Research Concepts & Skills
Volume 2: Skills

Text with self-test questions

D G Rossiter

Version 3.2
May 3, 2011

Contents

This second volume of the text *Research Concepts & Skills* presents some technical skills needed for successful MSc research, e.g. structured technical writing, searching and interpreting scientific literature, proper use of citations, abstracting, and graphical presentation.

1 Literature review, Citations, and List of references

This topic presents some aspects of the literature review, citations and list of references included in a research proposal, thesis or research paper:

- Information overload: how to find your way
- Purpose of a literature review
- Different types of sources and their reliability;
- Starting points for effective literature search.
- Why citations are used in a thesis or scientific paper;
- What must be included in a reference;
- Citing digital sources;
- Some common problems in citations;
- Formatting the list of references with EndNote.

1.1 Information

(This section written with Marga Koelen).

Key points

1. We live in an information age; so much is available, and so easily, leading to **information overload**;
2. To be **information literate**, the consumer needs to know why, when, and how to **use** all the available tools and **think critically** about the information provided (§1.1.1).
3. Information skills can be grouped in five categories: (1) awareness, (2) access, (3) evaluation, (4) use, and (5) generation.
4. These together make up your **personal information infrastructure** (§1.1.2).

"Knowledge is power. The more one knows, the more one will be able to control events."
– Francis Bacon, *Meditationes Sacrae* (1597)

We live in an age of abundant information, and the amount of information is increasing exponentially as the World becomes ever more connected via the Internet and ever more people are producing information. Information is highly **synergistic**: the more information one has, the more information one can generate.

Information is available from many sources and in many formats, such as:

- printed text (newspapers, magazines, journals, reports ...);
- television (increasingly international);
- videos;
- library databases; and
- web sites.

Much information of the first four types is increasingly available via the **Internet** as well as in traditional forms.

Information has become a vital resource for world economies and is certainly the basic component of any educational and research experience. The Internet is a common information and communication tool for secondary school or high school students, but at the university level other, more controlled, sources like bibliographic databases or article indexes are increasingly important.

All this information easily leads to **information overload**. There is so much information, and almost no control on who can present it and in what form. Therefore, the **information consumer** must **validate** and **assess** information to verify its reliability: the consumer must become **information literate**.

Q1 : *Name some positive and negative aspects of the Internet, with respect to information provision.* *Jump to A1* •

1.1.1 Information literacy

To be information literate, the consumer needs to know why, when, and how to use all the available tools and think critically about the information provided. Emphasis is on **critical thinking and analysis** about the information on offer.

Information Literacy is also about evaluating, conceptualizing, analyzing, **synthesizing and applying** information on one's own purposes. Information literacy is a set of skills that can be learned. This set of skills includes a certain attitude toward learning and doing research itself. Information skills are the most important skill in **lifelong learning** for an educated worker in the modern information economy, especially including scientific workers. During an ITC MSc programme, information skills are necessary to achieve the required level of **research competence.**

Information skills can be grouped in five categories:

1. awareness;
2. access;
3. evaluation;
4. use; and
5. generation.

Information **awareness** just means to realize that there is information available that should be sought and used; that is the purpose of this section of the text. Then, the first practical skill is to find information: this is **access.** The second practical skill is to choose among sources, to decide on the reliability and usefulness for your purpose of the information: this is **evaluation.** The third skill is to use the evaluated information for your purposes: this is **use.** Finally, the scientist **generates** new information for others to use.

These lead to the following skill set:

- to recognize the need for information;
- to access information from appropriate sources;
- to develop skills in using information technologies;
- to critically analyse and evaluate information;
- to organize and process information;
- to apply information for effective and creative decision making;
- to generate and effectively communicate information and knowledge to others.

The **ITC library** facilitates the information literacy program by different means: proving the student and researcher with a wealth of scientific information almost all accessible on the digital library pages but also with lectures, tutorials and supervised practicals.

During the **thesis proposal** stage, the MSc student's main information need is to find **relevant** and **accurate** information on the research topic, and then to relate these in a literature review. So the MSc student must be aware and know how to access, evaluate and use information.

1.1.2 Personal information infrastructure

As you try to navigate all the information to find what you need and pass rapidly over what you don't, you build up a set of strategies and sources that can be called a **personal information infrastructure**. This may include:

- A list of your favourite **reference materials** such as handbooks, user's guides and texts;
- A list of your favourite **journals** and book series;
- A list of the best **searchable databases** for your purpose (§1.5.2);
- A set of **saved searches** in the databases, also as e-mail or RSS **alerts** (§1.5.6).

This is a bit like having a personalized map of a very big city (Beijing, London, New York, Delhi, São Paulo, México ...) with your own annotation (favourite restaurants, budget hotels ...) and a list of resources you prefer to use to find new things (Time Out magazine, an information number for rail delays ...). With these you can find your way much more quickly without wasting effort or missing out on things that are interesting to you. You build up your "infrastructure" slowly and it becomes more useful and reliable with time.

1.1.3 Answers to self-test questions

A*1* : *Positive: information "at your fingertips", available from anywhere at any time; no dependence on the physical library. Negative: no regulation, no control on the quality and reliability of information provided.* *Return to Q1* •

1.2 Literature review

Key points

1. The main purpose of a literature review for MSc research is to establish its **originality** and to put the proposed research in **context** (§1.2.1).
2. The literature review also justifies **choice of research methods**.

1.2.1 Purpose of a literature review

A **literature review** is a structured discussion of a **selection** of documents relevant to the thesis topic, which contain "information, ideas, data and evidence written from a particular standpoint to fulfil certain aims or express certain views on the nature of the topic and how it is to be investigated and the effective evaluation of these documents in relation to the research being proposed" [24].

This can only be done in a **structured** and **systematic** way; see §1.5 "How to Search".

The main purpose within the context of MSc research is to establish its **originality**; that is, that the work proposed has not already been done. Almost always something related has been done; the review **organises** these, **discusses** them, and points out their **limitations**, some of which will be addressed in the research.

A second purpose is to place the proposed research in **context**, that is, to show its importance within a wider problem area. This must be established from the opinions of others, who define the context and identify important unsolved problems.

A third purpose is to compare **methodological approaches** to the research problem. There are almost always several ways to address a research problem, and here they are compared, in order to justify the approach to be taken in the proposed research. Note that this may combine aspects of several previous approaches.

Q2 : *Consider this statement from the literature review of the journal article by Gao Yan* et al. *[18] titled "Comparison of pixel-based and object-oriented image classification approaches - a case study in a coal fire area, Wuda, Inner Mongolia, China". and the literature reference that supports it:*

"An overview of the theory and case studies of detecting coal fires with remote sensing techniques can be found in the paper by Zhang et al., 'Detecting coal fires using remote sensing techniques' (Zhang et al. 2004)."

The cited reference is (note: using the original journal style):

ZHANG, J., WAGNER, W., PRAKASH, A., MEHL, H. and VOIGT, S., 2004, Detecting coal fires using remote sensing techniques. International Journal of Remote Sensing, 25, pp. 3193-3220.

Question: What is the authors' purpose in including this reference?

Jump to A2 •

Q3 : *From the same paper:*

"Pixel-based classification uses multi-spectral classification techniques that assign a pixel to a class by considering the spectral similarities with the class or with other classes (Casals-Carrasco et al. 2000). Although the techniques are well developed and have sophisticated variations such as software classifiers, sub-pixel classifiers and spectral un-mixing techniques, it is argued that it does not make use of the spatial concept (Blaschke et al. 2000)."

The cited references are:

BLASCHKE, T. and STROBL, J., 2001, What's wrong with pixels? Some recent development interfacing remote sensing and GIS. GeoBIT/GIS, 6, pp. 12-17.

CASALS-CARRASCO, P., KUBO, S. and BABU MADHAVAN, B., 2000, Application of spectral mixture analysis for terrain evaluation studies. International Journal of Remote Sensing, 21, pp. 3039-3055.

Question: What is the authors' purpose in including these references?

Jump to A3 •

Q4 :

From the same paper:

"Object-oriented classification was introduced in the 1970s (de Kok et al. 1999). The initial application was limited by

> *hardware, software, poor resolution of images and interpretation theories (Flanders et al. 2003)."*

The cited references are:

> *DE KOK, R., SCHNEIDER, T. and AMMER, U., 1999, Object based classification and applications in the Alpine forest environment. In Fusion of Sensor Data, Knowledge Sources and Algorithms, Proceedings of the Joint ISPRS/Earsel Workshop, 3-4 June 1999, Valladolid, Spain. International Archives of Photogrammetry and Remote Sensing, 32, Part 7-4-3 W6.*

> *FLANDERS, D., HALL-BAYER, M. and PEREVERZOFF, J., 2003, Preliminary evaluation of eCognition object-based software for cut block delineation and feature extraction. Canadian Journal of Remote Sensing, 29, pp. 441-452*

Question: What is the authors' purpose in including these references?

Jump to A4 •

1.2.2 Answers to self-test questions

A*2* : *This reference places the authors' specific work on comparing pixel-based and object-based methods for coal fire (see title) into the general context of remote sensing (of any kind) for detecting coal fires.* *Return to Q2* •

A*3* : *The key phrase here is "it is argued that"; the references support the authors' contention that they are doing something new, by going beyond pixel-based classifications that had been used up till then.* *Return to Q3* •

A*4* : *These give background on the object-oriented methods that the authors will use.* *Return to Q4* •

1.3 Types of sources

Key points

Sources may be categorized as:

1. Journal articles (§1.3.1)
   (a) ISI journals (§1.3.1)
   (b) Non-ISI journals (§1.3.1)
2. Conference papers (§1.3.2)
3. Book chapters (§1.3.3)
4. Textbooks (§1.3.4)
5. Technical reports (§1.3.5)
6. Web pages (§1.3.8)

These have different reasons for publication and different quality control methods.

Not all sources are equally valid. At one extreme, anyone can place any opinion on the Web, with no control. At the other extreme is a peer–reviewed paper in a highly competitive international scientific journal.

This section lists the principal types of sources for published scientific information, with some comments on their reliability. The following section (§1.4) explains **peer review** in detail and gives guidance on choosing among sources.

1.3.1 Journal Articles

This is an original contribution that appears in a published scientific journal.

These contributions have been **peer–reviewed** to ensure quality control (§1.4.1). However, not all peer–review is equally effective. In general, the more influential the journal (i.e. the more its work is cited and considered of top quality), the more likely that peer review has been rigorous. You are more likely to find reliable information in Nature than in some regional journal of development studies.

However, you should not take this argument from authority too far. Nothing done by humans is free from the possibility of error, illogical thinking, repayment of favours, or outright fraud.

There are several types of articles which may appear in a journal. Here are examples from peer–reviewed, internationally-circulated soil science journals.

Research Article — Describes an original investigation, method, or procedure. Specific and limited. Examples are Dobos *et al.* [11] and King *et al.* [27]:

- Dobos, E.; Micheli, E.; Baumgardner, M. F.; Biehl, L.; & Helt, T. 2000. *Use of combined digital elevation model and satellite radiometric data for regional soil mapping.* *Geoderma* **97**(3-4):367–391
- King, D.; Bourennane, H.; Isambert, M.; & Macaire, J. 1999. *Relationship of the presence of a non-calcareous clay–loam horizon to DEM attributes in a gently sloping area.* *Geoderma* **89**(1-2):95–111

Review Article — Summarises a set of research articles; surveying the state-of-art in a particular field. The title typically includes words like “review”, “summary”, or “overview”. Here the originality lies in the synthesis, not the investigation. Examples are by McBratney *et al.* [37] and Goovaerts [21]:

- McBratney, A. B.; Odeh, I. O. A.; Bishop, T. F. A.; Dunbar, M. S.; & Shatar, T. M. 2000. *An overview of pedometric techniques for use in soil survey.* *Geoderma* **97**(3-4):293–327
- Goovaerts, P. 1999. *Geostatistics in soil science: state-of-the-art and perspectives.* *Geoderma* **89**(1-2):1–45

Q5 : *What words in the titles of these two papers suggests that they are review papers?* *Jump to A5* •

Opinion — A scientific editorial, either by the journal editor or an invited contributor. An example is by Basher [3]:

- Basher, L. R. 1997. *Is pedology dead and buried?* *Australian Journal of Soil Research* **35**:979–94

ISI journals

The Institute for Scientific Information (ISI) is part of Thomson Scientific[1], a commercial scientific information provider, which includes the search engine and citation index Web of Science[2] and the Journal Citation Reports (JCR) (§1.5.9) available via the ITC library website (§1.5.8).

The ISI selects among all the journals a list of so-called **ISI journals** Many institutions and organizations, including ITC, only give credit to researchers for papers published in thee journals.

There are currently about 9 000 ISI journals in all fields, about 3 700 of them in the sciences. These are selected from a much larger group of candidate journals by expert editors by a holistic process[3], taking into

[1] http://scientific.thomson.com/
[2] http://isiwebofknowledge.com/
[3] http://scientific.thomson.com/free/essays/selectionofmaterial/journalselection/, accessed 26-March-2008

account:

- the journal's publishing standards (selectivity, review process);
- its editorial content (including reputation of its editorial board);
- the international diversity of its authorship; and
- its citation record in other journals (importance of its papers, see §1.5.9).

The selected journals are referred to as **ISI journals**, and are listed in the **master journal list**[4] and in the Journal Citation Reports (JCR) (§1.5.9). Papers in these journals are included in three **Citation Indices**[5], including the Science Citation Index, also from Thompson. These are the primary searchable databases to find references.

The practical point for the beginning researcher (e.g. in an MSc project) is that ISI journals have been evaluated as:

- **reliable** (especially their quality control and editorial policies), and
- **important** (i.e. their papers are cited by other researchers),

so articles in ISI journals should be used in preference to those in non-ISI journals if possible.

You can find out if a journal is ISI by searching the master journal list or Journal Citation Reports (JCR) (§1.5.9). For example, there are 15 ISI journals[6] listed in the category "remote sensing":

1. Canadian Journal of Remote Sensing (Canadian Aeronautics & Space Institute)
2. GIScience & Remote Sensing (Bellwether Publishing, USA)
3. GPS Solutions (Springer)
4. IEEE Geoscience and Remote Sensing Letters (IEEE)
5. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (IEEE)
6. IEEE Transactions on Geoscience and Remote Sensing (IEEE)
7. International Journal of Applied Earth Observation and Geoinformation (Elsevier Science)
8. International Journal of Remote Sensing (Taylor & Francis)

---

[4] http://scientific.thomson.com/mjl/
[5] http://scientific.thomson.com/products/sci/
[6] as of 12-April-2010

9. ISPRS Journal of Photogrammetry and Remote Sensing (Elsevier Science for ISPRS)
10. Journal of Geodesy (Springer)
11. Photogrammetric Engineering and Remote Sensing (American Society of Photogrammetry and Remote Sensing)
12. Photogrammetric Record (Wiley-Blackwell)
13. Photonirvachak (Indian Society of Remote Sensing)
14. Radio Science (American Geophysical Union)
15. Remote Sensing of Environment (Elsevier Science)
16. Survey Review (Commonwealth Association of Surveying and Land Economy)

You will likely recognize several of these titles as important journals for remote sensing. There are eight commerical journals (three Elsevier, two Springer, one Taylor & Francis, one Bellwether Publishing, one Wiley-Blackwell), three from national societies (India and USA), one from a research institute (Canada), one from an association, and three from an international society (IEEE).

In this case, all these journals are available in both printed and digital form from the ITC library.

Non-ISI journals

Recalling the standard to qualify as an ISI journal (selectivity, review process, reputation of its editorial board, international diversity of its authorship; importance of its papers) there are journals that fail some of these (e.g. not diverse, not important) yet still have good quality control. Thus there may be useful information in their papers.

> For example, as of 28-April-2009 there were 2 082 journals published by Springer, of which 794 were listed in JCR; this is a ratio of about 38%. So only a bit over 1/3 of the journals from this well-regarded publisher are ISI.

This especially true for non-ISI journals published by regional or national societies whose main reason for omission from the ISI list is their narrow geographic focus. Papers in these journals may include valuable information, especially about local conditions, for example in a study area.

Open Access journals

(This section written with Marga Koelen)

The idea of "Open Access" (OA) is simple: OA publications are those made freely available online to libraries and readers, anywhere, with no charges imposed for access. The attraction of this is obvious, especially for many countries where ITC is active. Yet, the commerical publishing model is well-established, and after all, someone has to pay for editing, reviewing, administering and publishing, whether commerical or academic. OA journals may be ISI or non-ISI.

History of OA The OA movement began with the *Budapest Open Access Initiative* at a meeting of the Open Society Institute (OSI) in 2001[7], with the stated objective of making research articles in all academic fields freely available on the internet. The next milestone was the *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities* of October 2003[8], a widely-supported call for public availability of publically-financed research results. This declaration clearly stated the social philosophy behind OA:

> "Our mission of disseminating knowledge is only half complete if the information is not made widely and readily available to society. New possibilities of knowledge dissemination not only through the classical form but also and increasingly through the open access paradigm via the Internet have to be supported. We define open access as a comprehensive source of **human knowledge and cultural heritage that has been approved by the scientific community**. In order to realize the vision of a **global and accessible representation of knowledge**, the future Web has to be sustainable, interactive, and transparent. Content and software tools must be openly accessible and compatible."

Over 260 organizations have signed the Berlin Declaration, and there are already almost 5,000 OA journals[9]; the Directory of Open Access Journals (DAOJ)[10] lists them.

The advantages of OA include:

1. OA promotes **international collaboration** between scientists. Free availability of research publications enables scientists in countries with fewer financial resources to access and use relevant scientific information. In the traditional (commercial) publishing model, they had no access due to restrictions from the international publishers. The OA movement helps to overcome the "digital divide", and al-

[7] http://www.soros.org/openaccess
[8] http://oa.mpg.de/openaccess-berlin/berlindeclaration.html
[9] as of 22-April-2010
[10] http://www.doaj.org/, also accessible from ITC Digital Library Website

lows all countries to share the research findings of the international scientific community.

2. Even in countries with more financial resources, many institutions can not afford to purchase journals outside their core fields, thus restricting opportunities for **inter-disciplinary collaboration**. OA allows scientists to freely find material outside their specialty.

3. OA ensures that public funds are spent for the benefit of the taxpayer, by promoting scientific progress. Most research is publicly funded, which implies that the results should also be publicly available. Under the commercial publishing model, scientific libraries, also mostly funded by the public, have to pay three times for their information:

   (a) Researchers executing their research and submitting their findings and results for publication are usually paid by the public sector;

   (b) The submitted manuscripts are peer reviewed by colleagues who are also usually paid by the public sector;

   (c) When the articles are published they must be purchased by the scientific library from the commercial publishers.

Implementing OA Librarians have identified two "roads" to OA:

1. The "Green" Road to OA: **repositories**

2. The "Golden" Road to OA: **Open Access journals**

Green Road

In the Green Road model, the scientific authors enter their articles into an electronic archive called a **repository**, either at their institution (e.g. university) or at a repository set up for the discipline. This is a form of self-archiving where all the articles are freely available via the internet. For the Green Road to be effective, libraries must offer an infrastructure in which this information can freely flow all over the world: i.e. searchable repositories and downloadable content. The University of Nottingham (UK) maintains a directory of institutional repositories: the Directory of Open Access Repositories (OpenDOAR)[11].

An example of an **institutional repository** is the "NERC Open Research Archive" of the Natural Environment Research Council of the UK[12]. This contains searchable full-text publications of all academic papers and reports resulting from NERC funding, including the British Geological Survey and the Centre for Ecology & Hydrology. However, in this and many

[11] http://www.opendoar.org/
[12] http://nora.nerc.ac.uk/

repositories, full text of articles published in commerical journals is not available.

The best-known example of a **discipline repository** is arXiv[13], hosted by Cornell University (USA), which allows open access to 599,465[14] e-prints in physics, mathematics, computer science, quantitative biology, quantitative finance and statistics.

Golden Road

In The Golden Road model, an article is published in a peer review journal and this journal is freely available worldwide. The underlying business model is different from the Green Road – here the journal has some other source of financing than subscriptions.

One example of a Golden Road OA journal is the *EARSeL eProceedings*[15], published by the European Association of Remote Sensing Laboratories (EARSeL)[16] with the cooperation of the BIS-Verlag of the Carl von Ossietzky University, Oldenburg (D)[17]. Although this is not an ISI journal (§1.3.1), it has an editorial board and is peer-reviewed. Financing is from the association, i.e. EARSeL, as part of its publically-funded mission.

Another example of a Golden Road OA journal is *Remote Sensing*[18], published by Molecular Diversity Preservation International (MDPI), which publishes many such journals. MDPI journals have a typical OA copyright[19], referencing one of the popular open-access publishing licenses:

> Articles published in Remote Sensing will be open-access articles distributed under the terms and conditions of the Creative Commons Attribution license. MDPI will insert following note at the end of the published text:
>
> "© 200...by the authors; licensee Molecular Diversity Preservation International, Basel, Switzerland. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/)."

Although this journal is also not ISI, its editorial board includes some well-known remote sensing specialists[20] and UT/ITC faculty[21]. Typical of new journals (first volume 2009), it is not indexed in Web of Science, but is searchable via DOAJ and Google Scholar.

[13] http://arxiv.org/
[14] as of 22-April-2010
[15] http://www.eproceedings.org/
[16] http://www.earsel.org/
[17] http://www.bis.uni-oldenburg.de/
[18] http://www.mdpi.com/journal/remotesensing/
[19] http://www.mdpi.com/journal/remotesensing/instructions/
[20] e.g. G.M. Foody, A.R. Huete, J.R. Jensen
[21] F.v.d. Meer, A. Vrieling

This journal has another Golden Road financing model: Article Processing Charges (APC), paid by the authors; these are typically paid by the authors' institutes or research funding bodies, as part of research grants. Current[22] rates are 300 – 1600 Swiss francs per article (about 220 – 1200), depending on the journal.

There are some ISI-listed Golden Road OA journals, e.g. Hydrology and Earth System Sciences[23].

Using OA sources A good starting point is the Directory of Open Access Journals (DAOJ)[24]. Here you can search for journal names by subject area, author, title and key words.

1.3.2 Conference Papers

This is an original contribution that was presented at a scientific meeting. In most cases these are **not** or minimally peer–reviewed. The conference organisers typically allow anyone who pays the registration fee to present whatever they want. In fact, the reason that scientists presents a paper at a conference is to inform their peers of their work, especially their new results and ideas which may not be fully "cooked" yet, and to get feedback. So it is correct to present work that could not pass the peer review process. A conference is somewhat like an open market or bazaar – don't believe everything the "seller" tells you!

However, in some fields (e.g. computer and information science, public administration) major conferences are highly-regarded, the papers are strictly peer-reviewed, and the published proceedings are good sources.

A practical problem with conference papers is that they may not be easily available from a publisher, in contrast to journals and books. Again, the more prestigious conferences may have their proceedings published in easily-accessible form. Some societies keep past conference papers on-line for easy access. The ITC library has access to the most important conference proceedings in its specialty field of GIS and remote sensing.

Conferences may publish their papers in several forms, in increasing level of reliability:

Conference Proceedings The original submissions, with no quality control; typically distributed at the meeting itself. "Published" by the conference organisers. Difficult to obtain after the conference is over. Avoid using this as a source if at all possible. An example is [44]:

[22] 22-April-2010

[23] http://www.hydrol-earth-syst-sci.net/

[24] http://www.doaj.org/

- ▷ Rossiter, D. G. 2006. *Classification of urban and industrial soils in the World Reference Base for Soil Resources.* In *18th World Congress of Soil Science*. Philadelphia, PA (USA): IUSS

Although this is a prestigious conference, there is no way to tell whether the paper was invited or volunteered, whether it's poster or oral presentation, and above all whether there was any quality control.[25]

Edited Proceedings A book from a publisher with some of the submissions, at least reviewed by a scientific editor to eliminate obviously wrong papers. An example is by de Gruijter & Marsman [23]:

- ▷ de Gruijter, J. J. & Marsman, B. A. 1984. *Transect sampling for reliable information on mapping units.* In Nielson, D. R. & Bouma, J. (eds.), *Soil spatial variability: proceedings of a workshop of the ISSS and SSSA*, pp. 150–163. Wageningen: PUDOC

Note the names of the editors and the title of the book in the reference, also the publisher – this is not just a "raw" conference paper.

Another example of a published proceedings is the EGOV (e-government) 2005 conference held in Copenhagen (DK); the peer-reviewed papers were collected in a book [56] published by Springer:

- ▷ Wimmer, M.; Traunmüller, R.; Grönlund, Å.; & Andersen, K. V. (eds.). 2005. *Electronic Government: 4th International Conference, EGOV 2005, Copenhagen, Denmark, August 22-26, 2005, Proceedings*, volume 3591 of *Lecture Notes in Computer Science*. Springer.

The full-text of individual papers from the conference, e.g. Scholl [48], can easily be found in this book and should be considered equally-reliable as a book chapter or peer-reviewed journal article:

- ▷ Scholl, H. J. 2005. *Organizational transformation through e-government: Myth or reality?* In Wimmer, M.; Traunmüller, R.; Grönlund, Å.; & Andersen, K. V. (eds.). 2005. *Electronic Government: 4th International Conference, EGOV 2005, Copenhagen, Denmark, August 22-26, 2005, Proceedings*, volume 3591 of *Lecture Notes in Computer Science*, pp. 1–11. Springer

Special issue of a journal Selected papers are sent for peer review; these should be considered journal papers for the purposes of citation and literature search, even if they were first presented at a conference. They have an editor and may be cited as a whole; in this sense they are like an edited book. An example of a special issue is by de Gruijter [22]; a contribution from this issue is by King *et al.* [27]. This work was

[25] In this particular case, the material from the talk was compiled into a peer-reviewed paper in an ISI journal [45].

originally presented at the Pedometrics '97 conference in Montpellier, but subsequently revised to a journal article. Then the article is cited as a journal article:

▷ King, D.; Bourennane, H.; Isambert, M.; & Macaire, J. 1999. *Relationship of the presence of a non-calcareous clay–loam horizon to DEM attributes in a gently sloping area*. *Geoderma* **89**(1-2):95–111

The entire issue may also be cited, if the works taken as a whole are relevant:

▷ de Gruijter, J. 1999. *Special issue: Pedometrics '97*. *Geoderma* **89**(1-2):1–400

Another examples is the paper by Rossiter [45], which was originally the conference paper mentioned above [44]:

▷ 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

Q6 : *Give three reasons why we should use this reference [45] in preference to the conference paper [44], if we have both available.*
Jump to A6 •

1.3.3 Book chapters

This is an original contribution that is collected into an edited book on a specific topic.

These are typically invited by the book editor and may undergo some peer review; certainly they are edited. Often they are review articles or summaries. Quality control is not as rigid as for journal articles. An example of such a book chapter is that of Skidmore [50]. The chapter is by Skidmore, an authority on the subject he was asked to review, and the book is edited by a group of well-known scientists. Still, this was not peer–reviewed in the same way as a journal article.

▷ Skidmore, A. K. 1999. *Accuracy assessment of spatial information*. In Stein, A.; Meer, F. v. d.; & Gorte, B. G. F. (eds.), *Spatial statistics for remote sensing*, pp. 197–209. Dordrecht: Kluwer Academic.

The entire book can also be cited if you want to make a summary statement about its contents:

▷ Stein, A.; Meer, F. v. d.; & Gorte, B. G. F. (eds.). 1999. *Spatial statistics for remote sensing*. Dordrecht: Kluwer Academic

1.3.4 Textbooks

This is a published book meant to introduce a subject for classroom teaching or self-study. It can treat a topic at any level (i.e. pre-requisites for understanding it), but given that level, it is intended as the first contact with the subject.

These are not peer–reviewed as such but are typically extensively edited and sent by the publisher to people who might use the text in teaching, to see if they find the book accurate and useful. Beware, not all publishers do this. The reputation of a publisher is important here. Among the good ones are Wiley, Springer, McGraw-Hill, Addison-Wesley, Oxford University Press. Others may be sloppier.

Some texts state this in their titles, for example: [31] and [12]:

- ▷ Dupriez, H. & de Leener, P. 1998. *Trees and multistorey agriculture in Africa : a textbook for agro - forestry*. Nivelles; Wageningen: Terres et Vie; Technical Centre for Agricultural and Rural Cooperation (CTA)
- ▷ Kutilek, M. & Nielsen, D. 1994. *Soil hydrology : textbook for students of soil science, agriculture, forestry, geoecology, hydrology, geomorphology or other related disciplines*. Geoecology paperback. Reiskirchen (D): Catena Verlag

Others do not explicitly say this, but an examination of the contents shows clearly that they are meant for teaching. For example, they have problems (exercises) and lists of further reading. Here are two examples, one elementary by Lillesand & Kiefer [33] and the other at advanced level, by Bishop *et al.* [4].

