity** to the model – perhaps one could conceive of *"nutrient concentrations"* directly affecting *"water transparency"*, even though the mechanism is via the *"algae and plants"*.

The main point is that choices made explicit in the conceptual framework can not be repaired – the research can not investigate what it doesn't conceptualize.

Q40 : *Why is there a dashed line around the concept* water fauna*?* *Jump to A40* •

4.1.6 Example: A dynamic concept

The final example is a dynamic (time-varying) concept, shown in Figure 4.6. Points to notice are:

- All the **terms** in the diagram, e.g., *reaction time*, *relaxation time*, *recurrence interval*, must be **defined**
- These must also enter into the **model** of system behaviour
- If this dynamic framework is **wrong** (e.g., if the system does not tend to an equilibrium after disturbances) the research is **invalid**

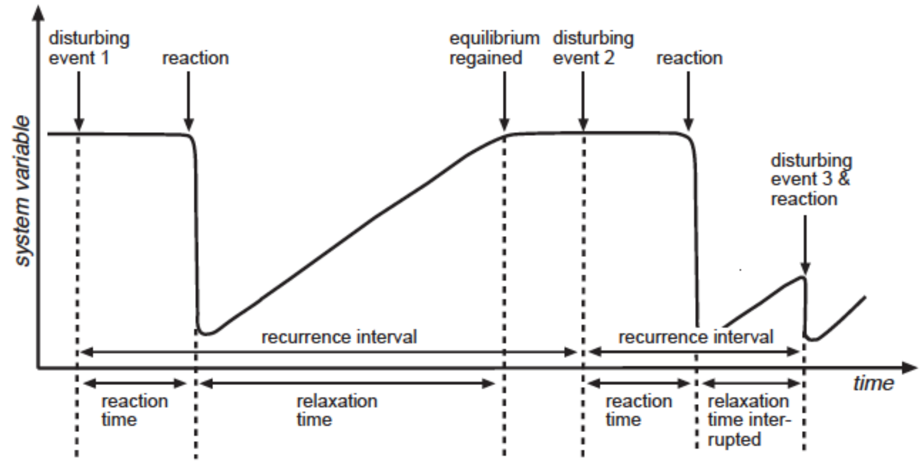

Fig. 1. Reaction and relaxation time after Graf (1988) and Brunsden and Thornes (1979).

Figure 4.6: Reaction and relaxation time (Source: [2])

4.2 Analytical frameworks

This limits the approach and defines the terms to be used in analysis, often of a system. It is often presented as a set of equations, and may be called a **mathematical framework**. These frameworks are common when a set of equations largely defines a problem and approach.

Figure 4.7 is an example.

Combining (1) with the continuity relation

$$\theta_s \frac{\partial s}{\partial t} = -\frac{\partial q}{\partial z} \qquad (2)$$

yields Richards' equation, which, after adding a sink term U representing water uptake by plants, is written as

$$\theta_s \frac{\partial s}{\partial t} = \frac{\partial}{\partial z}\left(k(s)\left[\frac{\partial \psi}{\partial z} + 1\right]\right) - U(s) \qquad (3)$$

where it is assumed that water uptake U is a function of s only.

Figure 4.7: Part of a mathematical framework (Source: [3])

The equations represents a **conceptual model** of reality:

- The terms of the equation, e.g. θ_s, s, q, z, t, U must be defined, along with **units of measure** – they are **symbolic** representations of mathematical **concepts**;

- These have **real-world** counterparts, e.g. $U(s)$ represents water uptake by plants;
- The **form** of the equation is the **concept** of how process operate in the real world,
  - e.g. $\theta_s \cdot \partial s/\partial t$ is the change in the water content of the soil with time, which (according to the equation) can be expressed as a proportion of the saturated water content;
  - The reader can decide if this in fact a reasonable representation of how the process really works, or if a different formulation (analytical framework) is required.

Equations can also be presented **graphically**, as in Figure 4.8.

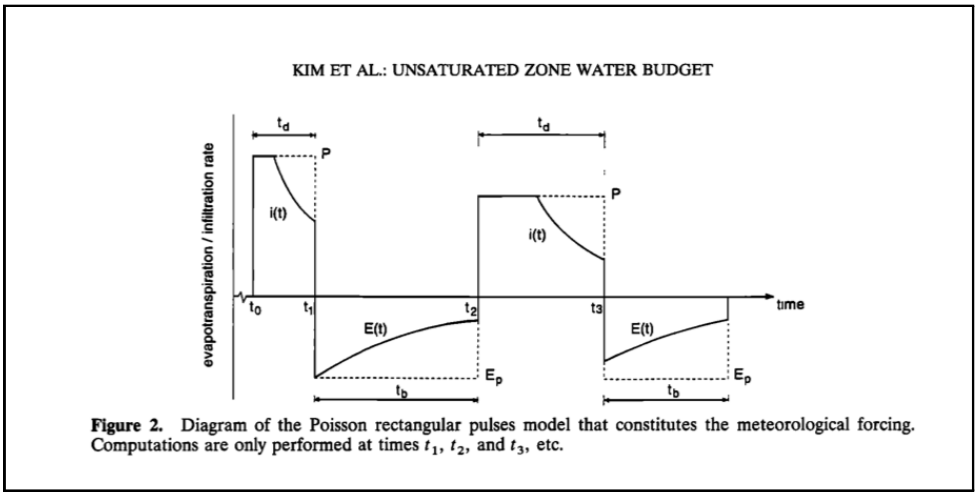

Figure 4.8: A graphic representation of equations (Source: [3, Fig. 2])

4.3 Process frameworks

These show the steps in a process or procedure, their order, and their dependence relations. They break down a complex process into manageable steps. These are not much used in research, rather in the description of a process to be carried out. Figure 4.9 gives an example of the process of risk management.

Notice in this figure:

- This shows a hierarchical (nested) processes, with three outer levels;
- “Management” is considered to contain “Assessment”, which in turn is considered to contain “Analysis”. In other words, the overall

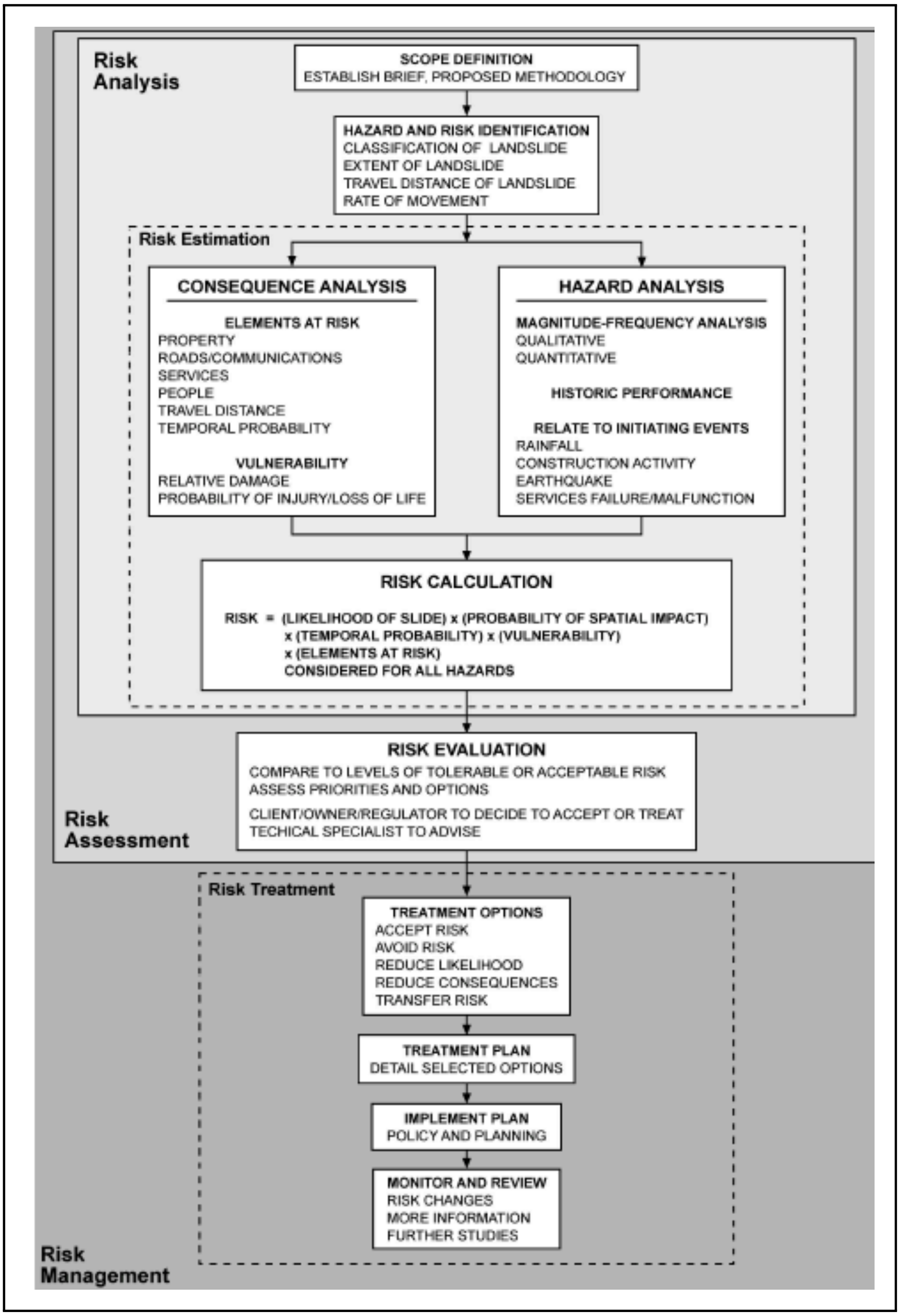

Figure 4.9: A process framework for risk management (Source: [2, Fig. 3])

process of risk management depends on a sub-process of risk assessment, which in turn depends on a sub-process of risk analysis;

- This implies: no analysis means no assessment, without assessment it is impossible to manage;
- Some process can run in parallel, for example in risk estimation, the processes "consequence analysis" and "hazard analysis".

Q41 : *What are the two processes that can run in parallel as part of the "Risk estimation" process? What are the three methods that could be used for "Hazard analysis"?* *Jump to A41* •

4.4 Research frameworks

These identify the components of the research and the flow of information between them:

- What needs to be done?
- What parts depend on what other parts?
- In which order are parts to be done?
- What external information is needed, and in which part of the research chain?

The framework can be presented graphically (as a flow diagram) and/or as structured text. A diagram gives a quick overview, text allows for more detail. Figure 4.10 is an example of a general MSc research framework.

Notice in this framework: (1) the sequence of major activities is shown by the flow; (2) tasks within each major activity are all of equal importance, here there is no sequence or hierarchy; (3) there is an adaptive step (which can be repeated), which is is emphasized by the feedback arrow.

Q42 : *What is the adaptive step?* *Jump to A42* •

Figure 4.11 shows the steps of a moderately complicated process.

Notice the decision points, where the later stages of the research will change according to the results of earlier steps. These are associated with adaptive feedback steps

---

Q43 : *What is the only output of the LISEM modelling step?* *Jump to A43* •

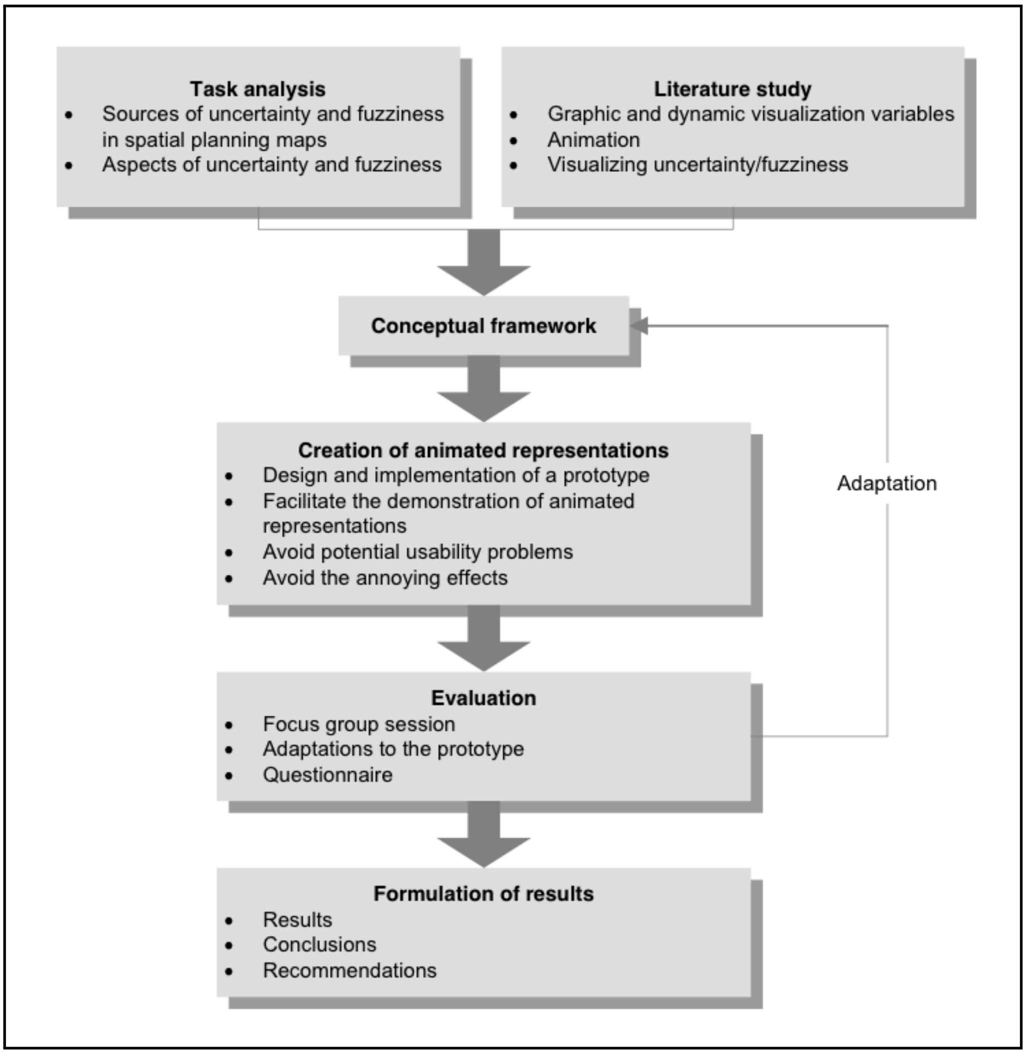

Figure 4.10: Example of a general MSc research framework (Source: [10, Fig. 1.1])

4.5 Use of frameworks in MSc research

There are several reasons to use frameworks in defining and reporting MSc research.

In a general sense, working with frameworks matches well with the structured approach to research; they are "frames" within which various concepts or steps can be developed. Specifically:

1. They help limit (circumscribe) the research; this simplifies the work and makes it feasible;
2. They force the author to clearly define all **concepts**;
3. A **conceptual framework** shows their inter-relation and information flows in a

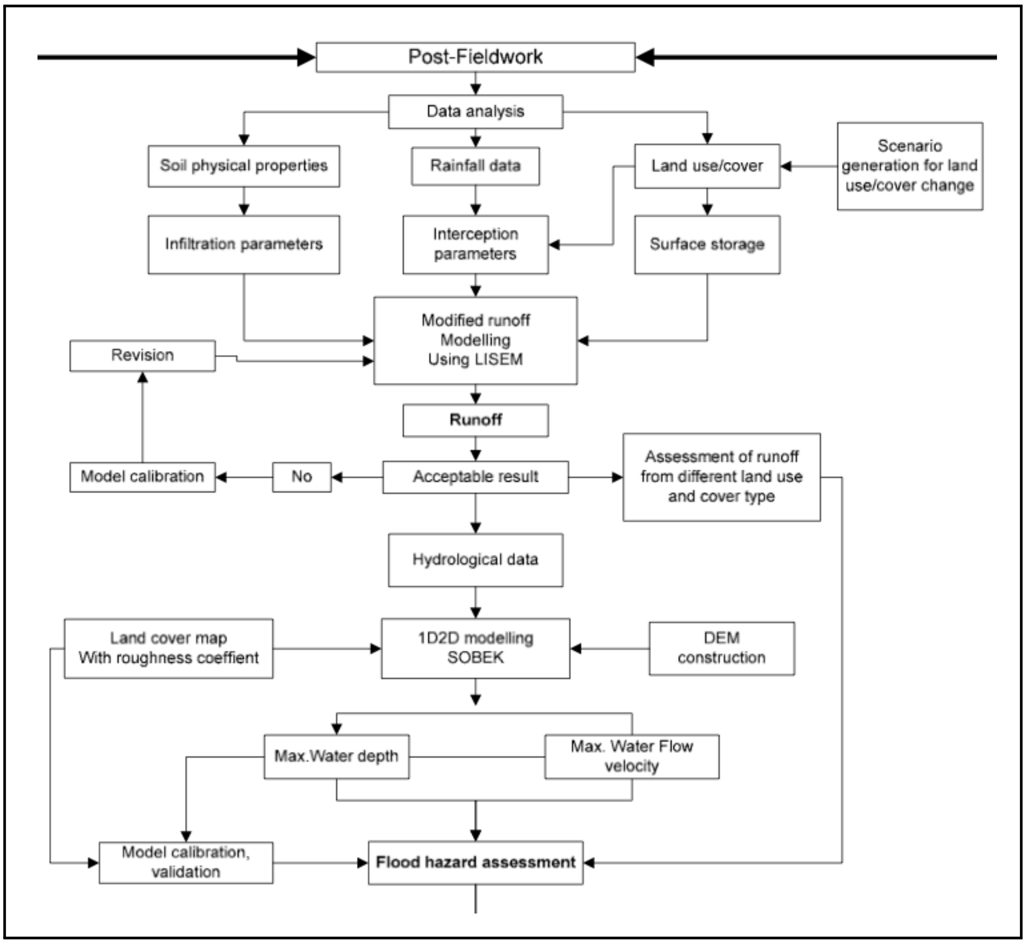

Figure 4.11: Example of a specific part of an MSc research framework (Source: [7])

4. A **research framework** shows the research plan as actions, required external information, information flow between stages, and time sequence;

5. If equations are used, they form an **analytical** (or, mathematical) framework

6. If the research studies processes, or produces a process as an output, these can be presented as a **process** framework.

Frameworks can be presented several ways in the proposal and thesis: as a diagram and/or text.

A **diagram** or **flowchart** is useful to help the reader quickly see the concepts/steps and their inter-relation. A diagram is not required, but inability to make one usually reveals the author's confusion about exactly what they intend to do or have done.

Each concept, step, and link must be further described in text in sufficient detail, as for any aspect of method description.

4.6 Other "frameworks"

4.6.1 EU Research Framework

A very common use of the term "research framework" within Europe refers to the "Framework Programmes for Research and Technological Development", abbreviation **FP**. A better term would have been "European Union Research Programme" It is the main mechanism by which the EU funds research. Each Framework Programme is a political compromise listing objectives, calls for proposals, and countries or regions where research must be carried out. The FP7 (2007-2013) budget is $\approx$ €50×10^9.

4.6.2 Logical framework

The so-called "Logical framework" (often called a "log-frame") is used for objectives-oriented planning. It is a tool used to design, monitor and evaluate projects [5], and was developed especially for international development work, where it is often required.

A logical framework uses a temporal logic (sequential) model of activities and outcomes

- "If these Activities are implemented, and these Assumptions hold, then these Outputs will be delivered"
- "If these Outputs are delivered, and these Assumptions hold, then this Purpose will be achieved"
- "If this Purpose is achieved, and these Assumptions hold, then this Goal will be achieved"

4.6.3 Answers to self-test questions

A*36* : *According to the diagram, only* life expectancy at birth *is used to measure the concept of* health *in the context of* human development*. This seems to be a very narrow idea of health, as it ignores the health status (sickly, healthy) of living persons.* *Return to Q36* •

A*37* : *No, according to this framework* error *is only for objects that are well-defined, only then can we say for sure that they are mis-placed on a map, for example.* *Return to Q37* •

A*38* : *Yes, it is impossible to* handle *uncertainty without knowing its* sources, *and these depend on their* nature *(type); this is the only possible logical flow between these three abstract concepts.* *Return to Q38* •

A*39 : The advantage is simplicity; we just need to find empirical relations between some factors and how land use changed. The disadvantage is that this is difficult to interpret: since always there are people (*actors*) who actually change the land use, in the* ***DF-C*** *model we have no way to know how the driver was translated to change.* *Return to Q39* •

A*40 : The concept* water fauna *is made up of three concepts* fish, insect, birds, *and a self-contained trophic network: birds eat fish and insects, fish eat insects. This can be modelled separately, given values of all the "outside" concepts – note that arrows only come into this box (e.g., from* temperature*); once these are known as boundary conditions, a model of trophic (feeding) relations could be built separately.* *Return to Q40* •

A*41 : "Consequence analysis" and "Hazard analysis" can be done in parallel; the latter has three methods: "Magnitude-frequency analysis", "historic performance", and "relate to initiating events". The diagram is too vague on how these are carried out, and how the keywords such as "rainfall" are to be interpreted; the text should explain the process steps in detail.* *Return to Q41*
•

A*42 : Depending on the* Evaluation *of the animated representation by focus groups and a questionnaire, the author may go back and modify the* Conceptual framework, *which then requires the* Creation *of a modified animated representation and another* Evaluation. *This is an example of a* ***prototype-test-evaluate*** *cycle common in software design.* *Return to Q42*
•

A*43 : Runoff. Note that LISEM can also model other hydrological outputs, but they are not used in this research.* *Return to Q43* •

4.7 References

Bibliography

[1] Hersperger, A. M.; Gennaio, M.; Verburg, P.; & Bürgi, M. 2010. *Linking land change with driving forces and actors: Four conceptual models. Ecology and Society* **15**(4):1 92

[2] Hufschmidt, G.; Crozier, M.; & Glade, T. 2005. *Evolution of natural risk: research framework and perspectives. Natural Hazards and Earth System Sciences* **5**(3):375–387 95, 97

[3] Kim, C. P.; Stricker, J. N. M.; & Torfs, P. J. J. F. 1996. *An analytical framework for the water budget of the unsaturated zone. Water Resources Research* **32**(12):3475–3484 95, 96

[4] Niemeijer, D. & de Groot, R. S. 2008. *A conceptual framework for selecting environmental indicator sets. Ecological Indicators* **8**(1):14–25 94

[5] NORAD. 1992. *The Logical framework approach (LFA): handbook for objectives-oriented planning.* NORAD. ISBN 9788275480819 URL http://www.norad.no/en/Tools+and+publications/Publications/Publication+Page?key=109408 101

[6] Oxford University Press. 2011. *Oxford English Dictionary.* URL: http://www.oed.com/, accessed 2013-April-10 88, 89

[7] Prachansri, S. 2007. *Analysis of soil and land cover parameters for flood hazard assessment : a case study of the Nam Chun watershed, Phetchabun, Thailand.* Msc, ITC. http://www.itc.nl/library/papers_2007/msc/aes/prachansri.pdf 100

[8] United Nations Development Programme (UNDP). 2010. *Human development index.* URL: http://hdr.undp.org/en/statistics/hdi/, accessed 2013-April-10 89

[9] Vullings, W.; de Vries, M.; & de Borman, L. 2007. *Dealing with uncertainty in spatial planning.* In *10th AGILE International Conference on Geographic Information Science 2007* 91

[10] Zhang, Q. 2008. *Animated representatation of uncertainty and fuzziness in spatial planning maps.* MSc, ITC. http://www.itc.nl/library/papers_2008/msc/gfm/qiuju.pdf 90, 99

5 Argumentation

Key points

1. To **argue** a point is to maintain its truth by reasoned debate, leading to a decision; this includes but is not limited to strict logic;
2. Argumentation often follows a **stereotypical structure** of claim, evidence, warrant and backing (§5.1);
3. Argumentation has several **styles**, including definitions, cause and effect, contributions and impacts, and analogy (§5.2);
4. Argumentation must be free of **flaws** (§5.3): material, verbal and logical.

