MSc Research Skills

Lecture: Formulating research problems, objectives and questions

D G Rossiter

University of Twente.
Faculty of Geo-information Science & Earth Observation (ITC)

May 23, 2013

Topics

Topic: The MSc research proposal

1. Logical structure
2. "Research", "Design", "Social", "Modelling" proposals

Common elements of a research proposal

A research proposal usually has a logical structure something like:

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

- The problem, objectives, questions and hypotheses are usually in one chapter called **Introduction**.
- This is usually followed by **literature review** and **methods** chapters.

  The **thesis** will then have several more chapters, covering:

  * results,
  * discussion,
  * conclusions and
  * recommendations.

Additional elements in the "research" thesis proposal

- **Study area**: location, map, description, suitability for answering the research questions

Additional elements in the "design" thesis proposal

A "research"-level design must have:

- 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.

The **"demand"** for a design replaces the "research problem" of the research thesis.

The "social" or "organizational" thesis proposal

Social analysis: the study of humans and human societies or their organizations.

- 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.
- Usually includes 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.

The "modelling" thesis proposal

The researcher builds a conceptual or (more commonly) computational **model** of a process.

This is usually applied to some real or synthetic **test data**.

- 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**.

Research problem

A general statement of **why** the research should be done.

This is something that is:

- **not well-understood** and
- can be addressed by **research**.

Research vs. Implementation (design)

Note that a **research** problem is not an **implementation**

- **Research**: something is not known or understood;
- **Implementation**: a known method needs to be applied to a practical problem.

However, it may not be known if an implementation will be successful . . . this can be a subject of research.

It may not even be known how to evaluate success . . . another research problem!

Problems in a social context

1. Why should anyone **care about the outcome** of this research?
   - What more will we know?
   - What practical problems will then be solvable?
2. Who would use the **results** of this research? and **for what?**
3. Why should anyone **sponsor** this research?

Categories of research problems

- **Social**: something (that might be) *wrong* with human society;
- **Environmental**: something (that might be) *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;

Social, environmental, management and technical problems often reveal an additional information or knowledge problem.

Example 1 – Naivasha – General statement

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

> "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."

Example 1 – Naivasha – social problems

- The destruction caused by wind erosion;
- The lack of effective intervention.

Example 1 – Naivasha – possible 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.

Example 2 – animation – general statement

> "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."

Example 2 – animation – social problem

Poor quality of planning decisions, because certain planning objects can not be well-visualized by planners.

Example 2 – animation – possible research problems

This was given at the end of the general statement:

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

So the problem is:

- It is not known how to integrate dynamic visualization methods into representations of uncertainty and fuzziness, in spatial planning applications.

Difficulties defining research problems

Confusion between a **social** problem and a **research** problem, e.g.,

The **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?

Possible research problems

(to be justified by literature review): all of these require **research** to solve.

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;
5. . . .

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.

There is usually a single **general** objective which is not operational.

This is broken down into a list of **specific** objectives which are then formulated as **research questions**, which are then operationalized as **research methods**.

Examples of general objectives

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."

Naivasha causes:

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

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."

Examples 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"

Examples of specific objectives – Naivasha causes

> "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"

Research outputs

In some thesis projects there may also be a **list of expected outputs**, 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 . . . ")

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.

Question words – simple

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

  (continued . . .)

- **"Can?"** (technique), in the sense of "Is it possible?"
  - "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)
  - "What is the accuracy of geo-referencing?"
- **"What?"** (is encountered in the field)
  - "What are the most common species of trees planted in domestic gardens?"
- **"How?"** (observational)
  - "How has water quality changed since the establishment of the irrigation project?"
  - "What, if any, are the change in water quality . . . ".

Question words – complex

- **"What is?"** (effects)
  - "What is the effect of increased grazing on vegetation density?"
- **"What is?"** (relation)
  - "What is the relation between increased grazing and vegetation density?"; this must be answered with a statistical model.
- **"Why?"** (causes)
  - "Why does increased grazing affect vegetation density?"; this must be answered with some proposed mechanism.
- **"How?"** (function)
  - "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?"

(continued . . .)

> "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?"

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?"

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 each objective.

Note that objectives are **declarative** sentences, whereas questions are **interrogative** sentences.

Example of 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?"

(continued . . .)

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?"

Hypotheses

Various definitions of this term, here we use:

Hypothesis: "[An] idea or suggestion that is based on known facts and is used as a basis for reasoning or further investigation" (Oxford Advanced Learner's Dictionary, 1995)

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**; another term for this is **proposition**.

Hypotheses should be as specific as possible

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)

> "Blow-outs and dunes caused by wind erosion can consistently be seen on SFAP at a nominal photo scale of 1:5 000."

> "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."

Example of hypotheses (animation)

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

Here the proposed design decision is stated; the truth of the statement must be tested during the research.

Statistical hypotheses

Another use of "hypothesis" is in frequentist statistical inference.

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.

Assumptions

These are **preconditions for research**:

- taken as true;
- **not questioned or verified during *this* research**;
- difficult to specify;
- made explicit and justified if questionable.

If an assumption is false, the research is (at least partly) invalid or infeasible.

Two kinds: **conceptual** and **logistical**

Conceptual assumptions

- **Laws of nature** (gravity, light, . . .): not stated
- **Laws within a discipline** (chemistry, soil science, . . . sociology?)
- **Facts taken as true**, not interfering with the factors being studied. Examples:
  - “Soils are fairly homogeneous in a 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 (*so we will survey kinship, not economic relations*).”
- This latter is called *ceteris paribus*, Latin for “with other things the same” (see next slide)

“Ceteris paribus”

- A **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;
- 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.

Logistical assumptions

These are also called “risks”: if these conditions are not met, the research will not be possible. Examples:

- “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 we need it”
- “ITC will acquire a license for a specialized computer program”

Be careful with these!!

Verifying assumptions

By definition these can not be verified, but their

- **plausibility** (concepts) and
- **feasibility** (logistics)

must be argued and **defended** in the research proposal.

Special care should be taken with *ceteris paribus*:

- have you identified other possible factors than the ones named in the hypotheses?
- can you argue that they are not important (i.e., constant) in your case?

Summary and next step

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.