- ▷ Lillesand, T. M. & Kiefer, R. W. 1994. *Remote sensing and image interpretation*. New York: John Wiley & Sons, 3rd edition
- ▷ Bishop, Y.; Fienberg, S.; & Holland, P. 1975. *Discrete multivariate analysis: theory and practice*. Cambridge, MA: MIT Press

1.3.5 Technical Reports

These are publications from an institution or project, and often contain primary data and maps which do not appear elsewhere. They are often difficult for others to obtain, but if they are the only source of information, they should be cited. They are not peer–reviewed; the quality control was only as good as the project.

Examples are:

▷ Center for Advanced Spatial Technologies (CAST). 1998. *AR-GAP final report: State-wide biodiversity mapping for Arkansas.* Report, Center For Advanced Spatial Technologies (CAST), Fayetteville, AR.

▷ Anonymous. 1985. *Soils and soil conditions, Kali Konto upper watershed, East Java.* Project Report ATA 206, Universitas Brawijaya (Malang), Agricultural University (Wageningen)

Note that the Kali Konto project was completed in 1984 (this date appears on the cover) but not published until August 1985 (this date appears in the publication information inside the report). Not only is no author given, but the project name isn't even given, so we have no alternative but to list the author as 'Anonymous'.

Some technical reports are published in the form of CDs or DVDs, but they are obtained and cited like books. For example:

▷ FAO. 1998. *Digital Soil Map of the World and derived soil properties (CD-ROM).* Land and Water Digital Media Series No. 1. Rome: FAO

Here the "information carrier" (physical object) is a CD-ROM; this is stated in the reference.

1.3.6 Peer-reviewed non-printed sources

Increasingly, peer-reviewed material is being supplied via the web, with no printed form. This includes several high-profile journals, the outstanding example being the peer-reviewed journals of the Public Library of Science (PLoS)[26], which concentrates on medicine and public health. Articles from these journals have the same status, and are cited in the same way as, printed peer-reviewed journals (§1.3.1).

For example:

> ' Bourne [6] provides hard-won advice for authors trying to get their work published in scientific journals. '

▷ Bourne, P. 2005. *Ten simple rules for getting published. PLoS Computational Biology* **1**(5):e57. http://dx.doi.org/10.1371/journal.pcbi.0010057

1.3.7 Web access to printed sources

The Web provides access to many printed sources of the types listed above, e.g. on-line versions of printed journals, technical reports, and books. In this case, the web source is simply a **copy** or a **differently-formatted** version of a printed source. The web alternative is easier to

[26] http://www.plos.org/

access, so this information can be mentioned in the corresponding entry in the reference list. For the purposes of literature review, the printed source is cited.

Here is an example of a published technical report that has been formatted for the Web (both HTML and PDF).

> 'In the USA, the National Gap Analysis Program (GAP) was carried out during the 1990's, in order to find a common language for discussing issues such as land cover mapping, vertebrate habitat characterisation, and biodiversity conservation. Most studies were carried out at state level, for example in Arkansas [10].'

▷ Center for Advanced Spatial Technologies (CAST). 1998. *AR-GAP final report: State-wide biodiversity mapping for Arkansas.* Report, Center For Advanced Spatial Technologies (CAST), Fayetteville, AR. URL: `http://www.cast.uark.edu/gap/`

Note that the reference is to the technical report, along with the organization's location (so that the reader could write to request a copy); the web address is given for the reader's convenience.

Almost all journal articles now have digital versions, which are accessed via their **Digital Object Identifier (DOI)**, for example:

▷ Dobos, E.; Micheli, E.; Baumgardner, M. F.; Biehl, L.; & Helt, T. 2000. *Use of combined digital elevation model and satellite radiometric data for regional soil mapping. Geoderma* **97**(3-4):367–391 DOI: `doi:10.1016/S0016-7061(00)00046-X`

The DOI system[27] allows the reader find the article on-line via the DOI resolver web site[28]. All peer-reviewed journal articles which are available on-line have a permanent DOI; the resolver finds the current location of the resource and then loads it into the web browser.

Note that these digital versions are for your **convenience**, they are *not* the primary reference. Thus the use of the DOI is not mandatory in the reference list; it may be added for the reader's convenience in finding the source.

So, the digital version of a journal article or technical report is *not* considered a web resource, rather a printed resource that can be found on the web.

[27] `http://www.doi.org/`
[28] `http://dx.doi.org/`

1.3.8 Web pages

Some information is only available via the web and is not permanent: tomorrow's version may be different from today's, it may move to another cyber-address, or it may even disappear. It is often difficult to determine the author or even a correct title. These are commonly known as **"web pages"**. These are:

- not peer-reviewed;
- not permanent;
- not 'published' in the traditional sense, or even the digital equivalent, e.g. using a DOI (§1.3.7)

The citation must include the **access date**, that is, when the content was viewed by the reviewer (i.e. you). This ensures that the information was at the given address; if it later is missing, it may be possible to find the version from the date in a search site's archives[29]. If the site changes over time, the citation should include the date the source was last modified; this is equivalent to the edition for a book. However, this information is often not available.

Here is an example of a reference to a web page:

> 'In many of the United States of America, soil conservation is aggressively promoted through attractive web sites [e.g. 53].'

▷ USDA Natural Resources Conservation Service. Un-dated. *North Carolina NRCS*. On-line document; URL: `http://www.nc.nrcs.usda.gov/`. Access date: 08-March-2008

Here is an example of a references to an unpublished PDF file available only on the web:

> 'A good introduction to ethnopedology is by Ettema [13].'

▷ Ettema, C. H. 1994. *Indigenous soil classifications*. On-line PDF document; URL: `http://www.itc.nl/personal/rossiter/Docs/Misc/IntroToEthnopedology.pdf`. Access date: 08-March-2008

To find a web page's title and author, refer to the page's *title* as given in the HTML `<title>` tag. This is shown as the window title in most browsers. The *author* is the organisation or individual who sponsors the page; layout alone is not considered scientific authorship. You may have to go up one or several levels in the hierarchy, possibly to the organisation's home page, to find out who actually sponsors the page.

[29] You might want to make your own local copy for backup.

§1.4.3 discusses how to evaluate the reliability of web pages.

1.3.9 Answers to self-test questions

A*5* : *McBratney: "overview"; Goovaerts: "state-of-the-art", "perspectives".* *Return to Q5* •

A*6* : *(1) peer-reviewed, so more reliable and complete; (2) newer, so necessarily more up-to-date; (3) easier for the reader of our literature review to obtain.* *Return to Q6* •

1.4 Quality of information sources

Key points

1. Sources have different degrees of **reliability**, generally in the order peer-reviewed journal articles (research and review); book chapters; conference papers; textbooks, technical reports, and web pages. (§1.3)
2. **Peer review** is the main quality control mechanism for scientific publications. (§1.4.1)
3. Not all sources are equally useful; references should be **relevant**, **reliable**, and **accessible**. (§1.4.2)
4. Web pages require special care; their reliability is often difficult to evaluate (§1.4.3).
5. There are only a few situations where a web page should be used as a source (§1.4.4).

1.4.1 Peer review for quality control

The peer review process attempts to ensure that what is published is reliable and important. The approval of the authors' peers means that someone who is not intimately familiar with the research being discussed (e.g. you as MSc candidates) can pretty much trust that the publication is methodologically-correct and honestly-performed. The conclusions are another matter; here you should still form your own opinion from the body of the paper.

Here's how it works:

1. The **authors** submit a draft of the article to a journal editor, or a book chapter to the book editor.

2. The **editor** checks that the subject matter is relevant for the journal or book, and that the paper or chapter meets the required format (length, figures, required sections etc.).

3. The editor sends the draft to several other scientists familiar with the subject matter; these are the **peer-reviewers.**

   Some peer review is **open**, where both parties know each other's identity. Some is **single-blind**, where the reviewer knows the author but not vice-versa, or the author knows the reviewer but not vice-versa. And some is **double blind**, where neither knows the identity of the other. There is much discussion of the best form, and this differs also between fields and journals.

4. The reviewers read the draft and advise the editor on what to do with the paper:

   (a) **Accept in present form.**

   If the article is accepted, it is typeset, sent to the authors for proofreading (not changes), and published in the journal.

   (b) **Accept with minor revisions** (specified by reviewers). Suggested revisions are typically of format or style and not of substance; the reviewer finds the work as such correct:

   - Better editing, revise language slightly, improve the English;
   - Re-consider a specific statement which the reviewer does not consider justified;
   - Reformat tables or graphics;
   - Explain something more clearly or illustrate with an example.

   If the article is accepted subject to minor revisions, the authors make the revisions, and re-submit to the editor, who checks that the revisions match what the reviewers recommended. If so, the paper is published. If the authors disagrees with the suggestions, they can argue the point with the journal editor, or withdraw the article.

   (c) **Reconsider if major changes are made.** Recommendation to revise and possibly re-submit can also be for several reasons:

   - Incorrect analysis, not suitable to the data (therefore the conclusions are not justified);
   - Unjustified conclusions, poor reasoning;
   - Work does not properly consider related work, i.e. does not compare its results to others.
   - Poor writing.

   If major revision is recommended, the authors are given the paper back and they have a chance to amend the paper and re-submit it to the same journal, or to submit it elsewhere.

   (d) **Reject.** Outright rejection can be for several reasons:

   - **Fraud**, e.g. plagiarism of others' work
   - Incorrect data collection or processing methods (therefore the data are not reliable);

- Work repeats what has already been done, nothing new is added to the existing literature;
- Work is too narrow ("light") to justify publishing, but could be incorporated into a bigger study;
- Work is not relevant for the proposed journal;

If the article is rejected, it may not be re-submitted.

1.4.2 Choosing among sources

Not all sources are equally useful to you or your readers. The references should be **relevant**, **reliable**, and **accessible.** Here are some considerations for choosing the best sources.

Relevance — Only cite material that directly bears on your work (i.e., is **relevant**). It should fit one of the categories listed above (§1.6). Superfluous references do not impress, they confuse.

Reliability — The order of **reliability** is approximately:

1. Peer–reviewed articles in international journals;
2. Book chapters in edited collections; Textbooks by well–regarded authors (i.e. with a publication record in peer–reviewed journals); Edited conference proceedings from international congresses;
3. Technical reports and digital documents with no printed equivalent from well–regarded institutions;
4. Peer–reviewed articles in national or regional journals; Textbooks by lesser authors;
5. Unedited conference proceedings; Edited conference proceedings from local congresses;
6. Technical reports and digital documents with no printed equivalent from unknown institutions.

ISI vs. non-ISI — Non-ISI journals should not be used for information that is also available in an ISI journal.

Accessibility — Cite the most **accessible** source among several that give similar information:

- easy to find in many libraries, or at least easily obtainable by interlibrary loan;
- written as clearly as possible;
- in English;
- the most recent synthesis, rather than an isolated report.

For example, a thesis that has later been turned into a book or article(s) in major journals is much easier for a reader in any country to find in these sources. Another example is an early study in a technical report or a minor journal that then is included in a synthesis (review paper or textbook).

Non-English language materials should be avoided if possible, because many readers of the thesis will not be able to read these articles. However, if it is the only source for relevant information for the thesis (e.g. description of secondary data), these sources should be used..

If a work has been presented both in a national language and then in English, use the English version, unless the original has more information or better quality (e.g. it is peer-reviewed, whereas the English-language source is not), in which case cite both.

National journals are usually in the national language, for example 遥感技术与应用 (*Remote Sensing Technology and Application*), published by the 中科院资源环境科学信息中心 (Chinese Academy of Sciences Natural Resources and Environment Centre). This journal contains papers specific to China, written in the standard Chinese language (except for an English-language translation of the title and abstract). An example is the paper of Zhang *et al.* [58] on vegetation change in Fujian province; this information is likely only available in this local journal.

Primary vs. secondary

For **primary** information, cite the **primary** source rather than an interpretation or summary, even if it is in a less accessible or reliable source. For example, the original census rather than an article on the results of that census. Sometimes both should be cited, since the summary may be more accessible.

Use of ITC theses

An ITC thesis should be cited if it contains the primary data and analysis, even if much of this was later included in an article; these theses are obtainable by anyone on request to the ITC library; many are in digital form on the library web pages.

1.4.3 Judging the reliability of web pages

Since web pages are not peer-reviewed (at least not in the same way as journal articles), the reader must use objective indications of reliability, including the **reputation**, **objectivity**, and degree of **control** of the organization or individual which publish it.

- **Reputation**: Judge this from the work done by the person or organization. Organizations such as NASA, NOAA, and the FAO have a reputation for reliability.

  On the other hand, it is difficult to find any independent verification for the fantastic claims on some websites, such as “Psychic 101”[30], nor any independent assessment of the reliability of the publisher of this web page.

The ease of web publishing has enabled many “hobby” or “amateur” sites, i.e. published by an enthusiast in a subject, not as part of any organization. Many of these are excellent, but since authors of such pages usually have no objective reputation, their reliability is difficult to judge.

- **Objectivity**: Judge this from the nature of the organization; an advocacy group such as the WWF[31] has a clear policy **agenda**, and their site will of course advance this. The facts presented may be correct as stated, but may be selectively presented.
- **Control**: a site with many contributors and little or no editorial control will in general contain a mix of reliable and unreliable information.

Wikipedia

A special case is Wikipedia[32]. There is no direct control on contributions, but for most entires there is an effective peer-control, where mistakes and violations of the neutrality policy are swiftly corrected by an army of contributors. Still, since the articles are unsigned (and unsignable) there is no way to judge their reliability.

Another way to evaluate web pages are the following five criteria[33]:

- **Accuracy**: As far as you can evaluate, is the factual information correct, are there grammatical, spelling and/or typographical errors?
- **Authority**: Who is sponsoring the page? Who wrote the material? What are the author’s qualifications? Who is the copyright holder? If you can’t even tell who sponsors the page or who wrote the text, that is a bad sign: no accountability implies no authority.
- **Objectivity**: Is the text advertising information or a public service?
- **Currency**: When was the page written? Is the information kept current? When is the information from any graphs and charts gathered? Which edition is online?
- **Coverage**: Is it completed or under construction? Is there a print equivalent (or at least some printed sources to back it up)? Is it an entire work or parts?

[30] http://www.psychic101.com/, accessed 09-May-2008
[31] http://www.panda.org/
[32] http://www.wikipedia.org/
[33] from Marga Koelen, ITC librarian

Several websites[34] [35] explain how to evaluate other websites. The Stanford University Web Credibility Research project[36] looks at the problem from the point of view of the site publisher: how to make a site appear credible. They list ten guidelines[37] to boost credibility. These are all verifiable by the user.

Q7 : *Look at one of the "how to evaluate" sites mentioned in the footnote and evaluate its reliability. What gives you confidence that this is reliable information? How do you know the evaluation criteria they present are themselves valid?* *Jump to A7* •

1.4.4 When to use a web page as a source

There are four situations when a web page is a good source; in almost all other cases, other information sources are preferred:

1. The page is set up specifically as the **primary information** for an organization; this is especially applicable when that information changes frequently. An example relevant to ITC is the NASA website describing its missions, for example Terra [38].

2. The **fact that the page exists** is itself interesting and will be discussed in the thesis; the information on the page is not used as reference as such. An example was already given:

   > ‘ In the USA, soil conservation is aggressively promoted through attractive web sites [e.g. 53]. ’

3. The **opinion** on the page is relevant to the thesis argument. For example:

   > ‘ According to the WWF, a new logging road in Riau Province, Sumatra threatens one of the world's largest carbon stores and a key tiger habitat [57]. ’

   This is clearly labelled as the WWF's opinion, and presumably the thesis author will comment on its validity.

4. The page is given as a **convenience to the reader**, as an entry point to other information sources. For example:

   > ‘ The R statistical computing environment is being actively extended to deal with spatial data; the many projects are listed by Bivand [5]. ’

[34] www.lib.berkeley.edu/TeachingLib/Guides/Internet/Evaluate.html
[35] www.library.jhu.edu/researchhelp/general/evaluating/
[36] credibility.stanford.edu
[37] credibility.stanford.edu/guidelines

This example could also fit the second reason given just above.

1.4.5 Answers to self-test questions

A7 : Both are published by highly-regarded US universities (UC Berkeley and Johns Hopkins). Both have full publisher and date information. The Hopkins page lists the author. Both are self-reflective: they implicitly encourage the reader to evalaute them. *Return to Q7* •

1.5 How to Search

Key points

1. There are good **starting points** for a search (§1.5.1);
2. A literature search through digital resources should follow a **search strategy** (§1.5.3).
3. Search results should be **evaluated** and the strategy should be adjusted if necessary (§1.5.5).
4. Once you have found some relevant literature, expand your search with the **"spider"** approach (§1.5.7);
5. There are excellent **digital resources** for searching (§1.5.8).
6. **Journal Citation Reports** are used to evaluate the overall impact of scientific journals; highly-cited journals tend to be the most reliable for MSc research (§1.5.9).
7. **Citation counts** can be used to evaluate individual papers (§1.5.10)

Time is limited, information is almost infinite. So, you must develop a **search strategy**: what information is needed, and how to find it. This can only be done in a **structured** and **systematic** way.

Finding relevant material, and especially the most important for your purpose, is not easy. It requires patience, detective skills, some luck, and continued hard work. Fortunately, in the digital age it is possible to make much more rapid progress than previously.

1.5.1 Starting points

Starting points include:

- Reference lists in lecture notes;
- Reference lists in textbooks;
- Review papers; these have the advantage that the literature is placed in context, and you can already have an idea of which references are most important;
- Keyword searches in electronic resources, both on- and off-line; see below (§1.5.3) for details;
- Recent issues of relevant journals. This can be intimidating because the articles tend to be specialised, but if the topic is interesting to you, you can often find more basic references in the article's Introduction;
- Reference lists in earlier theses.

Q8 : *Which of these starting points would enable you to make the most rapid progress in understand the topic and finding the most relevant sources?* *Jump to A8* •

1.5.2 Selecting databases

There are many on-line databases with which to find information. The ITC library has listed the most important on the "Digital Library" web page[38]. Each resource also has an "information button" which gives some background on it.

The most useful for most beginning searches are:

- Search and view full-text (if ITC has a subscription):
  1. Web of Science
  2. Elsevier Science Direct
  3. SpringerLike
  4. JSTOR
- Search abstracts, some links to full-text
  1. Scirus
  2. Google Scholar
- Subject-oriented bibliographic databases
  1. GEOBASE
  2. CAB Abstracts

Note that references found with Google Scholar are not in any standard form, so it is often difficult to recover the correct reference.

1.5.3 Search strategy for databases

(with Marga Koelen)

The advent of searchable digital databases has made it easy to locate large numbers of references; but how do you find the best ones?

First of all a clear **definition of your topic** should be made: define what is included and what is not included. Brainstorming about the research topic helps you choose good search terms before you begin to search the various databases. Write out a few detailed sentences about your topic

[38] http://www.itc.nl/library/digital_library.asp

and underline the main words in these sentences. From the description that you wrote you may create a list of search terms.

It is also possible that you create a list of search terms from an article that you already have about the topic you are going to investigate.

Having clearly decided on your search strategy you can start searching the different databases.

1. build **concept groups** from your research topic;
2. develop a **set of terms** for each concept group;
3. find **synonyms**;
4. decide which **Boolean logic** is needed, place brackets and combine concepts as appropriate;
5. decide whether to use **truncation** or not;
6. choose the **databases** and use proper search commands for each.

Note that each database has its own syntax and rules for searching, including Boolean operators and truncation. Take some time to become familiar with these.

Q9 : *Why not just start your literature search by typing relevant keywords into Google?* *Jump to A9*

•

1.5.4 An example search strategy

Topic: "Earth-observation or geo-information based system for real-time monitoring of forest/wildfires in the Mediterranean region to assist fire squads"

1. Concept groups:

   These are fairly independent concepts, which combine to define the topic. They are not yet search terms, just a category of knowledge.

   (a) (real-time) Monitoring

   (b) Forest fires/ Wildfires

   (c) Earth Observation / Geo-information

   (d) (Fire squads)

   (e) (Mediterranean region)

The “fire squads” and region concepts will not be further developed here; you can do so as an exercise.

2. Set of terms (keywords):

   Keywords are search terms that are used when discussing the concept; it is expected that these terms will be found in the title, keywords or abstract of the literature we are trying to find.

   These are most easily developed per-concept:

   (a) concept: (real-time) Monitoring
      - monitor/monitoring
      - control/controlling
      - detect/detection
      - real-time

   (b) concept: Forest fires/ Wildfires
      - forest fire
      - wildfire

   (c) concept: Earth Observation / Geo-information
      - earth observation
      - geo-information / GIS/ geographical information science / geographical information system

---

Q10 : *What keywords might be used for the concept “fire squads”?* *Jump to A10* •

3. Synonyms:

   Keywords may also have well-known synonyms, which may be used by some authors. Typical examples are geographic names:

   “Mediterranean region”: Spain, Italy, Algeria, Tunisia, . . .

   Some technical terms have variant forms or specializations:

   “Earth Observation”: remote sensing, satellite, . . .

   Even common-language terms have (near- or partial-) synonymns; these are given in many references [e.g. 28].

Q11 : *What are some (partial-)synonymns for the keyword “fire” in the concept “fire squad”?* *Jump to A11* •

4. Boolean logic:

   These combine keywords to broaden ("OR"), narrow ("AND") or exclude ("NOT").

   The usual strategy is to combine *concepts* (mostly independent from each other) with `AND`:

   > (concept 1) AND (concept 2) AND (concept 3) ...

   and within each concept combine *keywords* or variants with `OR`. So for example concept 1 might be expanded as:

   > (keyword 1.1) OR (keyword 1.2) OR (keyword 1.3) ...

   Some examples of using `OR` within concepts:

   (a) concept: monitoring:

       `monitor* OR control* OR detect OR real time`

   (b) concept: forest fires: `forest fire OR wildfire`

   (c) concept: Earth Observation: `earth observation`
       `OR geo-information`
       `OR geographical information science`
       `OR GIS OR geographical information system`

5. Truncation:

   Many databases allow searches for parts of a word; almost always at the end, as conventionally signalled by the `*` symbol:

   For example: `earth observation` would match just that phrase; whereas `earth obs*` would match "earth observation", "earth observations", "earth observer", "earth observatory", etc.

6. Databases:

   The exact syntax of search commands depends on the specific database; consult the help for the selected database.

   For example, phrases often have to be enclosed in quotation marks; if not they are taken as an implicit OR: `"earth observation"` vs. `earth observation`.

The final search might look like this:

```
(monitor* OR control* OR detect OR real time)
AND
(forest fire OR wildfire)
AND
```

```
(earth observation OR geo-information OR GIS OR geographical
information science* OR geographical information system*)
```

In Web of Science this search, in the “topic” (title, keywords, abstract) is formatted as:

```
TS=((monitor* OR control* OR detect OR real time) AND
(forest fire OR wildfire) AND (earth observation OR
geo-information OR GIS OR geographical information
science* OR geographical information system*))
```

Databases=SCI-EXPANDED Timespan=All Years

which results[39] in 78 hits. The first screen of results is show in Figure 1.1.

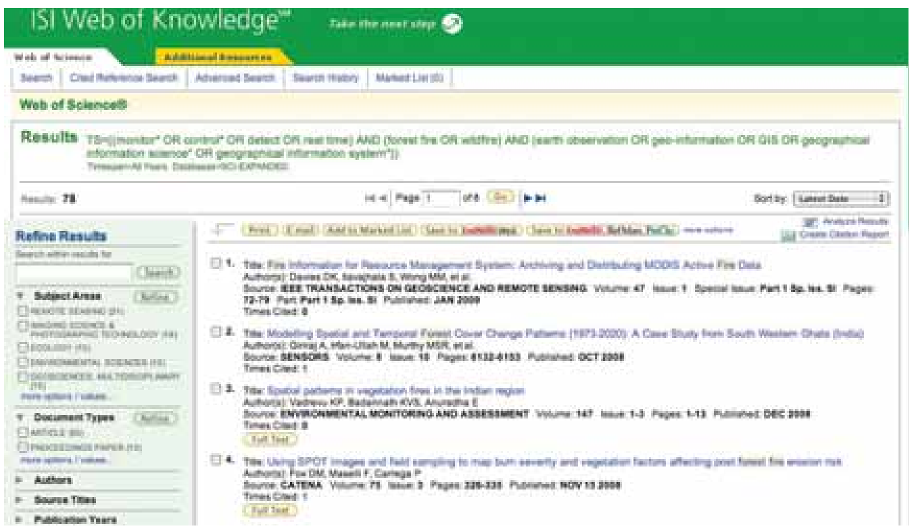


Figure 1.1: Results from keyword search

1.5.5 Evaluating search results

Don’t just take the first search results you get! Ask yourself if there are:

- Not **enough** references? Then **broaden** the search, or use a “spider” approach (§1.5.7) to find more from the few good references you have.
- Too **many** references? Then **narrow** the search by adding search terms with ‘and’.

[39] 26-March-2009

- Too **specific** references? Then use some **general terms**, or use a "spider" approach to look for general references cited in these specific references.
- Too **general** references?: Then use more **specific terms**, or look at the references cited in the general reference; these may be specific studies supporting the general ones.

1.5.6 Search tools

The various databases have many useful tools to help organize searches and find new information as it becomes available. The terminology and procedures differ somewhat among databases, but all have facilities to:

- **Save** the search for later use. This saves the search in the database's search language.
- Add an **alert** for the **search**, whereby you are informed by e-mail whenever new results that match the search are added to the database.
- Add an **alert** for a specific **journal**, **book series**, or **topic**, again with notification by e-mail. This is a good way to keep up with the contents of favourite journals.

Alerts can also be added as RSS (Really Simple Syndication) feeds to your newsreader.

To use these search tools you must **register** with the database and then **log in**. This service is free.

Figure 1.2 shows a search in Elsevier Science Direct; the button to save the search is circled. This alert can be named, as shown in Figure 1.3; it is then added to the list of alerts, which can also include topic and journal alerts, as shown in Figure 1.4.

A particularly good use of saved searches is to **periodically re-run** them; this is especially appropriate before writing the discussion and conclusions of a paper or thesis. In the interval between the thesis proposal and thesis submission (about nine months), much relevant work may have been done on your topic; this should be included in the discussion and also added back into the literature review.

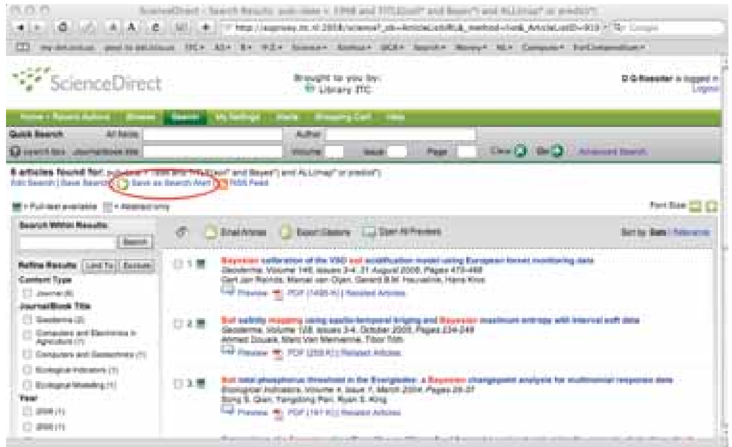


Figure 1.2: Results from keyword search, ready to save as an alert

Figure 1.3: Saved alert

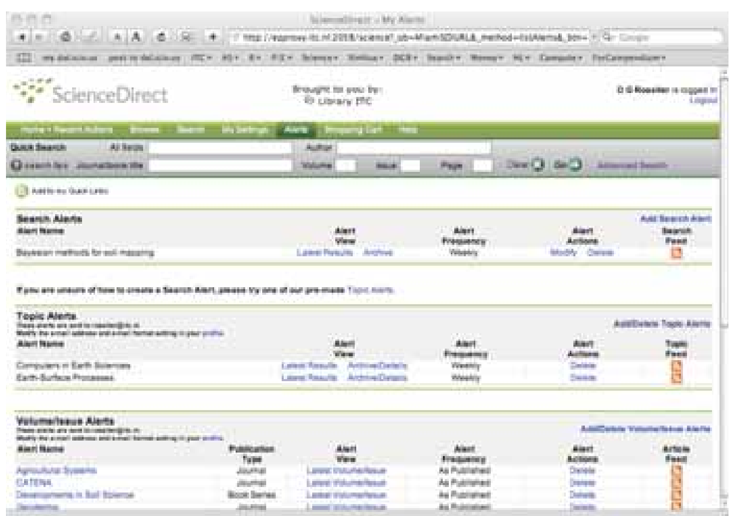


Figure 1.4: Saved alerts

1.5.7 The "spider" approach

Once you have found some relevant literature, you can often go further by searching for:

- Works that are **cited in** the papers you have found ("backwards spider"); ; note these have been put into context for you by the paper's authors;
- Works that **cite** the papers you have found ("forwards spider"); this is only possible by using the **forward search** of an electronic resource such as **Web of Knowledge**;
- Works listed as "related" in the search results ("sideways spider");
- Works by the **same author(s)**; very often the author continues working on related problems;
- Papers in the **same journal** or conference proceedings; a given journal tends to group papers that cover related areas.