Argumentation may be defined [5] as "methodical reasoning; debate". So, to argue a point (or position, assertion) is to maintain its truth by reasoned debate. But, what is then "reasoned" debate? What sorts of "reasonable" arguments are valid? How do we persuade others of our claims?

Argumentation it is not about "winning an argument", and certainly not or "arguing" (≈ "fighting verbally"). Rather, argumentation is about reasoning towards the best approximation to the truth. The usual aim to is take some sort of action based on the results of the argument, So, argumentation is a constructive debate to reach a solution [6].

There are cultural and social differences in acceptable argument in daily life. A good example of this is the weight given to the **argument from authority**:

In an authoritarian culture we might hear:

> ‘ I'm your boss, I'm telling you that this method is correct, so use it. ’

In a "flat" culture such as the Netherlands we might hear:

> ‘ I know I'm your boss, and I prefer this method, but if you have good evidence that your method is better, let's try your way. ’

However, in scientific argument there should be one standard: the closest approximation to the truth.

A good web resource for definitions and explanations related to argumentation is Straker [4].

5.1 Elements of an argument

Argumentation is often based on deductive and inductive logic; this is covered in §5.3.3. Certainly, a valid argument must be logical; but pure logic is not enough to make a good argument.

Argument can employ less rigorous methods. For example, it can be based on **weight of evidence** and human intuitions of **likelihood**. The aim of an argument is to build sufficient evidence for the claim in order to make a decision.

The English philosopher Toulmin [6] developed a stereotypical **structure** for argument:

1. a **claim** to be established; a proposition that you are trying to convince the listener to **accept** (agree to its truth) and, in general, **act on** (take some action because of that).

   The person proposing the claim can be referred to as the **claimant**.

2. **evidence** (also called "grounds" or "data") to support the claim.

   Note that the evidence must be accepted as true by the listener, otherwise it becomes another claim to be established, before we can continue with the present claim. So, the evidence is accepted, but then there must be some link between evidence and claim; otherwise the evidence is **irrelevant** to the claim.

3. a **warrant** or justification: the "since ..." which provides the link from data to claim. The warrant explains **why** the evidence presented implies that the claim is true.

   Warrants may be based on several ways of reasoning (Greek terms due to Aristotle):

   - *logos*: logical reasoning;
   - *ethos*: trust in the speaker, especially their reputation and credibility;

     > "My previous research projects were successful, so you should give me the resources to do this newly-proposed project."

Here the author makes no reference to what is contained in the new project and why it is important and feasible, only that the researcher's reputation is good.

- *pathos*: emotions, especially in relation to the listener's values.

  > "Agriculture is the soul of a country and must not be ignored in setting research priorities."

  Note that one could argue with logic about the importance of agriculture, but here an emotional appeal is made[1].

  Note: In a **research proposal**, *logos* is the main method of scientific argument. *Ethos* is also present, increasing the reader's trust and confidence that the researcher can carry out the proposed research by; *ethos* is improved by a well-crafted, carefully-argued proposal showing evidence of a thorough literature search and understanding of the topic. *Pathos* has no place in a scientific document; a research proposal or thesis must be based primarily on evidence, not emotion. Pathos may be an important part of policy documents or political arguments.

4. a **backing** that provides the context (not to be argued); this is often difficult to state precisely, and includes the entire context of the argument.

5. an optional **modal qualifier** that limits the extent of the claim; this can make it easier to accept because the claim is not exaggerated.

6. an optional **rebuttal** of anticipated counter-arguments; this preempts the listener's possible objections and strengthens the impression of the claimant's reliability.

   A rebuttal is a sub-claim and as such can have the same structure as the original claim.

This structure can be visualized in Figure 5.1. The "claim" in this diagram will often have a modal qualifier to limit its scope.

5.1.1 Example argument

1. **Claim** (what we want to convince the reader): "Crop yield forecasting using models coupled to daily weather satellite observations should be operationalised.";

2. **Evidence** (facts the reader will agree with):

   - "Current yield forecasting methods give poor results."
   - "Research results with these methods have shown good ability to predict yields."

[1] This is still the basis of much of European agricultural policy.

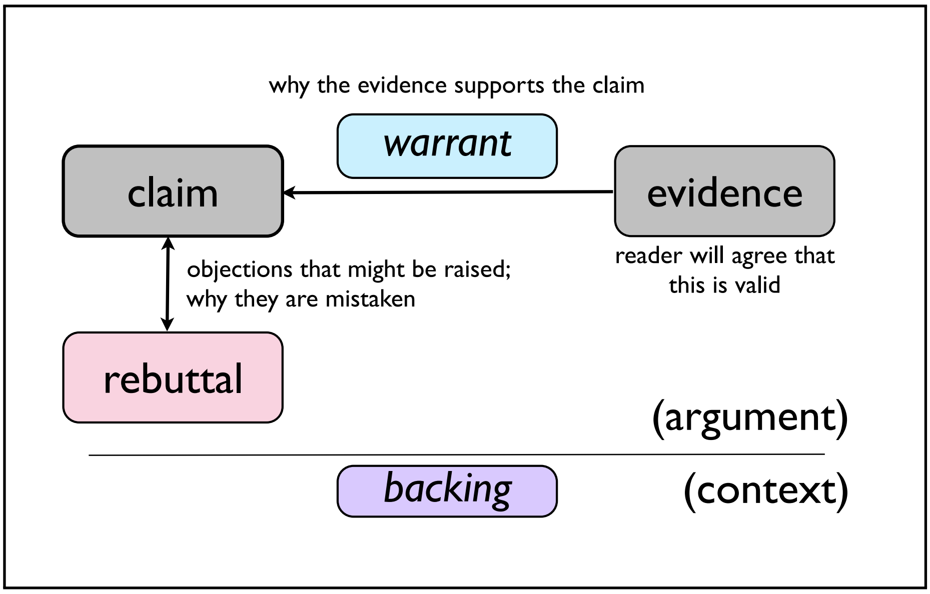

Figure 5.1: Argument structure

- "Import/export policy for staple crops has been erratic and not in line with actual supply."

3. **Warrant** (link between evidence and claim):
   - "Accurate yield forecasts are necessary for efficient agricultural markets and sound import/export policy."
   - "Proposed methods can provide these."
4. **Backing** (unstated assumptions behind the warrant):

   The entire context to the agricultural sector and its economy; the entire context of current remote sensing and information technology; for example "Efficient agricultural markets are desirable for society; improved weather satellites will continue to be developed, independently of agricultural applications; computers will continue to become cheaper and more powerful".

   Note that this sort of statement is usually not explicit, it is part of the shared knowledge of author and reader.
5. **Modal qualifier** (limiting the claim): "However, if yields are affected by extreme weather such as typhoons, yield forecasts from any method will be gross over-estimates."
6. **Rebuttal**:
   - Anticipated **counterargument**: Costs will be too high for routine use.

- Rebuttal: "Costs of acquiring daily weather images are very low and processing can be done on any desktop computer with free software."

An example of irrelevant evidence here is:

2. "New York City has the largest population of any US city."

This is certainly true, but what is the relevance to crop yield forecasting?

Now that this argument has been made explicit, we can look for flaws in the argument:

1. Is the reasoning (logic) correct as such? (see §5.3.3)
2. Is the evidence correct? Is it complete?
3. Is the warrant a sufficient justification?
4. Is the backing true, and does it contain all the relevant information?
5. Does the modal qualifier (if present) increase confidence in the main argument?
6. Does the rebuttal (if present) strengthen the argument?

Q44 : *Assuming the evidence given above is true (i.e. "Research results with these methods have shown good ability to predict yields"), is this by itself sufficient to establish the claim? If not, what else is needed?* *Jump to A44* •

Q45 : *What is the role of the backing in this argument?* *Jump to A45* •

Q46 : *How does this modal qualifier help the listener accept the argument?* *Jump to A46*
•

The critical examination of the argument to find weak points then leads to counter-arguments, which should be formulated as successive approximations to a final correct statement.

5.1.2 Extending the argument

The "since ..." **warrant** between **claim** and **evidence** may be accepted by some readers, but others may not have the background to understand it, or may object to the direct statement.

For example,

Claim — "In this study, soil loss does not vary linearly with rainfall intensity."

Evidence — A graph showing soil loss vs. rainfall intensity, from a group of experiments; linear regression diagnostic graphs.

Warrant — The graph shows a non-linear relation, and the regression diagnostics are show that the assumptions of a linear model are violated.

This argument would be easily accepted by a reader with a background in linear regression modelling. However, a reader without this background would not understand the link (warrant) between the graph and "non-linear". This reader would need the warrant itself to be proven:

Claim — (taken from the **warrant** of the higher-level argument): The graph shows a non-linear relation, and the regression diagnostics are show that the assumptions of a linear model are violated.

Evidence — Non-linear and linear relations and their regression diagnostics; or a theoretical development.

Warrant — Textbook discussion of the assumptions of linearity in regression models – this could be a citation.

So the **warrant** in one argument becomes a **claim** in a **sub**-argument. This is shown graphically in Figure 5.2.

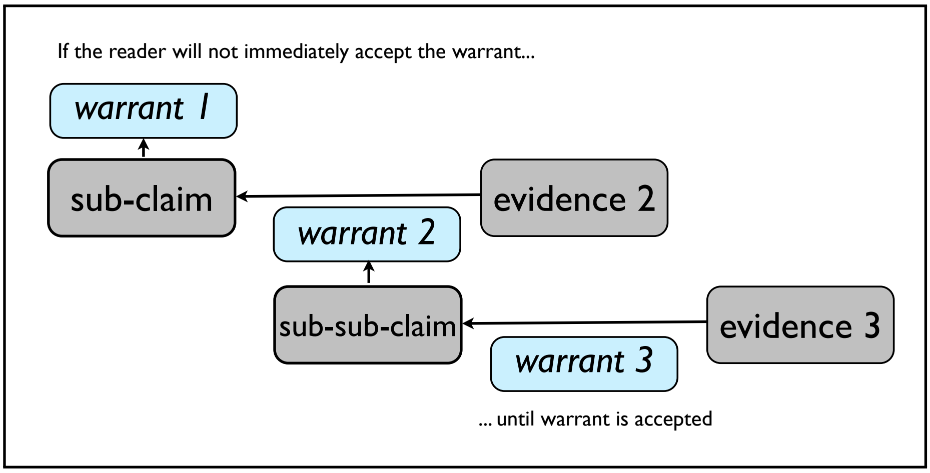

Figure 5.2: Extending the argument

5.2 Argumentation styles

There are many ways to argue a point. All require correct internal logic, but differ in how they present evidence and warrants.

1. From **definitions**, "define the problem away"; not too useful but may set up a more focused argument;

For example, the claim "Vegetarians are healthier than carnivores" can be defined away by narrowly defining what is meant by "healthier" to ensure that the available evidence supports the claim.

"A healthy person is one who does not eat animal protein. So vegetarians are healthier than carnivores."

2. From **cause and effect**, but these may be difficult to separate, and to distinguish from mere correlation; Note that these must occur in a time sequence (see next item), but a stronger argument (about processes) is needed;
3. From **time sequence**, a weaker form of cause and effect, evidence is that one thing always happens before another; beware of the logical fault of *post hoc ergo propter hoc* (§5.3).
4. From **contributions and impacts**, a weaker form of cause and effect, listing a number of contributing factors and observed results;
5. By **analogy** or **comparison** with similar cases; must establish similar context (geographic, social, environmental . . .) for the analogy to be valid; the argument must clearly state what is different in this case, and how it affects the argument;

5.2.1 Argument by analogy

Here is an example of argument by analogy:

1. Claim: "Community forestry (CF) should be introduced in [name your country]";
2. Evidence: Success of CF in Nepal;
3. Warrant: "What works there should work here";
4. Backing (implicit): "there are no relevant differences in society or environment between [here] and Nepal".

Putting it this way, it is clear that the backing is false. However, this provides a way to sharpen the argument, by identifying the relevant differences and modifying the argument to account for them and making the backing explicit. Some of the differences in this case might be:

- Social structure
- Administrative structure (government as a whole, forest sector)
- Infrastructure
- Economic, educational level, other social indicators
- Religion, beliefs

These differences can be handled in two ways:

- The differences can be identified but then "argued away", arguing that they don't have any relevance to CF;
- The claim can be modified; rather than adopt CF in its Nepalese form, modify it for local conditions. Claim: "CF as practised in Nepal, but with [list the modifications here], should be introduced ..."

5.3 Flawed argument

Argumentation depends on sound logic. This is a vast subject which has been a major branch of philosophy since Aristotle. Here we only consider some common flaws in reasoning.

Fowler [1] lists three classes of flawed argument:

1. **Material**: mis-statement of facts;
2. **Verbal**: wrong use of words;
3. **Logical** (formal): process of inference.

5.3.1 Material flaws

A material flaw is a mis-statement of fact.

Clearly, material flaws will generally lead to incorrect conclusions. If I incorrectly state that the sun rises in the west, I can draw many incorrect conclusions by sound logic, e.g. I should sleep in a room facing east to avoid the morning sun.

It is not always so clear that a stated fact is not true. There may be limited circumstances when it is false, even if it's generally true. Careful use of qualifiers may be required.

5.3.2 Verbal flaws

Fallacy of ambiguity

The most common verbal flaw is using a word in more than one sense in the same argument. This is the fallacy of **ambiguity** [2, p. 182], also called fallacy of **equivocation** (using words "equivocally") [1]. For example:

> ' Models have been used by engineers for many years to investigate the behaviour of full-scale systems before they are built. A well-known example is scale models of aircraft in wind tunnels. Therefore, a geo-database model of the soil-landscape of the is an appropriate method to investigate soil conservation practices. '

In the first part of the argument, “model” refers to “scale models”, i.e. physical analogues, while in the second part, it refers to “conceptual models” as in database design. There is no logical link, so the “therefore” is not valid.

5.3.3 Logic flaws

Logic deals only with the way **antecedent** statements and **inferential premises** are related to each other to reach conclusions, also called **consequents**. Logic does not deal with the truth of the antecedents or premises; those are material flaws (§5.3.1). Logic also does not deal with verbal flaws (§5.3.2); in the following we assume words are used consistently within an argument.

Humans seem to have a lot of trouble arguing with correct logic. Because of this, many logic flaws have been identified and analyzed, and some have received colourful names to help you remember them.

5.3.4 Invalid inferences

Hugh Gauch, in his thought-provoking text “Scientific method in practice” [2] identifies two invalid inferences that, superficially, resemble the correct application of *modus ponens*.

Modus ponens

> *Modus ponens*: If both fact A and the implication $A \to B$ are correct, we can conclude fact B.

Gauch gives the following example of a correct application of *modus ponens*:

1. Premise: implication $A \to B$: “If plants lack sufficient nitrogen, they become yellowish”;
2. Antecedent: fact A: “These plants lack sufficient nitrogen”;
3. Consequent: conclude B: “These plants become yellowish”.

This has the valid logical structure $p \longrightarrow q; p; \therefore q$.

Note that the correctness of the consequent depends on the truth of both the premise and the antecedent. The premise is an example of a typical scientific principle, established by experiment.

Another correct inference is *modus tolens*.

Modus tolens

> *Modus tolens*: If implication $A \to B$ is correct, and consequent B is false, then the antecedent A must also be false, because if it were true, so would be the consequent.

Continuing Gauch's example:

1. Premise: implication $A \to B$: "If plants lack sufficient nitrogen, they become yellowish";
2. Consequent: assert $\neg B$: "These plants are not yellowish";
3. Antecedent: conclude $\neg A$: "These plants do not lack sufficient nitrogen".

This has the valid logical structure $p \longrightarrow q; \neg q; \therefore \neg p$.

But consider these variants:

Asserting the consequent

First, the logical flaw "asserting the consequent":

1. Premise: implication $A \to B$: "If plants lack sufficient nitrogen, they become yellowish";
2. Consequent: assert B: "These plants are yellowish".
3. Antecedent: conclude A: "These plants lack sufficient nitrogen";

This has the invalid logical structure $p \longrightarrow q; q; \therefore p$.

Q47 : *What is the logical flaw here? Assuming both the premise and the consequent are true, why can't we assert the antecedent?* *Jump to A47* •

Denying the antecedent

Second, the logical flaw "denying the antecedent":

1. Premise: implication $A \to B$: "If plants lack sufficient nitrogen, they become yellowish";
2. Antecedent: assert $\neg A$: "These plants do not lack sufficient nitrogen";
3. Consequent: conclude B: "These plants will not become yellow".

This has the invalid logical structure $p \longrightarrow q; \neg p; \therefore \neg q$.

Q48 : *What is the logical flaw here? Assuming both the premise is true and the antecedent is false, why can't we assert that the consequent is also false?* *Jump to A48* •

5.3.5 *Post-hoc* reasoning

This is also called *post hoc ergo propter hoc*, which is Latin for "after this therefore because of this". If one event follows another (in time), the fallacy is concluding that the first event caused the second. Of course, this could be true, but more evidence than just time sequence is needed.

Compare these two arguments:

> ‘ The LANDSAT-4 earth observation satellite was launched in 1982, ten years after the original LANDSAT-1. The success of the LANDSAT-4 platform, in particular the TM sensor, would not have been possible without the experiences gained with the earlier LANDSAT platforms and their MSS sensors. ’

> ‘ In 1972 the USA sent the last two manned missions (Apollo 16 and 17) to the Moon. In 1982 the Falklands (Malvinas) War was fought between Argentina and the UK. The war was one of the many consequences of the moon landings. ’

Q49 : *Both of these arguments argue from time sequence 1972 → 1982. What is the difference between them? Which seems more valid?* *Jump to A49* •

5.3.6 Spurious correlations

Two things may be correlated ("co-related"), in that they occur together, but there is no connection between them, even no lurking variables (§5.3.8). The association could just be by chance. More evidence of correlation is needed than just common occurrence.

For example:

> ‘ In the past forty years the world population has increased steadily, while the earth's orbital period continues to slow slightly due to tidal friction. The statistical correlation between the two trends is quite strong. ’

Q50 : *Can we draw any valid conclusions from this strong correlation? Why or why not?* *Jump to A50* •

5.3.7 Correlation vs. causation

Two things may be correlated, but whether the first is a cause of the other, or vice-versa, or neither, must be argued from evidence beyond the correlation. Of course, without the correlation, causation is not possible; it is a pre-condition.

For example, a person's height and weight are positively-correlated: the taller a person is, the heavier they tend to be. Does the height cause the weight? Does the weight cause the height?

Causation must be argued from a time or process sequence. For example, if there is a correlation between surface soil compaction and the weight of tillage equipment used on fields, the compaction must be caused by the equipment, not the other way around. Physically, equipment presses on soil, but it is difficult to see how a compact soil could increase the weight of the equipment.

Q51 : *Identify the logical flaw in this argument. How can we decide if this flaw invalidates the argument?* *Jump to A51* •

Two correlated things may be caused by a third, as we now see.

5.3.8 Lurking variables

A common cause of correlation that is not the result of one of the two things causing the other is when a third factor explains both. This is called a **lurking variable**; it is "lurking" in the background, an explanation that is not clear until we go look for it.

The classic example is the correlation between the number of houses of worship (churches, synagogues, temples, mosques etc.) in a city and the number of violent crimes committed in that city in a given time span. In general, the more houses of worship, the more crime.

Q52 : *What is the lurking variable in this example?* *Jump to A52* •

Q53 : *Could you argue that the presence of houses of worship leads to crime? Could you argue the reverse?* *Jump to A53* •

5.3.9 Circular reasoning

This is also known as "begging the question" [2, p. 184]. In its obvious form it is easy to spot:

> ' Soil maps are important for land-use planning because soil maps are important for land-use planning. '

But a longer argument may disguise the circular reasoning:

> ‘ Soil maps are used in many planning offices. Planning offices are responsible for land-use planning. Soil maps are therefore important for land-use planning. ’

5.3.10 False dilemmas

This is when the author states that there are several options, all except one are false, so the remaining option must be correct [2, p. 183]. But, if there are other options, not mentioned by the author, this conclusion is false.

> ‘ Estimates of crop area can be obtained by manual interpretation of air photos or by supervised classification of multi-spectral satellite imagery. Manual interpretation takes too much time, so we choose the second option in this study. ’

Q54 : *What is the false dilemma set up here?* *Jump to A54* •

5.3.11 Fallacy of composition

This fallacy applies the properties of constituents to the whole. For example:

> ‘ Individual buildings can not be identified on 30 m resolution satellite imagery. Urban areas are composed of many individual buildings. Therefore, urban areas can not be identified on 30 m resolution satellite imagery. ’

Q55 : *What is the fallacy here?* *Jump to A55* •

5.3.12 Fallacy of division

This fallacy applies the properties of a whole to its parts. A general statement about the whole may not be true about the parts of which it is composed. For example:

> ‘ Per-capita income in India is much lower than per-capita income in Germany. Therefore every (or, a specific) Indian is poorer than every (or, a specific) German. ’

Q56 : *What is the fallacy here?* *Jump to A56* •

5.3.13 *Ad-hominem* argument

This is arguing by attacking the person who hold a contrary opinion, or associating the contrary opinion with an undesirable person.

> ‘ David Rossiter coordinated a proposal for urban soil classification [3]. He is an American, so this proposal should not be adopted. ’

An obviously undesirable trait of the author is being used to discredit the proposal.

Sometimes this is extended by a false logic:

> ‘ Stalin was an atheist; Stalin was evil; therefore atheism is evil. ’

Q57 : *What is the logical flaw here?* *Jump to A57* •

5.4 Answers to self-test questions

A44 : No, the evidence is not enough. What is still missing is the warrant, i.e. the "why" for the claim. In this example, even if the method is good, there is no reason to operationalised it unless there is some need, here the stated need for reliable information for efficient agricultural markets. *Return to Q44* •

A45 : The backing here is the implicit structure of the agricultural sector. For example, in a non-market economy, there would be no need for yield forecasting except on each farm separately. *Return to Q45*
•

A46 : The claimant is being clear on when the proposed method will not work, but also arguing that not only this method, but any other method, will not work in extreme circumstances. The listener feels that the claimant has understood the limitations, and that makes the listener more inclined to accept the argument. *Return to Q46* •

A47 : We are ***affirming the consequent*** *that the plants are yellow, rather than arguing it from the antecedent and premise. And then we use that consequent to conclude that the plants lack nitrogen. But there could be many other reasons for yellowish plant, for example a viral infection; the premise does* not *assert that lack of N is the* only *reason for yellowing.*

The farmer who makes this logical error will apply N to correct the yellowing; but this may not be the cause, so the problem may not be fixed and money may be wasted. Return to Q47 •

A*48* : *We are* ***denying the antecedent*** *that the plants lack nitrogen, but that does not imply that the plants can't turn yellow for some other reason (e.g. viral infection). As in the previous question, the premise does* not *assert that lack of N is the* only *reason for yellowing.* Return to Q48 •

A*49* : *In the case of the LANDSAT sensors, an additional link is made: the time sequence also represents an experience sequence. In the case of the moon landings and the war, it is unclear how the first could influence the second. Thus the first time sequence is a better argument.* Return to Q49 •

A*50* : *There is no physical relation at all between these, so we can't draw any conclusions. We can see this by imagining: what if a disease or war wiped out half of the earth's population? Would the orbital period be affected one way or the other? It seems unlikely.* Return to Q50 •

A*51* : *Yes. Perhaps the farmers take the amount of soil compaction into account when deciding how heavy the equipment for tillage must be. We should determine whether farmers take any account of soil compaction when making this decision, or if the equipment is chosen for other, independent, reasons.* Return to Q51 •

A*52* : *Population of the city: there is more of everything.* Return to Q52 •

A*53* : *There can be some interesting attempts to explain these, none very convincing.* Return to Q53 •

A*54* : *(Air photos and manual interpretation) vs. (multi-spectral imagery and supervised classification). It is also possible to manually interpret imagery or to used supervised classification on air photos. There are also other options not mentioned, such as object-based classification. So the argument for the chosen method is incomplete.* Return to Q54 •

A*55* : *In urban areas, buildings occur together in large groups, so that areas of 30 m by 30 m generally have an identifiable spectral signature even though many buildings may be located in a pixel.* Return to Q55 •

A*56* : *Per-capita income is an average applied to a whole, it says nothing about* *individuals.* *Return to Q56* •

A57 : *This is an example of the fallacy of composition (§5.3.11): the properties of one member of a group (Stalin as a representative of atheists) are extrapolated to the whole group. But the main purpose of this argument is to associate an undoubtably evil person (Stalin) with the position (atheism) which the author wishes to discredit. The fallacy is not important to the author; he just wants to juxtapose the words 'Stalin' and 'atheism' to cause an emotional reaction in the reader.* *Return to Q57* •

5.5 References

Bibliography

[1] Fowler, H. W. & Gowers, E. 1965. *A dictionary of modern English usage*. Oxford,: Clarendon Press, 2d edition 111

[2] Gauch Jr., H. G. 2002. *Scientific method in practice*. Cambridge: Cambridge University Press 111, 112, 115, 116

[3] Rossiter, D. G. 2007. *Classification of urban and industrial soils in the World Reference Base for Soil Resources*. *Journal of Soils and Sediments* **7**(2):96–100 117

[4] Straker, D. 2008. *Changing minds and persuasion – how we change what others think, feel, believe and do*. On-line document; URL: http://changingminds.org/. Accessed 2013-April-2013 105

[5] Sykes, J. (ed.). 1983. *The concise Oxford dictionary of current English*. Oxford, England: Clarendon Press, 7th edition 104

[6] Toulmin, S. E. 2003. *The uses of argument*. Cambridge: Cambridge University Press, updated edition 104, 105

6 Ethics & professionalism in science

In this chapter we examine the concepts of ethical behaviour as it applies to working scientists, as well as their professional obligations.

Key points

1. Scientific **ethics** are **rules of conduct** for carrying out scientific work (§6.1).
2. **Fraud** is any action which wilfully mis-represents the truth (§6.2); it has thee forms: **fabrication** (§6.2.1), **falsification** (§6.2.2), and **plagiarism** (§6.2.3).
3. **Plagiarism** is knowingly representing the work of others as one's own; this includes **text**, whether directly copied or paraphrased, **data** and **ideas**.
4. A **simple rule to avoid written plagiarism**: Everything you write outside of quotation marks must be the result of your own creative effort.
5. **Intellectual property** is any product of a creative effort; it may be protected by copyright, which must allow **fair use**, e.g. for comparison; other uses usually require **license agreements** (§6.4).
6. **Professionalism** refers to scientists' behaviour towards the society in which they live (§6.5).
7. Research is embedded into the wider **social context**; the scientist must make ethical decisions about choices of topics and their effect on society (§6.6).
8. The scientist is a **social animal** (§6.7), and has biases, values, and social context. These must be identified and accounted for.
9. Relations between researchers and their **human subjects** and **local populations** are subject to difficult ethical decisions (§6.7.2).

6.1 Ethics

In general, **ethics** refers to correct behaviour within some social setting.

In the narrow context of scientific procedure, ‘ethics’ refers to the **rules of conduct**: what is permitted. These rules have evolved along with science, both from more general codes of ethics such as religious value systems but also to aid scientific progress. The idea is that ethical behaviour isn’t just “right” in some abstract sense, but also that it ensures good science. It also ensures that scientists are properly rewarded for their work.

In the wider sense, ethics also includes the relation between researchers and society as well as the relation between researchers and research subjects or colleagues; these are explored in §6.6 and §6.7.2, respectively.

Scientific ethics in the narrow sense (internal to the scientific community) is organised around two main principles:

- **Honesty**: Science attempts to explain the natural world; technology attempts to manipulate the built world, including virtual ‘buildings’ such as computer systems. If we believe that there is an objective truth, we must be honest in reporting our observations of it. Otherwise our conclusions will be false, and either useless or harmful.
- **Credit for work performed**: This is the currency of the scientific world; personal advancement of the scientist or engineer depends on receiving credit (and taking blame!) for what has been actually done by that person.

6.2 Fraud

Fraud is any action which wilfully mis-represents the truth. This can be the truth as to who did something (i.e. not correctly crediting someone with their idea or data) or the truth as to what was actually seen (i.e. data falsification or manipulation). Fraud can be committed by *omission* (not saying something that should be said) as well as by *commission* (saying something false).

Intent to deceive

The key issue in fraud is the *intent to deceive* , in other words the *willful misrepresentation* of the facts (e.g. what was done, what was seen, who did what). When we read a piece of research, we may not accept the interpretations and conclusions of the author, but we expect that any statements of fact are indeed true, so that we can form our own conclusions or repeat the work.

The scientific enterprise responds harshly when cases of fraud are detected. This can be years after the fact (see for example Broad & Wade [5] for the case of British psychologist Cyril Burt who falsified and invented data for a series of very influential studies on identical twins) but is more likely to be sooner. Scientists are naturally suspicious and inquisitive, and will probe behind what is written to find out what is true. Supervisors, Professors, and external examiners are very good at identifying suspect parts of the thesis and asking about them.

Examples are:

- results that seem too good to be true;
- data that shows very regular patterns, consistent with the hypothesis;
- data points from a small dataset that lie very close to a good model fit (e.g. regressions, variograms);
- beautiful writing in the middle of an otherwise sloppy text; and
- data that should have taken a long time to collect but which are supposed to have been obtained in a short fieldwork.

Fraud may be classified in three divisions, roughly in order of seriousness:

1. **Fabrication**: making up data, lying about procedures;
2. **Falsification**: manipulating data (or not) to obtain a pre-determined conclusion;
3. **Plagiarism**: taking credit for someone else's work.

Fabricating or falsifying data is the **cardinal sin against science**, since only with true data can we make progress towards the truth. Plagiarism is an offence against another author, stealing his or her credit.

6.2.1 Data fabrication

> "False facts are highly injurious to the progress of science, for they often endure long; but false views, if supported by some evidence, do little harm, for everyone takes a salutary pleasure in proving their falseness and when this is done, one path towards error is closed and the road to truth is often at the same time opened."
>
> – Charles Darwin, 'The Descent of Man' (1871) [9]

Fabricating data is inventing data or lying about the procedures by which it was obtained. This is the **cardinal sin** in science, because it can never be un-done. A simple example is filling in survey sheets without actually making field visits, based on what is expected. A bit less obvious but still fabrication is over-interpreting a survey response ("He said he wasn't sure about when his family came here, but to me 1965 seems about right, so I'll enter that") or field observation ("I don't see any gravel in the subsoil here but there really should be, so I'll enter them on the form").

Without accurate primary data, the entire research is invalid. You can always interpret or manipulate the data (with appropriate justification, of course), but that is a separate step from the primary data collection.

Researchers should always **keep primary field records and logs**. They are the ultimate proof of what was actually done. Some researchers go so far as to have their logs notarized.

6.2.2 Data falsification

Falsifying data is manipulating actual data to obtain a pre-determined outcome, or not manipulating when it is required. It comes in several forms: omitting 'inconvenient' observations as well as changing data values to more 'reasonable' ones, without system or justification.

- Discarding data during **sampling** is possible but (1) when explicitly acknowledged and (2) based on clear criteria.

  Example: a planned soil fertility sample was found to be located in the middle of an irrigation ditch; this can be discarded because it's not representative of the population being sampled (i.e. agricultural soils). This must be on the basis of criteria defined **prior** to beginning the sampling.

  Example: a respondent in a household survey seems clearly to be mentally ill and deluded. Record his or her answers, but add a note about their mental state as you interpret it, and then state that this response was discarded for the reason that, in your opinion, the respondent was not reliable. Another researcher can still make use of your primary observation if they disagree with your assessment of the respondent's state.

- Discarding data during **analysis** is possible but (1) when explicitly acknowledged and (2) based on clear criteria.

  A typical problem concerns so-called *outliers*, that is, data points that don't fit the pattern. In any case, they must be reported. But you don't have to include them in the analysis (e.g. to compute a correlation coefficient) if you can argue convincingly that they are **not part of the population being analysed.** Some possibilities:

  - Poor technique (but how do you that know only *this* sample was affected?)
  - Poor record-keeping (reflects poorly on your technique, but at least you are admitting it);

  - From a markedly-different site that is *not* included in the population you are studying.

- An obvious recording error (e.g. missing decimal point) may be corrected with no further observation, but this change should be shown in the original field book with a note.
- Leaving out an 'inconvenient' observation with no comment and no justification is fraud.

Many of the advances in science come from researchers who rigorously pursued their data, or who noted anomalies in other researchers' data and tried to explain them. A classic example is the discovery of the microwave background radiation from the Big Bang by Wilson & Penzias.

Manipulating raw data It may be necessary to **adjust** raw data to correct for inconsistencies. For example:

- Different instrumentation or analytical methods have been used to measure the same thing, either within the same experiment, or due to a change in procedures over time. Examples are time-series of climate data taken with different instruments over the years, or soil analysis where procedures changed due to new instrumentation (e.g., particle-size analysis by Buyoucos hydrometer, pipette, or laser diffraction).
- Different operators (researchers) have measured the same thing – this is especially difficult in the case of subjective ratings, but even relevant when using instruments.

The aim is to achieve a consistent dataset for later analysis.

Manipulating data is permitted as long as:

- A **clear and consistent** methodology is applied **objectively**
  - can't "pick and choose"
  - **all** data items with a defined characteristic are adjusted in the **same way**
- The adjustment methodology is **documented** as part of the research
- The **original data** are available for inspection.

A typical method is to establish an empirical statistical model between samples that should be the same (or at least their means or medians should be the same), and use this as a calibration relation. The important *ethical* point however is just that the method should be documented, justified by evidence (such as the statistical model) and consistently-applied.

Not manipulating raw data when it is required A more subtle form of falsification is when raw data *should* be adjusted for known inconsistencies, e.g., different operators, different instruments, or different illumination conditions, but this is *not* done, in order to reach a *pre-determined conclusion*. We might call this "reverse falsification".

An interesting example of manipulating or not is the adjustment of climate time series for known shifts in climate station location (e.g., from coastal lowlands to inland slopes above cities), for known instrumentation changes, or for known local effects that are not related to the overall climate record. For example, the (New Zealand) National Institute of Water & Atmospheric Research (NIWA)[1] has adjusted its raw temperature time series (since the 1850's) according to documented and justified procedures[2]; in addition, the "cooked" series was checked against a raw series where the reasons for correction were not applicable. This procedure has been termed "falsification" (i.e., data manipulating to reach a pre-determined conclusion) by "climate-change skeptic" groups, e.g., the New Zealand Climate Science Coalition[3]. The NIWA responds that without the corrections the time series is in fact misleading, and that the skeptics group does not apply the corrections in order to reach their preferred conclusion. Thus both parties accuse the other of falsification.

6.2.3 Plagiarism

Plagiarism is defined and explained by many authors [e.g. 4, §11.5], more or less as follows [8, p. 3]:

Knowingly representing the work of others as one's own

This can occur many in several ways, for example:

1. Copying someone else's work;
2. Paraphrasing someone else's work, i.e. saying the same thing with slightly different words and phrasing;
3. Reporting someone else's work (e.g. fieldwork) as if it were your own;
4. Getting someone else to do your work for you ('ghostwriting');

[1] http://www.niwa.co.nz/
[2] http://www.niwa.co.nz/climate/nz-temperature-record
[3] http://nzclimatescience.net/

5. Using a particularly apt term or phrase which you didn't invent, without credit

Simple copying is easy to define, but some cases are not so straightforward. Here we go into detail on what is permitted and what is not, and the reasons for this. We start from some basic principles of honest writing:

Three golden rules

1. **Everything you *write* outside of quotation marks must be the result of your own creative effort**. Otherwise, you are taking credit for something you did not write.

   Note that "your own creative effort" does not mean that you can't incorporate ideas from others in your own thinking; in fact that is encouraged. It means you must creatively synthesize others' work and adapt it to your purpose.

2. **Every *idea* that is not your own must be credited** to the person(s) who conceived it. Otherwise you are taking credit for the other person's idea.

   Note that ideas that are common knowledge need not (should not) be credited; it is unique ideas that can be traced to a definite source that must be credited.

3. **Every *fact* that you did not yourself establish must be credited**. Otherwise you are claiming direct knowledge that you do not have. This includes field or lab. work actually done by others which you are reporting.

Plagiarism by direct quoting without attribution is a temptation for some researchers for several reasons:

- The researcher feels that the plagiarised author is an all-knowing authority, and their text should not be altered;
- The researcher feels that the author has explained matters perfectly, and their text can not be improved upon;
- The researcher is not a confident writer (perhaps because they are not used to writing in English) and prefers to use a ready-made text;
- It is very tempting to cut-and-paste from easily-available electronic documents (web pages, full-text journals ...).

Cut-and-paste plagiarism

The first reason is always false. The second may be true for the original author's purpose, but not for the purposes of the current research. The third may well be true, but paraphrasing is still plagiarism. Quoting is at least honest if lazy. The fourth (direct cut-and-paste) is **really stupid**; the same tools that the plagiarist uses to find the text will be applied by the examiner to find the text again and establish that it was plagiarised.

> **Note:** Plagiarism-detection software[4] is increasingly used to find suspected cases; however a skilled reader in tune with the writer's style can "feel" a change in style that signals plagiarism.

To be completely clear on this, here is an example of plagiarism by copying. First, from the original article by Bergsma [3]:

> 'Soil conservation is defined as the use of land, within the limits of economic practicability, according to its capabilities and the need to keep it permanently productive.'

Second, from an MSc thesis, *not* written by Bergsma:

> 'Soil conservation is defined as the use of land, within the limits of economic practicability, according to its capabilities and the need to keep it permanently productive.'

This is certainly plagiarism: straight copying. What if we add the citation?

> 'Soil conservation is defined as the use of land, within the limits of economic practicability, according to its capabilities and the need to keep it permanently productive [3]. '

This is not so bad, but it is still plagiarism. The author has credited Bergsma with the *idea* of this definition of soil conservation, but still implies that the actual *words* used are the author's interpretation, which they are not.

The correct way to use this exact definition and credit the author is to quote the exact text. This is correct but not very elegant:

> ' "Soil conservation is defined as the use of land, within the limits of economic practicability, according to its capabilities and the need to keep it permanently productive" [3].'

It is much smoother to put the relevant part of the quotation in a context, for example:

> 'Bergsma [3] defines soil conservation as "the use of land, within the limits of economic practicability, according to its capabilities and the need to keep it permanently productive".'

Or:

[4] For example, Ephorus (`http://www.ephorus.com/`)

> 'Soil conservation is defined by Bergsma [3] as "the use of land, within the limits of economic practicability, according to its capabilities and the need to keep it permanently productive".'

Note the use of quotation marks to set off the *exact words* of the original source.

If you only want to discuss the points made by Bergsma but the exact words don't matter, it's better to rephrase this to match the use you will make of these ideas, while giving credit for them. For example:

> 'Bergsma [3] emphasizes three aspects of soil conservation: using land according to its capability, permanent use, and economic feasibility. The present study is particularly concerned with the third aspect, because ...'

Unless you intend to discuss the exact definition or wording, it is better to synthesize with other sources or adapt to your own argument. An example in this case might be:

> 'The concept of soil conservation was originally aimed at the physical protection of the soil from erosion at any cost and for indefinite time [14], but the emphasis is now on measures that are economically practicable and in line with the land's capabilities to provide productive and ecological services [3].'

Here we use two sources to support an argument, and brings out the essence of what is meant by 'soil conservation', without plagiarizing either source. Bergsma is correctly credited with the emphasis on economics, and Hudson with the original concept.

Unnecessary plagiarism

Much of what is plagiarised is not really necessary for the thesis. Students sometimes plagiarise the bulk of their introductions and much of their literature reviews. Why define a GIS, for example, when it is so completely covered by other authors? Only if this author will proceed from that definition to something specific is it necessary. And in any case the definition should be phrased in this author's own way, or else the original should be quoted (not plagiarised).

Synthesis

In introductory material such as a literature review or the problem statement, it is common to make statements that are obviously not your own original ideas. If you have made a synthesis, that is, taken various ideas and facts and put them together to make your own argument or explanation, you have to give credit but you should not quote.

Quoting

It is almost always better to put things in your own words and argument rather than to quote. However, quoting is justified in these specific instances:

- Definitions that you will discuss;

  > 'A common-language definition of land is "the solid part of the earth's surface" [22]. However, when we use the term 'land' in when defining 'land evaluation', we have in mind a more specific meaning, following the FAO [10], ...'

  > 'Bergsma [3] defines soil conservation as "the use of land, within the limits of economic practicability, according to its capabilities and the need to keep it permanently productive". Thus the emphasis is on economic sustainability.'

- Direct statements that you will discuss;

  > 'Buol *et al.* [6] feel that there is widespread awareness of the existence of soils, calling them "objects of common experience and observation". We will argue that they are in fact not so widely perceived ...'

- Especially clever or unique sayings, aphorisms, literary references that are particularly appropriate to what you want to say.

  > 'As Yogi Berra[5] famously said, "You can observe a lot just by watching".'

Paraphrasing

A particular difficulty comes with *paraphrasing*:

- You say the same thing as as a single author;
- You say it in the same order or with the same argumentation;
- You use quite similar words or synonyms.

This is also plagiarism, although less egregious than out-and-out copying.