If the work is not recent, make sure to look at newer sources to see if it has been superseded, challenged, or revised. These can be by the same authors, or by others who cite the work and challenge it.

Figure 1.5 shows a record that has been found (by some search strategy) in Web of Science, and is relevant to the topic. Note at the top of the

Figure 1.5: Interesting record

abstract the number of **times cited** (by later works) and the number of **references** (earlier works used by this paper).

Figure 1.6 shows the earlier works (cited by this paper) as a list ("beackwards spider"). Figure 1.7 shows works by the same first author ("side-

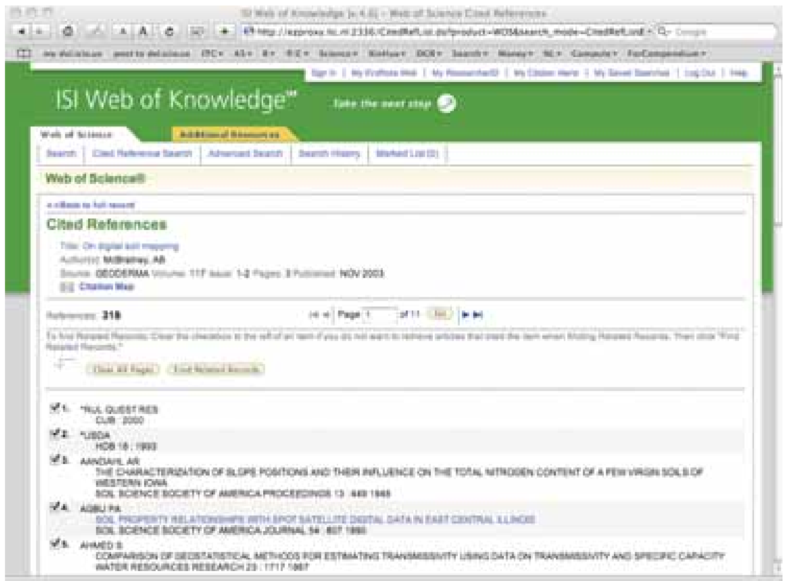


Figure 1.6: Articles cited by the interesting reference ("backwards spider")

ways spider"). Note in the original record (Figure 1.5) the right side of the screen: this shows references that cite this paper; these can be displayed in a separate screen (Figure 1.8). This is later work on the same topic ("forwards spider") – works that found this paper to be useful. Clearly, these are all useful ways to expand your information, based on your search strategy.

Q12 : *What is the added value of forward search?* *Jump to A12* •

1.5.8 Effective use of on-line search

The ITC library has excellent on-line access to a variety of sources; these are collected on the "Digital Library" web page[40].

Several scientific publishers maintain large on-line databases of abstracts and full-text articles. ITC has access to relevant titles via Elsevier's Sci-

[40] http://www.itc.nl/library/digital_library.asp

Figure 1.7: Articles by the first author of the interesting reference ("sideways spider")

Figure 1.8: Articles which cite the interesting reference ("forwards spider")

ence Direct[41], Blackwell's Synergy[42], SpringerLink[43], and Taylor & Francis[44], among others. There are also several independent databases; most notable is ISI Web of Science[45], from the company (Thomson) that maintains the science citation index.

Note that since ITC only has full-text access to certain journals from these publishers, the journals should be accessed via the ITC digital library, not via these home pages.

1.5.9 Using Journal Citation Reports to evaluate journals

How important is a given journal? That is a controversial question that has been motivating librarians at least since the 1920's. Once digital bibliographic databases of papers in reliable journals were compiled, it became possible to compute quantitative measures of impact. Thomson Scientific (then ISI) took the lead in inventing a journal "impact factor"[46] which is used for the annual Journal Citation Reports (JCR).

The impact factor of a journal is calculated by dividing the number of current year citations to the source items published in that journal during the previous two years. Although this factor has been criticized, and modified factors have been proposed and computed, it is still a useful measure of how widely a journal is read and used by other scientists to further their work.

The JCR covers about 9 000 of the most-cited journals in about 200 disciplines. Authors can use the JCR to identify journals in which to publish for maximum impact. At the level of MSc research, the student is advised to concentrate on the most important journals in the field; the citation report can help choose these.

The Journal Citation Reports is available from the ITC Digital Library. It is organized by subject area. For example, the 2008 report for subject "remote sensing" gives the list shown in Table 1.1, sorted by impact factor.

Q13 : *Which journal has the most impact? How far ahead is it of its nearest competitor?* *Jump to A13* •

---

[41] http://www.sciencedirect.com/
[42] http://www.blackwell-synergy.com/
[43] http://springerlink.metapress.com/
[44] http://www.tandf.co.uk/journals/
[45] http://isiwebofknowledge.com/
[46] http://scientific.thomson.com/free/essays/journalcitationreports/impact_factor/

Rank	Abbreviated Journal Title	2008 Cites	**Impact Factor**	Immediacy Index	2008 Items	Cited Half-life
1	REMOTE SENS ENVIRON	**15619**	**3.943**	**0.852**	330	6.8
2	IEEE T GEOSCI REMOTE	14614	3.157	0.489	364	7.1
3	ISPRS J PHOTOGRAMM	1235	2.293	0.283	46	6.3
4	INT J APPL EARTH OBS	398	1.947	0.233	43	3.8
5	PHOTOGRAMM ENG REM S	4100	1.846	0.102	108	9.4
6	IEEE GEOSCI REMOTE S	904	1.832	0.173	173	2.8
7	J GEODESY	992	1.689	0.345	55	5.7
8	GPS SOLUT	274	1.600	0.370	27	4.7
9	PHOTOGRAMM REC	239	1.417	0.167	24	6.2
10	CAN J REMOTE SENS	1175	1.315	0.065	77	6.1
11	RADIO SCI	3353	1.092	0.240	96	**>10.0**
12	INT J REMOTE SENS	9955	1.041	0.310	**420**	8.4
13	GISCI REMOTE SENS	54	0.707	0.077	26	
14	SURV REV	111	0.455	0.029	35	6.1
15	PHOTONIRVACHAK-J IND	11	0.049			

Table 1.1: Journal Citation Report, 2008, “Remote sensing”

Q14 : *Which two journals published the largest number of citable articles?* *Jump to A14*

•

There are two other interesting statistics in this report, a **immediacy index** ratio (roughly, how quickly an average paper is cited) and a **cited half-life** in years (roughly, how long an average paper remains cited). These can give some idea of which journals are reporting the “hottest” research, and which ones publish papers that continue to have an impact for a long time.

Note: The immediacy index is defined as the number of citations to articles published in a given year, divided by the total number of articles published that year in the same journal. Being a per-journal average, it tends to smooth out differences between large journals with many papers and small ones with few. But, it does not correct for number of issues in a year, nor for speed of editorial cycle.

So if by the end of the year, newly-published papers already cite papers from the beginning of the year in the same journal, the readers and authors of the journal are rapidly improving on each other’s work, or so is the theory. Papers have “immediate” impact on others. The problem with this measure is the long editorial cycle of some journals compared to others. Within a given field, however, it is a fair comparison.

Q15 : *Which journal is most immediate? Which is longest-lasting?* *Jump to A15* •

1.5.10 Using citation counts to evaluate papers

What makes a given paper "important"? An obvious answer is, if the paper is cited by many others, it probably has helped science progress[47]. For the MSc students, this indicates that the paper probably contains valuable information, and should be used by preference.

> **Note:** Note that a newly-published paper will not yet have any citations; and it takes several years for papers that cite this paper to themselves be published, the citation count isn't meaningful until about two years after a paper is published.

In **Elsevier ScienceDirect**, the citation count is given with the "Cited in" link; this shows all the papers which cite the paper, and the count. In **SpringerLink** this is called "Referenced by".

In the **ISI Web of Science**, the citation count is given with each paper's record. Search results can be sorted by citation count. For example, searching for papers on remote sensing of soil properties, but not soil moisture:

```
Title=("remote sensing")
AND Title=(soil)
NOT Title=(moisture or hydrology or water)
```

gives 74 papers; here is a part of this list, sorted by citation count, with the number of citations and the publication date **in bold**.

1. Title: EXPLORING A V-I-S (VEGETATION-IMPERVIOUS SURFACE-SOIL) MODEL FOR URBAN ECOSYSTEM ANALYSIS THROUGH REMOTE-SENSING - COMPARATIVE ANATOMY FOR CITIES
   Author(s): RIDD MK
   Source: INTERNATIONAL JOURNAL OF REMOTE SENSING Volume: 16 Issue: 12 Pages: 2165-2185 Published: **AUG 1995**
   Times Cited: 92

7. Title: Remote sensing of soil salinity: potentials and constraints
   Author(s): Metternicht GI, Zinck JA
   Source: REMOTE SENSING OF ENVIRONMENT Volume: 85 Issue: 1 Pages: 1-20 Published: **APR 25 2003**
   Times Cited: 23

[47] Although, it could have been cited many times as an example of bad science!

56. Title: Mapping within-field soil drainage using remote sensing, DEM and apparent soil electrical conductivity
Author(s): Liu JG, Pattey E, Nolin MC, et al.
Source: GEODERMA Volume: 143 Issue: 3-4 Pages: 261-272 Published: **FEB 15 2008**
Times Cited: 0

74. Title: USING REMOTE-SENSING TECHNIQUE TO STUDY SOIL SEDIMENTATION FLOW
Author(s): KOLAWOLE MO
Source: ENVIRONMENTAL MANAGEMENT Volume: 17 Issue: 1 Pages: 73-81 Published: **JAN-FEB 1993**
Times Cited: 0

Q16 : *Which of the first two listed papers (1 and 7 on the list) has the most impact, based on citation count and publication date?* *Jump to A16* •

Q17 : *The last two papers (56 and 74 on the list) have not been cited at all; which of these papers is definitely irrelevant to other scientists?* *Jump to A17* •

1.5.11 Answers to self-test questions

A*8* : *Review papers.* *Return to Q8* •

A*9* : *Much of the "information" will be irrelevant; it will be difficult to judge its quality; the sources are ordered by some index of web popularity, not scientific merit.* *Return to Q9* •

A*10* : *Some possible keywords are "fire fighter/fighting", "fire response", "fire brigade".* *Return to Q10* •

A*11* : *According to Kirkpatrick & Roget [28], some (partial-)synonymns for "fire" in the sense of "state of combustion" (as in "forest fire") are "flame", "blaze", and "incindiary".* *Return to Q11* •

A*12* : *Forward search finds the most up-to-date papers, i.e. the state-of-the-art, and makes sure we're not duplicating work that has been done since the* *publication of the source paper.* *Return to Q12* •

A*13* : Remote Sensing of Environment *has the most impact (factor* 3.943*).*
Return to Q13 •

A*14* : *The* International Journal of Remote Sensing *(420) and* IEEE Transactions on Geosciences and Remote Sensing *(364) published the largest number of citable articles in 2008.* *Return to Q14* •

A*15* : *Most immediate:* Remote Sensing of Environment, *far ahead of the others; most long-lasting:* Radio Science, *with* Photogrammetric Engineering and Remote Sensing *close behind.* *Return to Q15* •

A*16* : *As of the date these notes were compiled (April 2008) the Ridd* et al. *paper has accumulated 92 citations in 13 years and 8 months, a rate of* $92/164 = 0.56$ *citations per month. The Metternicht and Zinck paper has accumulated 23 citations in 4 years, a rate of* $23/48 = 0.48$ *citations per month. So the first-listed is ahead both on total count and rate.* *Return to Q16* •

A*17* : *Both papers have no citations; but the paper by Mo is 15 years old and has never been cited, thus is by definition irrelevant, whereas the Liu* et al. *paper is only two months old, so can not yet have been cited in a published paper.* *Return to Q17* •

1.6 Citations

Key points

1. Citations must be given for **definitions and concepts**; **opinions** of others; details of **methods**; and **facts** which are not established by the present research; and the source of direct **quotations** (§1.6.1).
2. In-text citations are placed near the information they support, to fit naturally with the text flow (§1.6.4)
3. A citation is **not** always needed (§1.6.3).
4. If at all possible, references cited must have been **read by the author** (§1.6.6), but in certain circumstances it may be necessary to use **indirect citation** ("cited in") (§1.6.7).

Science is a collective enterprise, with a history and a future. No one person can do all the work nor think up all the good ideas. The greatest scientists of all time, such as Newton and Gauss, explicitly acknowledged their intellectual debt to their predecessors. Furthermore, it would be impossible for one person to do all the experiments, collect all the primary data, or build all the information systems that have already been done by your colleagues.

Fortunately, science requires that we write down what we find and what we think. In a review of the literature we follow this historical trail, thereby saving us from having to duplicate previous work, and giving us the best possible basis for our own plans. It also saves a lot of writing, since you can just cite conclusions, and leave the detailed explanation for the original source.

1.6.1 Purpose of citations

In a thesis or other scientific writing, literature citations serve several purposes:

Definitions & concepts

They **present definitions and concepts** that are not yours, and **give proper credit** for them.

> 'Heuvelink [25] distinguishes two major conceptual models of soil spatial variability: the Discrete (DMSV) and Continuous (CMSV). The DMSV hypothesises that the variation in soil classes and properties across the landscape can be partitioned by sharp boundaries into homogeneous areas, whereas the CMSV ...'

> ‘There are two major conceptual models of soil spatial variability: the Discrete (DMSV) and Continuous (CMSV) [25]. The DMSV hypothesises that the variation in soil classes and properties across the landscape can be partitioned by sharp boundaries into homogeneous areas, whereas the CMSV . . . ’

Of these forms, the first is more explicit that Heuvelink actually invented these terms; the second form might just mean that Heuvelink’s article is a good review of the concepts.

> ‘Sahay & Woolshan [47], reporting on the implementation of GIS in a USAID-sponsored project in India, distinguished between what they termed “inhibiting” and “enabling” factors. ’

Q18 : *Why is it not proper to write these definitions without a citation?* *Jump to A18* •

Opinions

They **present opinions** that are not yours, **give proper credit** for them, and allow the reader to **verify your interpretations** of these works (i.e., the reader can go back to the original source and check if you correctly summarised it).

> ‘According to McBratney *et al.* [37], pedometric techniques are the future of soil survey *(Is that really a fair summary of that article?)*’

Q19 : *Why can’t we just write “According to Alex McBratney of the University of Sydney, pedometric techniques are the future of soil survey” without a citation? After all, we’re giving him full credit for the statement. Give two reasons.* *Jump to A19*
•

Quotes

They are used for **direct quotations** or paraphrases:

> ‘Authors are often tempted to show their supposed sophistication by adding some non-English text; they would do well to heed the advice of a master: “Those who use words or phrases belonging to languages with which they have little or no acquaintance do so at their peril” [17, article ‘foreign danger’].’

Data & results

They **present data and results** that are not from your own research, and allow readers of your work to find the original source if they wish.

'The Hungarian Environmental Monitoring System is a point–vector database containing 1236 soil profile descriptions [11].'

'About 48% of Africans, mainly in the centre and south, profess some form of Christianity, while about 41%, mainly in the north and Sahel, are Muslims [7, article "Afrika", p. 16]'

Note that the specific article in this reference work is mentioned in the citation; this is not strictly necessary, but it helps the reader find the information to verify it or see its context. The fact in this example could, for example, have been in an article about world religions; instead it happened to be in the article about Africa.

Q20 : *Which of the above two statements is more likely to be reported differently by a different source?* *Jump to A20* •

Introduction They **refer to previous work on your topic**, which you use in your introduction to motivate your study and place it in context:

'The first systematic study of soil map quality was by Webster & Beckett [55]. Somewhat later, a group at Cornell University worked for several years on aspects of soil survey adequacy, including accuracy assessment [16, 51]. At this same time, a group at the Staring Centre in the Netherlands developed methods for quantifying map unit composition and thematic quality [23, 36].'

Q21 : *Why are all these citations in the same paragraph?* *Jump to A21* •

Methods They refer to **standard methods**, so that you don't have to repeat them in your text. This is common in your 'Methods' chapter.

'Particle size distribution was determined by the pipette method with pre-treatment for organic matter but not for carbonates [41].'

Q22 : *How much detail must the cited reference provide about the methods?* *Jump to A22* •

Formulas They provide **detailed justification** of mathematical or statistical methods, so you don't have to derive or defend them:

'A formula for computing the variance $\sigma^2[\hat{k}]$ of the *kappa*

> map accuracy statistic was derived by Bishop *et al.* [4, §11.4.2] as: *formula follows* '

Note that I mention the section in this long book where this particular formula is derived. This is not strictly necessary, but may be a great aid to the reader in finding and verifying your interpretation. If the index of the book provides an easy way to find this (here, if there is an entry for '*kappa*, variance of'), it would not be necessary to mention the section here.

Q23 : *If you give this reference, why might you repeat the formula? (Two reasons).* *Jump to A23* •

Results They **refer to other studies related to your results**, with which you should compare, in your 'Results' chapter.

> 'This result appears to contradict that of Webster & Beckett [55], who found that only 10% of the area was unsuccessfully mapped using a similar method to the one in this study.'

> 'This successful clustering of the profiles by principal components analysis matches the results of Gobin *et al.* [20], who found that the first two components explained 64.7% of the total variance in a set of 72 pedons in southeastern Nigeria'

Q24 : *What might follow these statements in the Results chapter?* *Jump to A24* •

Further reading They give the reader material to go **deeper into a topic** than was necessary for your purposes. This is not needed for your work, but can be useful to some of your readers.

> 'A detailed description of the theory of atmospheric correction is given in the ATCOR user's manual [42]. '

1.6.2 Non-English language references

Cited material should be accessible to the readers of a paper or thesis. In English-language journals, and for ITC theses, it must be assumed that the only language common to all readers is English. References in languages other than English will not be understandable (accessible) to some of these readers.

Therefore, in an ITC thesis non-English language references should only be given when:

1. the reference is relevant and important;
2. the reference is not also available (perhaps in a modified form) in English.

Typical examples are primary data sources, method descriptions and previous research on a study area in a country where English is not the standard publication language.

For example Schraps *et al.* [49] is a project report of a mapping project of urban soils in Oberhausen city, Germany. Because it was prepared for local use, it is in the German language. If the thesis uses this map, it must be cited and used in the list of references. This report used methods also only described in the German language; these must also be cited.

> 'Soils of the study area had previously been mapped [49] according to German national standards for urban soil mapping [2]. '

▷ Schraps, V. W. G.; Kersting, A.; Pingel, P.; Schneider, S.; Baumgarten, H.; Burghardt, W.; Hiller, D. A.; Köppner, T.; Ohlemann, S.; & Metzger, F. 2000. *Stoffbestand, Eigenschaften un räumliche Verbreitung urban-industrieller böden – Ergebnisse aus dem Projekt Stadtbodenkartierung Oberhausen–Brücktorvertiel.* scriptum: Arbeitsergebnisse aus dem Geologischen Landesamt Nordrhein–Westfalen Heft 7, Geologischen Landesamt Nordrhein–Westfalen.

▷ Arbeitskreis Stadtböden. 1997. *Empfehlungen des Arbeitskreises Stadtböden der Deutschen Bodenkundlichen Gesellschaft für die bodenkundliche Kartierung urban, gewerblich, industriell und montan überformter Flächen (Stadtböden)*, volume Teil 1: Feldführer. Kiel: Sekretariat Büro für Bodenbewertung, 2. auflage

1.6.3 When not to use a reference

Not everything you say needs to be supported by a reference.

1. If it's your idea or result (then your report is the reference others will use);

   > 'On closer observation, it was obvious that the water samples all contained insect larvae ... '

2. If the fact is known to any person with a relevant (to the topic) education; this holds especially for general statements that will be developed further by argument;

> ‘Satellite remote sensing has an advantage over aerial photography: large areas are imaged all at once.’

Note that this statement would be backed up by a citation if it referred to a specific fact, e.g.

> ‘ The TM sensor of Landsat 4 and 5 images a swath 185 km wide [33, §6.8] ’

However, the general statement is well-known, easy to verify, and will be developed in further argument, so it needs no reference.

3. If the fact can be found in a standard secondary-school or general reference;

> ‘Since the area A of a circle is πr^2, we can compute . . . ’

4. If the fact is more or less fixed and can be verified in many ways;

> ‘Cuba is a Caribbean nation . . . ’

1.6.4 Placing citations in the text

A citation to support a direct quote or paraphrase is placed at the end of the sentence containing the quote or paraphrase, outside any quotation marks and before the period:

> ‘ Authors are often tempted to show their supposed sophistication by adding some non-English text; they would do well to heed the advice of a master: “Those who use words or phrases belonging to languages with which they have little or no acquaintance do so at their peril” [17, article ‘foreign danger’].’

In an author-date citation style the author’s name is in the citation; e.g., in the APA-5th style:

> ‘ “Those who use words or phrases belonging to languages with which they have little or no acquaintance do so at their peril” (Fowler & Gowers, 1965, article ‘foreign danger’). ’

If the author is named directly as part of the narrative, the citation is placed immediately following the author’s name; the name is not repeated in the citation.

> ‘ Heuvelink [25] distinguishes two major conceptual models of soil spatial variability: the Discrete (DMSV) and Continuous (CMSV). ’

In an author-date citation style the author's name has already been named, so only the date is used for the citation; e.g., in the APA-5th style:

> ' Heuvelink (1998) distinguishes two major conceptual models of soil spatial variability: the Discrete (DMSV) and Continuous (CMSV). '

If work is referred to in the text, the citation for this work comes at the end of the sentence:

> ' In the 1970s a group at the Staring Centre in the Netherlands developed methods for quantifying map unit composition and thematic quality [23, 36]. '

In an author-date citation style all citations are in the same list; e.g., in the APA-5th style:

> ' In the 1970s a group at the Staring Centre in the Netherlands developed methods for quantifying map unit composition and thematic quality (de Gruijter & Marsman, 1984; Marsman & de Gruijter, 1986). '

Q25 : *Where should you place references in the following text?*

"Satellite imagery is familiar to most people because of the common use of Google Earth images as backgrounds in television news broadcasts, and the use of weather satellite images as backgrounds for television weather forecasts. However, many people are unable to correctly interpret this imagery, for two reasons: (1) lack of general geographic knowledge and (2) mis-understanding of the colours and animations used."

Jump to A25 •

1.6.5 In-text citation style

The style of in-text citations goes along with the style of referene list entries; see §1.7.5 for a discussion.

1.6.6 Citations must have been read by the author

! → In general, **only cite material you have actually seen.**

Otherwise you can not be sure that it says what you are asserting that it does, or even that it really exists. *You are relying on someone else's interpretation* of what it says, which may well be wrong. You can not defend any interpretation of the material, since you haven't read it yourself .

! → **When you cite something, you are implicitly representing that you**

have read it.

The main exception to this rule is if the existence of the cited work is itself relevant to your study; for example, if you are writing a historical survey and need to refer to all works on a subject, even if you haven't been able to find it yourself. Another exception is if you can find the work but can not read its language.

See the next section for a solution if you absolutely must cite something you haven't seen.

Q26 : *What is the difference between these two statements (adapted from Lee [32, §3.8]) and corresponding references?*

"The problem we are going to consider has a long history, going back to Newcomb [39]"

"The problem we are going to consider has a long history, going back to Newcomb in 1881 [29]" *Jump to A26* •

1.6.7 Citing material you can't read

Always go back to **original sources** for facts that are established by them. In particular, **do not** cite a previous thesis for anything except its **own results.**

There are only two cases when you may cite something you haven't yourself read:

- You can't obtain the original source; or
- You can't read the original language.

Note that it may be sufficient for your purpose to cite the **secondary source** where you found the reference to these inaccessible sources, i.e. the "cited in" reference.

If absolutely necessary, use the 'cited in' approach: cite the original author, but the bibliographic reference is to the book you actually saw.

Example 1: You can't obtain the original source, but you have an abstract

This may be the case for conference proceedings, and works in minor journals, out-of-print books and reports. The abstract may be in a special abstracting publication (e.g. CAB Abstracts) or in an electronic database (e.g. GEOBASE). So, you can see the main conclusions of the work, and you want to cite it, but you can't see the full paper. In this case, place the notation "(Abstract)", with the name of the abstracting service, at the end of the reference:

▷ Oliver, M. A.; Webster, R.; & Slocum, K. 2000. *Filtering SPOT imagery by kriging analysis. International Journal of Remote Sensing* **21**(4):735–752. (GEOBASE Abstract)

Note that at ITC it should always be possible to obtain the original article, by request. But elsewhere you may not be able to do this.

A more common case is that you can only read the **English-language abstract** of a paper written in a language you do not understand. In this case, place the notation "(English Abstract)", for example (assuming you can't read French):)

▷ Gaultier, J. P.; Legros, J. P.; Bornand, M.; King, D.; Favrot, J. C.; & Hardy, R. 1993. *L'organisation et la gestion des données pédologiques spatialisées: Le projet DONESOL (English abstract). Revue de Géomatique* **3**(3):235–253

If possible, give the title in English with the name of the original language in parentheses. This assumes that someone has translated the title for you, or that it appears translated in the source. Journal and book names are not translated, because they are searchable by librarians.

▷ Gaultier, J. P.; Legros, J. P.; Bornand, M.; King, D.; Favrot, J. C.; & Hardy, R. 1993. *Organisation and management of soil data: the DONESOL project (in French) (English abstract). Revue de Géomatique* **3**(3):235–253

Example 2: You can't obtain the original source or its abstract

If the information in the original source still needs to be cited (typically so that the reader could find it), use the "Cited In" approach:

"Cited In"

'The Hungarian Environmental Monitoring System has been collecting detailed information since 1995 [54, cited in 11].'

▷ Várallyay, G.; Hartyáni, M.; Marth, P.; Molnár, E.; Podmaniczky, G.; I., S.; & Kele, G. 1995. *Talajvédelmi információs és monitoring rendszer. 1 kötet módszertan.* Technical report, Földmüvelésügyi Minisztérium, Budapest. (In Hungarian)

▷ Dobos, E.; Micheli, E.; Baumgardner, M. F.; Biehl, L.; & Helt, T. 2000. *Use of combined digital elevation model and satellite radiometric data for regional soil mapping. Geoderma* **97**(3-4):367–391

Here Várallyay *et al.* [54] has the information; I can't get this technical report, so I have to rely on the account in Dobos *et al.* [11], which is where I found out about this system. Also note I do not give the English title for Várallyay *et al.* [54], since the whole work is in Hungarian, and it has no English abstract (as far as I know).

> ‘In the South, many plantations faced problems of steadily-decreasing yields and economic ruin. Observant agriculturalists realised that different soils could sustain different levels of production, and recommended systematic soil surveys [46, cited in 8, p. 199].’

- ▷ Ruffin, E. 1832. *An essay on calcareous manures*
- ▷ Buol, S. W.; Hole, F. D.; & McCracken, R. J. 1989. *Soil genesis and classification*. Ames, IA: The Iowa State University Press, 3rd edition

Here I should cite Ruffin [46], even though I can’t find this old book, because I am referring to a specific historical situation. If I simply want to say that different soils sustain different levels of production, I should cite a modern textbook or reference work on soil fertility.

Example 3: You can’t read the language

> ‘Kubiëna [30, cited in 8] was the first taxonomist to make the fundamental distinction between *terrestrial* and *aquatic* soils.’

- ▷ Kubiëna, W. L. 1948. *Entwicklungslehre des Bodens*. Wien: Springer-Verlag

Here we cite Kubiëna [30] because we are going to use this important distinction, which he proposed. Even if I have seen the book, if I can’t read German, I must rely on the account in Buol *et al.* [8]. But unless your aim is to give a historical bibliography, you could just rely on the secondary source to establish the fact:

> ‘Kubiëna was the first taxonomist to make the fundamental distinction between *terrestrial* and *aquatic* soils [8].’

Example 4: You can read the language, but the reader may not be able to

If I can read German, but perhaps my reader can not, I should cite the original (since I myself read it) and a more accessible source (for my reader):

> ‘Kubiëna [30] was the first taxonomist to make the fundamental distinction between *terrestrial* and *aquatic* soils [see also 8, p. 199].’

1.6.8 Non-Latin scripts

How to cite sources written in extended Latin scripts (e.g. Czech, Polish, Icelandic, Turkish) or non-Latin scripts (e.g. Chinese, Devangari, Thai, Amharic, Cyrillic, Greek) is a huge topic, with good solutions for specialist literature, for example the Oxford University Press style manual [43, Ch. 11]. Scientific journals usually also have rules for these, or refer to a standard style manual.

Whatever solution is adopted must be used **consistently**.

Extended Latin scripts

These are scripts with many Latin letters but either some extra letters or diacritical marks, or both. Generally, the non-Latin characters are available in extended font sets; if they can be typed and printed, they may be used. There are generally keyboard layouts for each language where these characters are provided. English-speaking readers can make some guess (even if wrong) about the pronunciation of the extended characters.