Example of plagiarism by paraphrasing

Consider this passage:

> "People seem to have a natural tendency and urge to sort out and classify the natural objects of their environment. Soils are no exception, being objects of common experience and observation – undergirding agricultural production and supporting buildings and highways"
> – Buol *et al.* [6, p. 180]

[5] an American baseball player and folk hero well-known for his aphorisms

Here is a paraphrase that would certainly be considered plagiarism, *even if the citation is given*:

> 'As humans, we appear to have a built-in need to organise the things we find around us in the natural world. This is also true for soils, which everyone has seen, since soils are so necessary for agriculture and civil engineering [6, p. 180]'

Why is this still plagiarism? Because, although I have changed the words, the argument and sequence are the same. I have simply used synonyms and close paraphrases of the original:

Buol	**Paraphrase**
People	humans
natural tendency and urge	built-in need
to sort out and classify	to organise
natural objects	things . . . in the natural world
of their environment	we find around us
Soils are no exception	This is also true for soils
being objects of common experience and observation	which everyone has seen
undergirding	are so necessary for
agricultural production	agriculture
and supporting buildings and highways	civil engineering

Here is an acceptable compromise, where I give credit to the original authors and then extend their ideas with my own. I've also loosened up the style and argument. This is not plagiarism.

> ' Buol *et al.* [6] point out that, as humans, we seem to have a built-in need to organise the complexity of the natural world. Cognitive scientists such as Pinker [18] have even suggested that the desire to reduce complexity and form categories is 'hard-wired' into our brains by evolution. This tendency to classify extends to soils, at least to those properties that are readily perceived by soil users such as agriculturalists and engineers. '

6.3 Authorship

One of the two main principles of scientific ethics is "credit for work performed" (§6.1). The previous section (§6.2.3) discussed how to avoid taking credit for another's work by plagiarism. There is another aspect of giving credit where it is due, which is the **authorship** of original research when published publicly, e.g. in scientific journals, book chapters, or conference proceedings.

Scientific journals have written policies on authorship, usually included in their "Guide for Authors". These should be consulted by the corresponding author (usually the first author) before the manuscript is submitted, to ensure compliance with all journal policies. For example, Elsevier journals[6] have a common "publishing ethics" web page[7], which includes these guidelines (emphasis added):

> "Authorship should be limited to those who have made a *significant contribution* to the *conception*, *design*, *execution*, or *interpretation* of the reported study. *All those* who have made significant contributions should be listed as co-authors. Where there are others who have participated in certain substantive aspects of the research project, they should be *acknowledged* or listed as contributors.
>
> "The corresponding author should ensure that all appropriate co-authors and no inappropriate co-authors are included on the paper, and that *all co-authors have seen and approved the final version* of the paper and have agreed to its submission for publication."

Other publishers have quite similar requirements. Note the phrase **"significant"** contribution. This is of course a vague term, which can lead to disagreements.

UT/ITC Guidelines

At UT/ITC, guidelines for authorship of papers incorporating research results originating from MSc projects have been established by the Academic Board[8]. These are published by the ITC library[9] and explained below.

> **Note:** In case there is disagreement about the interpretation of these guidelines, a professor in the field concerned may be consulted, and in case the disagreement persists Head Research should be asked for a decision *prior* to submitting an article for publication.

Copyright law stipulates that copyright of MSc and PhD theses rests with the student. This is the case for all printed or electronically published material written by individual authors primarily as the result of their own initiative. However, UT/ITC retains the right to use the results of

[6] Journals from this publisher relevant to ITC research themes include *Remote Sensing of Environment*, *CATENA*, and *International Journal of Applied Earth Observation and Geoinformation*

[7] http://www.elsevier.com/wps/find/intro.cws_home/publishing

[8] approval 26-April-2010

[9] http://www.itc.nl/library/copyrightguide.aspx

the research done at UT/ITC or with its resources; this use can and often does result in scientific publication.

The following five cases may occur:

1. The advisor or supervisor uses **minor material** from an MSc thesis, e.g. a graph, table, or quote. A citation of the MSc thesis is required at the point where the material is used (i.e. normal citation practice), and an acknowledgement to the MSc student in the text or appendix is optional.
2. The article written by the advisor or supervisor has one or more **sections** that can be **directly traced to material from an MSc thesis**. In this case the MSc student should be invited to be co-author. Generally the MSc student will accept, and should review the article and comment on it before submission and during the journal review process.
3. If **several related MSc studies** are included in a paper, the supervisor or advisor assembles them into a coherent story, and is thus the first author. The MSc students whose work forms the basis for one or more sections are all co-authors. As in the previous case, they should review the article and comment on it before submission and during the journal review process.
4. If the material directly traceable to the MSc study makes up **more than half of the paper**, and if the MSc student takes the lead in authorship, the student is first author and the supervisor or advisor is a co-author. Again, all authors participate in the submission and review.

   > **Note:** The MSc student is free to publish a paper based on their MSc work as the sole author; however, the help and advice of the supervisor or advisor is generally helpful to a first-time author. Further, in general the MSc student has received substantial assistance during the thesis process from the supervisor or advisor, so that by the "significant contribution" rule (see above) they should be included as co-author(s).

5. If the MSc study is **substantially reworked** in concept (not just language) by the advisor or supervisor, these may be the first authors, and the MSc student a co-author.

The term "directly traceable" refers to the data, analysis and conclusions, not to the actual wording. Journal articles are typically reworked (often extensively) from reports such as MSc theses, since the reporting requirements are different. Similarly "more than half of the paper" does not refer to the number of words, but to the amount of intellectual content.

Examples

Following are examples of these cases[10], with supervisors as co-authors. The same rules apply to papers co-authored with PhD advisors.

1. Minor material from an MSc thesis is used:

   ▷ Bergsma, E., & Farshad, A. (2007). Monitoring erosion using microtopographic features. In J. de Graaff, J. Cameron, Sambran Sombatpanit, C. Pieri & J. Woodhill (Eds.), Monitoring and evaluation of soil conservation and watershed development projects (pp. 249-266). Enfield, UK: Science Publishers.

   This article includes the statement: "The sites were comparable in rainfall erosivity, general topography and soil (Table 4, basic data from Woldu, 1998)."; the cited thesis is:

   ▷ Hagos Dory Woldu. (1998). Assessment of the effect of present land use on soil degradation : a case study in Lom Kao area, central Thailand. Unpublished MSc, ITC, Enschede.

2. A substantial section, but less than half, from the work of an MSc student; the supervisor or advisor is the first author.

   ▷ Carranza, E. J. M., Hendro Wibowo, Barritt, S. D., & Prihadi Sumintadireja. (2008). Spatial data analysis and integration for regional - scale geothermal potential mapping, West Java, Indonesia. Geothermics, 37(3), 267-299.

   This uses material from:

   ▷ Wibowo, H. (2006). Spatial data analysis and integration for regional scale geothermal prospectivity mapping, West Java, Indonesia. Unpublished MSc, ITC, Enschede.

   Note the involvement in the article of additional authors from ITC (Barritt) and a research collaborator (Prihadi).

3. The work of several MSc students is synthesized by the supervisor or advisor.

   ▷ van Gils, H. A. M. J., Batsukh, O., Rossiter, D. G., Munthali, W., & Liberatoscioli, E. (2008). Forecasting the pattern and pace of Fagus forest expansion in Majella national park, Italy. Applied Vegetation Science, 11(4), 539-546.

   This has material from two MSc theses:

   ▷ Batsukh, O. (2007). Beech forest expansion : spatial environmental modelling for prediction, Majella national park, 1975-2003 Italy. Unpublished MSc, ITC, Enschede.

   ▷ Munthali, W. (2006). Beech expansion : patterns, process and prediction. Unpublished MSc, ITC, Enschede.

[10] Presented in the APA-6th bibliographic style

These were both supervised by the same ITC staff; there is also a contribution from a collaborator in the study area.

4. Mainly from one MSc thesis, written with the supervisor and/or advisor as co-author(s):

   ▷ Hengl, T., and Rossiter, D. G. (2003). Supervised landform classification to enhance and replace photo-interpretation in semi-detailed soil survey. Soil Science Society of America Journal, 67(6), 1810-1822.

   This is based on:

   ▷ Hengl, T. (2000). Improving soil survey methodology using advanced mapping techniques and grid based modelling : case study, Baranja, Croatia. Unpublished MSc, ITC, Enschede.

5. The MSc study is substantially reworked in concept by the supervisor or advisor:

   ▷ van Gils, H. A. M. J., & Loza Armand Ugon, A. V. (2006). What Drives Conversion of Tropical Forest in Carrasco Province, Bolivia? Ambio, 35(2), 81-85.

   This is based on:

   ▷ Loza Armand Ugon, A. V. (2004). Spatial logistic model for tropical forest conversion : a case study of Carrasco province, 1986 - 2002, Bolivia. Unpublished MSc, ITC, Enschede.

   This paper includes acknowledgements to other ITC staff who contributed to the work, but not enough to be co-authors:

   > "Dr. David G. Rossiter (ITC, Enschede) is acknowledged for his expert advice on modeling and reclassification of the soil-mapping units. ... Eduardo Westinga (ITC, Enschede) formatted the satellite images analysis (done in ILWIS) outputs into compatible ArcGIS input."

6.4 Intellectual property and fair use

The intellectual, intangible product of a creative effort, such as writing, music, or a computer program, is as much the property of the creator as is a tangible object such as a work of art or a machine. In some cases intellectual property is put into the public domain for free use, in other cases its use is restricted.

Misuse of intellectual property is easier than misuse of tangible property, but it is equally theft.

6.4.1 Copyright

Copyright (indicated by the © symbol) is the means by which an author asserts ownership of a work. Laws vary between countries, and there are international treaties. The basic idea is very simple: *the work 'belongs to the author*, who grants you certain *use rights*. If you obtain the work legally, you can use it for your own purposes (e.g. read it for pleasure or instruction). Other uses are made explicit, for example:

> "All rights reserved. No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording or by any information storage or retrieval system, without the written permission of the publisher, except for brief passages quoted by a reviewer." [21]

6.4.2 The concept of 'fair use'

In science or art we may want to compare our work with that of others, and to make our point we need to quote from the other work. This sort of use is recognised by copyright law as *fair use*: you obtained the work and you may use it for your professional purposes.

We may also want to make photocopies of printed matter. If we want the whole book, we are required to buy it. If we only want 'reasonable' excerpts, it is considered fair use to make a copy of these parts.

Fair use does not apply to **figures**, **drawings** and **photographs**; permission must be obtained to use these (unless the work is in the public domain). They must then be credited. Adaptations of figures are acceptable (with credit) if the original has been substantially changed.

6.4.3 License agreements

Some materials are made available only under the terms of a *license agreement*. That is, the person who obtains it in a legal manner, whether by purchase or free, must first agree how may be used. This is very common with *computer programs* and *digital data*.

For example, ITC has a license with certain publishers to allow on-line access to the full text of journal articles (usually as Adobe PDF files). The license allows full use within ITC, but it is *forbidden to supply a third party* with the file; they would need to obtain it under their own license.

Legal liability

You, your employer, or your educational institution (e.g. ITC) is *liable for your actions*. Remedies available to the copyright owner include expensive lawsuits and even criminal charges.

6.4.4 Restrictions on Datasets

Some digital data is supplied completely without restriction on what you can do with it, in particular data produced by the United States government. Most, however, is only supplied along with an *end-user license agreement* (EULA), to which you must agree.

You can not use data in your thesis which is not legally yours to use. This can be either via your own license, via your educational institution (e.g. ITC), or via some organisation of which you are considered part for licensing purposes.

6.4.5 Copyleft and open-source software

Some material is explicitly protected against theft, but its full use (including resale) is allowed if certain conditions are met. The most (in–)famous of these is the GNU General Public License (GPL) for certain open-source software[11], which requires that any new software that uses any code protected by GPL also itself be licensed under the GPL.

There are several similar licenses for written documentation, e.g. Creative Commons, Open Content license, Academic Free license, and the GNU Free Document license.

6.5 Professionalism

Professionalism refers to scientists' role in the society in which they work, i.e. not as individuals but as representatives of a profession. To be a "professional" means to gain one's living by carrying out a defined activity. The antonym is "amateur". This distinction is clear when we speak of musicians or athletes. For some professions, such as engineers, this status is recognized by a formal licensing procedure; in others there is only the requirement the the professional be practicing, and perhaps have received a certain academic qualification or have certain professional experience.

The term "professionalism" is used in a general sense to mean carrying out professional activities correctly, according to standards, and within a societal norms. All of these are ethical issues, tightly-linked to the role of the scientist within society.

6.5.1 Professional societies

Many professional groups have codes of behaviour. These include ethical standards within the profession, but also deal with how the professional should behave and act within the society at large. These are sometimes

[11] http://www.gnu.org/copyleft/gpl.html

called “professional codes of ethics” or “standards of professional conduct”. They may have legal standing in some countries.

For example, the Soil Science Society of America [20] includes the following:

“Members [of the Society] shall:

1. Uphold the highest standards of scientific investigation and professional comportment, and an uncompromising commitment to the advancement of knowledge;
2. Honor the rights and accomplishments of others and properly credit the work and ideas of others;
3. Strive to avoid conflicts of interest;
4. Demonstrate social responsibility in scientific and professional practice, by considering whom their scientific and professional activities benefit, and whom they neglect;
5. Provide honest and impartial advice on subjects about which they are informed and qualified;
6. As mentors of the next generation of scientific and professional leaders, strive to instill these ethical standards in students at all educational levels.”

Point (2) was already covered under Ethics, but the others are **socially-defined** values. Notice in particular the ethical standards for consulting covered by points (3) and (5), There is also attention to the explicit social role of the professional; science is not value-neutral! In particular, point (4) means that the social implications of research must be considered (e.g., a technology that favours capital-intensive farming will have implications for the survival of family farming). Point (6) has to do with inter-generational transmission of values.

These aspects of ethics go far beyond simple considerations of honesty. They may well be defined differently in different societies. Here social and religious values are indeed relevant.

6.5.2 Codes of conduct

There may be conflicts between certain “universal” scientific values and the socio-cultural context; these difficult issues are often addressed by national scientific societies. For example, in the Netherlands the association of universities has published a “Code of Conduct for Scientific Practice” [2], which all researchers in the Netherlands, including ITC, must follow. This has sections on:

1. **Scrupulousness**: Scientific activities are performed diligently, with care, resisting pressure to cut corners in order to “achieve”;
2. **Reliability**: The scientist makes every effort for their work to be accurate and thorough, thus reliable;
3. **Verifiability**: Any publication based on research must clearly state the basis for the data and conclusions, including the data source and analysis methods; all of this so that the reader can in principle independently verify the work;
4. **Impartiality**: In scientific activities, the scientist must have no other interest than science, and be prepared to prove this. This is most relevant when the scientist works for industry or has commercial interests;
5. **Independence**: Scientists operate in a context of academic freedom and independence from interference. If this is not possible for commercial, political or institutional reasons, this must be clearly stated and justified.

6.6 The social responsibility of the scientist

A wider ethical question than professionalism or a code of conduct is the role of the scientist in the wider social context – i.e. acting as a responsible member of society. This depends on the scientist’s personal values and society’s expectations.

The activities of a scientist (as any other member of society) have an effect on that society – there is no such thing as value-neutral research. Even by withdrawing from society into a basement laboratory, the scientist has made a social choice.

Science is now big business and an integral part of society.

6.6.1 Selection of a research topic

Important ethical decisions are made at the beginning of a research project, with the selection of a research topic.

- Would the results of the research be **useful** to society?
- Is the topic related to a **social problem** of importance?
- Would the results of the research be **socially valuable**, or at least not damaging?
- Are various sectors of society **marginalized** or even directly **harmed** by the research?

Many research topics pose ethical problems, for example:

- Any remote-sensing project by its nature (view from above) invades the privacy of individual land owners; it also violates the sovereignty of the country imaged.
- Any natural resources survey or land suitability evaluation project implies that knowledge of these will be given to people outside the affected area, who may make planning, investment or migration decisions that may not benefit the local population.

  For example, a biodiversity survey reveals a "hot spot" which is in danger of encroachment from expanding slash-and-burn agriculture. This information is used to justify expelling the nearby population.

  Another example: a watershed study establishes the spatio-temporal water contributions to a river system; this is used to site a large dam, which displaces the local population.
- A design thesis that builds on a specific computer program is implicitly endorsing that program and, if it is a commercial program, promoting the financial interests of the company that produced it (ESRI, Microsoft, ENVI ...). Conversely, use of an open-source program may reduce commercial opportunities but increase the overall productivity of the research community. Which side are you on?

Some topics pose ethical dilemmas, in that the results can be used for good or evil. A classic example is the rocketry research of German scientists [17][12]. It began (in Germany) and ended (in the USA) as an effort in space exploration, but was applied to ballistic missiles (in both countries) in between. Support for this research, in both countries, depended heavily on military interest and applications.

6.6.2 Trendiness and 'political correctness'

The modern research establishment runs on public (government) or charitable (foundations) funding. These have explicit agendas for research. A typical example is the European Commission's (EC) research frameworks, currently in the seventh round ("FP7")[13]. This has explicit social goals, which are translated into research priorities. For example, under the "Food, Agriculture and Fisheries; Biotechnology" theme the EC states explicitly:

---

[12] "Once the rockets are up, who cares where they come down?" – Tom Lehrer, *Werner Von Braun*

[13] http://ec.europa.eu/research/fp7/index_en.cfm

> "The advancement of knowledge in the sustainable management, production and use of biological resources (microbial, plant and animal) will provide the basis for safer, eco-efficient and competitive products and services for agriculture, fisheries, feed, food, health, forest-based and related industries. Important contributions to the implementation of existing and prospective policies and regulations in the area of public, animal and plant health and consumer protection are anticipated. New renewable energy sources will be supported under the concept of a European knowledge-based bio-economy."

Notice the policy-driven emphasis on "sustainable" management and "eco-efficient" production. Grants submitted under this theme must address these issues. So, research is not purely **curiosity-driven**, it is in support of an explicit social **agenda**.

So, scientists often have to trim their research to the prevailing winds to obtain funding. Current examples relevant to ITC are "global warming" or at least effects climate change, carbon credits and trading schemes, participatory approaches to natural resource management; these lead to research programmes and PhD projects[16].

A related issue is so-called **political correctness**: avoiding language or implications that might be considered offensive by some group with a self-identity, e.g. religious, national, ethnic, gender, age, caste, tribe, social status .... In the context of research, this means that some ideas are not acceptable for research, or even to mention in brainstorming sessions. Some topics are so sensitive ("hot button") that they are not to be mentioned, let alone considered for research.

However, these topics may address real problems, on which research can shed some light. Ignoring a problem does not make it go away.

An ITC-related example is the difference in spatial orientation skills and strategies between males and females, although this has been accepted as proper research topic for many years [7].

The problem often comes in interpretation of results, not in the results themselves.

6.7 The scientist as a social animal

Scientists are humans, with all the problems that implies.

6.7.1 Values, bias, subjectivity

Scientists can not escape from their upbringing and environment, but because of the self-reflective nature of science, they can perhaps be more aware of the implications:

Values

All humans have an internal ethical system, often (but not always) largely in agreement with some wider (e.g. religious) value system. Like all humans, the scientist feels that certain things are "right" and "wrong" ethical behaviour.

Most would agree, for example, that selecting interesting patients from asylums, studying them while alive, murdering them and then examining their brains, is somehow "wrong". But not so many years ago, precisely this was done by established researchers [13, Ch. 11]. Science itself has no answer for this; clearly this is part of human, not scientific, ethics.

Far less extreme examples may colour a researcher's choice of topic or approach. Consider a study of urban transport options. Does the researcher value social cohesion or individual opportunity more? In the first case the research will probably be oriented towards improved public transport, in the second towards increased automobile use.

Bias

Humans do not approach problems as a "blank slate"; they build up biases (prejudices) as part of their life experience. Some of these are quite useful in survival, but can interfere with objective research.

For example, one of the main arguments against allowing women to vote in elections was the bias (by men in the power structure) that women are more irrational than men and would be easily-convinced by smooth-talking or good-looking candidates, and would not listen seriously. This bias made it almost impossible to investigate the truth of this claim, it was taken as an assumption. Similar biases continue and perhaps are as strong as ever regarding ethnic groups, tribes, religions, social class, age and educational level.

Researchers must strive to **uncover their own biases** and consider how these are affecting the objectivity of their research.

Subjectivity

By definition humans are "subjects" (in the grammatical sense of "I") when they interact with the world, so their viewpoint is unavoidably **subjective**.

The opposite of subjectivity is **objectivity**: the ability to see and report according to what is really found, rather than with pre-conceived notions.

Objectivity should be easy for reporting **results** of experiments or observations. However, even the process of **measurement** can be subjective:

- Where is the vertical boundary between two soil horizons?
- Where is the spatial boundary between undisturbed and exploited forest?
- Which of the answers in a questionnaire corresponds best to the respondent's answer?

Note these are not measurement uncertainties (e.g. reading the meter on an instrument), they are subjective interpretations.

Objectivity is more difficult when drawing **conclusions**. Here a sound., logical argument using multiple lines of evidence is required.

6.7.2 Interactions

Most scientists interact with colleagues, both in and out of their own institution. In any research which includes fieldwork, scientists also interact with local populations in the study area. These relations are a matter of ethics as well as professionalism.

This is particularly (but not only!) relevant in qualitative social sciences research, where interaction with research subjects is precisely the point of the research. Miles & Huberman [15, §11] provide a useful list of "thinking points".

Interactions with colleagues

Much of the interaction with **colleagues** is governed by narrowly-defined scientific ethics as outlined above (§6.1), particularly the rules for assigning credit for work performed. However, there are often cultural differences (both general and scientific) in working methods, expectations of roles and responsibilities, priorities, attitudes towards authorities, and communication style which can hinder scientific progress. Economic and status differences between colleagues can exacerbate these cultural differences. Awareness, sensitivity, communication, flexibility and common sense go a long way towards achieving a good working relationship.

Interactions with human subjects

Although ITC does not generally work with individuals as research subjects (e.g. in medical research), individuals can become involved in ITC research in ways that may affect their life for better or worse.

> For example, the ITC researcher may take a photograph of a farmer's field or a slum dweller's home, and publish the photo

in the thesis or a paper to illustrate a point. This photo may be seen and recognized by the government, or neighbours, or NGO's and used to the detriment (or benefit) of the individual.

The basic principle of working with individuals is **informed consent**: they should know what information is being collected about them or their environment, and what will be done with it.

Interactions with local populations

Professional societies whose members do research with and in **local populations** have had extensive debate about the relation between researcher and subject; an example is by the American Anthropological Association [1]. This is also dealt with in texts on social science research methods [e.g. 11, 12] and in the context of recent research on participatory GIS [19].

The basic problem here is that some humans (researchers) are studying other humans (local population) or at least infringing on their territory. There are inherent differences in status, economic power, priorities between the two groups of humans.

Here are some examples of ethical questions raised by such research:

- How should local people be approached? What information about the research purpose should be given?
- Will the results of the research be 'returned', and if so, in what form?
- What to do if the research is not in the benefit, or even to the detriment, of local populations? Example: studying soil erosion vs. farming practices, this may lead to a ban on certain crops or management on certain lands (e.g. steep slopes), which is a short-term economic loss to the farmers?
- If surveys are to be performed, what information about them is given to the participants? Should they be paid or otherwise rewarded?
- What are ethical methods of asking questions or making observations? Can subjects be "tricked" with false promises or pretexts?
- How intimate should the researcher be with the population? Does the researcher sacrifice neutrality or objectivity by identifying too closely with the subjects or target group?
- How should researchers balance their own cultural values with those of their subjects?

- How to extract reliable information within cultural limitations? Example: It is considered improper in the local context for a male researcher to talk directly with a female subject; should the researcher trust a male relative's interpretation of what the female says?

Research is never neutral – someone (maybe the researcher?) benefits more than others; the results may be used for political ends, and so forth. Researchers are themselves always biased and have their own cultural references. These must be made explicit, at least to the researcher.

6.7.3 Who benefits?

One way to consider ethics in the widest sense is to ask: Who benefits from your scientific activity?

1. You: Producing a thesis advances your scientific career; you are able to do interesting work; it allows you to satisfy your curiosity, feed your ego, or whatever else makes you tick;
2. Your family;
3. Your educational institution (e.g. ITC) receives credit for having facilitated and supervised your work, which shows its ability to train students and do research, so it continues to attract students and support;
4. Your sponsor (home organization, funding agency): they get what they paid for;
5. The scientific enterprise in general (more is known);
6. Future employers (they get a capable worker);

These surely benefit from a successful MSc thesis. There are also indirect beneficiaries:

1. Society as a whole
2. The individuals or communities who helped you or made your research possible.

Society may be presumed to benefit from increased knowledge; but in the specific social context in which your research will be used, is there a benefit? For all parts of society? Or are some empowered and some enfeebled?

What goes back to the individuals or communities who helped you?

6.8 References

Bibliography

[1] American Anthropological Association. 1998. *Code of ethics*. Arlington, VA: American Anthropological Association. URL: http://www.aaanet.org/committees/ethics/ethcode.htm 144

[2] Association of Universities in the Netherlands (VSNU). 2004, revised 2012. *The Netherlands Code of Conduct for Scientific Practice: Principles of good scientific teaching and research*. The Hague: VSNU URL http://www.vsnu.nl/ 138

[3] Bergsma, E. 1996. *Terminology for soil erosion and conservation*. Wageningen; Enschede: International Society of Soil Science; International Soil Reference and Information Centre (ISRIC); ITC 128, 129, 130

[4] Booth, W. C.; Colomb, G. G.; & Williams, J. M. 1995. *The craft of research*. Chicago: University of Chicago Press 126

[5] Broad, W. & Wade, N. 1982. *Betrayers of the truth*. New York: Simon and Schuster 122

[6] Buol, S. W.; Hole, F. D.; & McCracken, R. J. 1989. *Soil genesis and classification*. Ames, IA: The Iowa State University Press, 3rd edition 130, 131

[7] Coluccia, E. & Louse, G. 2004. *Gender differences in spatial orientation: A review*. *Journal of Environmental Psychology* **24**(3):329–340 141

[8] Cornell University. 2000. *The code of academic integrity and acknowledging the work of others*. Ithaca, NY. URL: http://web.cornell.edu/UniversityFaculty/docs/AI.Acknow.pdf 126

[9] Darwin, C. 2004. *The descent of man : and selection in relation to sex*. Penguin classics. London ; New York: Penguin. (original publication 1871) 123

[10] FAO. 1976. *A framework for land evaluation*. Soils Bulletin 32. Rome, Italy: FAO 130

[11] Frankfort-Nachmias, C. & Nachmias, D. 1996. *Research Methods in the Social Sciences*. New York: St. Martins Press, 5th edition 144

[12] Frankfort-Nachmias, C. & Nachmias, D. 2000. *Research methods in the social sciences*. New York: Worth Publishers, 6th edition 144

[13] Gratzer, W. 2000. *The undergrowth of science: delusion, self-deception and human frailty*. Oxford: Oxford University Press 142

[14] Hudson, N. 1981. *Soil conservation*. Ithaca, NY: Cornell University Press, 2nd edition 129

[15] Miles, M. B. & Huberman, A. M. 1994. *Qualitative data analysis : an expanded sourcebook*. Thousand Oaks, CA: Sage Publications, 2nd edition 143

[16] Minang, P. 2007. *Implementing global environmental policy at local level : community carbon forestry perspectives in Cameroon*. ITC Dissertation;144. Enschede: ITC. http://www.itc.nl/library/papers_2007/phd/minang.pdf 141

[17] Neufeld, M. 2007. *Von Braun : dreamer of space, engineer of war*. New York: A.A. Knopf 140

[18] Pinker, S. 1997. *How the mind works*. New York: W. W. Norton 131

[19] Rambaldi, G.; Chambers, R.; McCall, M.; & Fox, J. 2006. *Practical ethics for PGIS practitioners, facilitators, technology intermediaries and researchers*. *Participatory Learning and Action* **54**:106–113 144

[20] Soil Science Society of America. 1999. *Statement of ethics*. On-line HTML document; URL: http://www.soils.org/files/about-society/soils-statement-of-ethics.pdf. Accessed 2013-April-10 138

[21] Strang, G. 1986. *Introduction to applied mathematics*. Wellesley, MA: Wellesley-Cambridge Press 136

[22] Thompson, D. (ed.). 1995. *The concise Oxford dictionary of current English*. Oxford: Oxford University Press, 7th edition 130

7 Statistical inference for research

In this chapter we examine some of the concepts behind the use of statistical inference to "prove" or at least reach conclusions as part of research.

Key points

1. **Quantitative** statements are generally more useful than **qualitative** ones.
2. Statistical **inference** is used to quantify the **certainty** of quantitative statements (§7.1).
3. A clear distinction must be made between **populations** and **samples**; the population that a sample represents must be unambiguously specified (§7.2).
4. There are two interpretations of probability: **frequentist** and **Bayesian** (§7.3).
5. **Bayesian** probability is the degree of rational belief that a statement is true; Bayesian inference works by updating **prior** to **posterior** probabilities, based on new observations. (§7.4).
6. **Frequentist** probability is the proportion of time an event would occur, should the experiment that gives rise to it be repeated a large number of times. Observations represent a sample from a population that has some fixed but unknown parameters (§7.5).
7. Frequentist **hypothesis testing** calculates the probability that **rejecting** a given **null hypothesis** is an incorrect decision. This involves the concepts of **significance levels**, **Type I and Type II errors**, and **confidence intervals** (§7.5.1).
8. Inferences are based on **statistical models**: their **functional form** and **parameters** (§7.6). The aim is to model the **structure**, and not the **noise**.
9. A clear distinction is made between model **calibration** ("*post*diction", parameter estimation) and model **validation** ("*pre*diction") (§7.6.5).
10. **Correlation** and **regression** are often used uncritically and inappropriately; distinctions must be made between **fixed** and **random** predictors, and between **descriptive** and **predictive** models. (§7.7).
11. **Correlation** does not necessarily imply **causation**; this link requires **meta-statistical** reasoning.
12. When **law-like relations** are to be modelled, **structural analysis** should be used instead of regression (§7.7.5).
13. Models should be **parsimonious**; this avoids fitting noise rather than structure (§7.7.7).

7.1 The inferential paradigm

Most research questions should be posed so that the answer is **quantitative**; this leads to deeper understanding and better information on which to base decisions. Kelvin made the definitive statement about the value of numerical measurement:

> "In physical science the first essential step in the direction of learning any subject is to find principles of numerical reckoning and practicable methods for measuring some quality connected with it. I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of Science, whatever the matter may be." [12, 1:73]

A similar sentiment has been the inspiration for the development of **inferential statistics**, which seeks to quantify the plausibility of statements about the world. These inferences are the main result of scientific research. Davis [6, p. 11] puts it nicely:

> "Statistics ... may best be considered as the determination of the **probable** from the **possible**."

A similar view is given by Shalizi [19]:

> "Statistics is the branch of applied mathematics that studies ways of **drawing inferences** from **limited and imperfect data**. ... We have some data ..., but we know that our data are incomplete, and experience tells us that repeating our experiments or observations, even taking great care to replicate the conditions, gives more or less different answers every time. It is foolish to treat any inference from only the data in hand as certain."

Here are some examples of statements we might like to make in the conclusions of a research project:

- "The projective transformation can successfully georeference a small-format air photo (SFAP) from ten ground control points measured with a single-receiver GPS".
- "In villages where a participatory GIS was developed there was less conflict between government plans and local goals."
- "Shifting cultivation systems have expanded in the past ten years,

mainly at the expense of primary forest."

These statements need to be quantified; in particular we would like to give a precise meaning to words like "successfully", "less", "expanded", "mainly".

Some statements already are a quantitative statement, based on some observations, so seem adequate as they stand:

- "Primary forest covers 62% of the study area."
- "On 10–September-2000 Lake Naivasha contained $8.36 \cdot 10^9 m^3$ of water."
- "Twice as many boys as girls attend secondary school in District X."

Yet unless these are made by exhaustive sampling (visiting every pixel, pumping the contents of Lake Naivasha though a water meter, counting every school child), they are **uncertain**, so we'd like to give some **range** in which we are fairly sure the true value lies.

We then use the **inferential paradigm**:

- We have a **sample** which **represents** some **population**;
- We want to make a quantitative statement about the population;
- This requires us to **infer** from sample to population.

7.2 Basic concepts

7.2.1 Descriptive and inferential statistics

The term "statistics" is used in two main senses:

Descriptive: A numerical summary of a dataset

For example: "Two hundred computer simulations were run on randomly-produced 1000-node graphs; the minimum run time was 0.3 s and the maximum 3.5 s, with a median of 1.2 s."

This is just a statement of fact, where 200 numbers (the simulation times) are summarized as the minimum, median and maximum, which are interesting for interpretation.

This is often used in common language: "The population statistics of X district are shown in Table Y".

Inferential: A number representing some characteristic of a population, inferred from a sample (see §7.2.2, just below).

For example: "This algorithm has a mean running time of 1.1 $\pm$0.2 s on 1000-node graphs". This statement is about all possible 1000-node graphs, inferred from tests on some representative sample.

7.2.2 Populations and samples

Any inferential statement refers to a **population**, which is the set of objects about which we want to make this statement. Most of these objects have not been observed, yet we would like to make some statement about them. It can be surprisingly difficult to precisely specify the population.

In the small-format air photo (SFAP) example, the population might be:

- All small-format air photos (SFAP) that were taken in this project
- All SFAP that could have been taken under 'similar' situations (how 'similar'? only in the study area? in 'similar' areas?);

The **sampling frame** is that portion of the population from which the sample will be selected, i.e. which might be observed.

In the small-format air photo (SFAP) example, the sampling frame must be the photos that were actually taken; a sample will be selected from these.

The **sample** is that portion of the population that we have observed.

In the small-format air photo (SFAP) example, the sample is the photos that were actually taken and then selected for geo-referencing.

7.2.3 Sampling design

The relation between population and sample is the **sampling design**. This is specified by the researcher, who must argue that the sample **represents** the population for inferences to be correct.

This is the basis of statistical inference – the researcher must be able to:

1. explicitly identify the population of interest – the inferential statement is made about **this population** and no other;
2. argue that the **sampling frame** represents the **population**;
3. describe the relation between the actual **sample** and the **sampling frame** – in particular, the *a priori* **probability** of selecting each potential sampling unit.

Note that the third step comes from the nature of the sampling design, but the others require **meta-statistical** arguments.

Q58 : *Consider a research project to determine whether high-resolution* *imagery can be used to assess post-disaster reconstruction in urban areas. The researcher would like to determine the proportion of reconstructed buildings that could be successfully extracted from the imagery by different image-processing techniques.*

(a) What is the population of interest?

(b) What could be a reasonable sampling frame? *Jump to A58* •

If some locations in geographic or feature space were purposely not sampled, it is difficult to argue that they be included in the population about which statements can be made. Typical reasons for not sampling include:

- Inaccessibility
- Lack of permission
- Uninteresting for purposes of the study (e.g. rock piles in a study of soil moisture)

In the first two cases, it might be possible to argue that the unsampled areas are similar to ones that were sampled, but this would have to be a convincing argument that the reasons for lack of access has no relation to the phenomenon being sampled. This is unlikely if, for example, a border village is not included in the sample because it is "too unstable".

7.2.4 Statements of fact vs. inference

The word "statistics" by itself simply refers to mathematic summaries of a set of data. These are just statements of fact:

- "The median sigma of georeferencing of 14 photos was 5.16 m".
- "Participants in the workshop had from two to ten years of formal education."
- "Twelve of the 40 crop fields surveyed in 2004, with an area of 6.3 ha out of the 18 ha total crop land surveyed, were covered by primary forest in 1990".

It is another step entirely to draw **inferences** from such statements. This must be from a sample (what has been observed) to a population (what could have been observed), being careful to specify the population:

- "The median sigma of georeferencing with the projective transform is no greater than X m"
- "Small farmers in the district have from X to Y years of formal education."

- “X% of the crop fields active in 2004 and Y% of their area were covered by primary forest in 1990”.

To make such statements, we use statistical inference.

For example, here is a set of the sigma values from the projective transform applied to 14 small-format air photos from a particular study [16]:

```
4.36 3.63 6.01 3.78 7.58 8.36 5.18
4.77 4.80 7.18 5.79 5.14 5.42 3.81
```

It is just a statement of fact to say that the minimum of this set is `3.63`, the maximum is `8.36`, the median (half greater, half less) is `5.16`, and so forth. It is even a statement of fact to report the sample mean and standard deviation, computed according to the standard formulas, as `5.145` and `1.453`. But what we’d really like to say is how successful this procedure would be (or would have been) when applied in “similar” circumstances. This has several related meanings:

1. If, now, photos have already been taken and GPS points collected in the study area but have not yet been processed;
2. If, hypothetically, a different set of photos had been taken and GPS points collected in the same area during the same mission;
3. If, in the future, more photos are taken and GPS points collected in the same area and under the same conditions;
4. As these three, but in other areas.

We can not avoid this sort of “meta-statistics”, and it leads us to consider the **plausibility** (not provability!) of each. In all cases we are arguing that the data in hand are a **representative sample** of the larger **population**.

The first statement seems the most secure: the population is then all photos and GPS points that were obtained. That is, the success with 14 photos can be used to predict the success with the rest of them. The second is similar but deals with a hypothetical population: all the photos that could have been taken; those that were are a sample of what was possible. The third is of interest if we want to repeat the study in the same area and the fourth if we want to extend it.

7.3 Frequentist and Bayesian interpretations

Meta-inference, that is, what an inferential statement really means, is still a contentious topic. There are two principal interpretations [10]:

- **Frequentist**, also called **classical** or **British-American**; and
- **Bayesian**.

The two approaches begin from quite different ideas about what is meant by "probability" and then carry these differences over to methods of inference.

Historically, the frequentist approach was developed under the leadership of R A Fisher, a statistician working at the Rothamstead Experimental Station in England, and was propagated in his highly-influential works *Statistical methods for research workers* (first appearing in 1925) and *The design of experiments* (1935) and by his disciples such as Neyman and Pearson. He worked in the 1930's at the Iowa State University (USA) and there influenced well-known workers such as Snedecor, Cochran, and Yates. Because of the close historical connection with field and laboratory research, and the well-developed theory of inference promulgated in many texts [e.g. 3] and computer programmes, the frequentist approach is the most common in practical work today.

The Bayesian approach is named for the English nonconformist minister and mathematician Thomas Bayes (1701–1761) but he did not develop it; he is however responsible for the first statement of **Bayes' Rule** of probability (§7.4), published posthumously in 1763. Inspired by Bayes' ideas about the meaning of inference, a group of statisticians, including Laplace, Jeffreys, de Finetti, Wald, Savage and Lindley, developed another view of statistical inference, known (somewhat misleadingly) as **Bayesian** inference [13]. This approach can reproduce the frequentist interpretation, but can also be extended to a much richer set of inferences where frequentist methods fail.

7.4 Bayesian concepts

The Bayesian viewpoint begins from a subjective definition of **probability**: it is the **degree of rational belief** that that something is true. The restriction to "rational" beliefs means that certain rules of consistency must be followed; I can't simply state a belief with no evidence. In this viewpoint, all probability is conditional on evidence, and can be updated in view of new evidence.

The Bayesian goes further and asserts a frankly **subjective** view of probability: any parameter that we are trying to estimate is not fixed, i.e. some hypothetical "true" value, but instead is something we want to develop a personal **probability distribution** for. Naturally, the distribution must be **consistent** with the available evidence, but there is no attempt to narrow down the estimate to some hypothetical but ultimately unknowable "true" value. Also, there are limits on the subjective distribution: it must in some sense agree with distributions estimated by others with similar subjective beliefs.

7.4.1 Types of probability

- **Prior** probability: before observations are made, with previous knowledge;
- **Posterior** probability: after observations are made, using this new information;
- **Unconditional** probability: not taking into account other events, other than general knowledge and agreed-on facts;
- **Conditional** probability: in light of other information, specifically some other event(s) that may affect it.

The distinction between conditional and unconditional probability depends on one's standpoint with respect to the possible conditioning event.

7.4.2 Simple form of Bayes' rule

In its simplest form, **Bayes' Rule** is used to update a **prior** probability $P(A)$, based on new information that an event B with prior probability $P(B)$ has occurred, and knowing that the conditional probability $P(B|A)$ of B given A, to a **posterior** conditional probability $P(A|B)$ [3, 1.3.5]:

$$P(A|B) = P(A) \cdot \frac{P(B|A)}{P(B)} \tag{7.1}$$

The last factor is the proportion by which the prior is updated, sometimes called the **likelihood function**.

Equation 7.1 is derived by reformulating the definition of intersection probability from conditional probability:

$$\begin{aligned} P(A \cap B) &= P(A|B) \cdot P(B) \\ &= P(B|A) \cdot P(A) \end{aligned} \tag{7.2}$$

Equating the two right-hand sides and rearranging gives the rule.

This rule can be used for diagnosis. For example, suppose we have a fever (event B) and therefore suspect that we may have malaria (event A); we would like to calculate the probability that we in fact have malaria, so that we can take the appropriate medication. To compute this, we need to know:

1. The conditional probability of a person with malaria having a fever, $P(B|A)$, which we estimate as, say, 0.9 (some people who are infected with malaria don't have a fever);
2. The unconditional probability $P(A)$ of having malaria, i.e. the proportion of the population that has it, say 0.2; this is the **prior** probability of having malaria before looking at our symptoms;

3. The unconditional probability of having a fever from whatever cause, say $P(B) = 0.25$.

These probabilities can be estimated from a large random sample of people, independent of their health, where they are tested for malaria to give $P(A)$ and observed for fever to give $P(B)$, and together to give $P(B|A)$, the presence of fever in those that tested positive for malaria. Note that this prior would be quite different in different locations. Then the **posterior** probability that, given that an individual has a fever, that they have malaria is $P(A|B) = 0.2 * (0.9/0.25) = 0.72$. The probability of malaria has been greatly increased from the prior (0.2) because the presence of fever is so closely liked to the disease. The likelihood function was thus $0.9/0.25 = 3.6$; the odds increased by 3.6 times in the presence of the information about the symptom.

If fever were more prevalent overall in the population, or if a smaller proportion of malaria sufferers showed a fever, the updated probability would be different. For example if $P(B|A) = 0.5$ (fever less symptomatic), then $P(A|B) = 0.2 * (0.5/0.25) = 0.4$; if $P(B) = 0.5$ (fever is more common overall), then $P(A|B) = 0.2 * (0.9/0.5) = 0.45$; in both cases it is less likely that our symptom (fever) indicates the disease (malaria). If malaria were less prevalent overall in the population, the posterior probability will be reduced proportionally; this is because fever from other causes is now more likely.

7.4.3 General form of Bayes' rule

Bayes' rule has a general form which applies when a sample space A of outcomes can be divided into a set of mutually-exclusive outcomes $A_1, A_2, \ldots$. Then the conditional probability of any of these outcomes A_i, given that event B has occurred, is [3, 1.3.6]:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)} \tag{7.3}$$

An example here is the land cover class at a particular location. This is one of the possibilities given by a legend. The **prior** probability $P(A_i)$ of a location belonging to class i is estimated from prior knowledge of the area to be mapped, perhaps a previous map or even expert opinion. The conditional probability $P(B|A_i)$ of some event (such as an aspect of a spectral signature) in for all possible land must also be given either from theory or statistical estimation. Then we can compute the **posterior probability** that a given location is in fact in the given class. This is precisely what "Bayesian" image classification algorithms do.