Examples:

- España (Spanish)
- utilização (Portuguese)
- der Schluß, der Schlüssel (German)
- Provençal, épater, l'Hôpital (French)
- Łodz (Polish)
- İzmir, Diyarbakır, Erdoğan (Turkish)

Some may require special fonts, or typesetting in LaTeX:

- tiẽng Việt (Vietnamese)

Some of these languages have standard **transcriptions** into Latin script; for example, in Vietnamese it is common to write for European journals without the diacritical marks; in German there are standards for replacing umlauts with a phonetic equivalent; in Turkish the two i's (*ı* and *i*) are often both written as the Latin (dotted) *i* in lower case, and the Latin (undotted) *I* in upper case (e.g. *Diyarbakir* for *Diyarbakır*, *Izmir* for *İzmir*).

So there are two choices; whichever is adopted must be used consistently:

- Use the extended Latin script;
- Use a standard full or partial transcription to Latin script, as specified in a style manual.

Completely non-Latin scripts

Since ITC does not deal with linguistic, cultural or historial subjects, and all ITC theses are written in English, the use of completely non-Latin scripts is not allowed in the main thesis text or references, because English-speaking readers will not even be able to pronounce the text.

> **Note:** It is possile to transcribe the reference with a **romanization** method, i.e. a way to represent the text in Latin script. An example is Hànyǔ Pīnyīn for Chinese characters. If you can read the language in question, you should know the corresponding standard romanization.
>
> Transliteration is not too helpful to those who do not know the language; they may make some attempt at pronunciation but can not understand the words.

For scientific documents written to be read by English-speakers, the best solution (required at ITC) is to **translate** the title, and indicate the source language in parentheses.

For example, here is the original reference for a paper [58] written in the standard Chinese language:

▷ 张春桂,潘卫华,陈惠,黄朝法. 2007. 利用多时相中分辨率卫星影像监测福建省植被覆盖变化. 遥感技术与应用 第22卷 第05期

This journal has defined its own English title: *Remote Sensing Technology and Application*, and the article has an English abstract and title, so the translation can use these:

▷ Zhang, C.; Pan, W.; Chen, H.; Huang, C. 2007. Monitoring vegetation coverage change of Fujian using multi-temporal remote sensing image (in Chinese; English abstract). Remote Sensing Technology and Application **22**(5).

> **Note:** It is important to include the source language, so that a reader who does not read that language can avoid the trouble of looking up the paper, which will be (to him or her) unreadable. In this case (in Chinese); the reader is also informed that there is an English-language abstract.

The translation of the title and journal name may not be official (i.e. it may be done by the student), if it is not already translated in the journal. In this case, the authors translated the title; but this may be a poor translation, which you can correct if you wish.

The above reference implies the thesis author has read the Chinese original. If only the English abstract was read, this must be indicated by placing the text "(Abstract)" after the title; this is the same format as if only the abstract is available (§1.6.7).

▷ Zhang, C; Pan, W.; Chen, H.; Huang, C. 2007. Monitoring vegetation coverage change of Fujian using multi-temporal remote sensing image (English abstract) (in Chinese). Remote Sensing Technology and Application **22**(5).

1.6.9 Corporate vs. Individual Authorship

This is sometimes difficult to determine, especially for project reports, technical manuals, or reference works. The rule is to credit an individual author or editor (or, several individuals) if it is clear from the work that they are largely responsible for its contents, even if the work is sponsored or published by an organisation. Simply compiling a group of papers, or summarising a discussion or project, is not enough for authorship.

In the case of the *Oxford Advanced Learners Dictionary* [26], Hornby is considered to be the overall editor, who was responsible for ensuring a consistent style, therefore the work is cited under his individual name.

In the case of the Brockhaus encyclopedia [7], there is an editor listed inside, but he is only named as the 'editorial leader', which sounds more like a co-ordinator than an editor, therefore the work is cited under the organisation.

In some cases, no author is given, even if one author wrote the work, because it was done under contract. An example is FAO Forestry Paper 48 [14], written by Anthony Young, although he is nowhere mentioned in the report. Then there is no choice but to cite the work under the organisation.

Some project reports have tens or scores of contributors or participants in a workshop. If there is an editor listed, the work is cited under the individual name as editor. If there is just a long list of contributors, the work is cited under the organisation.

1.6.10 Multiple sources for the same fact

Sometimes you have looked at several sources, all of which support a synthesis that you want to make. If you don't refer to the individual contributions for other reasons, cite all of the sources in one list, using language that makes it clear that you are referring to all of them. In the following example, I am referring to all of the works together:

> 'Soil map quality has been studied by several groups over the last thirty years [16, 23, 36, 51, 55]. The general conclusion is that we are a long way from making routine assessments of quality; indeed there is no agreement on the concept of 'quality' when applied to soil maps.'

1.6.11 Answers to self-test questions

A18 : *If no citation is given, the implication is that the author (you) have invented these concepts yourself.* *Return to Q18* •

A19 : *The citation is needed so the reader can (1) verify that is indeed what McBratney says, and (2) read McBratney's support for the statement.* *Return to Q19* •

A20 : *The proportion of Africans with different religions is more difficult to establish exactly, so other sources may well give different figures. The number of soil profiles in a database is a much simpler piece of information to determine, and it reasonable to assume that Dobos has this right.* *Return to Q20* •

A21 : *They form part of a single narrative, the history of soil map accuracy assessment.* *Return to Q21* •

A22 : *Sufficient detail for the laboratory method to be reproduced by another worker. Certainly it must explain the "pipette method" and the named pretreatments.* *Return to Q22* •

A23 : *(1) As a convenience to the readers, so they don't have to find the original text; (2) because you are going to discuss some aspect of it, e.g. "Note that the second term depends on . . . ".* *Return to Q23* •

A24 : *A discussion of* ***why*** *the present study agrees or disagrees with the previous one.* *Return to Q24* •

A25 : *The first sentence needs no reference, it is common knowledge. Second sentence: "unable to interpret the imagery" needs a citation; there must be a study to show that (if not, that could be the research topic: "it is suspected that . . . "); the "two reasons" clause needs two references, one for each reason.* *Return to Q25* •

A26 : *The first form implies that the author has read the 1881 paper by Newcomb [39], the second that the author relied on the 1981 account by Knuth [29].* *Return to Q26* •

1.7 List of references

Key points

1. The **list of references** in a proposal, thesis or article is an appendix listing the sources used (§1.7.1).
2. The list must contain every reference in the text, and vice-versa (§1.7.2).
3. Every item in the list of references should be easy for a competent librarian to **locate and obtain** (§1.7.3).
4. **Any consistent citation and reference list style** is acceptable for an ITC thesis; each journal has its own rules (§1.7.5).
5. EndNote or another **reference management software** provides consistent formatiing, as long as the bibliographic database entry must be correct (§1.7.6).

1.7.1 Purpose of the list of references

The **list of references** is an appendix to a thesis, paper or report which lists the sources that are cited in the text.

This is different from a **bibliography**, which is a (usually categorised) list of all sources consulted (whether cited or not), or a comprehensive list of relevant sources for some topic. In the ITC thesis and research papers, it is the list of references that is required, unless one of the purposes of the study is to compile a bibliography.

Q27 : *Why might you want to compile a bibliography?* *Jump to A27* •

1.7.2 What items must appear in the reference list?

It is quite simple:

- **Every citation** in the text must appear in the **reference list.**
- **Every reference** in the reference list must be cited in the **text.**

Q28 : *Why must every reference in the text appear in the reference list?* *Jump to A28* •

1.7.3 What must appear in the reference list entry?

The key point about citations is that there must be sufficient information given for readers to find the same source, if they wish.

> Items in the reference list should be
> **easy for a professional librarian to find.**

This implies that there can be **no personal communications or unpublished materials.** If these must be used, they are cited in the main text (see §1.7.4, below),

This requirement implies what must be included in the reference as a **minimum**; other **optional** information is sometimes added for the reader's convenience.

Examples:

Journal article — **Required**: Journal name or standard abbreviation; journal volume number; first page number
Optional: title, authors,full page range; if available in electronic form: DOI

Book — **Required**: Author (or editor); title; edition; publisher; city
Optional: number of pages, ISBN

Book chapter — **Required**: Author of chapter; chapter number, further as for book
Optional: chapter title, page range; further as for book

Further details may be found in style manuals [e.g. 9].

Q29 : *What does the "professional librarian" mentioned above imply about the level of detail required in the reference list?* *Jump to A29* •

1.7.4 Personal communications

As explained above (S1.7.3), personal communications or unpublished materials do not belong in the list of references. Instead, the information is given in the text with the notation "(personal communication)" and the details of the source. For example:

> "In 2008 the park office registered 4,120 visitors, of whom 1,120 were foreign nationals (personal communication, Ms. Mary James, Park Liason Officer, interview 12-Sept-2009)."

Data from unpublished sources is mentioned as such. For example:

> "Daily rainfall has been recorded by a local observer since the 1930's. These logbooks are stored in the District Collector's office (personal communication, J Reddy, District Conservation Officer). The researcher photographed pages corresponding to 1990–2010 from the station logbook and manually entered the data in spreadsheet format."

1.7.5 Reference list and citation style

Formatting of in-text citations and entries in the list of references varies greatly among journals and book publishers.

At ITC **any consistent style is acceptable** as long as sufficient information is present in the reference list entry, and the reference list entry can be uniquely identified from the in-text citation; see above §1.7.3.

Many good sources of styles are available; especially recommended is the Council of Biology Editors' Style Manual [9] available in the ITC library. There are some style manuals available on-line, for example from the American Society of Agronomy *et al.* [1]. These two style manuals use an **author-date** system. The *APA-Published* style in EndNote is a reasonable implementation of this. **Numbered** styles take up less space in the text and do not distract the reader; you can still use the author's name in the text as a proper name if appropriate. This is the style used in these notes.

Here are some examples of different styles:

Journal article

APA 5th:

▷ Nijland, W., Addink, E. A., de Jong, S. M., & van der Meer, F. D. (2009). Optimizing spatial image support for quantitative mapping of natural vegetation. Remote Sensing of Environment, 113(4).

Geoderma

▷ Nijland, W., Addink, E.A., de Jong, S.M. and van der Meer, F.D., 2009. Optimizing spatial image support for quantitative mapping of natural vegetation. Remote Sensing of Environment, 113(4).

Chicago Manual of Style, 15th edition

▷ Nijland, W., E.A. Addink, S.M. de Jong, and F.D. van der Meer. "Optimizing Spatial Image Support for Quantitative Mapping of Natural Vegetation." Remote Sensing of Environment 113, no. 4 (2009).

IEEE

▷ [1] W. Nijland, E. A. Addink, S. M. de Jong, and F. D. van der Meer, "Optimizing spatial image support for quantitative mapping of natural vegetation," Remote Sensing of Environment, vol. 113, 2009.

Book

APA 5th:

▷ Barrett, E. C., & Curtis, L. F. (1999). Introduction to environmental remote sensing (4th ed.). Cheltenham, Glos., UK: Stanley Thornes Publishers.

Geoderma

▷ Barrett, E.C. and Curtis, L.F., 1999. Introduction to environmental remote sensing. Stanley Thornes Publishers, Cheltenham, Glos., UK, xxi, 457 pp.

Chicago Manual of Style, 15th edition

▷ Barrett, E. C., and L. F. Curtis. Introduction to Environmental Remote Sensing. 4th ed. Cheltenham, Glos., UK: Stanley Thornes Publishers, 1999.

IEEE

▷ [1] E. C. Barrett and L. F. Curtis, Introduction to environmental remote sensing, 4th ed. Cheltenham, Glos., UK: Stanley Thornes Publishers, 1999.

Chapter in an edited book

APA 5th:

▷ Foody, G. M. (2004). Sub-Pixel Methods in Remote Sensing. In S. M. D. Jong & F. D. v. d. Meer (Eds.), Remote Sensing Image Analysis: Including the Spatial Domain (pp. Ch 3). Dordrecht: Kluwer.

Geoderma

▷ Foody, G.M., 2004. Sub-Pixel Methods in Remote Sensing. In: S.M.D. Jong and F.D. v.d. Meer (Editors), Remote Sensing Image Analysis: Including the Spatial Domain. Remote Sensing and Digital Image Processing, Vol. 5. Kluwer, Dordrecht, pp. Ch 3.

Chicago Manual of Style, 15th edition

▷ Foody, G M. "Sub-Pixel Methods in Remote Sensing." In Remote Sensing Image Analysis: Including the Spatial Domain, edited by S. M. De Jong and F. D. Van der Meer, Ch 3. Dordrecht: Kluwer, 2004.

IEEE

▷ [1] G. M. Foody, "Sub-Pixel Methods in Remote Sensing," in Remote Sensing Image Analysis: Including the Spatial Domain, S. M. D. Jong and F. D. v. d. Meer, Eds. Dordrecht: Kluwer, 2004, p. Ch 3.

Web site

APA 5th:

▷ NASA. (2008, 12-March-2010). TERRA, the EOS flagship. Retrieved 13-April, 2010, from http://terra.nasa.gov/

A consistent entry in a reference manager can be re-formatted in hundreds of styles. You can fairly easily create or modify reference manager styles. There is no ITC house style. However, the correct formatting is only as good as the information in the bibliographic database. Thus, items in the reference manager's database must have accurate information in the correct fields.

Using **EndNote** [35] or another reference information manager ensures that:

1. All citations in the text are in the reference list, and vice-versa;
2. All citations in the text are formatted correctly;
3. All references in the reference list are formatted according to a consistent style.

Note: Besides EndNote there are other commercial programs, e.g. Reference Mananger[48], and some open-source alternatives such as Zotero[49] and JabRef[50]. The important point is to use one of these programs consistently and correctly.

1.7.6 EndNote tips

1. Select the correct **reference type** for each new reference. Some choices are: journal article, book, book chapter, and electronic source. You will notice that the fields in the record are different for each type.
2. Don't trust any **import filter**. Review each downloaded record to make sure the reference type is correct and that information is in the correct fields.
3. **Review the formatted references**, i.e. read your own list of references! If something seems out of place, it may be that the output style you are using doesn't use a certain field, or it may be that you have the wrong reference type for the reference, or it may be that the fields are not correctly filled in.
4. If you use the **automatic link from EndNote to Word**, never make a change in the formatted Word document; always change the EndNote reference or style. This ensures that all items of the same type will have the same format.

[48] http://www.refman.com/
[49] http://www.zotero.org/
[50] http://jabref.sourceforge.net/

5. Check the author names in your EndNote entry; there should be one **author name per line**, with no punctuation at the ends of lines. Using the format "LastName, First I." for authors with European names is most reliable.

Corporate authors

Enter **corporate authors** with a single comma after the complete name; EndNote will not try to abbreviate it:

- "National Resources Conservation Service,"

If you forget the final ",", in some styles EndNote will present this as *Service, N.R.C.* which is certainly not what you want.

Non-European author names

If an author's name does not follow the European convention of personal and family names, write the entire name **with a comma after it**; EndNote will not try to force it into a European format and will never try to abbreviate it. Examples are Chinese, Vietnamese, Indonesian and Ethiopian names:

- "Liu Xuehua," [34] (Chinese)
- "Nguyen Sinh Cung," (Vietnamese)
- "Sudibyakto," (Indonesian)
- "Atkilt Girma," (Ethiopian)

Sometimes an author from one of these groups will Europeanize their name, especially if they follow a career outside their country; it can be difficult to recognize this if you don't know the language and thus what are typical family and given names. For example, the Chinese author listed above followed the European convention while completing her PhD, and thus wrote the family name "Liu" second, as "Xuehua Liu", in her thesis and papers. To determine which order has been used, you have to be familiar with the language, and in this case know that "Liu" (刘) is a family name and "Xuehua" (雪花) a female's given name to recognize this shift[51].

In this case the entry in EndNote would be in the standard format, i.e. "Last Name, First Name":

- "Liu, Xuehua"

[51] In the case of Chinese names, seeing the name written in Chinese characters is the only foolproof method, since transliteration into the Pinyin or Wade-Giles romanization does not distinguish between the many homophonic characters. For example both "Gao Yan" (高炎) and "Yan Gao" (颜高) are valid names.

which EndNote could abbreviate in some styles as “Liu, X.”.

Spanish surnames are written three ways depending on the degree of formality, and you may encounter any of them:

- “José Antonio Navarrete Pacheco”: formal, with both father’s family name (*Navarrete*) and mother’s family name (*Pacheco*); this will be used in formal documents and theses; enter in EndNote as “Navarrete Pacheco, José Antonio”.
- “José Antonio Navarrete P.”: less formal, with father’s family name written out and mother’s family name abbreviated; enter in EndNote as “Navarrete P., José Antonio”.
- “José Antonio Navarrete”: every-day usage, with only father’s family name; this may also be used if the Spanish-surname author publishes in an English-language journal; enter in EndNote as “Navarrete, José Antonio”, which EndNote could abbreviate in some styles as “Navarette, J. A.”.

Q30 : *What should you do if you have several references by the same author, but using a different form of their name at different points in their career?* *Jump to A30* •

1.7.7 Answers to self-test questions

A*27* : *To provide the reader a comprehensive list of sources on a topic, whether you used them or not. This would be one objective and one output of the research project.* *Return to Q27* •

A*28* : *So that readers can check the source for themselves.* *Return to Q28* •

A*29* : *A professional librarian is trained to find materials with various indices and reference works, for example, standard lists of journal abbreviations.* *Return to Q29* •

A*30* : *Use the name as given in each reference. If it is important to link them, write some text to explain.* *Return to Q30* •

1.8 References

Bibliography

[1] American Society of Agronomy; Crop Science Society of America; & Soil Science Society of America. 2004. *Publications Handbook and Style Manual*. Madison, WI: American Society of Agronomy. URL: https://www.agronomy.org/publications/style/; accessed 23-March-2009 63

[2] Arbeitskreis Stadtböden. 1997. *Empfehlungen des Arbeitskreises Stadtböden der Deutschen Bodenkundlichen Gesellschaft für die bodenkundliche Kartierung urban, gewerblich, industriell und montan überformter Flächen (Stadtböden)*, volume Teil 1: Feldführer. Kiel: Sekretariat Büro für Bodenbewertung, 2. auflage edition 51

[3] Basher, L. R. 1997. *Is pedology dead and buried? Australian Journal of Soil Research* **35**:979–94 10

[4] Bishop, Y.; Fienberg, S.; & Holland, P. 1975. *Discrete multivariate analysis: theory and practice*. Cambridge, MA: MIT Press 19, 50

[5] Bivand, R. S. 2008. *R spatial projects*. URL: http://sal.uiuc.edu/csiss/Rgeo/, accessed 29-March-2008 29

[6] Bourne, P. E. 2005. *Ten simple rules for getting published. PLoS Computational Biology* **1**(5):e57. http://dx.doi.org/10.1371/journal.pcbi.0010057 20

[7] Brockhaus GmbH, F. A. 2000. *Der Brockhaus in einem Band*. Leipzig: F. A. Brockhaus GmbH, 9. edition 49, 59

[8] Buol, S. W.; Hole, F. D.; & McCracken, R. J. 1989. *Soil genesis and classification*. Ames, IA: The Iowa State University Press, 3rd edition 56

[9] CBE Style Manual Committee. 1994. *Scientific style and format : the CBE manual for authors, editors and publishers*. Cambridge: Cambridge University Press, 6th edition 62, 63

[10] Center for Advanced Spatial Technologies (CAST). 1998. *AR-GAP final report: State-wide biodiversity mapping for Arkansas*. Report, Center For Advanced Spatial Technologies (CAST), Fayetteville, AR. URL: http://www.cast.uark.edu/gap/ 21

[11] Dobos, E.; Micheli, E.; Baumgardner, M. F.; Biehl, L.; & Helt, T. 2000. *Use of combined digital elevation model and satellite radiometric data for regional soil mapping. Geoderma* **97**(3-4):367–391 9, 49, 55

[12] Dupriez, H. & de Leener, P. 1998. *Trees and multistorey agriculture in Africa : a textbook for agro - forestry*. Nivelles; Wageningen: Terres et Vie; Technical Centre for Agricultural and Rural Cooperation (CTA) 19

[13] Ettema, C. H. 1994. *Indigenous soil classifications*. On-line PDF document; URL: `http://www.itc.nl/personal/rossiter/Docs/Misc/IntroToEthnopedology.pdf`. Access date: 08-March-2008 22

[14] FAO. 1984. *Land evaluation for forestry*. Forestry Paper 48. Rome, Italy: Food and Agriculture Organization of the United Nations 59

[15] FAO. 1998. *Digital Soil Map of the World and derived soil properites (CD-ROM)*. Land and Water Digital Media Series No. 1. Rome: FAO

[16] Forbes, T. R.; Rossiter, D.; & Van Wambeke, A. 1982. *Guidelines for evaluating the adequacy of soil resource inventories*. SMSS Technical Monograph 4. Ithaca, NY: Cornell University Department of Agronomy, 1987 printing edition 49, 59

[17] Fowler, H. W. & Gowers, E. 1965. *A dictionary of modern English usage*. Oxford,: Clarendon Press, 2d edition 48, 52

[18] Gao Yan; Mas, J. F.; Maathuis, B. H. P.; Zhang Xiangmin; & Van Dijk, P. M. 2006. *Comparison of pixel-based and object-oriented image classification approaches; a case study in a coal fire area, Wuda, Inner Mongolia, China*. *International Journal of Remote Sensing* **27**(18):4039 - 4055 6

[19] Gaultier, J. P.; Legros, J. P.; Bornand, M.; King, D.; Favrot, J. C.; & Hardy, R. 1993. *L'organisation et la gestion des données pédologiques spatialisées: Le projet DONESOL (English abstract)*. *Revue de Géomatique* **3**(3):235–253

[20] Gobin, A.; Campling, P.; Deckers, J.; & Feyen, J. 2000. *Quantifying soil morphology in tropical environments: Methods and application in soil classification*. *Soil Science Society of America Journal* **64**:1423–1433 50

[21] Goovaerts, P. 1999. *Geostatistics in soil science: state-of-the-art and perspectives*. *Geoderma* **89**(1-2):1–45 10

[22] de Gruijter, J. 1999. *Special issue: Pedometrics '97*. *Geoderma* **89**(1-2):1–400 17

[23] de Gruijter, J. J. & Marsman, B. A. 1984. *Transect sampling for reliable information on mapping units*. In Nielson, D. R. & Bouma, J. (eds.), *Soil spatial variability: proceedings of a workshop of the ISSS and SSSA*, pp. 150–163. Wageningen: PUDOC 17, 49, 53, 59

[24] Hart, C. 1998. *Doing a literature review : releasing the social science research imagination*. London: Sage Publications 6

[25] Heuvelink, G. B. M. 1998. *Error propagation in environmental modelling with GIS*. Research Monographs in GIS. London: Taylor & Francis 47, 48, 52

[26] Hornby, A. S. (ed.). 1995. *Oxford advanced learner's dictionary*. Oxford: Oxford University Press 59

[27] King, D.; Bourennane, H.; Isambert, M.; & Macaire, J. 1999. *Relationship of the presence of a non-calcareous clay–loam horizon to DEM attributes in a gently sloping area*. *Geoderma* **89**(1-2):95–111 9, 17

[28] Kirkpatrick, E. M. & Roget, P. M. 1998. *Roget's thesaurus*. London: Penguin, new edition 34, 45

[29] Knuth, D. E. 1981. *The art of computer programming, Volume 2: Seminumerical algorithms*. Reading, MA: Addison-Wesley 54, 60

[30] Kubiëna, W. L. 1948. *Entwicklungslehre des Bodens*. Wien: Springer-Verlag 56

[31] Kutilek, M. & Nielsen, D. 1994. *Soil hydrology : textbook for students of soil science, agriculture, forestry, geoecology, hydrology, geomorphology or other related disciplines*. Geoecology paperback. Reiskirchen (D): Catena Verlag 19

[32] Lee, P. M. 2004. *Bayesian statistics: an introduction*. London: Arnold, 3rd edition 54

[33] Lillesand, T. M. & Kiefer, R. W. 1994. *Remote sensing and image interpretation*. New York: John Wiley & Sons, 3rd edition 19, 52

[34] Liu, X. 1997. *Analysis of factors influencing the giant panda's distribution using GIS: a case study in Wolong nature reserve, China*. MSc, ITC 66

[35] Maas-Prijs, P. 2002. *How to use EndNote 5.0*. Handout, ITC 65

[36] Marsman, B. A. & de Gruijter, J. J. 1986. *Quality of soil maps : a comparison of soil survey methods in a sandy area*. Soil survey papers 15. Wageningen: Soil Survey Institute 49, 53, 59

[37] McBratney, A. B.; Odeh, I. O. A.; Bishop, T. F. A.; Dunbar, M. S.; & Shatar, T. M. 2000. *An overview of pedometric techniques for use in soil survey*. *Geoderma* **97**(3-4):293–327 10, 48

[38] NASA. 2008. *TERRA (EOS AM-1)*. URL: `http://terra.nasa.gov/`, accessed 29-March-2008 29

[39] Newcomb, S. 1881. *Note on the frequency of use of the different digits in natural numbers*. *American Journal of Mathematics* **4**:39–40 54, 60

[40] Oliver, M. A.; Webster, R.; & Slocum, K. 2000. *Filtering SPOT imagery by kriging analysis*. *International Journal of Remote Sensing* **21**(4):735–752. (GEOBASE Abstract)

[41] van Reeuwijk, L. P. 1995. *Procedures for soil analysis*. ISRIC Technical Paper 9. Wageningen: ISRIC 49

[42] Richter, R. 2006. *Atmospheric/Topographic correction for satellite imagery (ATCOR-2/3 User's Guide)*. Wessling (D): DLR, 6.2 edition 50

[43] Ritter, R. M. (ed.). 2003. *The Oxford manual of style*. Oxford: Oxford University Press 57

[44] Rossiter, D. G. 2006. *Classification of urban and industrial soils in the World Reference Base for Soil Resources*. In *18th World Congress of Soil Science*. Philadelphia, PA (USA): IUSS 16, 18

[45] 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 17, 18

[46] Ruffin, E. 1832. *An essay on calcareous manures* 56

[47] Sahay, S. & Woolshan, G. 1996. *Implementation of GIS in India: organizational issues & implications*. *International Journal of Remote Sensing* **10**(4):385–404 48

[48] Scholl, H. J. 2005. *Organizational transformation through e-government: Myth or reality?* In [56], pp. 1–11 17

[49] Schraps, V. W. G.; Kersting, A.; Pingel, P.; Schneider, S.; Baumgarten, H.; Burghardt, W.; Hiller, D. A.; Köppner, T.; Ohlemann, S.; & Metzger, F. 2000. *Stoffbestand, Eigenschaften un räumliche Verbreitung urban-industrieller böden – Ergebnisse aus dem Projekt Stadtbodenkartierung Oberhausen–Brücktorvertiel*. Scriptum: Arbeitsergebnisse aus dem Geologischen Landesamt Nordrhein–Westfalen Heft 7, Geologischen Landesamt Nordrhein–Westfalen. ISSN 1430-5267 51

[50] Skidmore, A. K. 1999. *Accuracy assessment of spatial information*. In Stein, A.; Meer, F. v. d.; & Gorte, B. G. F. (eds.), *Spatial statistics for remote sensing*, pp. 197–209. Dordrecht: Kluwer Academic 18

[51] Soil Resources Inventory Group. 1981. *Soil resource inventories and development planning: Proceedings of workshops at Cornell University 1977-1978*. Soil Management Support Services (SMSS) Technical

Monograph 1. Washington, DC: Soil Conservation Service, USDA 49, 59

[52] Stein, A.; Meer, F. v. d.; & Gorte, B. G. F. (eds.). 1999. *Spatial statistics for remote sensing*. Dordrecht: Kluwer Academic

[53] USDA Natural Resources Conservation Service. 2000. *Welcome to the USDA Natural Resources Conservation Service homepage for North Carolina*. On-line document; URL: http://www.nc.nrcs.usda.gov/. Access date: 29-March-2008 22, 29

[54] Várallyay, G.; Hartyáni, M.; Marth, P.; Molnár, E.; Podmaniczky, G.; I., S.; & Kele, G. 1995. *Talajvédelmi információs és monitoring rendszer. 1 kötet módszertan*. Technical report, Földmüvelésügyi Minisztérium, Budapest. (In Hungarian) 55

[55] Webster, R. & Beckett, P. H. T. 1968. *Quality & usefulness of soil maps*. *Nature* **219**:680–682 49, 50, 59

[56] Wimmer, M.; Traunmüller, R.; Grönlund, Å.; & Andersen, K. V. (eds.). 2005. *Electronic Government: 4th International Conference, EGOV 2005, Copenhagen, Denmark, August 22-26, 2005, Proceedings*, volume 3591 of *Lecture Notes in Computer Science*. Springer. ISBN 3-540-28466-4 17, 71

[57] WWF. 2008. *APP irregularities threaten massive climate and tiger impact*. URL: http://www.panda.org/index.cfm?uNewsID=128041, accessed 29-March-2008 29

[58] Zhang, C.; Pan, W.; Chen, H.; Huang, C. 2007. *Monitoring vegetation coverage change of Fujian using multi-temporal remote sensing image*. *Remote Sensing Technology and Application* **22**(5). In Chinese, with English abstract 27, 58

2 Critical Reading and Abstracting

2.1 Critically reading a research paper

Key points

1. The research paper is the **primary unit of scientific production**;
2. Research papers are written for **specialists** with a strong background in the subject matter; as a student your background may not be as strong;
3. Your goal is the **extract** the information you need from the paper;
4. At the same time, you must **evaluate the reliability** of the paper;
5. A reliable research paper has **solved certain problems** that you do not have to duplicate in your own work;
6. However, no paper answers all question; the unsolved questions in the paper can **motivate your own research.**

2.1.1 What is a research paper?

A research paper presents **original work** done by the authors, along with a claim for the work's **novelty** and **significance**. It attempts to answer one or more **research questions**, identified by the authors, which have not previously been answered by others.