To take a simple example, consider a legend with three classes: open water (A_1), grassland (A_2), and forest (A_3) with prior probabilities $P(A_1) =$

0.1, $P(A_2) = 0.4$, $P(A_3) = 0.5$; these must of course sum to 1. That is, we expect the final map to have 10% open water. The event which is used to update this prior could be an NDVI < 0.2, for which we could estimate $P(B|A_1) = 0.95, P(B|A_2) = 0.02, P(B|A_3) = 0.05$; that is, in known pixels of water (from a training set), 95% of them had NDVI < 0.2; for grassland and forest there were only 2% and 5%, respectively with such low NDVI. Now, if we observe a pixel with NDVI < 0.2, we compute its posterior probability of in fact being open water as:

$$P(A_1|B) = \frac{0.95 \cdot 0.1}{0.95 \cdot 0.1 + 0.02 \cdot 0.4 + 0.05 \cdot 0.5} = 0.7422$$

The information that this pixel has a low NDVI has increased the probability that it represents open water from 0.1 (in the absence of spectral information) to 0.7422; the likelihood function was thus 7.422. The probability of being grassland or forest are similarly calculated as 0.0625 and 0.1953, respectively; note that the three probabilities add to 1 as expected. You might be surprised by the fairly high probability that the pixel is forest (nearly 20%); but recall that we expect half the map to be forest, and an appreciable proportion (5%) of pixels from these areas have low NDVI.

These formulas are not at all controversial in case the prior absolute and conditional probabilities are known. However, even if they're not, we may have some idea about them from previous experience, and that should give us better results than simply accepting the **non-informative** priors, i.e. that all outcomes $P(A_j)$ are equally-likely. In the above example, if we didn't know anything about the overall proportion of land covers in the area, we'd take $P(A_1) = P(A_2) = P(A_3) = 0.\bar{3}$, and compute the posterior probability of water, given low NDVI, of:

$$P(A_1|B) = \frac{0.95 \cdot 0.\bar{3}}{0.95 \cdot \bar{3} + 0.02 \cdot \bar{3} + 0.05 \cdot \bar{3}} = 0.9314$$

with the probability of being grassland or forest being 0.0196 and 0.0490, respectively. These probabilities are much lower than computed above (because the prior probability of these classes is lower); thus we see the major influence of prior probability. This has led to criticism of this approach as being **subjective**. But in image classification we often do have estimates of the proportion of various land uses or covers, either from previous studies or just reconnaissance; all classes are not *a priori* equally likely at each pixel in the classified image.

Bayesians argue that we are rarely in a state of ignorance about the object of study, and it makes sense to take account of what we already know. The medical diagnosis example supports this: doctors would be foolish not to take into account the difference between **a priori** rare and common diseases, even if they can not put a precise number on the relative occurrence. It's much more likely that someone with a fever in Yaoundé has malaria than someone with a fever in Enschede, and the doctors in those two places should not reason otherwise.

7.4.4 Practical problems with the Bayesian approach

For a single condition there is no problem. But of course diseases have many symptoms, and land covers give rise to many spectral conditions, and these are often not completely independent. So Bayes' rule can't simply be applied sequentially, symptom-by-symptom, it has a much more complicated form when there are conditional probabilities between conditions.

7.5 Frequentist concepts

To a frequentist, the **probability** of an event is the proportion of time it would occur, should the experiment that gives rise to the event be repeated a large number of times.

This is intuitively-appealing in the case of throwing dice, for example; we can imagine throwing dice in the same way many times. It is less appealing if we think of an agricultural yield trial; in this case we're imagining that the trial could have been done in many similar locations, in many similar years. Yet since we can't repeat the same conditions, the interpretation of 'frequency' becomes difficult; there is always a hypothetical aspect to the argument. Does this have any meaning with events such as the probability of a large meteor hitting the Earth within the next ten years, or the probability that the human species will make itself extinct within the next ten years?

In this view, the observed data from an experiment represent a sample from a population that has some fixed but unknowable parameters. For example, we have evaluated the transformation sigma of a set of air photos with a set of GPS measurements; if we decide these represent all possible photos and GPS readings that could have been taken on the day, this is the population about which we'd like to make some statement, for example, how successful is the projective transformation in the study area.

Frequentist and Bayesian approaches agree exactly in some situations:

1. Uninformative prior probabilities of various outcomes; or
2. Exactly-known (objective) prior probabilities of various outcomes.

7.5.1 Frequentist hypothesis testing

A common use of frequentist inference is to decide whether a hypothesis is probably true or false. More strictly, the frequentist can give the probability that rejecting a given hypothesis is an incorrect decision. This has a clear interpretation for the decision-maker: it's the chance of making a wrong decision. It also has an interpretation for the scientist: the chance of making an incorrect statement about nature.

7.5.2 The null and alternate hypotheses

Frequentist reasoning distinguishes the null and alternate hypotheses:

- The **null** hypothesis H_0: Not rejected until proved otherwise ("innocent until proven guilty"); if the evidence is not strongly against this, we can't reject it.
- The **alternate** hypothesis H_1: Something we'd like to prove, but we want to be fairly sure

A classic example of a null hypothesis is that a new crop variety does not have a higher yield than the currently-grown variety. The alternative in this case is that it does; note that this is a **one-tailed** alternate hypothesis because we don't care whether or not the new variety is worse.

On the other hand, we might have an **informative** null hypothesis; this is where some ideas from the Bayesian viewpoint are incorporated. For example, many studies may have shown that wood from hardwood species are denser than softwoods, so if we are repeating the study in a new area, we'd be quite surprised if the softwoods turned out to be denser. The null hypothesis then would be that the hardwoods are denser, unless proven otherwise; we might even use a specific numerical difference as the null hypothesis.

7.5.3 Significance levels and types of error

In frequentist tests we need to quantify the risk of making an incorrect inference. These are of two types:

- α is the risk of a **false positive**: rejecting the null hypothesis when it is in fact true; this is called **Type I** error;
  - "The probability of convicting an innocent person" (null hypothesis: innocent until proven guilty)
- β is the risk of a **false negative**: not rejecting the null hypothesis when it is in fact false), this is called **Type II** error.
  - "The probability of freeing a guilty person"

- The quantity $(1 - \beta)$ is called the **power** of a test to detect a true positive.

The following matrix shows how these kinds of error arise from the decision which we take and the truth of the matter (which of course we don't know):

	Null hypothesis H_0 is really ...	
Action taken	True	False
Reject	Type I error committed	success
Don't reject	success	Type II error committed

Note that in strict frequentist thinking we can never "accept" it; all we can say is that we don't have sufficient evidence to reject it. We can never say that it's probably true, only that it's probably not false.[1]

7.5.4 Deciding on a significance level

In frequentist inference, α is set by analyst, whereas β depends on the form of the test, the true difference, and the variance of the data:

- inherent in the phenomenon (uncontrollable), and
- due to imprecise measurements (controllable).

These must be balanced depending on the **consequences** of making each kind of error. For example, if the null hypothesis is that a new crop variety is no better than the current one:

- The cost of introducing a new crop variety if it's not really better, and the lost income in case the new crop is in fact worse (Type I error), vs. ...
- The lost income by not using the truly better variety (Type II error)

This reasoning is mirrored in concepts of law. The British-American legal system is heavily weighted towards low Type I errors (to keep innocent people out of prison, even if some criminals are walking free), whereas the Napoleonic system accepts more Type I error in order to lower Type II error (to keep criminals off the street, even if some innocent people are sitting in prison).[2]

[1] This sort of convoluted reasoning is frequently cited by Bayesians as evidence that the frequentist approach is misguided.
[2] Or maybe the British and Napoleonic systems have opposite null hypotheses about human nature.

Significance levels

Often α is set by convention, or several are reported with conventional levels:

- "Marginally Significant" : $\alpha = 0.1$
- "Significant" : $\alpha = 0.05$
- "Highly Significant" : $\alpha = 0.01$
- "Very Highly Significant" : $\alpha = 0.001$

This can roughly be equated to "sure", "very sure", "extremely sure" that a Type I error is not being committed. Which level we choose to accept is subjective – and here we see that Bayesian 'subjectivity' is not absent from frequentist inference.

7.5.5 Examples of frequentist inference

In the frequentist paradigm, there is one true value of a population parameter, and we try to estimate it from the sample. We compute the "best guess" estimate by some procedure which we justify from the **assumed characteristics of the underlying population**.

The most common inferences are **point estimates**, to infer the true value of a single parameter, such as a population mean or a correlation between two variables. Since we only are estimating from a sample, we can't pin such an estimate down exactly, so we also compute a **confidence intervals**, which is a range having a **known probability** of containing the true value, again under our **assumptions**.

A simple example of point estimation is of the population **mean** or centre of gravity. If we can assume that the n observations we make are from a single population, with (unknown) **identically–** and **independently**–distributed (abbreviation "IID") errors of observation, then the most likely ("expected") value of the true mean is given by the well-known formula:

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{7.4}$$

The notation $\hat{\mu}$ means that we are **estimating** the true population mean μ; whereas $\bar{x}$ is simply shorthand for the right-hand side. So $\bar{x}$ is not an inference, but $\hat{\mu}$ is.

The interval which has probability $(1 - \alpha)$ of containing the true value is:

$$(\bar{x} - t_{\alpha/2,n-1} \cdot s_{\bar{x}}) \;\le\; \mu \;\le (\bar{x} + t_{\alpha/2,n-1} \cdot s_{\bar{x}}) \tag{7.5}$$

where $t_{\alpha/2,n-1}$ is Student's t with $n-1$ degrees of freedom at confidence level $\alpha/2$ and $s_{\bar{x}}$ is the standard error of the mean:

$$s_{\bar{x}} = \frac{1}{\sqrt{n}} \cdot \left[\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2\right]^{1/2} \tag{7.6}$$

Note that the confidence level α, say 0.05, is halved, say to 0.025 for each side of the interval, because this is a two-sided interval. The t-distribution must be used because we are estimating both the mean and variance from the same sample; for reasonably-large sample sizes the normal distribution itself can be used.

The null hypothesis here is that the true mean μ is the value estimated from the sample $\hat{\mu}$. The alternate hypothesis is that the true mean is not this value; outside the confidence interval we can be fairly confident in rejecting the null hypothesis.

Using the geometric correction example above, recall we had 14 values of transformation sigma:

```
4.36 3.63 6.01 3.78 7.58 8.36 5.18
4.77 4.80 7.18 5.79 5.14 5.42 3.81
```

from which we compute the sample mean `5.145` and sample standard error of the mean is `0.403`. These are not yet inferences about the population, only statements about the sample. Then we find the required value of t ($\alpha = 0.025$, 13 degrees of freedom) is 2.160.[3] Then the confidence interval that we assert covers the true mean with only 5% chance that we are wrong is:

$$\begin{aligned}(5.145 - 2.160 \cdot 0.403) \quad \leq \quad & \mu \leq (5.145 - 2.160 \cdot 0.403) \\ 4.274 \quad \leq \quad & \mu \leq 6.015\end{aligned}$$

Now we make the **inferential statement** "With only 5% chance of being wrong, I assert that the mean transformation error is at most 6.015 m".

7.5.6 The variability of small samples

Figure 7.1 shows an example of inference: four samples of size 30 were drawn from a known normal distribution[4] and then we attempted to infer the true mean and standard deviation, which in this case was known. The four random samples gave estimates from 177 to 184.3 for the true mean (180) and 16.5 to 20.0 for the true standard deviation (20). This is typical of inferences from small samples. In this simulation we can draw as many samples as we wish, but in a field experiment where we are again assuming a true mean and standard deviation, and even assuming the distribution of the variable, we can not easily repeat the experiment, and certainly not in the exact same conditions.

[3] R code: `qt(.975,13)`
[4] R code: `rnorm(30, 180, 20)`

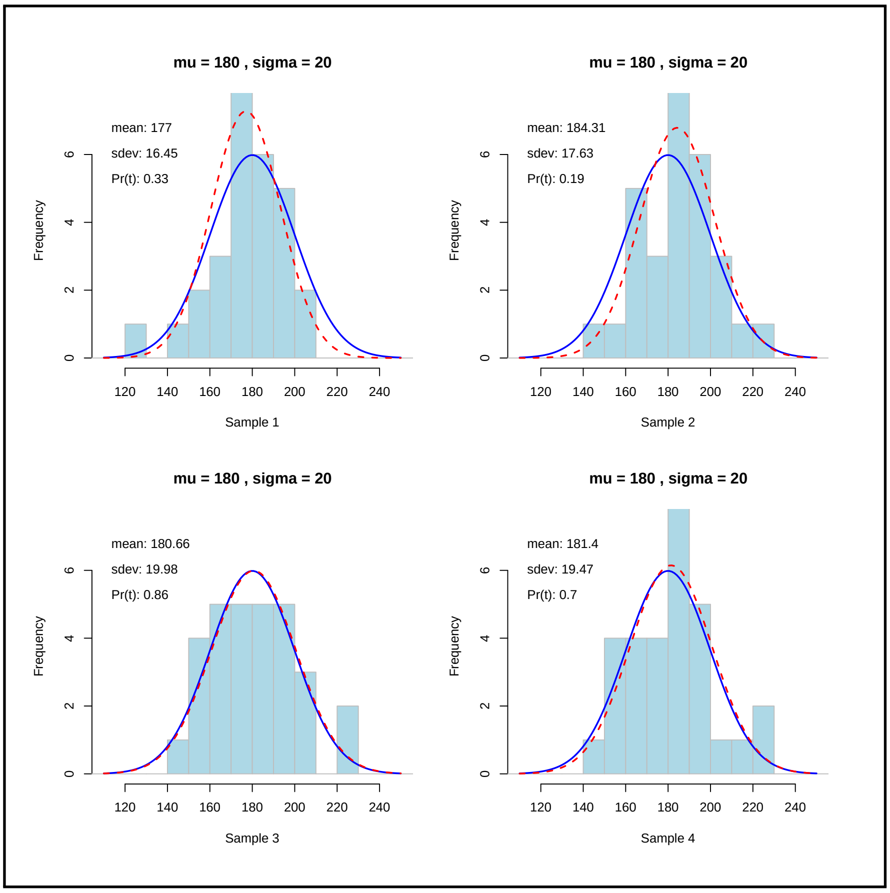

Figure 7.1: Four estimates of the parameters of a known population

7.5.7 What do we really mean by a 'confidence interval'?

In the frequentist view, the confidence interval for a parameter is said to **cover** the its true value with some probability p. This means that if we would (or could) repeat the procedure many times, in that proportion of the cases the realised confidence interval would contain the true value of the parameter. For example, if $\alpha = 0.05$, in 95% of the hypothetical repetitions of the sampling the computed interval would in fact contain the true (but unknown) value. Looking at this from the other side, the one realised confidence interval we have from our one sample has probability $(1 - p)$ that it **does not** contain the true value; that is the risk of a Type I error.

Map accuracy assessment

An appropriate statement for map accuracy assessment based on a binomial test of ground truth vs. mapped class might be:

> "This land cover map was made primarily by manual and automatic interpretation of a satellite image, so that most locations have not been visited. However, a representative sample of locations was visited to assess the thematic accuracy

> of this map, which is reported as the proportion of 15x15 m ground locations areas where the land cover as reported on the map agrees with the actual dominant land cover. Under the assumption that the ground truth locations were representative of all the possible samples that could have been chosen, we used the binomial distribution to calculate a 90% 'confidence interval', which gives a minimum and maximum accuracy. These intervals as reported here have a nine-in-ten chance of containing the true accuracy. If we had been able to take a large number of similar samples, 90% of the confidence intervals calculated from these would have contained the true accuracy. We have no way of knowing whether the one sample we did take is one of those 90% or one of the 10% where the computed confidence interval, reported here, does not contain the true accuracy. So, there is a one-in-ten chance that the true accuracy is outside the interval we report here."

Specifically for this example, given an unbiased sample of size n, with n_t successes, the proportional accuracy and its standard deviation are estimated as parameters of the binomial distribution by [18]:

$$\begin{aligned} p &= n_t/n && (7.7)\\ s &= \left[\frac{p \cdot (1-p)}{n}\right]^{1/2} && (7.8) \end{aligned}$$

If the sample size is large enough[5], the confidence interval of the estimate may be approximated as:

$$p \pm \left[s \cdot Z_{1-\alpha} + \frac{1}{2n}\right] \tag{7.9}$$

where $Z_{1-\alpha}$ is the two-tailed normal score for the probability of non-coverage α; this can be obtained from tables or computed in software. The factor $1/(2n)$ is a small-sample correction The lower and upper limits as computed by Equation 7.9 are truncated at 0 and 1, respectively, if necessary.

To be specific, suppose we have $n = 163$ total ground truth samples, of which $n_t = 86$ are correctly classified. Then $p = 0.5276$ and $s = 0.0391$. To limit the probability of non-coverage to 5%, the corresponding area under the normal curve is $\Pr = 0.95$, which is obtained for the two-tailed test with $Z = 1.96$[6], so that the 95% confidence interval for p is $[0.4479 \ldots 0.6073]$. This is interpreted to mean that if we had repeated the same sampling scheme a large number of times, in 95% of these samples the observed accuracy would be somewhere between 44.8% and

[5] For small samples, especially if p is near 0 or 1, the confidence interval must be determined directly from the binomial distribution as explained by Rossiter [18].

[6] R code: `qnorm(.975)`

60.7%. We are taking a 5% risk that the true proportion is $< 44.8\%$ or $> 60.7\%$.

(Note that we can narrow the confidence interval at the expense of a higher risk of non-coverage. For example, increasing this risk to 10%, we obtain $Z = 1.64$[7] and an interval for p of $[0.4602 \ldots 0.5950]$, i.e. about 2.5% narrower. Increasing the risk to 20%, i.e. a one in five chance of the true value being outside our calculated interval, we obtain $Z = 1.28$ and an interval for p of $[0.4744 \ldots 0.5808]$, now 5.3% narrower.)

7.6 Building a statistical model

7.6.1 A provocative example

Before investigating the nature of statistical modelling in depth, we review a paper that, we hope, was presented by its authors as a cautionary example. The reader may also want to come back to this example after reading further.

Figure 7.2 is taken from an article in *Nature*, generally considered one of the top two general scientific journals[8], titled “Momentous sprint at the 2156 Olympics? Women sprinters are closing the gap on men and may one day overtake them” [21].

The article uses sophisticated statistical methods[9], with extensive internal error checks, to show that women sprinters (at the 100 metre distance) are improving faster than men sprinters, and further argues that women will overtake men, most likely at the 2156 Olympic games.

Figure 7.2 is a scatterplot of winning time vs. year of competition. The blue points are men’s past performances, the blue solid line is the best-fit statistical model to past performances, and the blue dashed line is this line extrapolated to the future; the lines also have confidence intervals shown by black dashed lines, within which 95% of the true values are expected to fall. Red points and lines show the corresponding information for women. The statistical methods are sound (follow best practices), and many alternate models were investigated (for example, second-order or negative exponential models that would show a change in the rate of improvement), without changing the statistical inference: a separate linear fit for men and women, as shown.

> Here are some facts that may or may not be of use in answering the following questions:
>
> - No games were held in 1916, 1940 or 1944.

[7] R code: `qnorm(.95)`
[8] The other is *Science*
[9] including a contribution by a current ITC staff

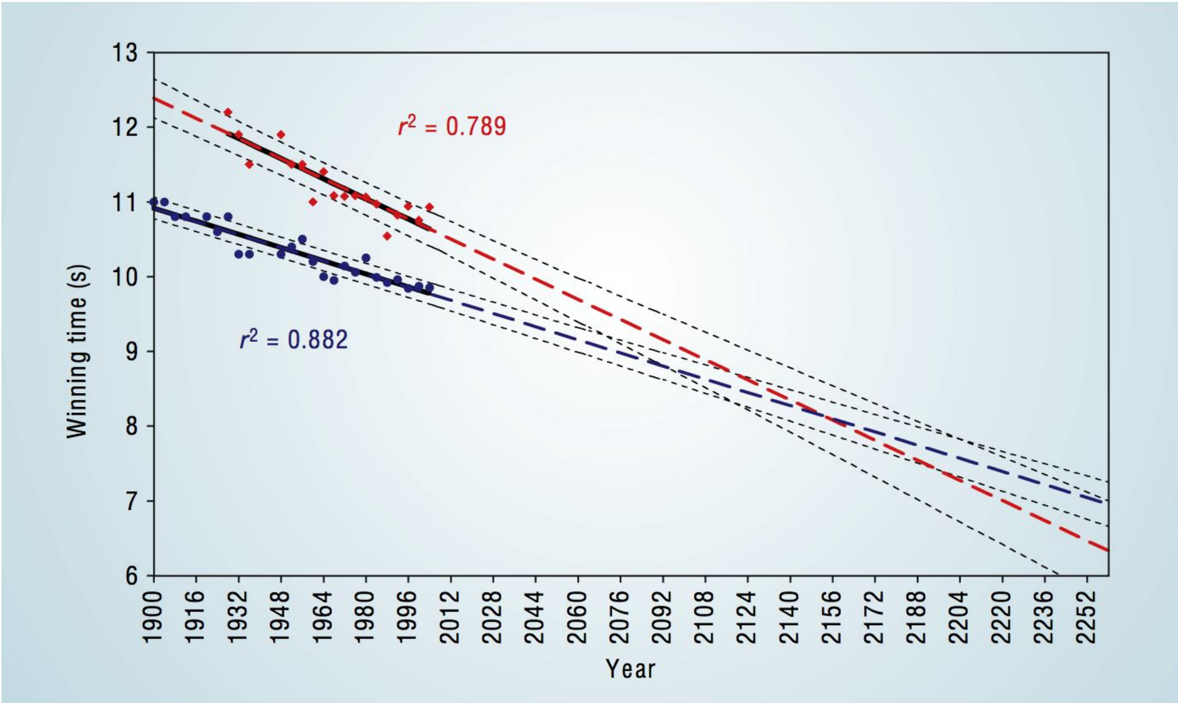

Figure 1 The winning Olympic 100-metre sprint times for men (blue points) and women (red points), with superimposed best-fit linear regression lines (solid black lines) and coefficients of determination. The regression lines are extrapolated (broken blue and red lines for men and women, respectively) and 95% confidence intervals (dotted black lines) based on the available points are superimposed. The projections intersect just before the 2156 Olympics, when the winning women's 100-metre sprint time of 8.079 s will be faster than the men's at 8.098 s.

Figure 7.2: "A momentous sprint" (from[21])

- Women did not compete at 100 m until 1928.
- Cinder tracks were used until the Mexico City games of 1968.
- Electronic timing with a precision of 0.01 s was introduced in 1968; until then hand times were rounded up to the nearest 0.1 s. Hand times were converted to equivalent electronic times, compensating for the well-known delay at the start.
- The 1968 games were held at high altitude (Mexico City) which is known to favour sprinters.
- Doping (performance-enhancing drugs) are generally acknowledged to have been widespread from the 1970's (especially "east-block" athletes) through the 80's and 90's; much stronger doping controls were devised before the 2004 games, although there have been revelations of acknowledged users who were not caught;
- Some remarkably-fast performances: Men: Eddie Tolan (1932) and Jesse Owens (1936), both 10.3; "Bullet" Bob Hayes (1964) 10.0; Women: Wilma Rudolph (1960): 11.0; Florence Griffith (1988) 10.54.
- Some remarkably-slow performances: Men: none; Women: Fanny Blankers-Koen (1948) 11.9[10], Nesterenko (2004) 10.93.

Q59 : *The statistical model shows an improvement for men of* 0.044 ± 0.003 s *per games; for women* 0.067 ± 0.009 s. *What could be some reasons for the overall improvement in both?* *Jump to A59* •

Q60 : *The women's improvement is provably ("significantly") better than the men's. What could be some reasons for this difference?* *Jump to A60* •

Q61 : *The authors state "Should these trends continue, the projections will intersect at the 2156 Olympics, when . . . the winning women's 100-meter sprint time of 8.079 seconds will be lower than that of the men's winning time of 8.098 seconds."*

[10] A Dutch national hero from Hengelo, she won four gold medals at these games

Given your previous two answers, do you expect "these trends to continue" to the point where the projections ever intersect? Explain why or why not. *Jump to A61* •

Q62 : *Were the authors justified in presenting Figure 7.2? Was the journal (reviewers and editors) justified in accepting this paper?* *Jump to A62* •

7.6.2 The nature of statistical models

Every inference we make is based on an underlying statistical model. For example, an inference about a population mean depends on the assumed distribution of the variable (normal, log-normal, Poisson, Weibull ...). There are four steps:

1. Selecting a **functional form**, i.e. the model to be fitted;
2. Determining the **parameters** of the model; this is called **calibration** or **parameter estimation**;
3. Determining how well the model describes reality; this is called **validation**.
4. **Criticising** (re-examining) the assumptions and possibly **iterating** from step 1.

Figure 7.3 shows the basic paradigm of statistical modelling. Note the feedback from first results to model adjustment; this is mostly a **criticism** of the model form.

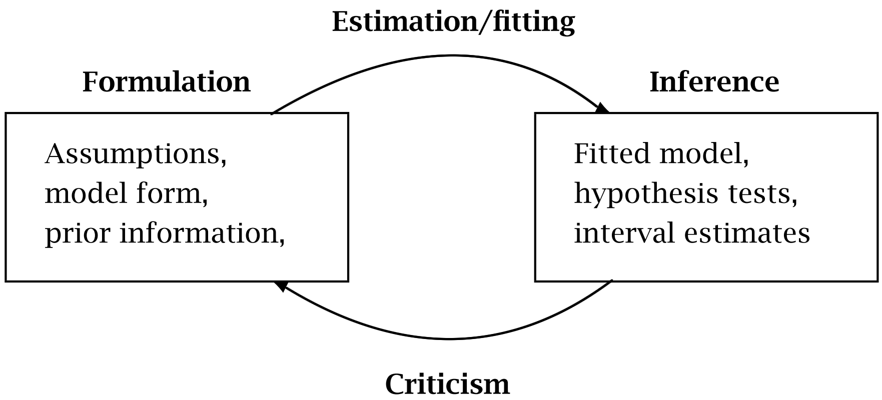

Figure 7.3: Schematic outline of modelling (Cook & Weisberg [4, Fig. 1.2.1])

7.6.3 Structure and noise in statistical modelling

The following conceptual equations show the inferences we are making:

- Observations = f(Structure, Noise)
- Observations = f(model, unexplained variation)
- Observations are a subset of Reality, so ...
- Reality = f(Structure, Noise)
- Reality = f(deterministic processes, random variation)

The aim is to match our **model** with the true **deterministic process** and match our estimate of the **noise** with the actual **random variation**. It is equally an error to model the noise (**over-fit** the model) as to not model the process (**under-fit** the model).

7.6.4 Evidence that a model is suitable

For most datasets a numerical solution can be computed for many functional forms. The question naturally arises as to whether it should be. In other words, is a model **meaningful** or **applicable**?

There are two levels of evidence:

1. **external** to the model:
   (a) what is known or suspected about the process that gave rise to the data; this is the connection to the reality that the model is trying to explain or summarise;
   (b) how well the model fits further data from the same population: success of **validation** against an independent dataset
2. **internal**: from the model itself:
   (a) how well the fitted model meets the **assumptions** of that functional form, e.g. by examination of regression diagnostics (§7.7.6).
   (b) how well the model fits the data (success of **calibration**, i.e. parameter estimation);

For example, the set of errors associated with georeferencing a satellite image from control points identified on a topographic map would seem to conform to the model of many small, independent errors[11] that we know (from theory) give rise to a normal (Gaussian) distribution. So it makes sense to estimate the standard deviation (so-called "sigma") of that distribution, to evaluate the average size of these errors and therefore the quality of the transformation.

[11] map compilation and printing, image distortion, map registration to digitiser, ...

However, even in this example we may find evidence that the errors are not independent:

- the distribution of individual errors across the image does not seem to be random → georeference sections of the image separately?
- the distribution of individual errors does not seem to be fitted by a normal distribution → use a different transformation? exclude some points? (but on what basis?)

This last point highlights the **assumption** underlying the Gaussian model: errors are all the result of small, random processes. If we make a gross error (e.g. mis-identify a road intersection on the image with one several km away) this is a different kind of error, which violates the model, and that is why we are justified in eliminating it, once it is identified.

Another example is a time-series of the area occupied by various land uses in a study area. A first look may suggest a steady de-forestation: a **linear** function of area vs. time. The parameter estimation is then to determine the linear rate of change. However, the deforestation may be speeding up (model form a **higher-order polynomial**, or even **exponential**), or it may be slowing down (same forms, but with reverse signs), or there may have been some **threshold point** where the underlying causes changed (stricter land use regulations? establishment of a protected zone? in-migration?) so that some **piecewise** function is more appropriate.

If the functional form is not appropriate, the following steps are invalid.

7.6.5 Model calibration vs. model validation

Once a functional form is chosen, the process of fitting a model to observed data is **calibration**, that is, the model parameters are adjusted ('calibrated') to best fit the available experiments. This is also called **parameter estimation**. In the case of regression, this is part of developing the equation. This yields a goodness-of-fit measure such as R^2 (the **coefficient of determination**), which expresses how well we were able to match the model to the data. This is the complement of the **residual sum of squares** (RSS) as a proportion of the **total sum of squares** (TSS):

$$\begin{aligned} R^2 &= 1 - \frac{\mathrm{RSS}}{\mathrm{TSS}} \\ RSS &= \sum_{i=1}^{n} (z_i - \hat{z}_i)^2 \\ TSS &= \sum_{i=1}^{n} (z_i - \bar{z})^2 \end{aligned}$$

where $\hat{z}_i$ is the predicted (modelled) value and $\bar{z}$ is the sample mean. If there are no residuals, RSS = 0 so that $R^2 = 1$; if the model explains nothing, RSS = TSS so that $R^2 = 0$. However, this only measures how well the model fits the data set, i.e. how well it is calibrated to this data.

Once a functional form is selected, we estimate its parameters by formulas that were developed **assuming** the functional form is correct, e.g. maximum-likelihood estimators. For example, having decided on a simple linear regression, we must estimate the slope and intercept of the best-fit line; the maximum-likelihood method if all errors are independent and identically-distributed is least-squares. This is **model calibration**, also known as **parameter estimation**.

Postdiction vs. prediction

Another name for calibration is **postdiction** (as opposed to prediction), from the Latin 'post' (after) and 'dicere' (to say). This allows us to use the past (already observed) to make probabilistic statements about the how well the observations are explained by the calibrated model. If the observations were representative of a population, we would expect to obtain the same parameters, within experimental and observational error, in similar repeated studies. However, there is no way to be sure that, because we can't in general re-do the study. We can compare the predicted and actual values of our one sample, to see how well they match; this is the **goodness-of-fit** with respect to the sample. This tells us how well the model can match the sample, but it says little about how well it would match other similar samples. An example is the reported coefficient of determination (R^2) from a regression; this is a measure of the success of **calibration** (postdiction).

If we have a second independent sample, we can compare its values with what the model predicts. Note that the model calibration procedure did not use these observations, so this is an independent test, which can fairly be termed **validation.**

There are several measures of validity:

- **Root mean squared error** (RMSE) of the **residuals**: the **actual** (observed) less the **estimate** (from the model) in the **validation** dataset; lower is better:

$$\mathrm{RMSE} = \left[\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \right]^{1/2}$$

- **Bias** or mean error (ME) of estimated vs. actual mean of the **validation** dataset; should be zero (0):

$$\mathrm{ME} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)$$

- **Gain** of the least-square fit of estimated vs. actual data; this should be 1, otherwise the estimate does not increase at the same rate as the actual data.

These can be visualised by plotting fitted vs. actual values on the same scale, with a 1:1 line (Figure 7.4). The residuals are the vertical distances from this line:

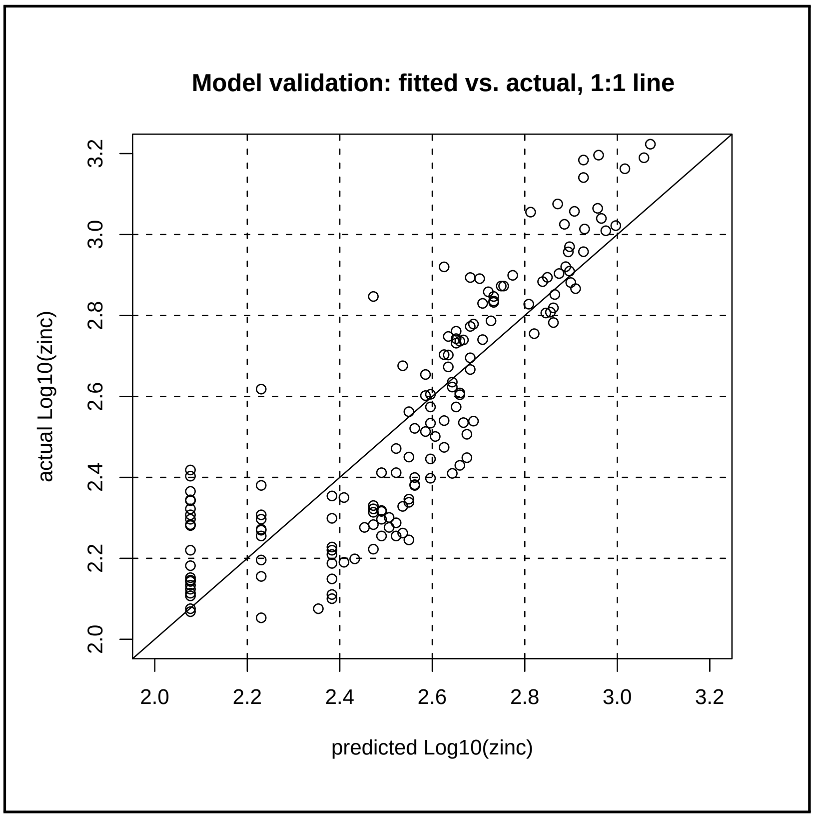

Figure 7.4: Validation with an independent dataset

The null hypothesis of no bias (intercept 0) or gain (slope 1) (i.e. the model is valid) can be tested by fitting a regression model to the actual vs. fitted values and simultaneously testing the two regression parameters of this model with an F-test [15]. A simpler approach is to consider the tests of each parameter separately in the regression analysis of variance table.

7.6.6 Unimodal vs. multimodal populations

We can always compute a sample mean; this is just a summary statistic. But we typically do so to **infer** the mean of the population of which the sample is representative. But, how do we know our sample comes from a population with only one central tendency?

For example, is it helpful to speak of “the mean” of the 400-observation sample whose histogram is shown in Figure 7.5?

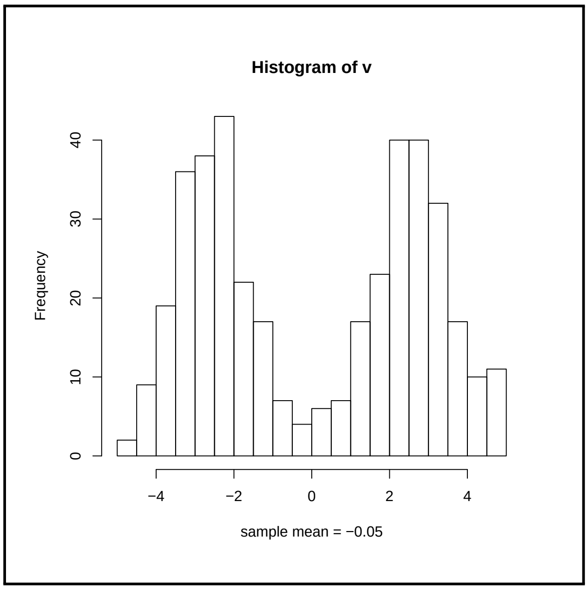

Figure 7.5: Histogram of a bimodal sample

It seems likely that there are two distinct populations in this1 sample[12],so that we'd like to estimate are the means of the two populations, not the (meaningless) overall mean.

This becomes even clearer if we imagine calculating a confidence interval for the population mean.

7.7 Conceptual issues in correlation and regression

Correlation is some measure of the co-relation of two or more things; in statistics this is an **inter-dependence** of two or more variables. Correlation is a **symmetric** relation: the several variables being examined are considered equally.

Regression is a term invented by Galton [8] in 1885, to describe the tendency of the mean value of an inherited characteristic (in his example, height of children) to be closer to the overall mean of the group (e.g. heights of all children) than to the value of the parents. He termed this "regression" [in the sense of 'going backwards'] "to the mean": tall parents have children who are, on average, shorter and vice-versa[13]. This first application of the word shows clearly the **non-symmetric** nature of regression: one of the variables (in this example, height of the children) is being analyzed in terms of the other (in this example, height

[12] In fact there *are* two populations; this sample was created with the R code `v<-c(rnorm(200,-2.5,1),rnorm(200,2.5,1))`

[13] So, every generation regresses to the mean, but the overall distribution does not change.

of the parents). The two variables are often called **dependent** and **independent**; these are statistical terms, not (without further argument) causative.

Q63 : *In Galton's example, why is the dependent variable caused by the independent, and not the other way around, or not by some common cause?* *Jump to A63* •

The techniques invented by Galton were soon applied to many other areas besides genetics. The simple linear modelling used by Galton was extended, so that today the term **regression** is used as a near-synonym of any form of linear modelling or even the analysis of any form of dependence of one random variable on others.

Correlation and various kinds of regression are often misused. There are several articles that explain the situation, with examples from earth science applications [14, 23]. A particularly understandable introduction to the proper use of regression is by Webster [24], whose notation we will use.

7.7.1 Correlation vs. causation

A fundamental distinction must be made between two of concepts:

1. The **relation** between two or more variables, often described mathematically as the **correlation** ('co-relation');
2. The **causation** of one variable by another, often described by **regression** techniques.

This second is a much stronger relation than the first. The issue of causation also includes some **conceptual model** of how the two phenomena are related. Statistics can never prove causation; it can only provide evidence for the strength of a causative relation supported by other evidence. Thus we must always make a **meta-statistical** argument about causation.

7.7.2 Description vs. prediction

Regression analysis can be used for two main purposes:

1. To **describe** a relation between two or more variables, whether the relation is supposed to be causative or not;
2. To **predict** the value of a variable (the **predictand**, sometimes called the **dependent** variable or **response**), based on one or more other variables (the **predictors**, sometimes called **independent** variables.

These can lead to different inferential procedures.

A statistical model that does not assume causation can still be useful for prediction. For example, the prevalence of two plant species may be correlated, so that we can develop a model to predict the presence of one from the presence of the other, without having to make any argument that the presence of one in any way "causes" the presence of the other. (In fact, we are more likely to argue that these have a common cause.) So we can have a regression equation that we use for prediction, not at all based on any notion of causation.

7.7.3 Types of models

A **simple** correlation or regression relates two variables only; a **multiple** correlation or regression relates several variables at the same time. Modelling and interpretations are much trickier in the multivariate case, because of the inter-relations between the variables.

A **linear** relation models one variable as a linear function of one or several other variables. That is, a proportional change in the predictor results in a proportional change in the predictand or the modelled variable. Any other relation is **non-linear**.

Non-linear relations may be **linearisable** by means of a transformation of one or more variables, but in many interesting cases this is not possible; these are **intrinsically non-linear**.

7.7.4 Fixed vs. random variables

A distinction is made between predictors which are known without error, whether fixed by the experimenter or measured, and those that are not. Webster [24] calls the first type a "Gauss linear model", because only the predictand has error, and the predictor a **mathematical** variable, as opposed to a **random** variable which is measured with error. The regression goes in one direction only, from the mathematical predictor to the random response, and is modelled by a **linear model with error**:

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

There is no error associated with the predictors x_i, only with the predictand y_i. Thus the predictors are assumed to be known without error, or at least the error is quite small in comparison to the error in the model. An example of this type is a designed agricultural experiment where the quantity of fertiliser added (the predictor) is specified by the design and the crop yield is measured (the predictand); there is random error ε_i in this response.

An example of the second type is where the crop yield is the predictand, but the predictor is the measured nutrient content of the soil. Here we are modelling the relation as a **bivariate normal distribution** of two random variables, X and Y with (unknown) population means μ_X and μ_Y, (unknown) population variances σ_X and σ_Y, and an (unknown) correlation ρ_{XY} which is computed as the standardised (unknown) covariance $\mathrm{Cov}(X, Y)$:

$$
\begin{aligned}
X &\sim \mathcal{N}(\mu_X, \sigma_X) \\
Y &\sim \mathcal{N}(\mu_Y, \sigma_Y) \\
\rho_{XY} &= \mathrm{Cov}(X, Y)/\sigma_X\sigma_Y
\end{aligned}
$$

In practice, the distinction between the two models is not always clear. The predictor, even if specified by the experimenter, can also have some measurement error. In the fertiliser experiment, even though we specify the amount per plot, there is error in measuring, transporting, and spreading it. In that sense it can be considered a random variable. But, since we have some control over it, the experimental error can be limited by careful procedures. We can not limit the error in the response by the same techniques.

7.7.5 Structural Analysis

The regression of two variables on each other depends on which variables is considered the predictor and which the predictand. If we are predicting, this makes sense: we get the best possible prediction. But sometimes we are interested not in prediction, but in **understanding** a relation between two variables. This so-called **structural analysis** is explained in detail by Sprent [20] and more briefly by Webster [24] and Davis ([5, pp. 214–220] and [6, pp. 218–219]).

In structural analysis we are trying to establish the best estimate for a **structural** or **law-like** relation, i.e. where we hypothesise that $y = \alpha + \beta x$, where both x and y are mathematical variables. This is appropriate when there is no need to predict, but rather to understand. This depends on the prior assumption of a true linear relation, of which we have a noisy sample.

$$
\begin{aligned}
X &= x + \xi \\
Y &= y + \eta
\end{aligned}
$$

That is, we want to observe X and Y, but instead we observe x with random error ξ and y with random error η. These errors have (unknown) variances σ_ξ^2 and σ_η^2, respectively; the ratio of these is crucial to the analysis, and is symbolised as λ:

$$
\lambda = \sigma_\eta^2/\sigma_\xi^2 \tag{7.10}
$$

Then the maximum-likelihood estimator of the slope, taking Y as the predictand for convention, is:

$$\hat{\beta}_{Y.X} \quad = \quad \frac{1}{2s_{XY}}\left\{(s_Y^2 - \lambda s_X^2) + \sqrt{(s_Y^2 - \lambda s_X^2)^2 + 4\lambda s_{XY}^2}\right\} \qquad (7.11)$$

Equation 7.11 is only valid if we can assume that the **errors in the two variables are uncorrelated**. The problem is that we don't have any way of knowing the true error variance ratio λ, just as we have no way of knowing the true population variances, covariance, or parameters of the structural relation. We estimate the **population** variances σ_X^2, σ_Y^2 and covariance σ_{XY} from the sample variances s_x^2, s_y^2 and covariance s_{xy}, but there is nothing we've measured from which we can estimate the **error** variances or their ratio. However, there are several plausible methods to estimate the ratio:

- If we can assume that the two error variances are equal, $\lambda = 1$. This may be a reasonable assumption if the variables measure the same property, use the same method for sampling and analysis, and there is an *a priori* reason to expect them to have similar variability (heterogeneity among samples).
- The two **error** variances may be estimated by the ratio of the **sample** variances: $\lambda \approx s_y^2/s_z^2$. That is, we assume that the ratio of variability in the measured variable is also the ratio of variability in their errors. But, these are two completely different concepts! One is a sample variance and the other the variance of the error in some random process.
- The variance ratio may be known from previous studies.

Figure 7.6 shows the large difference that may result from viewing one variable as a function of the other or vice versa, compared to the structural relation between two correlated variables.

7.7.6 Selecting the correct regression model

A classic example is provided by Anscombe [1], who developed four different bivariate datasets, all with the exact same correlation coefficient $r = 0.81$ and linear regression equation $y = 3 + 0.5x$ (Figure 7.7).

The question is whether the linear regression model, i.e. that the value of y depends linearly on x, is applicable. Second, whether the least-squares estimate of the regression coefficients gives a correct summary of the relation.