Some journals explicitly label research papers as such (to differentiate them from review papers, book reviews, opinion etc.), but in most journals it is asssumed that papers are original research, unless otherwise identified. Figure 2.1 shows a typical table of contents where the research articles are clearly marked as such.

2.1.2 Why read research papers?

Any topic you study is reported in a variety of ways:

- Popular articles
- Textbooks

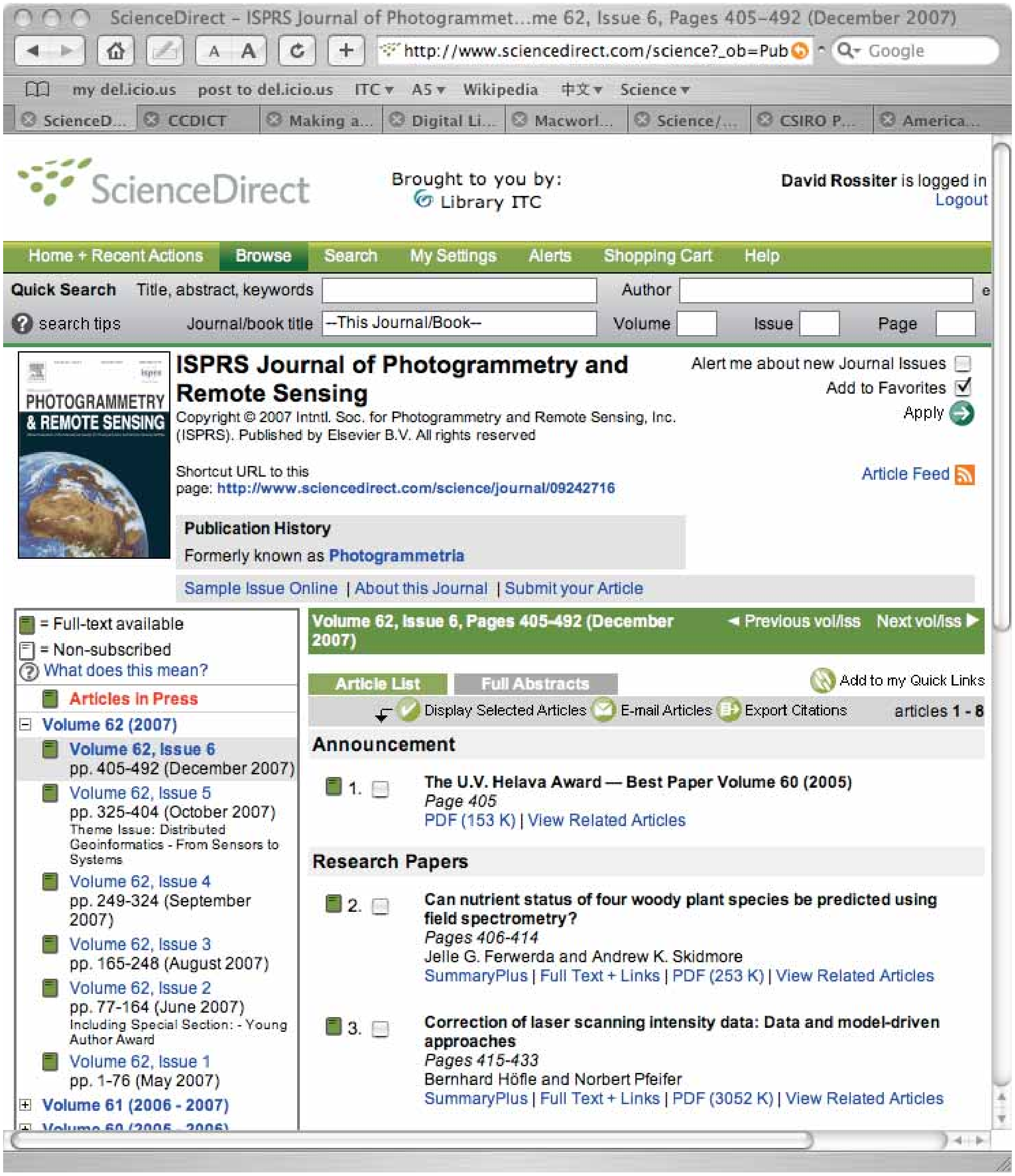


Figure 2.1: Journal Table of Contents

- Review articles
- Technical reports

    - Published
    - Unpublished (so-called “gray literature”)
- Conference papers and presentations
- Research papers

! → See the lecture notes on literature review for a more complete discussion of these sources and their reliability.

In this list, the research paper has a special value:

- It is the **primary unit of scientific production** (“minimum publishable unit”);
- It is the minimum unit that is presented to the reader as **reliable** and **self-contained**;
- It has been **peer-reviewed** for quality control;
- It is what the researcher wrote, rather than someone else’s summary or interpretation.

Thus, most of your time in the **literature review** should be spent with research papers. However, research papers are written for specialists with a strong background in the subject matter. The authors are requested to write in a concise style, assuming that the readers are familiar the vocabulary of a field and with previous studies. As a student your background may not be as strong, so it can be a difficult task for you to understand a research paper.

2.1.3 Difficulties reading research papers

It can be intimidating to read research papers, especially for the beginning graduate student:

- The **vocabulary** may be unfamiliar, especially if the paper is specialised;
- The **concepts** may be unfamiliar; this is a deeper problem than vocabulary, because it is harder to understand an abstract concept than a specific vocabulary item;
- The **methods** may be unfamiliar; especially since modern research generally uses sophisticated methods, well beyond most textbooks;
- The **writing** tends to be compact, especially in describing the methods, because the audience is assumed to be experienced research scientists with a knowledge of the field covered by the journal;

- A good paper will be fairly **comprehensive** ("deep"), with sophisticated argument;
- A good paper will refer to a large amount of **other work** in the field, which may be unfamiliar to you, and which you need to understand before you can fairly evaluate this paper's claims.

Eventually *you* will be an experienced research scientist in the field covered by the journal (but even then, you will often read papers that touch on your field but are not fully in it). Meanwhile, how do you approach a paper when you are still an apprentice researcher?

2.1.4 How to approach a research paper

The **overall aim** of reading a research paper is to determine the scientific innovation the authors claim to be making, and to evaluate that claim.

However, *your* aim is generally to **extract what is useful to you** in your own research. Thus you do *not* have to understand everything in the paper, but you do need to understand how it fits together.

So, **skim** and then **go deep** as necessary.

One approach to extracting the information you need is:

1. Read the **title**;
2. Read the **abstract**;
3. Skim the paper for its **structure**;
4. Read the **introduction** to get the context and purpose of the research;
5. Identify the research **objectives** and **questions**;
6. Read the **discussion and conclusions**, compare with objectives and questions;
7. Read the **methods**: how was the research carried out?;
8. Read the **conclusions**: what are the implications? How does this fit with other work?

During all this you may need to:

- Decide what parts of the paper you need to **understand in depth** and what parts you can safely **skim over**;
- Identify the **vocabulary** you don not understand and learn it from references;

- Follow the **references** (citations) for explanations in depth (especially for methods);
- Critically evaluate the **claims** in the paper.

2.1.5 What can you get out of a research paper?

Fong [3] divides the reading process into three steps:

1. **comprehension** of what the authors are saying;
2. **evaluation** of their claims;
3. **synthesis** and motivation for your own research.

Comprehension

The first step is to understand what the **authors** intend with the paper, i.e. what they claim. Later you will evaluate these claims, here you just have to decipher what they say.

1. Why was this research done?
   (a) What is the **motivation** for this research?
   (b) Is the paper mainly to develop or improve **methods** or to answer some **question** about the "real world"?
      - If about methods, what is wrong with existing methods?
      - If about the "real world", what was not (or imperfectly) known?
   (c) What is the **research problem**? I.e. something not known that this research will address.
2. What does this paper claim to do about the problem? In other words, what is **new**?
   - Does the paper propose an **entirely new research field** or **paradigm**? This is unlikely, but it does happen; an example from soil science is Phillips [7].
   - Does the paper ask a **new question** (no one has asked it before)?
   - Does the paper propose a **new approach** to answering the question?
   - Does the paper propose an **improvement** on **existing methodologies**?
   - Does the paper propose a **better answer** to the question, compared with answers from previous studies?

- Does the paper propose a **new synthesis** or **conceptual framework** of previous studies?

3. What **methods** do the authors use to address the problem?
   - Are there observations? If so, systematic (how?) or opportunistic?
   - Are there designed experiments?
   - Are computer simulations or models applied?
   - Do the authors propose new mathematical or statistical methods or computer algorithms?
   - Is there a comparison or analysis of documents?
4. What was the **result** of applying these methods?
   - For a field study, what information was collected?
   - For a simulation, model, processing etc., what was the result of the process?
   - For a synthesis, what is the new conceptual framework?
5. Are there any **case studies**? If so, what is the relation of the case study to the preceding theoretical development?
6. What makes the claims **scientific** (based on facts and logic), as opposed to being mere opinions? Or as Fong so nicely puts it "what makes it a research paper rather than science fiction?"
7. How do the authors **substantiate** (back up) their claims?
8. What **conclusions** do the authors draw from their results?
   - Is the result claimed to be of general interest or only applicable to the author's own case?
   - Do the authors identify problems that they could not solve, and which could be the subject of future research?
9. What could we get out of the paper for our **own work**?
   - According to the authors, should we change our **methods**? or at least adapt them in certain circumstances? If so, which?
   - According to the authors, are there open **research problems** which we should address?

Evaluation

The second step is to evaluate how successful the authors were. We've seen (we hope) what they claim, now what do we think of that? Not all research is equally significant, well-carried out, relevant or important. Of course, you may need wider experience in the field to fairly evaluate the paper. This is where a lack of background may require additional reading, especially of this paper's references.

1. How **significant** is the research **problem**?
    - Is the work just dealing with some small part of a problem (Fong: "scratching minor itches")?
    - Is the problem real, or artificial? Fong asks, do the authors put up some "strawman" which they can then easily knock down?
    - Is the problem well-known and unsolved?
    - Do we really care if the problem is solved?
2. How **significant** is the **contribution** to solving the problem?
    - Are the authors aware of the relation of their work to existing literature? If not, it's likely they are just repeating previous work.
    - Are the results surprising, i.e. seriously challenging previous results or received wisdom?
    - Does the work give us new or significantly improved tools?
    - Does the work tell us something really new about an application or about the "state of nature"?
    - Does the work have practical applications or has that been left for someone else to work out?
3. Are the claims **valid**?
    - Has the work been carefully done? Have important details been omitted (or not reported)?
    - If there is an experimental design, is it suitable for the problem?
    - Are the methods correct for the problem?
    - Are the mathematics, statistics, algorithms etc. correct?
    - Do the results support the author's conclusions?
    - Are there confounding factors that the authors should have considered, but didn't?
    - If the authors are comparing their work to previous work, are they making fair comparisons? That is, is like being compared with like?

- If the authors make generalizations from their study, are these valid? Do the authors sufficiently consider the differences between their situation and the more general one?
- Are the claims exaggerated or too modest?

Remember, the authors (and reviewers) are humans; the scientific enterprise is also a human enterprise. So, do not believe everything you read at face value. Always read critically.

Some work requires **extra skepticism**:

- Work that claims to be completely novel.

  > "Be very sceptical of work that is so 'novel' that it bears no relation to any existing work, builds upon no existing paradigm, and yet addresses a research problem so significant that it promises to transform the world. Such are the signs that the author might not be aware of existing literature on the topic. In such a case, the authors could very well be simply repeating works that have already been done decades ago." [3]

- Work with results that are in almost perfect agreement with the authors' hypotheses. The authors may be guilty of wishful thinking (at best).
- Work that claims to overturn a large body of previous results. It may well be true, but it is more likely that the authors are at best self-deceived.[1]

  As Carl Sagan said (paraphrasing the Scottish philosopher David Hune), "precisely because of human fallibility, extraordinary claims require extraordinary evidence."[2]
- Finally, be aware that some **fabrications** and **lies** may sometimes slip through the peer-review process.

Synthesis

The third step is to put the paper in context, see where it fits in the overall research agenda, and determine whether we can do better. This is where critical thinking is especially important.

Some questions you can ask yourself:

- What, finally, is the essential research problem that was addressed, and how well was it addressed; how completely was it solved?

[1] See for example the controversy on the "System of Rice Intensification", [8].
[2] Interview with WGBH Boston, 1996

- Could the research be improved? Deepened? Extended? If so, how?
- Are there other approaches to the research problem? Did the authors pick the most suitable, or can you think of another approach that might be more fruitful?
- Did the authors get full value from their approach and data? Could a deeper analysis have been done, to gain additional insight into the problem?
- Do the authors make a good argument to back up their claims, or can you think of a stronger argument?
- Can you make an argument against the case made by the authors? I.e. can you think of a counter-argument that would explain their results but with a different explanation? Can you think then of some way to determine which argument is correct?
- Are the research results valid in a wider context? Is more research needed to determine this? What aspects of the research might be different in another context?
- What are the unsolved problems related to this research? These may be identified by the authors or they may occur to you while reading.

2.1.6 Critical reading examples

We now discuss each of the steps of §2.1.4 "How to approach a research paper", using as examples two recent papers [6, 9] in the field of geo-information science and earth observation and one [2] in the field of geo-information management.

The first two are available on-line at the following addresses, using the permanent Digital Object Identifier (DOI), resolved by the DOI server at http://dx.doi.org/:

- Möller *et al.* [6]

  http://dx.doi.org/10.1016/j.jag.2006.10.002

- Wu *et al.* [9]

  http://dx.doi.org/10.1016/j.watres.2007.05.018

The last-named paper (Bekkers & Homburg [2]) is available via the ITC library at:

http://intranet.itc.nl/papers/ext2000/bekkers_myths_2007.pdf

In each section there are **self-study questions**, to test your understanding. If you are reading this on line, you can jump directly to the answer using the hyperlink written like this, and then back to the question.

Read the title

A well-written title gives the topic of the paper in a few words. It may be obvious already from the title that the paper is of no interest to you – but beware! It is is difficult to find a good title, so that some titles are vague or even misleading.

Here are the examples:

- "The comparison index: A tool for assessing the accuracy of image segmentation" [6]
- "Concurrent monitoring of vessels and water turbidity enhances the strength of evidence in remotely sensed dredging impact assessment" [9]
- "The myths of E-government: Looking beyond the assumptions of a new and better government" [2]

Looking at just the title:

Q31 : *Which of these papers proposes a new methodology?* *Jump to A31* •

Q32 : *Which of these papers is using geo-information technology to answer a question about the real world?* *Jump to A32* •

Q33 : *Which of these papers proposes a new conceptual framework?* *Jump to A33* •

Q34 : *Do any these papers contain a case study? If so, where?* *Jump to A34* •

Q35 : *List the technical words or phrases in the titles that you do not understand.* *Jump to A35* •

Read the abstract

As explained in the next section (§2.2), a well-written abstract is designed to present the most important information in the paper; it is the "paper in miniature", including the main claims to originality and the most important results. It should help the reader decide whether to read the paper in detail. So it makes sense to read the abstract first.

Here are the three abstracts, listed as single sentences to help you read them more easily.

First, from Möller *et al.* [6]:

1. Segmentation algorithms applied to remote sensing data provide valuable information about the size, distribution and context of landscape objects at a range of scales.
2. However, there is a need for well-defined and robust validation tools to assessing the reliability of segmentation results.
3. Such tools are required to assess whether image segments are based on 'real' objects, such as field boundaries, or on artefacts of the image segmentation algorithm.
4. These tools can be used to improve the reliability of any land-use/ land-cover classifications or landscape analyses that is based on the image segments.
5. The validation algorithm developed in this paper aims to: (a) localize and quantify segmentation inaccuracies; and (b) allow the assessment of segmentation results on the whole.
6. The first aim is achieved using object metrics that enable the quantification of topological and geometric object differences.
7. The second aim is achieved by combining these object metrics into a 'Comparison Index', which allows a relative comparison of different segmentation results.
8. The approach demonstrates how the Comparison Index CI can be used to guide trial-and-error techniques, enabling the identification of a segmentation scale H that is close to optimal.
9. Once this scale has been identified a more detailed examination of the CI-H diagrams can be used to identify

precisely what H value and associated parameter settings will yield the most accurate image segmentation results.

10. The procedure is applied to segmented Landsat scenes in an agricultural area in Saxony-Anhalt, Germany.
11. The segmentations were generated using the 'Fractal Net Evolution Approach', which is implemented in the eCognition software.

Q36 : *Classify these sentences into the the standard categories: (1) Rationale; (2) Hypothesis and objectives; (3) Methods; (4) Results; (5) Conclusions.* *Jump to A36* •

Second, from Wu *et al.* [9]:

1. Remotely sensed assessment of dredging impacts on water turbidity is straightforward when turbidity plumes show up in clear water.
2. However, it is more complicated in turbid waters as the spatial or temporal changes in turbidity might be of natural origin.
3. The plausibility of attributing turbidity patterns to dredging activities would be greatly enhanced when demonstrating association between dredging infrastructure and water turbidity.
4. This study investigated the possibility to strengthen the inference of dredging impact while simultaneously monitoring vessels and water turbidity in the northern Poyang Lake, China, where dredging was first introduced in 2001 and rapidly extended onwards.
5. Time-series of Landsat TM and MODIS images of 2000–2005 were used to estimate the distribution and number of vessels as well as water turbidity.
6. MODIS images revealed a significant increase in water turbidity from 2001 onwards. Landsat TM image analysis indicated a simultaneous increase in the number of vessels.
7. Regression analysis further showed a highly significant positive relationship (R2 = 0.92) between water turbidity and vessel number.

8. Visual interpretation of ship locations led to the conclusion that clear upstream waters developed turbidity plumes while passing the first cluster of vessels.
9. We concluded that dredging caused the increase in water turbidity, and simultaneously monitoring the water turbidity and vessels enhanced the strength of evidence in remotely sensed dredging impact assessment.

Q37 : *Classify these sentences into the the standard categories: (1) Rationale; (2) Hypothesis and objectives; (3) Methods; (4) Results; (5) Conclusions.* *Jump to A37* •

Notice how both of these abstracts begin their second sentence with the qualification "However, ...". This is a conventional formulation: first sentence to show the importance of the work, second sentence to show what was missing and is addressed in this work.

Third, from Bekkers & Homburg [2]:

1. In general, rhetoric and myth play important roles in policy-making.
2. Myths may inspire collective action but may also mystify and blur views on reality.
3. In this article we identify, analyze, and reflect on the myths underlying the e-government programs of Australia, Canada, the United Kingdom, Denmark and the Netherlands.
4. We found that in all national policies myths of technological inevitability, a new and better government, rational information planning, and empowerment of the intelligent citizen can be discerned.
5. Although the mobilizing powers of these myths are acknowledged, we conclude that existing empirical studies have generated little support for the inescapable telos[3] of these myths, which makes canvas cleaning effects of e-government initiatives less likely.

Q38 : *Attempt to classify these sentences into the the standard categories: (1) Rationale; (2) Hypothesis and objectives; (3) Methods; (4)*

[3] OED: "End, purpose, ultimate object or aim"

Results; (5) Conclusions. If you have difficulty, state why. Jump to A38
•

Skim the paper for its structure

Most papers have numbered sections, and perhaps some un-numbered subheads, that reveal the structure. The HTML versions of papers available on the web, as well as the tables of contents of PDF documents, provide a convenient way to see these.

Looking at the on-line HTML version at `http://dx.doi.org/10.1016/j.jag.2006.10.002`, we see that Möller *et al.* [6] article has the following structure:

1. Introduction
2. Methods
   2.1. Image segmentation
   2.2. Segmentation validation
      2.2.1. Object metrics
      2.2.2. Integral consideration of object metrics
      2.2.3. Local validation
      2.2.4. Global validation
3. Study area
4. Results
   4.1. Reference objects and segmentation
   4.2. Local validation
   4.3. Global validation
5. Conclusions

References

Q39 : *What is the first-level sequence? Do you agree with the order of topics?* Jump to A39 •

Q40 : *Do the subsections of the Methods match those of the Results?* Jump to A40 •

The PDF of the second article [9] includes a table of contents in the bookmarks pane (Figure 2.2).

Q41 : *What is the first-level sequence? Do you agree with the order of topics?* Jump to A41 •

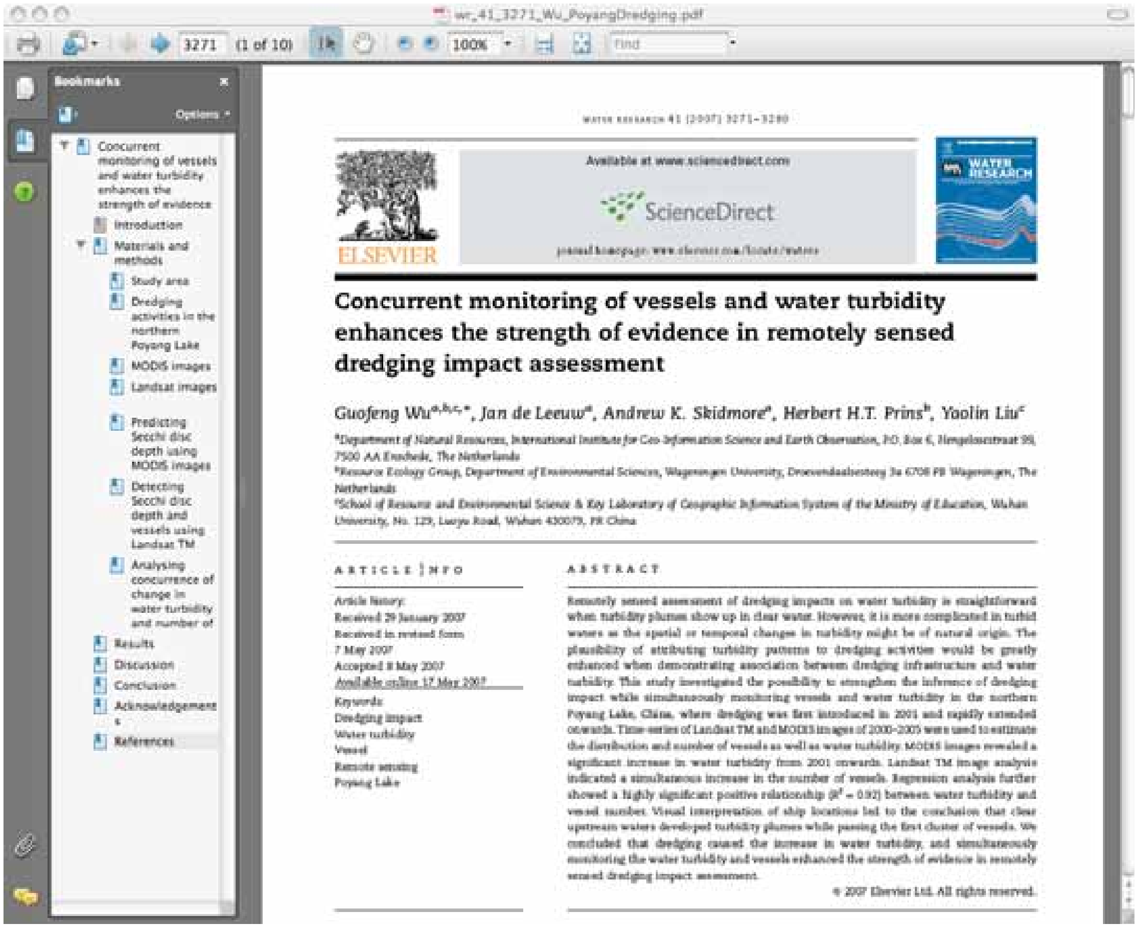


Figure 2.2: PDF Table of Contents

Q42 : *Do the subsections of the Methods match those of the Results?* *Jump to A42* •

Q43 : *What is the second-level sequence of the Methods? Do you agree with the order of topics?* *Jump to A43* •

The conceptual paper by Bekkers & Homburg [2] has quite a different structure[4]:

1. (Introduction) – implied, not a section title
2. The concept of e-government
3. The role of myths in policy processes

[4] the numbers here are for easy reference, they are not used in the paper

4. Myth 1: A new and better government
   4.1 Reconstruction of the myth of a new and better government
   4.2 Reflections on a new and better government
5. Myth 2: The myth of technological progress and instrumentality
   5.1 Reconstruction of the myth of technological progress and instrumentality
   5.2 Reflections on technological progress and instrumentality
6. Myth 3: The myth of e-government as rational information planning
   6.1 Reconstruction of the myth of rational information planning
   6.2 Reflections on as rational information management
7. Myth 4: The myth of citizen as empowered consumer
   7.1 Reconstruction of the myth of citizen as empowered consumer
   7.2 Reflections on the citizen as consumer
8. Conclusion

Q44 : *What is the first-level sequence? Does it seem logical?* *Jump to A44* •

Q45 : *Each of the four "myths" has two sub-sections. What is the structure?* *Jump to A45* •

Read the introduction

This should give the background to the work. It also discusses previous work on the subject, and establishes the **orginality** of the research.

Here is the first paragraph of the introduction from Möller *et al.* [6]:

> "Image segmentation algorithms such as those contained within eCognition™ are increasingly popular for a wide range of image processing tasks, and the advantages of working with image segments, rather than individual pixels are widely recognized ([Fortin et al., 2000] and [Shi et al., 2005]). However, there is a wide range of variables to manipulate, whereas segmenting an image and identifying an 'optimal' result can be difficult. The tools developed in this paper aim to make this process more objective and rigorous."

Q46 : *What do the authors identify as their original contribution?* *Jump to A46* •

Q47 : *How do the authors justify their statement that "the advantages of working with image segments, rather than individual pixels are widely recognized"?* *Jump to A47* •

Here are the two references mentioned in the paragraph:

Fortin *et al.*, 2000 M.-J. Fortin, R. Olson, S. Ferson, L. Iverson, C. Hunsaker, G. Edwards, D. Levine, K. Butera and V. Klemas, *Issues related to the detection of boundaries*, Landscape Ecol. 15 (2000), pp. 453–466.

Shi *et al.*, 2005 W. Shi, M. Ehlers and M. Molenaar, *Uncertainties in integrated remote sensing and GIS*, Int. J. Remote Sens. 26 (14) (2005), pp. 2911–2915.

Q48 : *In what journals are these papers? Are these both well-known peer-reviewed journals?* *Jump to A48* •

Q49 : *Are the titles of the two papers related to the statement to be justified?* *Jump to A49* •

Here are two paragraphs from the Introduction to Wu *et al.* [9]:

> "Indeed, it is easy to infer dredging from remotely sensed turbidity patterns when a plume of turbid water appears in a clear water environment, because the plume describes the impacted area while the tip localizes the dredging vessel. However, in turbid waters, it is much more difficult to access the impact of dredging, as sediment plumes (if detectable) are not as easily recognizable as in clear water. Moreover, highly turbid waters do not necessarily imply dredging impacts, because the high turbidity could reflect natural variability as well.
>
> "Thus, it would be a challenge to improve the ability to discern dredging impacts in waters with high and variable turbidity. Osenberg et al. (1994), exploring techniques to assess environmental impacts from field assessments, remarked that it was challenging to isolate effect from the noise introduced by natural variability in a system. Schmitt et al. (1996) argued

> that it was difficult to demonstrate the effects of environmental impacts with traditional scientific approaches, which rely solely on experiments with replicated treatments. In order to overcome these problems, an alternative before-after (BA) design approach that compares the state of the environment before and after an impact, or before-after control-impact (BACI) design, comparing control and impact sites, was used. So far, such study designs have, to our knowledge, not been employed to infer the dredging impacts from remote sensing images."

Q50 : *According to the authors, what is the main difficulty in detecting dredging activities from remotely-sensed imagery?* *Jump to A50* •

Q51 : *How do they propose to overcome this difficulty?* *Jump to A51* •

Here is the second reference from this section:

Schmitt *et al.*, 1996 R.J. Schmitt, C.W. Osenberg, W.J. Douros and J. Chesson, The art and science of administrative environmental impact assessment. In: R.J. Schmitt and C.W. Osenberg, Editors, Detecting Ecological Impacts: Concepts and Applications in Coastal Habitats, Academic Press, UK (1996), pp. 281-293.

Q52 : *Other studies (Schmitt et al., 1996) have already proposed the BACI design; so what is new about the present study?* *Jump to A52* •

Q53 : *To what extent does the present paper propose a new methdology?* *Jump to A53*
•

Here is an extract from the first paragraph of Bekkers & Homburg [2]:

> "E-government – or electronic government – is one of the buzzwords in the discussions on modernizing public administration (*4 references*). Modern information and communication technologies (ICTs), especially Internet and web technologies, are seen as enhancing the access, transparency, efficiency, and quality of public administration. . . . Not withstanding the intuitive appeal of these claims, studies have shown that the actual implementation of e-government initiatives has been disappointing (*7 references*)."

Q54 : *How does this introduction, along with the title*[5] *set the stage for the body of the paper?* *Jump to A54* •

Identify the research objectives and questions

Somewhere in the paper the authors identify these, often at the end of the introduction. Both of the papers used here as examples are quite clear on this, which makes it easy to later assess how successful they were.

Möller *et al.* [6] state:

> "The main objectives of this study are:
>
> "(1) the development of a user-friendly evaluation procedure to visualize and quantify segmentation inaccuracies . . .
>
> "(2) the assessment of segmentation results on the whole, by which we want to achieve an optimal parameter setting of the applied segmentation method . . . "

The last paragraph of the introduction of Wu *et al.* [9] declares the research objective:

> "This paper investigates the possibility to corroborate the strength of inferring dredging impact on water turbidity, while simultaneously monitoring the number and distribution of vessels and water turbidity using time-series Landsat TM and MODIS images in the northern Poyang Lake."

Read the methods

There are two main reasons to read the methods:

1. To evaluate whether appropriate methods were used;
2. You may want to use some of them in your own work.

The methods usually have a lot of detail, and require a strong background in previous studies or texts. On first reading, you need to identify what methods were used, without necessarily understanding them. When you evaluate the paper, then you need to determine if the methods were appropriate and applied correctly.

Here is a description of the methods for "object-based segmentation validation" from §2.2.1. "Object metrics" of Möller *et al.* [6]; the mentioned figure is reproduced here as Figure 2.3.

[5] "The myths of E-government: Looking beyond the assumptions of a new and better government"

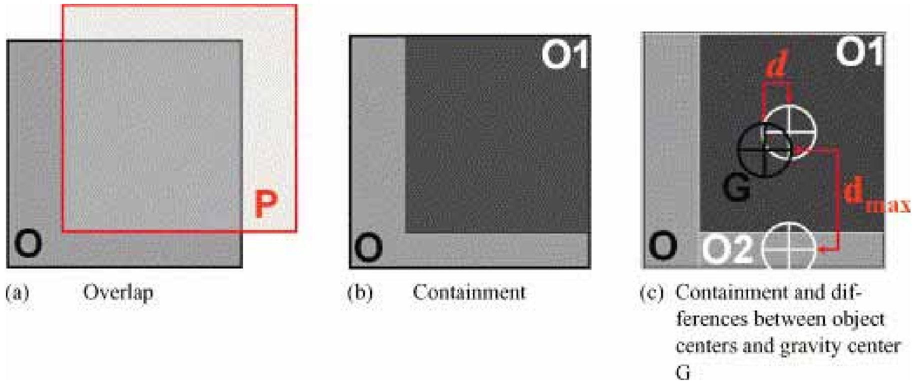


Figure 2.3: Figure 1 from Möller *et al.* 2006

"Object-based segmentation validation can be described as 'the problem of matching objects' (Zhan et al., 2005) where at least two hierarchical object-levels have to be considered. Object differences can be specified by, (1) topological and (2) geometric relationships ([Molenaar, 1998], [Ragia and Winter, 2000], [de Bruin et al., 1999] and [Zhan et al., 2005]):

"(1) the topological relationships of interest are 'containment' and 'overlap' (Fig. 1). In Fig. 1a object P overlaps object O. Both object levels are not hierarchically connected. If a topological overlay GIS-operation was carried out, the resulting object O1 is contained in the primary object O (Fig. 1b). O1 and O are hierarchically coupled if a 'part of'-relation was created (de Bruin et al., 1999). As a result, O1 is a sub-object of the superior or super-object O. Metrics of containment arise from the comparison of object sizes between related super- and sub-object;

"(2) geometric object differences can be determined by the comparison of object positions. Common metrics arise from distances between the gravity centers or skeletons as well as super- and sub-objects ([Ragia and Winter, 2000] and [Zhan et al., 2005]). Fig. 1c shows the gravity center G of the super-object O and the centers of sub-object O1 and O2."

Q55 : *According to the authors, what are the two methods to compare a object identified by a segmentation algorithm with a reference (true) obect?* *Jump to A55* •

Here is an extract from §2.6 “Detecting Secchi disc depth and vessels using Landsat TM images” of Wu *et al.* [9], describing how the number of vessels involved in dredging was determined:

> “During a field visit to Poyang Lake, we noted that barges transporting the sand passed by in regular order (i.e. one following the other, separated by a distance of one to a few hundred metres). Vessels were easily recognizable in Landsat TM imagery, as they manifested themselves as a sequence of linearly arranged dots. While exploring six Landsat TM bands, we noted that the barges were well discriminated by band 7, a short-wave infrared band (2.08–2.35 um) with strong water absorption and strong reflectance of soil and rock[6]. The following processes were used to detect and count vessels from each Landsat TM image. We used unsupervised classification to classify band 7 into ten classes. A visually established threshold was used to discriminate platforms and barges from water. The resulting binary image was transformed into polygons with vector format, and the number of polygons was considered to represent the number of vessels.”

Q56 : *What processing steps were used to count the vessels?* *Jump to A56* •

Q57 : *How do the authors justify that this method gives an accurate count?* *Jump to A57* •

Q58 : *Do you see any problem with their method?* *Jump to A58* •

The methods used by Bekkers & Homburg [2] are quite a bit “softer”:

> “We reflect on the cleavage [between perception and reality in e-government] in a cultural, narrative sense, by reading against the assumptions embodied in policy documents . . . We do so by analyzing e-government policies and technologies as myths . . . [which] we define . . . as hymns to progress, and as utopian visions or promises unfilfilled or unfulfillable . . . [P]olicy documents of the Netherlands, the United Kingdom, Denmark, Australia and Canada are scrutinized. . . . In order to confront the rhetoric with the reality of e-government, we analyze a

[6] `http://web.pdx.edu/~emch/ip1/bandcombinations.html`, accessed 16 October 2006

> range of assumptions in the policy documents ... First, we identify assumptions with respect to the goals and ambitions behind e-government initiatives ... "

Q59 : *What is the only source of information that the authors will use in their study?* *Jump to A59* •

Read the results and discussion

This is usually the "meat" of the paper, where the authors state what they believe their research shows. Here is where you must be especially critical.

Here is a paragraph from Möller *et al.* [6], discussing the results of five different segmentation approaches, named S1-S5; the mentioned figure is reproduced here as Figure 2.4.

> "The resulting CI–H-diagrams show differences regarding the 'segmentation accuracy' CI_A (Fig. 6). CI_A values vary from 88.4 (S5) to 90.2 (S3). Accordingly, the results of S3 version show the best topological and geometric similarity. This means that the optimal scale parameter setting for detecting agricultural field boundaries in the study area is H_{opt} of 56, and this scale parameter should be used in combination with a w_{shape} parameter of 0.4 and a w_{compt} parameter of 0.3 (which were the settings used for the S3 segmentation)."

Q60 : *According to the authors, which segmentation method gave the best overall results for detecting agricultural field boundaries?* *Jump to A60* •

Q61 : *Do they suggest that the same settings for the various classification parameters would be found for other types of objects?* *Jump to A61* •

Here is a paragraph from Wu *et al.* [9], discussing the relation of Secchi disc depth and day number; the mentioned figure is reproduced here as Figure 2.5.

> "Fig. 5 displays the variation of Secchi disc depth in the northern Poyang Lake from 2000 to 2005. The following regression model explained 56% of the variation of Secchi disc depth:

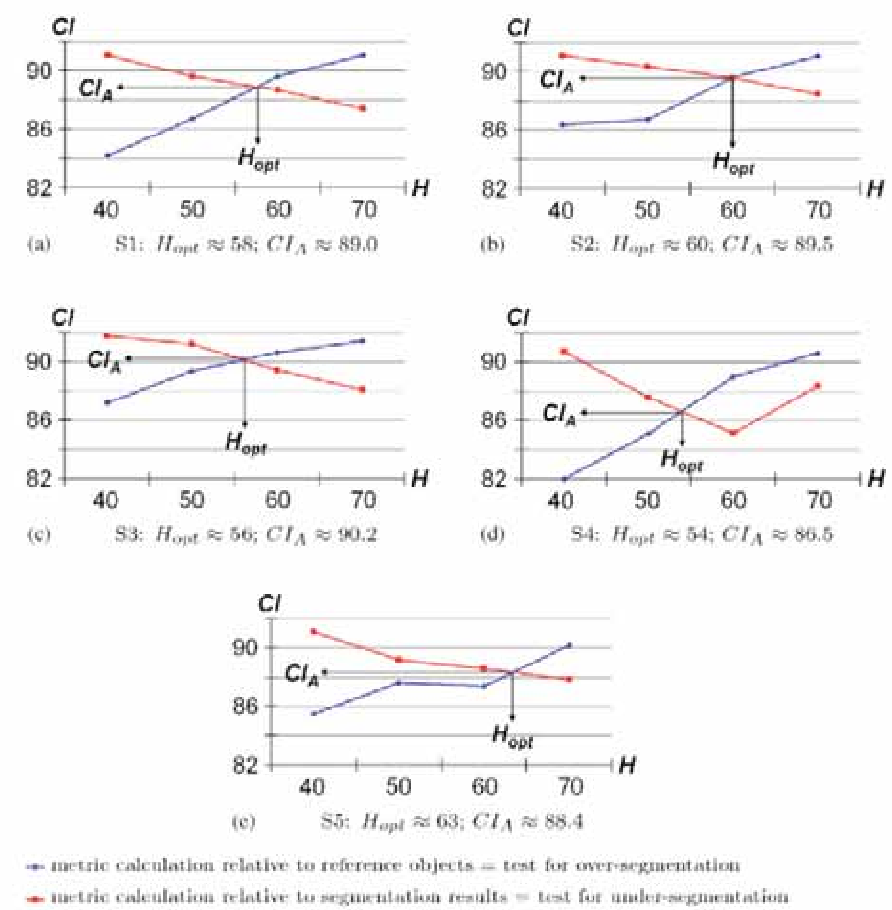


Figure 2.4: Figure 6 from Möller *et al.* 2006

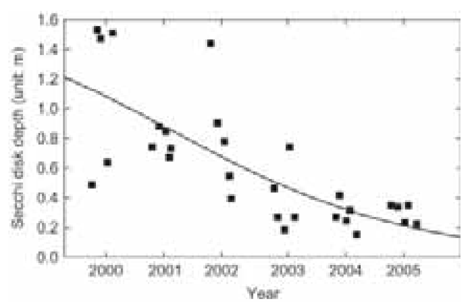


Figure 2.5: Figure 5 from Wu *et al.* 2007

$$\text{SDD} = 1.7 / \left(1 + e^{-0.9098+0.0014\text{D}}\right)$$

> where SDD represents Secchi disc depth and D the day number since 1 January 2000. The slope of this regression equation differs significantly from zero (b=0.0014, s.e.=0.00029, d.f.=29, t=4.79, $P < 0.001$). Hence, we rejected the null hypothesis which states that there is no change, and accepted the alternative, namely, a significant decline in Secchi disc depth."

Q62 : *What proportion of the variation in Secchi disc depth could be explained by the given regression model?* *Jump to A62* •

Q63 : *On what basis is the null hypothesis of no relation rejected?* *Jump to A63* •

The results of Bekkers & Homburg [2] are in the form of a narrative analysis of four "myths" in turn; each "myth" is first described ("reconstructed") and then evaluated ("reflected"). Here is a selection of the results for "Myth 2: The myth of technological progress and instrumentality":

> (Reconstruction) "In the UK the various promises of ICT are written in the imperative: 'ICT will . . . ,' for instance, 'make our life easier'. . . . ICT as an exogenous driving force is also evident in Danish documents . . . Dutch programs . . . show a strong belief and trust in the potential of modern ICT . . . In Australia . . . there is hardly any sphere of activity that could not be improved by online government . . . The Canadians also see a changing landscape in which distance perishes and a picture of ubiquitous computing dawns . . . "
>
> (Reflection) "In the various national policy documents, there is a strong belief and trust in the potential of ICT's. Optimism prevails in the descriptions of the progress the information society and Internet technology will bring."

Q64 : *How are the authors presenting their results?* *Jump to A64* •

Read the conclusions

Finally, the authors wrap up the paper with conclusions and implications for future research.

Here is the first paragraph of the conclusions of Möller *et al.* [6]:

> "The main objective of this paper was the determination of an optimal parameter setting of the FNEA-segmentation method in eCogntion™ based on the assumption that an optimal parameter setting is reached when over-and under-segmentation are balanced. The approach developed in this paper demonstrates how a two-stage process can be used to identify a segmentation scale H that is close to optimal using trial-and-error tests in combination with the CI metric. Once this scale has been identified a more detailed examination of the CI-H-diagrams can be used to identify precisely what H value and which w_{shape} and w_{compt} values will yield the most accurate image segmentation results."

Q65 : *Do the authors conclude that their objectives have been met? Do they suggest that it should be generally adopted?* *Jump to A65* •

Here is the final paragraph of Wu *et al.* [9]:

> "We argue that there is reason for concern about the possible environmental impacts on the ecology of Poyang Lake given the experiences on dredging impacts from abroad and recent concerns expressed in the Chinese media. So far, the nature and magnitude of these ecological impacts remain unknown. We suggest that it would be worthwhile to explore these possible impacts in order to develop scientific knowledge to support the decisions that need to be made by the responsible authorities, as they need to balance the pros and cons of dredging when deciding on how to rationally manage this unique lake ecosystem."

Q66 : *According to the authors, what further research is needed, and why?* *Jump to A66* •

Here is part of the Conclusion of Bekkers & Homburg [2]:

> "Our analysis shows that there is indeed a dominant, powerful mythical component to many e-government policies . . . an inescapable telos suggestion that technology by itself enables

> or even causes public sector agencies to transform themselves from self-centered conglomerates to citizen-oriented administrative apparatuses . . . [T]he chasm between the ambitious goals . . . and the rather disappointing pace of implementation . . . raises questions about the usefulness of the myths."

Q67 : *How does this conclude the analysis? What further steps are implied?* *Jump to A67* •

2.1.7 Answers to self-test questions

A_**31**_ : *The first-listed paper [6].* *Return to Q31* •

A_**32**_ : *The second-listed paper [9]. Note the phrase "remotely sensed"; this is the GI technology used for the mentioned "concurrent monitoring".* *Return to Q32* •

A_**33**_ : *The third-listed paper [2]. The "myths" of the title are used as the organizing principle in a conceptual framework.* *Return to Q33* •

A_**34**_ : *The titles give no clue about case studies; the papers may contain them but the titles do not say.* *Return to Q34* •

A_**35**_ : *Some words and phrases that might be difficult:* image segmentation, concurrent monitoring, turbidity, impact assessment, myth *(this common word is clearly being used in a technical sense)*, E-government. *Return to Q35* •

A_**36**_ : *(1) Rationale: sentences 1–4; (2) Hypothesis and objectives: sentence 5; (3) Methods: sentences 6–7, 10–11; (4) Results: not stated here; (5) Conclusions: sentences 8–9.* *Return to Q36* •

A_**37**_ : *(1) Rationale: sentences 1–3; (2) Hypothesis and objectives: sentence 4; (3) Methods: sentence 5; (4) Results: sentences 6–8; (5) Conclusions: sentence 9.* *Return to Q37* •

A_**38**_ : *(1) Rationale: sentences 1–2; (2) Hypothesis and objectives: not stated; it is implied that the myths of e-governance have not been satisfactorily analyzed, so the objective is to do this; the implied hypothesis is that there are common myths in the named countries, with some specific differences that can lead to a comparitive analysis; (3) Methods: sentence 3 (although, the actual method of 'compare' is not stated); (4) Results: sentences 4; (5) Conclusions: sentence 5.* *Return to Q38* •

A_**39**_ : *Introduction → Methods → Study area → Results → Conclusions*

Yes, this is logical. *Return to Q39* •

A_**40**_ : *The subsections of the Results seem to match the methods as follows:*

2.1. Image segmentation ⇔ 4.1. Reference objects and segmentation
2.2.3. Local validation ⇔ 4.2 Local validation
2.2.4. Global validation ⇔ 4.3 Global validation

Return to Q40 •

A41 : *Introduction → Materials & methods → Results → Discussion → Conclusion*

Yes, this is logical. *Return to Q41* •

A42 : *The Results does not have subheadings to match the methods.* *Return to Q42* •

A43 : *Study area → Dredging activities in the area → MODIS images → Landsat images → Predicting Secchi disc depth from MODIS → Detecting Secchi disc depth and vessels from Landsat → Analysing concurrence between turbidity and vessels.*

Yes, this is logical. We first see where the study is located, and learn about the activities that may be causing the problem. We then find out about the imagery that will be used. The first-named imagery is used for one purpose, the second-named for another. Finally we put the two together to try to answer the main research question.

Return to Q43 •

A44 : *Conceptual background → four myths → Conclusions (presumably considering all four myths).*

The four myths seem to be of equal value; the order seems arbitrary. *Return to Q44* •

A45 : *Each myth has first a "Reconstruction" (review and synthesis) and then a "Reflections" (evaluation). This parallel structure emphasizes that each myth is being analyzed in the same way.* *Return to Q45* •

A46 : *An objective and rigorous process for identifying 'optimal' image segmentation.* *Return to Q46* •

A47 : *The papers by Fortin et al. and Shi et al..* *Return to Q47* •

A48 : Landscape Ecology *and* International Journal of Remote Sensing, *both* *well-regarded peer-reviewed journals.* Return to Q48 •

A*49* : *Not obviously.* Return to Q49 •

A*50* : *In normally-clear waters, there is no problem: any turbidity is clearly due to human activity. However, in turbid waters, where suspended sediments come from many sources, turbidity can be detected easily enough, but the link to dredging activities is tenous.* Return to Q50 •

A*51* : *They propose a "before-after control-impact (BACI)" experimental design.* Return to Q51
•

A*52* : *The present study applies this concept to a concrete situation: determining the effect of dredging activities on turbidity in Poyang Lake by means of remote sensing.* Return to Q52 •

A*53* : *The general outline of the method is already known; its applicability to this situation, and possible adaptations for this situation, are not.* Return to Q53 •

A*54* : *"E-government" is clearly the topic of the paper – it is even the first word in the Introduction. The second sentence states the prevailing view (overall "myth") that ICT will improve government. The last sentence repeats the views of many authors that the reality has not lived up to the promise.*

This leads the reader to expect that the authors will somehow "solve" the discrepency, or at least explain it. Return to Q54
•

A*55* : *(1) Topological overlay of reference and classified object, followed by an comparison of object sizes; (2) distance between the centres of gravity of overlapping reference and classified objects.* Return to Q55 •

A*56* : *(1) Unsupervised classification of TM band 7 into ten classes; (2) thresholding into a binary image vessel/no vessel; (3) transform classified image to polygons; (4) count the polygons.* Return to Q56 •

A*57* : *Visual evaluation of TM image, compared to field visit, which showed that the pattern of vessels (linear in the main channel) could also be seen on the image; visual evaluation of unsupervised classification of TM band 7 and* *subsequent thresholding.* *Return to Q57* •

A58 : *The main problem is that the field visit is not declared to coincide with TM date. There is no suggestion that the actual vessel locations on the imagery date were recorded; thus we have to trust the inference process of the researchers, as well as their selection of a threshold.* *Return to Q58* •

A59 : *They will examine written policy documents from the named countries. There are no interviews, focus groups, or objective measures of success of e-government. This paper is only about examining myths.* *Return to Q59* •

A60 : *S3, with the given parameters.* *Return to Q60* •

A61 : *No; the specific wording implies that other parameters would be optimal under other circumstances.* *Return to Q61* •

A62 : *56% (just over half); the rest is unexplained.* *Return to Q62* •

A63 : *The slope of this regression equation differs significantly from zero, as revealed by a* t*-test. A zero slope is expected by the null hypothesis of no association.* *Return to Q63* •

A64 : *They list the expression of the myth from each country in the first subsection, and then generalize in the second.* *Return to Q64* •

A65 : *Yes, and they give a specific method for achieving an optimum classification, which is implied to be applicable in all circumstances.* *Return to Q65* •

A66 : *This paper established a relation between dredging and water turbidity, but the environmental effects of this turbidity are unknown. Without this, the responsible authorities do not have sufficient information to balance the pros and cons of the economically-important sand dredging.* *Return to Q66* •

A67 : *After examining the four myths, the authors generalize to this concluding statement. The gap ("chasm") between myth and reality calls into question whether the current myths are useful. They imply that the myths should be re-examined and either re-formulated or replaced.* *Return to Q67* •

2.2 Abstracting a research paper or thesis

Key points

1. The abstract is often **the only part of your work that will be read**; either because it is all that is available or because the reader is in a hurry.
2. The abstract is **the paper in miniature**; everything that is important in the paper must be included (in abbreviated form) in the abstract.
3. The abstract does not usually contain **citations** or **detailed reasoning**; there is not enough room to prove your case as you do in the thesis or paper.

2.2.1 Purpose of abstracting

Your thesis work fits into the larger enterprise of scientific progress. Others want to know what you have accomplished and what you have discovered, so they can verify or extend your work, or just use its results. The abstract is the only part of your work that most people read, and sometimes the only part they have available to them.

The abstract is meant to be read by **fellow-researchers**, either in the same field or in related fields. The first group will likely read the entire paper if they can, the second not.

The abstract is not intended for **policy makers**, even if scientifically-literate; they need executive summaries. Nor is the abstract intended for the **general public**; they need popular-science news articles. Both of these types of communication are quite important for the scientific enterprise; but here we are concerned with the scientific abstract for fellow researchers.

To quote American Society of Agronomy *et al.* [1, p. 8]:

> "An abstract has two typical uses. Printed at the head of a scientific paper, helps readers decide whether to delve into the paper as a whole; abstracts are also published separately in outlets such as Web sites and secondary and indexing journals. Thus, **the abstract will be seen and read by many more people than will read the paper**.
>
> "With this in mind, a basic rule emerges: **Everything that is important in the paper must be reflected in the abstract.** Let the abstract call attention to new techniques, observations, or data. Be specific.

The abstract is the paper in miniature

> In essence, an informative abstract . . . presents **the paper in miniature**, complete within itself. It moves from an introductory statement of the rationale and objectives or hypotheses, through materials and methods, to the results and conclusions."

Different journals have different rules; when you write for a journal consult their guide for authors. In general, the the abstract must be written as:

- one continuous paragraph;
- with a limit of 250 to 300 words; thus fitting on one page.

Some journals label sections of their abstracts, to force the author to structure the abstract. For example, the *Journal of Applied Ecology* author guidelines[7] give a formula for the "Summary" (replacing the "Abstract" of other journals):

> "The Summary should follow a formula in which point 1 sets the **context and need for the work**; point 2 indicates the **approach and methods used**; the next 2-3 points outline the **main results**; and the last point identifies the wider implications and **relevance to management or policy**. The final summary point carries the subheading **'Synthesis and applications'** and is the most important of all in maximising the impact of the paper. It should synthesise the paper's key messages as widely as possible: it should be generic, seminal and accessible to non-specialists. The whole Summary should be readily understandable to all the Journal's readers, and must not exceed 350 words."

As another example, the *Journal of Soils and Sediments* author guidelines[8] require an abstract of 150 to 250 words, divided into four named paragraphs: (1) Purpose (stating the main purposes and research question); (2) Methods; (3) Results; (4) Conclusions.

2.2.2 Where are abstracts found?

- At the head of the paper in the **printed journal**
- At the head of the paper in the **on-line journal**
- In **on-line abstract databases**, e.g. Web of Science

[7] http://www.journalofappliedecology.org/view/0/authorGuideline.html
[8] http://www.springer.com/environment/soil+science/journal/11368

Figure 2.6 shows the abstract from Wu *et al.* [9] in the abstracting database Web of Science[9].

Figure 2.6: Abstract in on-line searchable database

2.2.3 The "Paper in miniature" abstract

The abstract can usefully be structured exactly as the thesis, with one or more sentences for each section:

1. Rationale
2. Hypothesis and objectives
3. Methods
4. Results

[9] Part of the ISI Web of Knowledge, http://isiwebofknowledge.com/

5. Conclusions

These words do *not* appear in the abstract itself; they are listed here just to show you the structure.

Since space is so limited, writing must be compact. There is little need in such a short piece for connective text. Every word should count. Some sentences from the introduction and conclusion may be taken over almost verbatim, but usually they must be shortened.

Balance

This depends on the main purpose of the paper. The main new information should be emphasized. This might be methods (if the main purpose is to develop a new method), results (if the method is well-known but not the application), or discussion (if the main purpose of the paper is to compare with other works).

Some style points for the abstract

- The abstract should not contain citations unless they are necessary to understand the work; an example is if the main purpose of the paper is to follow up someone else's work.
- There is not enough room in the abstract for detailed reasoning; you are not expected to to prove your case as you do in the thesis or paper.
- Pick a voice (active or passive) and stick to it; do not change voices. Because it is assumed that the text refers to your work if not explicitly mentioned otherwise, passive voice is acceptable.
- Do not refer to the main body of the text, with phrases such as "will be discussed" or "as shown in Table 2"; the abstract must stand alone; the reader may not have access to the full paper.
- Phrases such as "The results show", "The analysis reveals" etc. are rarely needed, for the same reason: these are usually wasted words.
- Don't include equations unless they are the main finding of the study.
- Limit the use of non-standard abbreviations, and define the ones you do use.
- Don't refer to the authors in the third person – you are the authors.

2.2.4 An example

Here is an example, written by me and loosely based on the MSc. thesis by Hengl [4][10] with the different sections labelled. Note that I've tried to put in as much *specific* information as possible.

Rationale	"Semi-detailed soil survey is a costly and time-consuming activity, requiring subjective expertise for photo-interpretation of soil landscapes.
Objective	"The aim of this study was to replace as far as possible subjective photo-interpretation with a more rapid and objective procedure.
Method	"Conventional photo-interpretation maps were prepared for five sample areas totalling 111km^2, representing 10.5% of a survey area in Baranja County, eastern Croatia. Landform parameters (slope gradient, wetness index, relative elevation, plan curvature, profile curvature, and relative annual solar radiation) extracted from a digital elevation model (DEM) were used to train several supervised classifiers, and these were then used to extrapolate over the entire area.
Results	"Using all six predictor variables resulted in a reasonable classification ($\hat{k} = 0.4$) of soil-landscape units inside the training areas. Greatly improved accuracy was obtained by separately classifying high ($\hat{k} = 0.7$) and low ($\hat{k} = 0.5$) relief areas, using only units known to occur in each area. In both landscapes, a reduced predictor set of three variables, as well as the first three principal components, provided almost as much accuracy as the full set. Some photo-interpretation classes, representing about 15% of the area, were consistently mis-classified.
Discussion	"These were mostly in flatter areas where the DEM had insufficient vertical resolution. The extrapolation maps show fine details not achievable by photo-interpretation. Some of these may be artifacts but others represent small areas of contrasting soils.
Conclusion	"The procedure worked well and can be operationalised. However, the automated procedure must be guided by a soil surveyor with good knowledge of the local soil-landscape relations."

[10] This thesis formed the basis for a journal paper [5]

Balance

The balance in this example is:

- Rationale: 20 words (7.5%)
- Objective: 22 words (8.2%)
- Method: 67 words (25.0%)
- Results: 93 words (34.7%)
- Discussion: 41 words (15.3%)
- Conclusion: 28 words (10.4%)

This seems about right; the innovative methods and results of applying them take up about 60% of the abstract, the discussion of the results another 15%, and the context (rationale, objective, conclusion) 25%.

Final form of the example abstract

Here is the final form, as one long paragraph of 268 words. It presents the entire article in miniature, and stands on its own as information:

> "Semi-detailed soil survey is a costly and time-consuming activity, requiring subjective expertise for photo-interpretation of soil landscapes. The aim of this study was to replace as far as possible subjective photo-interpretation with a more rapid and objective procedure. Conventional photo-interpretation maps were prepared for five sample areas totalling 111km^2, representing 10.5% of a survey area in Baranja County, eastern Croatia. Landform parameters (slope gradient, wetness index, relative elevation, plan curvature, profile curvature, and relative annual solar radiation) extracted from a digital elevation model (DEM) were used to train several supervised classifiers, and these were then used to extrapolate over the entire area. Using all six predictor variables resulted in a reasonable classification ($\hat{k} = 0.4$) of soil-landscape units inside the training areas. Greatly improved accuracy was obtained by separately classifying high ($\hat{k} = 0.7$) and low ($\hat{k} = 0.5$) relief areas, using only units known to occur in each area. In both landscapes, a reduced predictor set of three variables, as well as the first three principal components, provided almost as much accuracy as the full set. Some photo-interpretation classes, representing about 15% of the area, were consistently mis-classified. These were mostly in flatter areas where the DEM had insufficient vertical resolution. The extrapolation maps show fine details not achievable by

> photo-interpretation. Some of these may be artifacts but others represent small areas of contrasting soils. The procedure worked well and can be operationalised. However, the automated procedure must be guided by a soil surveyor with good knowledge of the local soil-landscape relations."