Q64 : *For each of the four Anscombe relations shown in Figure 7.7, state whether (1) a linear regression is appropriate; (2) the least-squares*

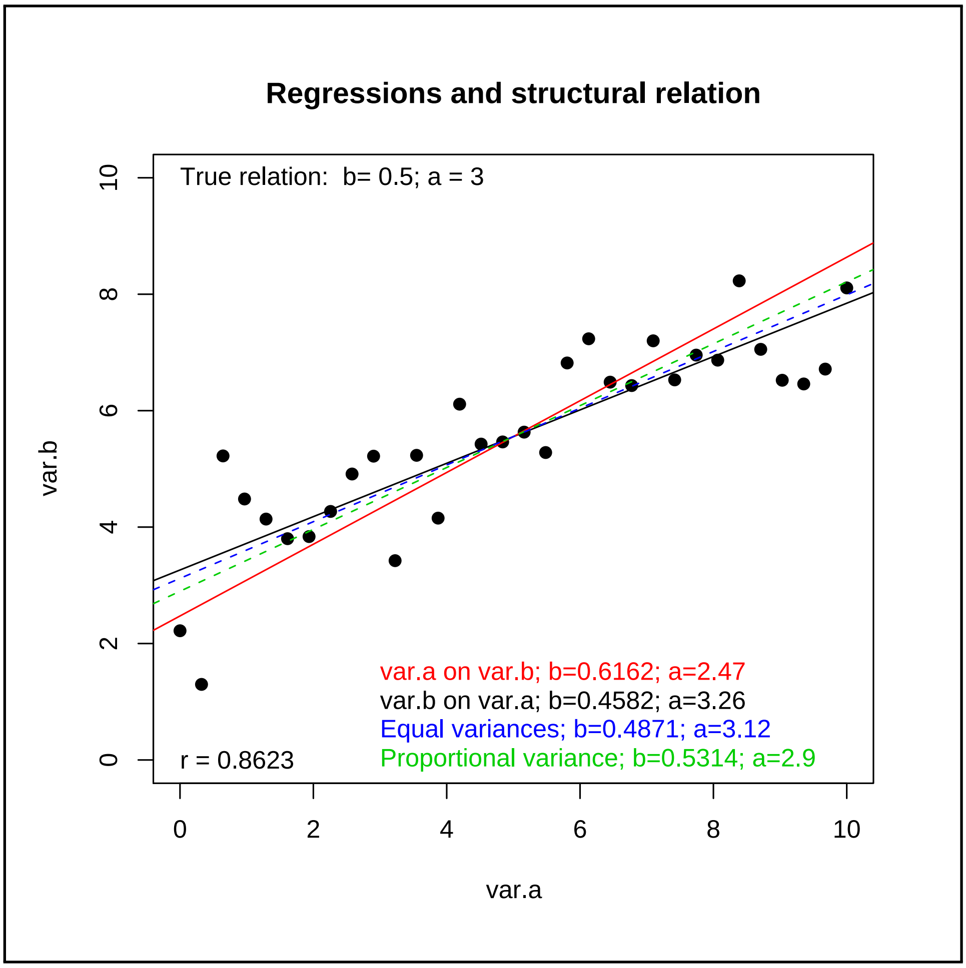

Figure 7.6: Regression in two directions compared to structural relations

estimate of the regression coefficients gives a correct summary of the relation. *Jump to A64* •

How do we know that the chosen model is appropriate?

1. From *a priori* knowledge of the process;
2. From **internal evidence** when we try to fit the model.

In the second case there are many so-called **regression diagnostics** with which we can evaluate how well the model satisfies its assumptions. A common set of diagnostics examines the **residuals**, that is, the discrepancy of each fitted point from its observation. If any are unusually large, it may be that the observation is from a different population, or that there was some error in making or recording the observation. If large residuals are associated with large values, this is evidence of **heteroscedasticity** (i.e. variance is not constant across the range of the predictor). Texts on regression [e.g. 7] explain these in detail.

Figure 7.8 shows an example of a regression diagnostic for the Anscombe

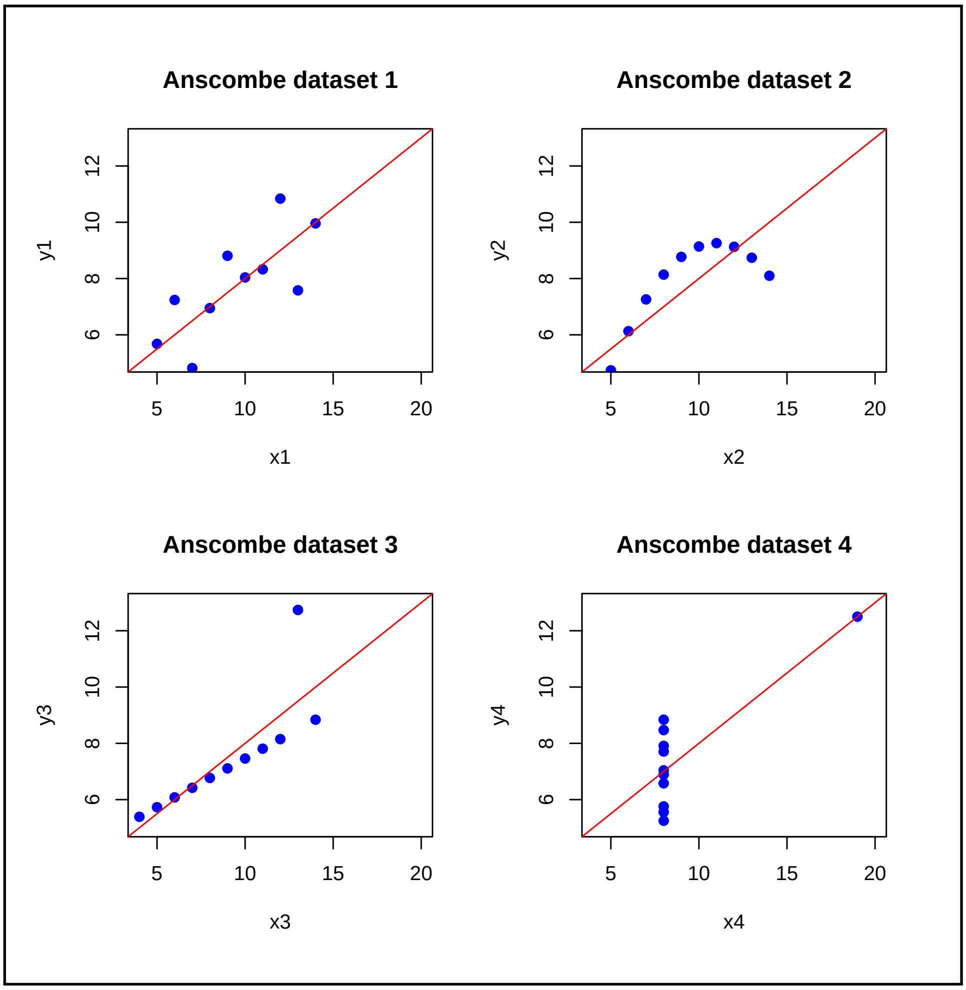

Figure 7.7: Anscombe's bivariate regression examples

data. The 'diagnostic' here is that the residuals should show no relation to the fitted value; we can see that is the case in regression 1 (the 'normal' case) but violated badly in all the others. This gives evidence that the selected model was not correct.

7.7.7 Parsimony

This is a technical term used in statistics to express the idea that the simplest relation that explains the data is the best. Gauch Jr. [9] gives an accessible introduction to this concept. It is especially applicable in **multiple regression** models, where the model can be made increasingly complex, apparently explaining more and more of the dataset (as measured by the unadjusted R^2).

However, after a certain point the more complex model is explaining the **noise** (experimental error), not the **relation**. Even with only one predictor, it is always possible to fit n data points exactly by using a polynomial

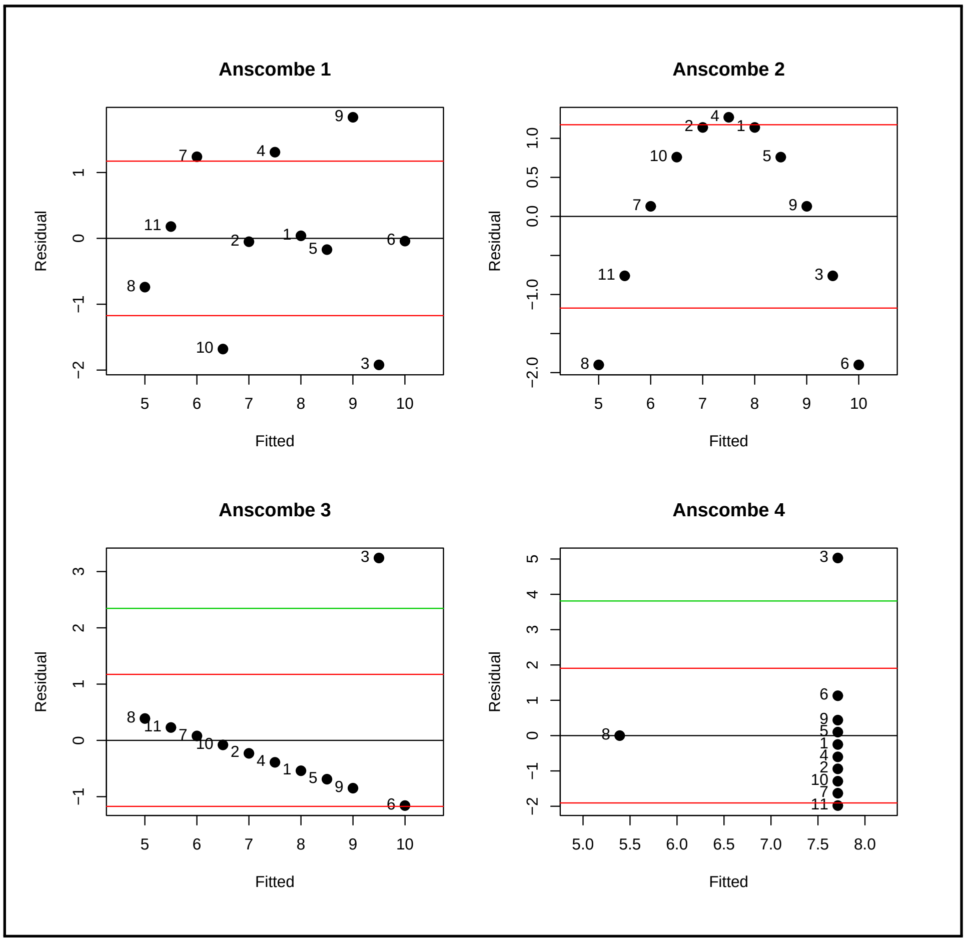

Figure 7.8: Anscombe's bivariate regression examples; residuals vs. fits

of degree $n-1$. This effect is shown in Figure 7.9. The points should all lie on the dashed line (the true relation), but because of experimental error they deviate from it with error mean 0 and standard deviation 3; each experiment will have a different error. The best fits to two different sets of points for increasing polynomial degree are shown. Note that the underlying relation is the same. Also note that the lower-order (linear) fits are similar for both noisy datasets, but the higher-order fits differ greatly, as each fits its own noise, rather than the structural relation.

One measure, which applies to the standard linear model, is the **adjusted** R^2. This decreases the apparent R^2, computed from the ANOVA table, to account for the number of predictive factors:

$$R^2{}_{\text{adj}} \equiv 1 - (1 - R^2)\frac{n-1}{n-k-1}$$

That is, the proportion of variance not explained by the model $(1 - R^2)$ is **increased** with the number of predictors k. As n, the number of observations, increases, the correction decreases. A more general measure, which can be applied to almost any model type, is Akaike's **An Informa-**

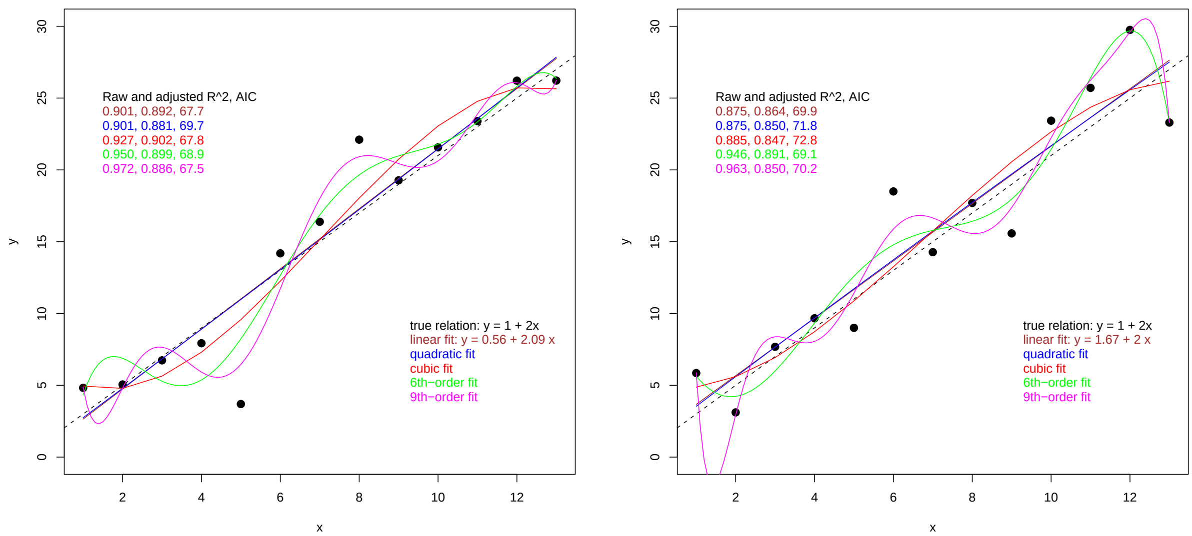

Figure 7.9: Fitting noise with higher-order polynomials

tion Criterion, abbreviated AIC. The lower value is better.

7.8 Answers to self-test questions

A58 : *(a) The population certainly includes all the damaged and then reconstructed buildings in a defined study area, after a particular disaster (tsunami, earthquake, flood ...), But, it would be more interesting to be able to make a more general statement, so the population could be all damaged and then reconstructed buildings anywhere after any disaster; or limited to one kind of disaster; or limited to one type of study area.*

(b) If there is a list of reconstructed buildings (from a municipal council or a disaster relief agency), this would be a suitable sampling frame. The researcher would selected some buildings from this list (using some sampling design) and determine whether the image processing methods could identify them.

Another sampling frame might be all the buildings, whether affected or not; the researcher would draw a sample of these, assess their reconstruction status (either from the municipal list or field survey) and then assess these by image processing. *Return to Q58* •

A59 : *Among the possible causes are: improved facilities (better tracks),* *improved and more intensive training methods, improved biomechanics, improved peaking methods, improved nutrition.*

Another possible cause is the increased participation and professionalism. Early games were restricted informally to well-to-do European amateurs. As the games became more popular, other groups participated: African- and Jewish-Americans (famously in 1936), Caribbean nations (well-known for their sprinting skills), and people who were not wealthy and could make a living as professional athletes. As the world economy improved after the Second World War, and as travel became more common, more potential athletes were identified and were able to participate.

It is also possible that humans are steadily getting faster by evolution; however within a time of only 104 years it seems unlikely. Return to Q59 •

A*60* : *All the factors from the previous answer played a role, and it is easy to argue that they were more serious for women than men.*

Women started later (28 fewer years to improve); training for women was actively discouraged until the rise of socialist sports and improved programmes at historically-Black colleges[14] *in the USA, both in late 1950's and early 1960's; coaches had little experience with female physiology, biomechanics, and training methods; in the era of drug use, synthetic male hormones probably have more effect on women's sprint performances than on men's. Professionalism for women athletes came later than for men, although in socialist countries they were equally-supported.* Return to Q60 •

A*61* : *The fundamental flaw is identified by the authors themselves: "should these trends continue ...". Many of the trends have already run their course: women's participation in sprinting and their training is probably equal to men's.*

Certainly the trends can not continue indefinitely (to zero or negative times!) but perhaps there are new factors that may allow continued improvement, and even differential improvement for men and women. For example, performance-enhancing drugs and biomechanical surgery could surely improve performances, were they allowed. Return to Q61

•

A*62* : *The authors were not justified in presenting a logical absurdity that is not supported by non-statistical evidence. In particular, for both men and women, there must be a physiological limit set by biomechanics and energy conversion in muscles; and there is no evidence that speeds of* ≈ 12.35 m s^{-1} *(*44.4 km hr^{-1}*) are achievable by humans over any distance, let alone 100 metres.*

Certainly there is a logical limit of zero seconds, and the equations predict that too!

[14] Most notably Tennessee State College of Nashville

We can only suppose that the journal accepted the paper specifically to provoke this kind of criticism, and give a warning about "lying with statistics".

The message is "prediction is dangerous, especially about the future". In statistical terms, extrapolation is difficult to justify, unless there is a strong ***external*** *(meta-statistical) argument that past trends will continue.* *Return to Q62* •

A*63* *: Parents (independent variable) have children, not the other way around. The second question is trickier: we must know about genetics and inheritance. There is a mechanism by which height is inherited that is internal to the process and not dependent on anything outside.*

Or is inheritance strictly an internal process? What about the influence of environment? Don't both parents and children generally live in the same environment, so that, for example, poor nutrition or diseases of parents might be associated with poor nutrition or diseases of children?

Separating the effects of 'nature' and 'nurture' has been a main task of biostatisticians since the re-discovery of Mendelian genetics in 1900. *Return to Q63* •

A*64* *:*

1. *Yes and yes, the data seem well-fitted by a line, and errors are equally-distributed around it;*
2. *No and no, the data seem to fit another functional form perfectly;*
3. *Yes and no, all the data except one perfectly fit a substantially-different line,* $y = 4 + 0.346x$[15]
4. *Yes and yes; except we are quite uncomfortable with the best estimate, because we suspect that if we took more observations at* $x = 19$ *we would see a similar spread to the observations at* $x = 8$*, and we have no way of knowing where the single observation is in within this distribution.*

Return to Q64 •

[15] **Robust** regression methods [2, 7, 22] can successfully fit this relation.

7.9 References

Bibliography

[1] Anscombe, F. 1990. *Residuals.* In Eatwell, J.; Milgate, M.; & Newman, P. (eds.), *The new Palgreave: time series and statistics*, pp. 244–250. New York: W.W. Norton & Company 178

[2] Birkes, D. & Dodge, Y. 1993. *Alternative methods of regression.* Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons, Inc. 184

[3] Casella, G. & Berger, R. L. 2002. *Statistical inference.* Pacific Grove, CA: Duxbury, 2nd edition 155, 156, 157

[4] Cook, R. & Weisberg, S. 1982. *Residuals and influence in regression.* New York: Chapman and Hall 169

[5] Davis, J. C. 1986. *Statistics and data analysis in geology.* New York: Wiley 177

[6] Davis, J. C. 2002. *Statistics and data analysis in geology.* New York: John Wiley & Sons, 3rd edition 150, 177

[7] Fox, J. 1997. *Applied regression, linear models, and related methods.* Newbury Park: Sage 179, 184

[8] Galton, F. 1886. *Regression towards mediocrity in hereditary stature.* *Journal of the Anthropological Institute* **15**:246–263 174

[9] Gauch Jr., H. G. 1993. *Prediction, parsimony & noise.* *American Scientist* **81**:468–478 180

[10] Gauch Jr., H. G. 2002. *Scientific method in practice.* Cambridge: Cambridge University Press 154

[11] Ihaka, R. & Gentleman, R. 1996. *R: A language for data analysis and graphics.* *Journal of Computational and Graphical Statistics* **5**(3):299–314

[12] Kelvin, Lord (William Thompson). 1889. *Popular Lectures and Addresses.* London and New York: Macmillan and Co. Lecture to the Institution of Civil Engineers, 3 May 1883 150

[13] Lee, P. M. 2004. *Bayesian statistics: an introduction.* London: Arnold, 3rd edition 155

[14] Mark, D. M. & Church, M. 1977. *On the misuse of regression in earth science.* *Mathematical Geology* **9**(1):63–77 175

[15] Mayer, D.; Stuart, M.; & Swain, A. 1994. *Regression of real-world data on model output: an appropriate overall test of validity*. *Agricultural Systems* **45**:93–104 173

[16] Nagelhout, A. 2001. *Performance analysis of small format aerial photography SFAP in assessing current status and trends in wind erosion : a case study in the Longonot – Kijabe hill area, Naivasha district, Kenya*. MSc thesis, International Institute for Aerospace Survey and Earth Sciences, Enschede (NL) 154

[17] Rossiter, D. G. 2003. *Introduction to the R Project for Statistical Computing for use at ITC*. Enschede (NL): International Institute for Geo-information Science & Earth Observation (ITC). URL: http://www.itc.nl/personal/rossiter/teach/R/RIntro_ITC.pdf

[18] Rossiter, D. G. 2004. *Statistical methods for accuracy assesment of classified thematic maps*. Technical note, International Institute for Geo-information Science and Earth Observation (ITC). URL: http://www.itc.nl/personal/rossiter/teach/R/R_ac.pdf 165

[19] Shalizi, C. 2010. *The bootstrap*. *American Scientist* **98**(3):186–190 150

[20] Sprent, P. 1969. *Models in regression and related topics*. Methuen's monographs on applied probability and statistics. London,: Methuen 177

[21] Tatem, A.; Guerra, C.; Atkinson, P.; & Hay, S. 2004. *Momentous sprint at the 2156 Olympics? Women sprinters are closing the gap on men and may one day overtake them*. *Nature* **431**:525 166, 167

[22] Venables, W. N. & Ripley, B. D. 2002. *Modern applied statistics with S*. New York: Springer-Verlag, fourth edition 184

[23] Webster, R. 1989. *Is regression what you really want?* *Soil Use & Management* **5**(2):47–53 175

[24] Webster, R. 1997. *Regression and functional relations*. *European Journal of Soil Science* **48**(3):557–566 175, 176, 177