2.2.5 Keywords

In addition to the abstract, many journals also require keywords. These are used by abstracting services, such as ISI Web of Knowledge to help with searching for relevant articles easier. Some journals have standard lists of keywords from which the author must choose; others allow authors to select their own keywords.

With increased computer power and full-text searches keywords are less important than they once were, but they can still be useful to categorize the paper.

An ITC thesis does not need to specify keywords. They are only needed if the thesis will be reworked into a journal article. There are no ITC standards for keywords; the author should use the standards for a target journal. See "instructions to authors" for the journal.

For example, the author instructions for Geoderma[11] says:

> "Authors should provide 4 to 6 keywords. These must be taken from the most recent American Geological Institute GeoRef Thesaurus and should be placed beneath the abstract."

[11] http://www.elsevier.com/wps/find/journaldescription.cws_home/503332/authorinstructions, accessed 21-March-2009

2.3 References

Bibliography

[1] American Society of Agronomy; Crop Science Society of America; & Soil Science Society of America. 2004. *Publications Handbook and Style Manual*. Madison, WI: American Society of Agronomy. URL: https://www.agronomy.org/publications/style/; accessed 23-March-2009 103

[2] 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 81, 82, 85, 87, 90, 93, 96, 97, 99

[3] Fong, P. W. L. 2004. *How to read a CS research paper*. URL: http://www2.cs.uregina.ca/~pwlfong/CS499/reading-paper.pdf; accessed 27-November-2007 77, 80

[4] Hengl, T. 2000. *Improving soil survey methodology using advanced mapping techniques and grid-based modelling*. MSc thesis, International Institute for Aerospace Survey & Earth Sciences (ITC) 107

[5] Hengl, T. & Rossiter, D. 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. URL: http://soil.scijournals.org/cgi/content/full/67/6/1810 107

[6] Möller, M.; Lymburner, L.; & Volk, M. 2007. *The comparison index: A tool for assessing the accuracy of image segmentation*. *International Journal of Applied Earth Observation and Geoinformation* **9**(3):311–321. URL: http://dx.doi.org/10.1016/j.jag.2006.10.002 81, 82, 83, 86, 88, 91, 94, 97, 99

[7] Phillips, J. D. 1998. *On the relations between complex systems and the factorial model of soil formation (with discussion)*. *Geoderma* **86**:1–42. URL: http://dx.doi.org/10.1016/S0016-7061(98)00054-8 77

[8] Surridge, C. 2004. *Rice cultivation: Feast or famine?* *Nature* **428**(6981):360–361. URL: http://dx.doi.org/10.1038/428360a 80

[9] 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 81, 82, 84, 86, 89, 91, 93, 94, 97, 99, 105

3 Technical Writing

An MSc thesis is a **written** document, meant to be **read** by others who want to know **what** research was done, **why** it was done, **how** it was done, and what **conclusions** you drew from it. Readers can only understand this from the written document. If it's not **written down**, it might as well not have been done; if it's not written **clearly**, the reader will be confused.

This topic deals with two aspects of writing:

1. **structured technical writing** ('macro-writing');
2. **technical English** ('micro-writing');

The structure of the argument itself (i.e. how to convince your audience of your position) is dealt with in Volume 1 "Concepts", topic "Argumentation"; here we just deal with writing.

Good writing is difficult, even for professional writers. So, there are many "how to write" books. One of the earliest (first edition 1918) and perhaps still the best is Strunk & White [12], full of practical advice for clear, vigorous, and well-organized writing. It is highly recommended that the writer use this or a similar guide, together with a modern English dictionary.

3.1 Structured technical writing

> **Key points**
>
> 1. The purpose of technical writing is to **communicate information** to the reader. One aspect of communication is **structure**.
> 2. One way to structure technical writing is to **write from an outline** (§3.1.1).
> 3. A **paragraph** is a set of sentences that work together to make one point (§3.1.2).
> 4. One way to write paragraphs is with **topic sentences** (§3.1.3).

The MSc thesis is the **'story'** of a research project and must be written clearly, concisely and attractively, so other scientists (including your examiners) can determine what you have done and how well you have done it.

The purpose of technical writing is to **communicate information** to the reader in a compact, clear, and efficient manner. A major aspect of communication is the **structure** of the document.

Q68 : *What information does the document **structure** communicate to the reader?* *Jump to A68* •

3.1.1 Structuring a document by outlining

It is intimidating to most researchers (and even some professional writers!) to begin with a blank piece of paper. There are several techniques to get started and to keep going. This is highly personal. Some people make great progress with a structured technique, whereas others find their creativity blocked. If you have a system that works, fine. But if not, the **outlining** technique presented here may be attractive.

One approach which is especially appealing to those with structured minds (such as many scientists and engineers) is **outlining**, that is, working from the overall structure of the document in a **hierarchical** manner to arrive at the specifics.

A major advantage of this method is that you are sure all the pieces of the story will be in place before you have to write. Also, you see their inter-relation, in particular, that the order of argumentation is clear.

MS-Word

MS-Word directly supports outlining, with its 'View | Outline' command, LaTeX

which uses heading styles to organise the outline. It allows you see the document structure at different levels and to easily re-organise the document. LaTeX uses heading commands such as `\section` for document organization, and some text editors can use this to show the document outline (e.g. Emacs `outline-minor-mode` within `latex-mode`).

The thesis structure already provides an outline, namely the chapter titles. For example, one stereotypical research thesis has these chapters:

1. Introduction
2. Materials & Methods
3. Results
4. Discussion
5. Conclusions

> **Note:** This is not the only possible thesis organization, see the topic "Research Structure" in this lecture series.

These are then broken down into sub-topics. For example, 'Materials & Methods' could be divided as follows:

1. Introduction
2. **Materials & Methods**
   - **2.1. Sampling design**
   - **2.2. Field methods**
   - **2.3. Data processing**
   - **2.4. Data analysis**
3. Results
4. Discussion
5. Conclusion

These sub-topics in turn may be broken down into sub-sub-topics, and so forth. For example, 'Field methods' could be broken down as follows:

1. Introduction
2. Materials & Methods
   - 2.1. Sampling design
   - **2.2. Field methods**
     - **2.2.1. Infiltration and saturated water content**
     - **2.2.2. Soil profile description**
     - **2.2.3. Bulk density**
   - 2.3. Data processing
   - 2.4. Data analysis
3. Results
4. Discussion
5. Conclusion

This process can continue indefinitely; however, more than four levels is unusual.

3.1.2 Paragraphs

Finally the level of the **paragraph** is reached. A paragraph is a set of sentences that work together to make one point. The limit between a paragraph and a section is not sharp. Sections go deeper into a topic, and often require several related paragraphs.

Fowler's "Modern English Usage" [4], defines a paragraph as:

> "a unit of thought, not of length; it must be homogeneous in subject matter and sequential in treatment".

Note the emphasis on homogeneity, also called **coherency**, and a **sequential development** of the idea.

Although paragraphs do not have titles, they can be considered as a **final level** in the hierarchy of an outline. So, you can decide on the purpose of each paragraph as a final step of outlining. Before writing any sentences, you can write a brief description, using key words, of what the paragraph is intended to communicate.

Coherency

A paragraph must be coherent, i.e. the various sentences must belong together.

Here is an example of a coherent paragraph:

> Reaching the University of Twente (UT) in Enschede from anywhere in the world is easy. Flights from hundreds of cities all over the world arrive at Amsterdam's Schiphol airport, from which the traveller can take a comfortable train every two hours direct to Enschede station, from early morning till almost midnight. Trains on the other half-hours require one change of trains, either in Henglo or Amersfoort. The train takes a bit over two hours from Schiphol to Enschede. From Enschede station the ITC building of the UT is an easy five-minute walk; to the main campus of the UT on the outskirts of Enschede there is frequent bus service from the bus terminal directly in front of the train station.

Q69 : *What is the function of the first sentence in this paragraph?* *Jump to A69* •

Q70 : *Why do we call this paragraph "coherent"?* *Jump to A70* •

Q71 : *Why are the sentences presented in this order?* *Jump to A71* •

By contrast, here is an example of an **incoherent** paragraph:

> Reaching the University of Twente (UT) in Enschede from anywhere in the world is easy. The ITC faculty of the UT specializes is the application of geoinformation and earth observation to pressing societal problems, especially in developing countries. The causes of underdevelopment have been debated for years, and recently a new critique of the traditional development aid approach has been presented to the Dutch parliament for its consideration. Remote sensing has become increasingly important for rapid assessment of earth system processes that may result in natural hazards; ITC is specialized in remote sensing.

Q72 : *Why is this paragraph incoherent?* *Jump to A72* •

Linking words and phrases

A common way to show the flow of ideas within a paragraph is with linking words or phrases, also called **connectives**. These explicitly draw the reader's attention to the connection between sentences.

Compare these two paragraphs:

First, without links:

> The guitar is ubiquitous in popular music. This was not always the case. The guitar has a long history. Until the early part of the 20th century it was hardly used. Popular music was accompanied by the piano. The early jazz bands relied on the banjo for rhythm. Two technical breakthroughs made it possible for the guitar to grow into its present prominance. German immigrants to the USA, notably C F Martin, developed a strong bracing system that allowed steel strings to be used on a thin dense-grained wood top on large body. The "dreadnought" guitar was first produced in 1916. The volume and projection of the acoustic guitar was dramatically improved. It could replace the piano in backing up popular songs. The Rickenbacker and Gibson companies introduced the electric guitar in the early 1930s. With amplification the guitar could compete with loud acoustic instruments in jazz orchestras. The revolution was complete when Charlie Christian joined

> the Benny Goodman orchestra in 1939. This band made radio broadcasts and recordings. Bands in the USA and Europe immediately added guitars as a main instrument for backing rhythm and lead.

Second, with links (here **highlighted**; in actual text they are not):

> The guitar is ubiquitous in popular music. **However**, this was not always the case. **Although** the guitar has a long history, until the early part of the 20th century it was hardly used. **Instead**, popular music was accompanied by the piano, and the early jazz bands relied on the banjo for rhythm. Two technical breakthroughs made it possible for the guitar to grow into its present prominence. **First**, German immigrants to the USA, notably C F Martin, developed a strong bracing system that allowed steel strings to be used on a thin dense-grained wood top on large body. **This in turn** allowed the development of the "dreadnought" guitar, first produced in 1916. **In this way** the volume and projection of the acoustic guitar was dramatically improved, **so that** it could replace the piano in backing up popular songs. **Second**, the Rickenbacker and Gibson companies introduced the electric guitar in the early 1930s. **Obviously**, with amplification the guitar could compete with loud acoustic instruments in jazz orchestras. The revolution was complete when Charlie Christian joined the Benny Goodman orchestra in 1939. **Because of** this band's radio broadcasts and recordings, bands in the USA and Europe immediately added guitars as a main instrument, **both** for backing rhythm and lead.

Q73 : *What are the linking words and phrases in the second version?* *Jump to A73* •

Q74 : *Have any facts been changed from the first to the second version?* *Jump to A74* •

Q75 : *In what way is the version with links easier to understand?* *Jump to A75* •

Linking words can also apply to a paragraph's topic sentence (§3.1.3), to show the relation of the current paragraph to the previous.

Table 3.1 shows some common linking words and phrases, grouped into functional groups:

In addition	Also	Similarly	Further(more)
By contrast	However	Despite	Even though
Thus	In this way	Therefore	Hence
On the one hand	On the other hand	First(ly)	Second(ly)
Initially	Later	During	Finally
Because (of)	As a consequence (of)	Since	As a result
Assuming that	Presuming that	Supposing that	Consequently
With respect to	With regard to	Considering	Regarding
Fortunately	Unfortunately	By coincidence	Incidentally
Still	Nonetheless	And yet	Nevertheless
In short	In summary	In conclusion	To summarize
Surprisingly	To our surprise	As expected	Unsurprisingly

Table 3.1: Some linking words

There are of course many others. These should not be over-used, but when used naturally they can greatly help the reader to follow the author's argument.

Length

Paragraph length is one of the more controversial or idiosyncratic elements of writing. Editorial styles differ between languages (e.g. German usually tolerates much longer paragraphs) but also within English among presses. The Oxford Manual of Style [10] states (p. 18)

> "No absolute rules regulate a paragraph's length, since its size is a function of the arrangement and flow of the text it contains."

The Oxford manual recommends "to avoid losing the reader's attention … you may need to divide … or splice …", So it's all about flow. If the subject of a proposed paragraph can be divided into two or three major parts, each with its own topic sentence (§3.1.3), and perhaps a linking phrase, this should be done. If the division is less severe, or there is less to say about each part, the paragraph should be kept as one.

Fowler [4] gives the reader's view of a paragraph:

> "The purpose of paragraphing is to give the reader a rest. The writer is saying to him [or her]: 'Have you got that? If so, I'll go on to the next point.' "

We illustrate this with a paper by Leisz *et al.* [9] on the distribution of farming systems in Vietnam's northern mountain region. This paper has the overall structure:

1. Introduction
2. Methods and mapping
   2.1. Defining the farming system types
   2.2. Land cover mapping
   2.3. Land cover pattern analysis
3. Identifying general farming system types by commune
4. Results
5. Discussion
6. Conclusions and perspectives

Q76 : *What is the purpose of the Introduction?* *Jump to A76* •

The Introduction is divided into six paragraphs. Their purposes can be described as follows:

1. Problem: degradation of Vietnam's natural resources; especially in the uplands;
2. Swidden agriculture as a major cause of this;
3. An alternative swidden system exists that does not cause problems;
4. The actual spatial extent this of alternative is not known;
5. Authors propose a new method to map farming systems;
6. The method depends on understanding process-pattern linkage at farm level.

Put this way, we can see the overall message of the introduction, and how each paragraph contributes to the message.

3.1.3 Writing paragraphs by topic sentences

Even knowing what should be communicated in each paragraph, we are still confronted with the difficulty of actually writing it!

The **topic sentence** technique is often used to define paragraphs. The idea is to write a sentence that introduces the topic of the paragraph, and leave the details of that paragraph for later. Remember, many readers will skim a document exactly this way: they will read the first sentence and then decide whether to read the fuller explanation.

Example of writing with topic sentences

Here is a paragraph adapted from the second introduction paragraph of Leisz *et al.* [9], where a topic sentence is used to introduce the main idea in the paragraph, and then supporting sentences are added to complete the argument.

(Note: citations have been removed for ease of reading.)

> ‘ In large areas of the Vietnamese uplands, swidden agriculture (also known as slash-and-burn or shifting cultivation) is thought to constitute the most serious threat to the natural environment. This viewpoint is found in many places throughout the world and across all of Southeast Asia, causing it to be outlawed at various times in almost every country in the region. The Vietnamese government has subscribed to this belief and has repeatedly attempted to prohibit its practice through a major program intended to “sedentarize” upland populations. Despite heavy expenditures, it is believed that this program has enjoyed little success, because it is unable to provide the swiddeners with alternative methods of earning livelihoods that are commercially viable, culturally acceptable, and ecologically sustainable. ’

Q77 : *Which is the topic sentence? What ideas does it introduce?* *Jump to A77* •

Q78 : *What is the second sentence? How does it add to the ideas of the topic sentence?* *Jump to A78* •

Q79 : *What is the third sentence? How is this linked to the second sentence? How does it add information?* *Jump to A79* •

Q80 : *What is the final sentence? How is this linked to the third sentence? How does it add information? Is it a good conclusion to the paragraph?* *Jump to A80* •

Q81 : *What do you expect as the topic of the next paragraph?* *Jump to A81* •

3.1.4 Resources

A classic short guide to clear technical writing is by Katzoff [8]; Gopen & Swan [5] present a psychology-based approach to effective scientific writing, with several worked examples.

3.1.5 Answers to self-test questions

A*68* : *The structure shows the main topics (by the section headings) and their relations:* ***sequential*** *(at the same level) and* ***hierarchical*** *(sub-headings under main headings). Thus they show the reader how the parts of the document are related.* *Return to Q68* •

A*69* : *The first sentence ("Reaching the University of Twente (UT) in Enschede from anywhere in the world is easy") is the* ***topic*** *of the paragraph: how to reach the UT.* *Return to Q69* •

A*70* : *All the sentences explain how to reach the UT; this is the only subject of the paragraph.* *Return to Q70* •

A*71* : *After the topic sentence, the order is from furtherest distance (the world) to the closest (from the Enschede station to the UT).* *Return to Q71* •

A*72* : *The topic sentence is the same as the previous (coherent) paragraph. But there is nothing else in the paragraph on how to reach the UT. Instead, an aspect of the UT is mentioned (faculty ITC); this leads to a discussion of development aid, and then to the value of remote sensing, and finally back to ITC. Many issues are raised and they are all somehow related to UT/ITC (reaching it, mission, development, remote sensing) but the relation between the issues is not clear.* *Return to Q72* •

A*73* : *However; Although; Instead; First; These developments; Thus; so that; Second; Obviously; Because of.* *Return to Q73* •

A*74* : *No facts are changed.* *Return to Q74* •

A*75* : *The linking words and phrases lead the reader through the story, and emphasize the relation between facts.* *Return to Q75* •

A*76* : *The Introduction gives the background of the study: the problem, proposed solutions, previous work, and the proposed contribution of this work.* *Return to Q76* •

A*77* : *Topic sentence: "In large areas of the Vietnamese uplands, swidden* *agriculture (also known as slash-and-burn or shifting cultivation) is thought to constitute the most serious threat to the natural environment."*

We are introduced to the area (uplands), the land use being discussed (swidden agriculture), and the supposed problem (threat to environment).

Note the phrase "thought to constitute", this is key to the topic, since the authors will challenge this view. Compare the topic if this were "constitutes"; in that case the authors would be taking the statement as true. *Return to Q77*
•

A*78* : *Second sentence: "This viewpoint is found in many places throughout the world and across all of Southeast Asia, causing it to be outlawed at various times in almost every country in the region."*

"This viewpoint" refers directly back to the topic sentence's assertion that swidden agriculture is a primary cause of environmental degradation. The mentioned bans support the idea that swidden agriculture is widely considered problematic. *Return to Q78*
•

A*79* : *Third sentence: "The Vietnamese government has subscribed to this belief and has repeatedly attempted to prohibit its practice through a major program intended to 'sedentarize' upland populations."*

There is no explicit link; it is the position following the second sentence that implies that the two are related.

This third sentence fills out the second sentence's information on bans, by making it local (Vietnam) and by mentioning a specific measure taken locally to implement the ban. *Return to Q79* •

A*80* : *Final (fourth) sentence: "Despite heavy expenditures, it is believed that this program has enjoyed little success, because it is unable to provide the swiddeners with alternative methods of earning livelihoods that are commercially viable, culturally acceptable, and ecologically sustainable."*

The link is here explicit: 'Despite heavy expenditures . . . '; the additional information is about the failure of the mentioned programme.

This is a good conclusion: we started by describing a practice, then the problems it is thought to cause, then the actions taken by the government, and finally their perceived failure.

Note the sequence of filling sentences in the paragraph:

viewpoint ⟶ ***action*** ⟶ ***consequences*** *of that action.*

These build the argument and prepare the reader for the next paragraph (see next answer). *Return to Q80* •

A*81* : *The next paragraph will probably propose some new solution to the swidden cultivation problem. Indeed, its topic sentence is: "The composite swiddening agriculture system (CSA) is an alternative farming system that appears to overcome these problems."* *Return to Q81*

•

3.2 Technical English

Key points

1. The key skill in technical writing is **writing clearly** (§3.2.1).
2. Write for your **intended audience.**
3. All good authors **revise and rewrite** extensively (§3.2.2).
4. Problems are often encountered with **verbs** (tense and voice) (§3.2.3) and **punctuation** (§3.2.4).
5. **Non-native writers** of English have particular difficulties, depending on their mother tongues (§3.2.6).

3.2.1 Writing clearly

Now you've organised, outlined, organised your argument into paragraphs, and decided what to say. But, how do you actually write readable, coherent text?

Don't try to be literary or clever! Scientific communication should be clear, direct, unambiguous and forceful ("plain speaking"), with no extraneous material. This was best expressed by Strunk [11, §III.13]:

> "Omit needless words. Vigorous writing is concise. A sentence should contain no unnecessary words, a paragraph no unnecessary sentences, for the same reason that a drawing should have no unnecessary lines and a machine no unnecessary parts. This requires not that the writer make all his sentences short, or that he avoid all detail and treat his subjects only in outline, but that every word tell."

Booth *et al.* [2, §11.4.1] explain two styles of drafting, which appeal to different types of writers:

- Write as fast as possible, correct later;
- Write carefully, don't leave any problems.

You have to experiment to find the one which fits your style of work. An intermediate form is to write fast, but keep a separate page with notes to yourself of things to check on later.

Q82 : *What are the advantages of each of these two methods?* *Jump to A82* •

Author's notebook

Many authors and other creative people keep a **notebook** with them at all times (even by their bed) so that they can write down ideas whenever they occur.

Everyone has a time of day or days of the week when they are at their most **creative**. Schedule the creative phase of your writing for these times. Other times are suitable for detailed work such as checking spelling and details of the argument.

Re-read your text with a fresh eye

In a separate step, **re-read your text as if it were written by someone else**, and ask yourself:

- Is the **argument** clear? Does it say what you want it to? Remember, *you* know what you wanted to say, but can the *reader* find that in your text?
- Does it read **smoothly**? Would it benefit from more explanation or connectives?
- Is it as **short and direct** as possible? Are there redundant words or phrases? Have you said the same thing twice? Can the text be shortened?
- Is it **too short** so that the reader will have difficulty understanding the argument? Should some explanation be added?
- **Does it read like English?** Writers who are not accustomed to writing English often use constructions typical of their native language. If what you are reading doesn't sound like it was written by a native, most likely it is because of such problems.

Write for your intended reader

There are many styles of writing. A good criterion for selecting a style is to think about the **intended reader**. For a thesis, the most important readers are the members of the examining committee, especially the external examiners; they represent the wider scientific community who (you hope) will read your thesis.

Shorter is (usually) better

Make your point in as few words as possible. A shorter text is easier on the reader. But …

Don't sacrifice length for clarity

Use all the words you need to make your point clearly. In particular, use **connectives** to link parts of your text and bring out their relation. For example, look at the use of the connective 'in particular' in the previous sentence.

Use a dictionary

You should always work with a standard dictionary [e.g. 13] to be clear on word meaning and usage.

Use a style manual

Technical points of style are dealt with in standard English-language references, such as the Chicago manual [6]. Each field has its own difficulties regarding nomenclature, units of measure, etc.; these are covered in specialised style manuals. An example for soil science is American Society of Agronomy *et al.* [1, Ch. 2].

Use the spelling checker

Use your word processor's **spelling checker**, but remember that just because a word is correctly spelled, that does not mean it is the word you want!

> "Infiltration measurements were located on **bear** soils ..." (Are these soils where bear like to walk? Usually we prefer **bare** soils.)
>
> "All the evidence points to one **confusion** ..." (That's unfortunate; I was hoping that you could come to a **conclusion**. But maybe you really **do** mean 'confusion'!)
>
> "We had to decide **weather** or not to include **boarder** trees."

Pick the right word

Be especially alert for words that seem right, but aren't. Common errors are with the pairs (*imply, infer*), (*much, many*), and (*affect, effect*).

Q83 : *What are the differences between these pairs? (If you are in doubt, consult a dictionary.)* *Jump to A83* •

Avoid barbarisms

Avoid barbarisms, i.e. words that seem to be English but which are not, for example *to impact* (correct: *to affect, to have an impact*). This is a common error of sloppy native speakers, especially provincial newspapers in the USA.

In all this, if you are unsure, use a dictionary with good definitions and usage examples.

Use the grammar checker

MS-Word's **grammar checker** is surprisingly good, but make sure its advice is correct for your situation. There are also various open-source grammar checkers, e.g. LanguageTool[1].

3.2.2 Revising and re-writing

All good authors revise extensively.

When you have finished a piece of text, put it aside for a day, and then **re-read it.**

You will easily see the obvious mistakes in spelling and punctuation; when in doubt use your dictionary and style manual.

More importantly, you will see your writing with a fresh eye:

- Does it say what you intended?

[1] `http://www.languagetool.org/`

- Are all the points clear?
- Is it at the right technical level for your intended reader?
- Is it at the right language level (style, vocabulary) for your intended reader?

3.2.3 Verbs: tense and voice

English verb constructions are very rich and can cause problems both for native and non-native speakers. Here we present only some aspects of verb use that cause problems in a thesis project; consult a grammar for more information.

Tense

This is the time to which the verb refers. It should present few difficulties if you place yourself in the position of the reader and consider the time to which the statement refers **at the time it is written.**

Future This is used for events in the future when the document was written. It is often used in the "Methods" section of the research proposal:

> 'Fifty plots **will be** sampled.'

It is also used for speculative statements in the "Conclusions" section of the thesis:

> 'Sub-pixel classification **will become** a routine technique for forest inventory.'

Past This is used for events already in the past when the document was written. It is often used in the "Methods" section of the thesis:

> 'Fifty plots **were** sampled.'

Note that only the tense had to change once the work was done.

It is also used for a result that is specific to our study:

> 'Ground control with single-receiver GPS survey **was** able to adequately rectify 1:5 000 scale small-format aerial photography to map accuracy standards.'

This is simply a statement of fact about our study, and does not imply that we think the method is **in general** successful. If we believe that, we should use the present (see below).

Present This is used for statements that are always true, according to the author, for some continuing time period.

> 'Sub-pixel classification **is** a new technique for forest inventory.'

This statement may be false some years from now, but it's true at the time of writing and for some time thereafter.

It is also used for a result that is widely-applicable, not just to our study:

> 'Ground control with single-receiver GPS survey **is** adequate to rectify 1:5 000 scale small-format aerial photography to map accuracy standards.'

This statement implies that we think the method is **in general** successful; our data show this and we are confident it is everywhere true.

Several other tenses are less common:

Past perfect This is used for events already in the past when another event in the past occurred.

> 'Fifty plots **had been** sampled as part of a previous project.'

This implies a context such as "...when we decided how many more plots to sample."

Future perfect This is used for future events that will have been completed when another event in the future occurs.

> 'Fifty plots **will have been** sampled by this project.'

This implies a context such as "...before we arrive to sample our plots."

Voice

The two voices are **active** and **passive**. In the first case the subject of the sentence controls the verb:

> 'Pests damage crops.'

In the second case there is no explicit subject, only (perhaps) one implied in the complement.

> 'Crops are damaged by pests.'

Note that "by pests" is not necessary for a grammatical sentence. We could have said "Crops are damaged in the spring".

Passive voice

The passive voice should be used when the object is more important than the subject, especially when the identity of the subject doesn't matter:

> 'Nelson Mandela is widely respected.'

This implies "by everyone"; the sentence could be re-written:

> 'Nelson Mandela is respected by everyone.'

It could also be re-written in the active voice:

> 'Everyone respects Nelson Mandela.'

However, this puts the main information at the end of the sentence, thereby making it weaker. But if this sentence were paired with a contrasting sentence with a different subject, the active voice would bring out the contrast:

> 'Everyone respects Nelson Mandela, but very few know his middle name[2].'

Avoiding "I"

One use of the passive is to avoid calling attention to oneself; this is considered egotistical or conceited:

> 'I selected five representative villages.'

as opposed to:

> 'Five representative villages were selected.'

The second form is **ambiguous** unless the entire context implies the actor. It could be made explicit:

> 'Five representative villages were selected by me.'

but that sounds awkward. It sounds better if referring to a third party:

> 'Five representative villages were selected by the project team.'

and these forms are grammatically equivalent.

When the active voice is required

The passive voice is thus often preferred for courtesy, but it can lead to serious problems of interpretation. This is nicely discussed by Webster [14] (emphasis added):

> "There are several actions that I very definitely want reported in the active voice: they are ones of **assumption**, **decision** and **choice.** When I read 'It is assumed that', 'It was decided to', and 'Sites were chosen', I immediately ask myself who assumes?, who decided?, and who chose? – the author?, or his

[2] Rolihlahla

boss?, his client?, or some overseeing committee? These actions lie at the heart of original research, and results depend crucially on them ...

"A second reason why I want to know who did what is that we scientists are human, and we are fallible therefore. We make mistakes, we take foolish decisions, we choose unrepresentative specimens, we overlook uncomfortable results, and we misinterpret what we see. We do so not necessarily, not usually, wilfully, but we are responsible, and readers must know that we are."

Here are some examples of Webster's categories:

- **assumption**: 'It is assumed that': who does?
  - 'We asssume that ... ' or 'The previous survey assumed that ... '
- **decision**: 'It was decided to': who decided?
  - 'The authors decided to ... ' or 'The local government authority decided to ... '
- **choice**: 'Sites were chosen': who chose?
  - 'An experienced soil surveyor chose the sites' or 'The local extension agent directed us to cooperative farmers, who were convinced by the agent to allow use of a small portion of their fields, of the farmers' choosing'.

Use of "the author"

One way to avoid "I" while still being clear about who is the actor is to use the euphemism "the author" in the third person:

> 'The author selected five representative villages.'

> 'The authors, basing themselves on the general geomorphic map, partitioned the study area into three sampling strata: the lake plain, the volcanic plain, and the isolated volcanic hills.'

Another way to avoid excessive use of "I" is to introduce a paragraph with the active voice, and then write the rest in the passive; the reader will correctly infer that the all the actions mentioned in the paragraph were the active actions of the author:

> ' We designed this study to avoid bias. Sites were chosen so that ... Care was taken in sampling ... Samples were placed immediately in a thermally-isolated container ... '

Q84 : *Compare these two descriptions:*

"Fifteen erosion plots were established in the watershed. Their locations were selected to be representative of landform and land-use conditions."

"Prior to this work, researchers from the Soil Conservation Research Institute had established ten erosion plots in the watershed, based on their subjective opinion of representative sites, and the willingness of farmers to collaborate. In this work we set up an additional five plots, also located subjectively but all on slopes steeper than any of the original ten plots."

Question: What is the added value of the second description? *Jump to A84* •

3.2.4 Punctuation

Punctuation is used to break words into groups with a related function, so that it is easier for the reader to understand the intent of the author, and to mark emphasis or question. The rules for the full stop or period ("."), semi-colon (",") and colon (":") are fairly standard, but the comma (",") is used in a more varied manner. When in doubt, consult a style manual. Punctuation has changed over the years and also can vary substantially between countries or even styles within a country (journalistic, popular, academic, formal, . . .).[3]

Use of the period (".")

This completes a sentence, which must be grammatically-correct. A speaker would come to a full stop at this point when reading the document aloud.

> 'No man thinks more highly than I do of the patriotism of the very worthy gentlemen who have just addressed the House.'

Use of the exclamation point ("!")

This also ends a sentence, but indicates surprise or emphasis.

> 'I repeat it, sir, we must fight!'

[3] The texts in this section are adapted from the speech given by Patrick Henry, 23–March–1775 to the Second Virginia Convention, commonly referred to by its closing statement "Give me liberty or give me death".

Use of the question mark ("?")

This also ends a sentence, but indicates a question which could be answered, either by the reader or the author.

> 'Is life so dear, or peace so sweet, as to be purchased at the price of chains and slavery?'

Here the reader is expected to answer "No!"[4]

Use of the semi-colon (";")

This joins two complete sentences where the second is closely related to the first. A speaker would come to a pause at this point when reading the document aloud. The semi-colon can be repeated.

> 'We have petitioned; we have remonstrated; we have supplicated; we have prostrated ourselves before the throne, and have implored its interposition to arrest the tyrannical hands of the ministry and Parliament.'

Each sentence above is complete by itself. The author could have written:

> 'We have petitioned.
> We have remonstrated.
> We have supplicated.
> We have prostrated ourselves before the throne, and have implored its interposition to arrest the tyrannical hands of the ministry and Parliament.'

The author wanted to draw attention to the close relation between the four actions. Note that the last sentence must use a comma instead of a semi-colon because the phrase "and have implored . . . " does not have an independent subject and so can not be an independent sentence.

Use of the colon (":")

The colon is placed at the end of a complete sentence to introduce examples that are usually a list but can also be a phrase or sentence.

> 'I have but one lamp by which my feet are guided: the lamp of experience.'

Note that the part before the colon form a complete sentence; the part after expands the thought but is not necessary to complete the grammar.

[4] The author answers his own question: "Forbid it, Almighty God!"

Use of the comma (“,”)

In American usage, the comma is placed where the speaker would make a brief pause to show the relation between words. One reason would be because the words form a list, and the speaker must clearly show that the items are separated from each other:

> ‘I am willing to know the whole truth, to know the worst, and to provide for it.’

Here the speaker has three closely-related items which are in a list:

1. to know the whole truth,
2. to know the worst, and
3. to provide for it.

In American usage there is a comma before the last clause, to avoid ambiguity (are there two or three items in the list?). In British usage the final comma is omitted.

Another use is to join an introductory clause to the main part of the sentence, at the point where a speaker would take a breath. It adds nothing to the understanding of the sentence.

> ‘For my own part, I consider it as nothing less than a question of freedom or slavery.’

3.2.5 Some matters of style

There are many excellent style manuals [e.g. 6]; here I only point out some common style errors made by ITC students.

Repeated words

> ‘We interviewed ten local farmers. Local farmers said that ...’

> ‘We interviewed ten local farmers, who said that ...’

Superfluous words

> ‘As a result of the field measurements, it could be observed that the average steady-state infiltration rate of the soils was 1.2 cm hr^{-1}. ’

> ‘The average steady-state infiltration rate was 1.2 cm hr^{-1}. ’

In this example, it should be clear from the previous sections of the paper or thesis, specifically the ‘methods’, that infiltration was measured in the field. So it is not necessary to repeat this in the results.

Wordiness

> 'The results show that after computing the correlation matrix between the NIR and IR bands the correlation was found to be 0.95 for the LANDSAT TM7 and 0.96 for the ASTER images. '

> 'The NIR and IR bands were highly correlated (LANDSAT TM7 $r = 0.95$, ASTER $r = 0.96$). '

Adjectival phrases

As part of its Germanic heritage, English allows noun clusters to function as adjectives:

> ' We examined the user data information communication requirement. '

Here there is one object noun (*requirement*); all the other nouns are functioning as adjectives. This is grammatically-correct (from Germanic grammar) but difficult to read. Take advantage of the Latin heritage (via French) in English to move some of these to a dependent position:

> ' We examined the user's requirement for communication of information on data. '

References to agents

A common mistake in English is an incorrect reference to the agent of an action within a sentence:

> ' Arriving at the field site, the plots were observed to be heavily eroded. '

Who is arriving here? Presumably the observer, but since the main sentence is in the passive voice, there is no reference possible to the first clause.

The active voice solves this:

> ' Arriving at the field site, we observed heavy erosion in the plots. '

Now it is clear that "we" arrived and then observed. This is equivalent to:

> ' When we arrived at the field site, we observed heavy erosion in the plots. '

This can be problem even in the active voice:

> ‘ Being a soil surveyor, terrain modelling is an important new tool. ’

Grammatically this is equivalent to “terrain modelling is a soil surveyor, and also an important new tools.” This incorrect reference can be corrected by putting a referrent in the main sentence:

> ‘ Being a soil surveyor, I consider terrain modelling an important new tool. ’

3.2.6 For non-native speakers

Non-native speakers have special problems with writing English. Here we briefly discuss a few of these. For some languages there are specialist books available that discuss common pitfalls for that particular language (e.g. Burrough-Boenisch [3] for Dutch speakers).

If you are not fully comfortable with the English language, and you are serious about improving your skills, not just surviving the thesis-writing phase, a learner’s English-English dictionary is highly recommended (e.g. the Oxford Advanced Learner’s Dictionary [7]). These have extensive notes on correct usage, discussions of difficult topics such as use of articles and prepositions, and suggestions to make your writing idiomatic (i.e., as if it were written by a well-educated native speaker).

Don’t write in your native language

Some students like to write in their native language and then translate. This almost always is more work than **thinking and writing in English** from the beginning. It’s OK to write notes to yourself in your own language, but not full texts.

Avoid Latinisms

For students whose native language is a modern dialect of Latin (e.g., Portuguese, Spanish, French), there are words with cognates in English that have a very different meaning in English. A well-known examples is *actually*, meaning “in fact” (rather than “now, currently”). More serious are using full Latin constructions, e.g. “Could not be sampled the soil” (correct: “The soil could not be sampled”).

Latin-speaking students often try to write long, complex sentences with many dependent clauses. Short, simple sentences are easier to write and to understand.

Avoid non-English constructions

Other native languages give other typical problems in English. Some languages (e.g. Slavic) don’t use articles, and many do not use articles the same way as English. So a student may write “Soil was red”, which in English means that soil, in general, has a red colour, rather than “The soil was red”, meaning that the specific soil in question was red. Dutch and German (as well as modern Latin) speakers often use gendered pronouns where English uses “it”, e.g. “We sampled the soil. *He* [correct: *it*] was red”. English does not use both the noun and pronoun in the same phrase, e.g. “The soil he was red” (correct: omit “he”).

On-line tools

Popular languages have increasingly powerful on-line tools; a good example is the MDBG Chinese-English dictionary[5].

Machine translation

Machine translators such as Altavista’s BabelFish[6] are interesting ideas, and can often cope with simple declarative sentences. They may be used to give a rough translation, but always check the results.

Be especially careful of automatic translations of words with several meanings in either the target or source language. For example, the verb ‘to bear’ in English may be translated into Spanish as:

- ‘sostener’ (to support, e.g. ‘The beam can bear a heavy load’),
- ‘dar’ or ‘producir’ (to yield, e.g. ‘The tree bears fruit’),
- ‘llevar’ (to carry, e.g. ‘The wise men came bearing gifts’),
- ‘tener’ (to have, e.g. ‘He bears a distinguished name’),
- ‘soportar’ (to stand, e.g. ‘I can’t bear his jokes’),
- ‘odiar’ (to hate, e.g. ‘I can’t bear spiders’),
- ‘parir’ (to give birth, e.g. ‘She bore a child’),
- ‘resistir’ (to stand up to, e.g. ‘That idea will not bear close inspection’), and
- various other idiomatic uses.

- In addition, the noun ‘bear’ (the animal) is translated as ‘el oso’ (or ‘la osa’).

Similar lists can be made when translating into English. This is difficult enough for a human translator, and well beyond the capabilities of computers, although they can give alternatives from which you can choose.

3.2.7 Answers to self-test questions

A*82* : *Writing fast: creativity is not blocked, ideas can flow.*

Writing carefully: the first draft has a greater chance of being complete and correct. *Return to Q82* •

A*83* : *Imply is active and infer is passive. An author presents some evidence that implies a conclusion; the reader takes the evidence and infers a conclusion.*

[5] http://www.mdbg.net/
[6] http://babelfish.altavista.com/

Much is an adjective for uncountable nouns, many for countable nouns. For example "much effort" but "many students".

As verbs: Affect means for the subject to influence the object ("This new law will affect many citizens"), whereas effect means to bring about or accomplish something ("The new orientation was effected by a series of laws."). Affect is not a common noun; an effect is a result of some action. *Return to Q83* •

A*84* : *The second description tells who set up the plots; in fact there were two sets, with different objectives, as well as different people making the subjective decisions about where to place the plots.* *Return to Q84* •

3.3 References

Bibliography

[1] American Society of Agronomy; Crop Science Society of America; & Soil Science Society of America. 2004. *Publications Handbook and Style Manual*. Madison, WI: American Society of Agronomy. URL: https://www.agronomy.org/publications/style/; accessed 23-March-2009 125

[2] Booth, W. C.; Colomb, G. G.; & Williams, J. M. 1995. *The craft of research*. Chicago: University of Chicago Press 123

[3] Burrough-Boenisch, J. 1998. *Righting English that's gone Dutch*. The Hague: Sdu Uitgevers 134

[4] Fowler, H. W. & Gowers, E. 1965. *A dictionary of modern English usage*. Oxford,: Clarendon Press, 2d edition 114, 117

[5] Gopen, G. D. & Swan, J. A. 1990. *The science of scientific writing*. *American Scientist* **78**(November-December):550–558 119

[6] Grossman, J. (ed.). 1993. *The Chicago manual of style : the essential guide for writers, editors and publishers*. Chicago: University of Chicago Press, 14th edition 124, 132

[7] Hornby, A. S. (ed.). 1995. *Oxford advanced learner's dictionary*. Oxford: Oxford University Press 134

[8] Katzoff, S. 1964. *Clarity in technical reporting*. Technical Report NASA SP-7010, National Aeronautics and Space Administration (NASA) 119

[9] Leisz, S. J.; Nguyen thi Thu Ha; Nguyen thi Bich Yen; Nguyen Thanh Lam; & Tran Duc Vien. 2005. *Developing a methodology for identifying, mapping and potentially monitoring the distribution of general farming system types in Vietnam's northern mountain region*. *Agricultural Systems* **85**(3):340–363 117, 118

[10] Ritter, R. M. (ed.). 2003. *The Oxford manual of style*. Oxford: Oxford University Press 117

[11] Strunk, W. 1918. *The elements of style*. Ithaca, New York: Privately Printed. http://www.bartleby.com/141/ 123

[12] Strunk, W. & White, E. B. 2009. *The elements of style*. New York: Pearson Longman, 50th anniversary edition 111

[13] Thompson, D. (ed.). 1995. *The concise Oxford dictionary of current English*. Oxford: Oxford University Press, 7th edition 124

[14] Webster, R. 2003. *Let's rewrite the scientific paper. European Journal of Soil Science* **54**(2):215-218. http://www.blackwell-synergy.com/links/doi/10.1046/j.1365-2389.2003.00515.x/ 128

4 Graphic presentation

by: D G Rossiter & Wim Feringa

Key points

1. Tables and figures (graphs, illustrations, maps, images) are **graphic presentation**, i.e. not part of the running text.
2. Graphics are used when **text is not effective** in presenting the message.
3. General issues for all types of graphics are positioning, captioning, referencing, resolution, typography and use of colour (§4.1).
4. There are five types of graphical elements: **tables**, **graphs**, **illustrations**, **maps**, and **images**, each with **design principles** which should be respected for proper communication with the reader (§4.2).
5. There are many **tools** for preparing graphics (§4.3); proper layout and design is an art.

Tables and figures are all types of **graphic presentation** – i.e. anything that is not running (narrative) text.[1] Well-designed graphics communicate information effectively ("one picture is worth a thousand words"); conversely, confusing graphics at best irritate the reader and at worst hide or even distort the message. Therefore, journal style guides [e.g. 1, 5] pay close attention to graphic presentation. It is equally important in a thesis.

A constant theme of this chapter is "simpler is better": if all required information is present, legible and logical, the reader is informed rather than distracted or mislead.

> **Note:** This text only deals with graphic design for a printed thesis; if you will prepare web pages other rules apply; see for example the Web Style Guide[2].

Effective use of tables, illustrations, images and maps is a subject with a long history and well-developed best practices, based on the psychology

[1] Text used within a graphic is not running text, and is equally a "graphical element", as explained later.
[2] http://www.webstyleguide.com/wsg3/index.html

of human perception; here we can only give some of the most important points. You are encouraged to consult the references given in each section for more ideas and examples of bad and good graphical presentation. Among the most important general references are:

- The works of Tufte on graphical presentation [39, 40];
- For statistical graphs, the works of Cleveland [6, 7];
- For social science, the source books of Miles and Huberman [13, 22].

Authors often spend much more effort on the text than the graphics, although graphics are often the elements of the thesis or paper that the reader spends more time on! So good graphic design is well-worth the required investment.

Here is some sound advice from Tantu [37][3] (emphasis added):

> "Treat graphics as **first-class citizens** of your papers. They deserve as much time and energy as the text does. Indeed, the creation of graphics might deserve even more time than the writing of the main text since **more attention will be paid to the graphics** and they will be **looked at first**.
>
> Plan **as much time** for the creation and revision of a graphic as you would plan for text of the same size.
>
> Difficult graphics with a high information density may require even more time. Very simple graphics will require less time, but most likely you do not want to have 'very simple graphics' in your paper, anyway; just as you would not like to have a 'very simple text' of the same size."

☺ ☹

Throughout this chapter we use the ☺ symbol to show good practice, and the ☹ symbol for bad practice.

4.1 General graphical issues

This section discusses issues that are relevant for all graphical elements.

[3] creator of the PGF and Ti*k*Z graphics languages

Key points

1. Graphical elements are usually positioned in a document as **floating blocks**; these are referenced in a List of Tables and List of Figures, and must be **referenced in the text** (§4.1.1)
2. Each graphical element must have a self-explanatory **caption.**
3. Graphic elements may be **scalable** or **non-scalable** (§4.1.2)
4. Use scalable ("**vector**") graphics for line drawings and text; use non-scalable (**"raster"**) for intrinsic images, e.g. photographs, satellite images, scans, and screen dumps.
5. When scanning (§4.1.3) or otherwise using non-scalable graphics, use the proper **resolution** (§4.1.4) throughout the production process.
6. Use **colour** only as a **visual variable** to add information, or (less commonly) as an **aesthetic element** to make information more attractive (§4.1.6).
7. Use proper **typography** for readable text within graphics (§4.1.7).

4.1.1 Floating elements in a document

Graphical elements are usually positioned in a document by text processing programs as **floating blocks**, each with its own **caption** (explanation), which must be **referred to** in the main text. The block is "floating" since its position is not fixed by the author, but determined by the typesetting engine of a "desktop publishing" program[4]. Since the graphical element may be separated from its related text, **each table or figure must be referenced in the text.** Also, since these are often consulted alone, they must have **captions**, i.e. small titles which explain their purpose.

Figures are generally placed in the text as close as possible to their first reference (i.e. when they are mentioned in the text); if this is not possible on the same page the figure should float to the top of the next page. Colour figures may be grouped on their own pages (often called "plates") to reduce printing costs.

Here is an example of running text and a floating figure, with a proper caption and reference, and legible text in the diagram. Note that the LaTeX typesetting engine has in this case elected to position this just after the paragraph in which the reference occurs.

> "A mass spectrometer is an instrument for determination of the elemental composition of a sample. Figure 4.1 shows

[4] LaTeX, SGML/XML interpreters, or (poorly) MS-Word

schematically its three modules: an ion source, a mass analyzer to separate ions by their masses, and a detector, which measures the abundances of each ion in the sample."

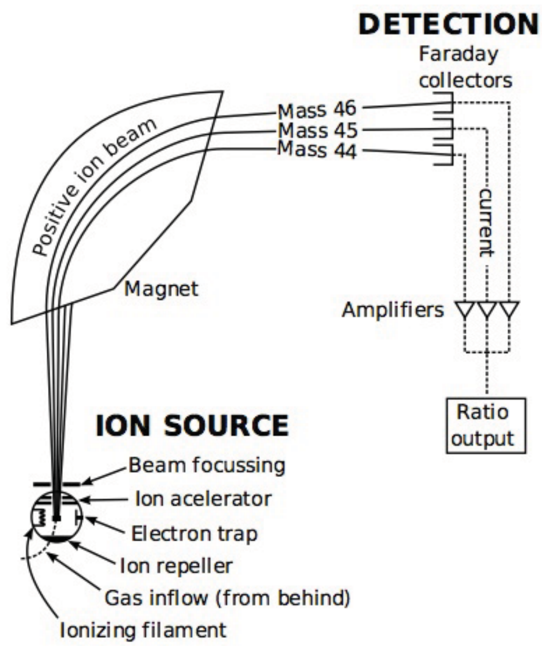


Figure 4.1: Mass spectrometer (schematic); Source: [28]

Because tables and figures do not necessarily occur next to the text that describes them, and because they are often the most interesting parts of the document, they are often listed in a separate **List of Figures** and **List of Tables**.

4.1.2 Scalable and non-scalable graphics

A fundamental distinction in computer graphics is between **scalable** and **non-scalable**:

- **Scalable** formats store mathematical descriptions of their elements;
- **Non-scalable** formats store a grid (matrix) of **pixels** ("picture elements"), each with a colour or grayscale value; these are also called **rasters**, another word for "grid".

We discuss these in turn.

Scalable graphics

Vector graphics are **scalable** and can be printed or viewed at any resolution with an appropriate output device. Examples are **encapsulated**

postscript graphics (EPS), Adobe **PDF** graphics (proprietary postscript), and **Windows metafile** (WMF, EMF) graphics.

Figures, including labels, legends and annotation, must be **legible**. For scalable graphics (including PostScript fonts, also as used in PDF documents) printed on a high-quality printer (e.g. a 1200 DPI or finer laser printer) the main limitation is the resolution of the human eye at normal reading distance.

Line drawings can be scaled to the appropriate dimension without sacrificing quality. This is an advantage, but also a problem. All elements scale, including line thickness. When the drawing is scaled up (enlarged), the lines become thicker, and the text font size becomes larger; when scaled down (reduced), the opposite. So the lines and fonts may not match thetypeset part of the document.

So, the best practice is to first decide on the size, line thickness, fonts and colours (also referred to as "specifications") of all drawings in the publication in order to get a consistent design. When creating a drawing without these specifications, the document becomes messy and inconsistent.

Apart from an inconsistent design another problem arises with scalable graphics. When scaling down a drawing to, for example, 50%, line thicknesses are also reduced by 50%. So, a line 0.1 mm thick in the original becomes 0.05 mm thick in the result. A computer screen can display this very thin line, since 1 pixel is the minimum the monitor can and will show. A laser or inkjet printer can also show it, because they also have a minimum thickness they can and will show. However, when the output device is an offset-printing machine, the 0.05 mm line will be transferred to the press, but the line will disappear when printing.

Line drawings (including text) are best prepared in vector formats. Figure 4.2 shows an example of scalable graphics, in this case using the PDF format.

Clearly, the smallest image is not suitable for printing; however, if you are viewing this document on-screen, you can enlarge ("zoom in") to the smallest figure and see all the detail of the largest[5].

When down-scaling a drawing in your document, the content of the drawing will not change, so the graphic information must be denser. The text and symbols will become smaller, and at some reduction they will not be legible. This is against the basic rule of creating an illustration: **"all content must be legible"**.

[5] Go ahead, try it!

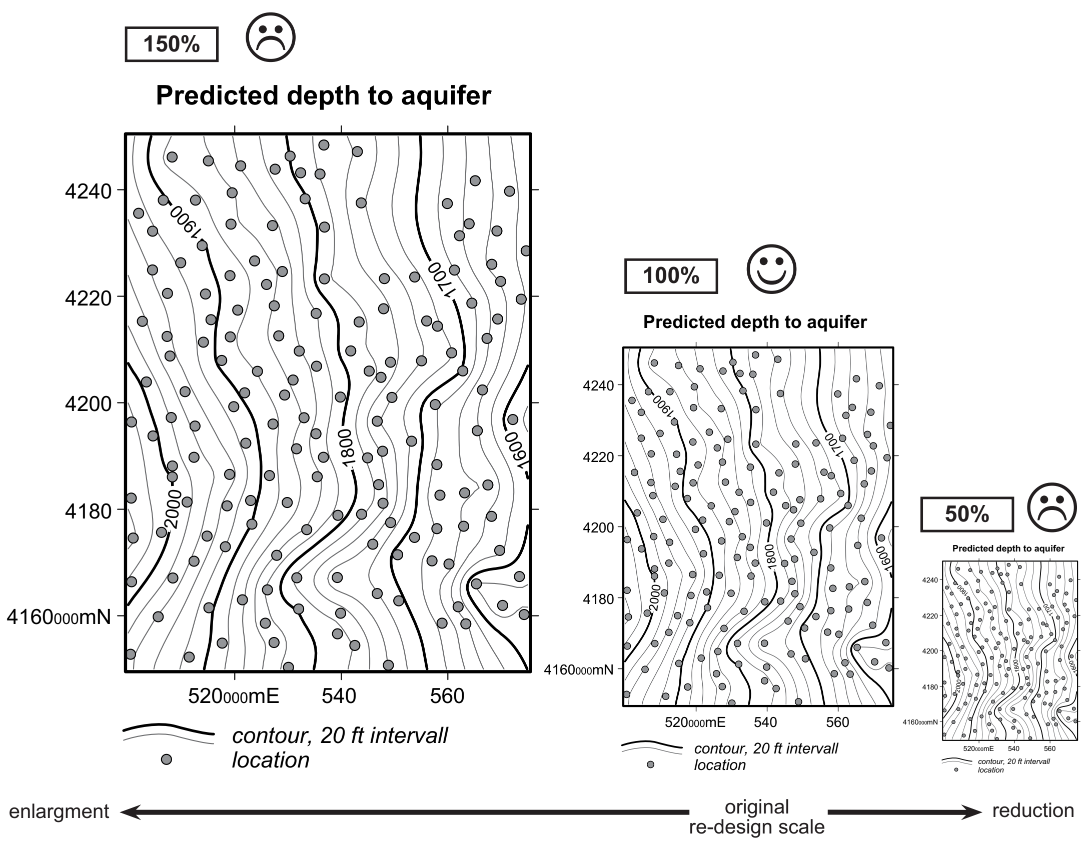


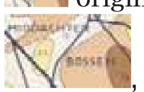
Figure 4.2: Example of scalable graphics

Non-scalable formats may be **compressed** to save storage space. These conceptually store a full grid, but in fact use various data-compression algorithms; the compressed file is automatically decompressed on output. There are two types of compression:

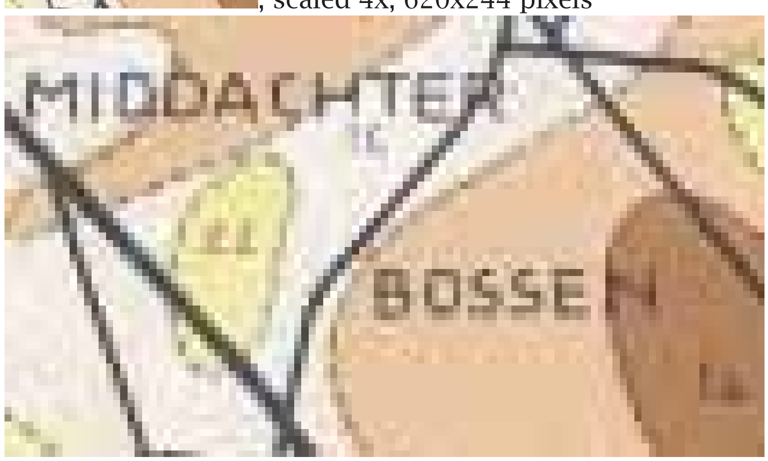
- **Lossless**: The original raster can be recovered exactly; an example is **Portable Network Graphics** (PNG) format;
- **Lossy**: The original raster can not always be recovered exactly; an example is **Joint Photographic Experts Group** (JPEG) format.

Figure 4.3 shows an example of non-scalable graphics, in this case using the JPEG format. The maximum resolution is the bitmap; any enlargement just enlarges the pixels. If you zoom in on the smallest image, eventually it will look like the largest one[6]. The width of the unscaled bitmap, 105 pixels, is displayed with a width of 9 mm (= 0.355”) on the default output device (here, A4 paper, 210 x 297 mm). This matches the scan resolution of 300 DPI with which the bitmap was created.

 original bitmap, 105x61 pixels

, scaled 2x, 310x122 pixels

, scaled 4x, 620x244 pixels



scaled 12x, 1360x732 pixels

Figure 4.3: Example of non-scalable graphics; Source: 300 DPI scan of [43]

[6] Go ahead, try it!

The legibility of non-scalable graphics is limited by their `resolution`, as well as by the design size of their components.

Thus the first rule is to **use scalable graphics whenever possible** (§4.1.2) and let the output device convert it as necessary; this is the great advantage of using PS or PDF for the final document. It is becoming common practice to have two versions of a document, one for **printing** and one for **screen display**. Scalable vector graphics such as PDF can be displayed at any resolution, so they should be used in on-screen documents for graphics that are not inherently rasters. However, the print and screen versions of non-scalable graphics must be different. That is, illustrations must be produced for a specific output device, not only for the resolution and density of information, but also the colour space (RGB for monitor and CMYk for print, see §4.1.6).

In some situations non-scalable graphics must be used, e.g. satellite imagery, digital elevation models, pixel-based maps, and photographs. Then sufficient resolution is necessary for visualization, but excessive resolution just wastes (digital) space. It also gives uncontrollable screen output. A screen has a fixed resolution, generally 72 or 96 dpi. Images of higher resolution than the screen will be resampled by the computer or monitor's graphics card: the number of pixels is reduced, and colour is recalculated. The result is often blurry and in any case out of the control of the author.

It is preferable to control the output by designing it for the known characteristics of the output device, using a specialized image editing software[7]. The non-scalable graphic must be prepared with the required resolution, which depends on the output device type and size. Specialized graphics programs allow the selection of a resolution when storing figures. These are often expressed as "dots per inch" (DPI) or "pixels per inch"[8] (PPI).

4.1.3 Scanning for publication

A common task in thesis production is to scan a printed document, such as paper map, and reproduce all or part of it as an illustration in the thesis. This requires careful planning.

Figure 4.4 shows the scan of a fragment of the soil map of the Netherlands (publication scale 1:50 000) [36].

Scan resolution was 400 ppi. The left side of the figure shows an enlarged fragment, where clearly the pixel can be distinguished: resolution is too low. The right side of the figure shows the reduced scan, giving a

[7] For example, Adobe Photoshop or Adobe Fireworks

[8] 1" = 2.54 mm exactly

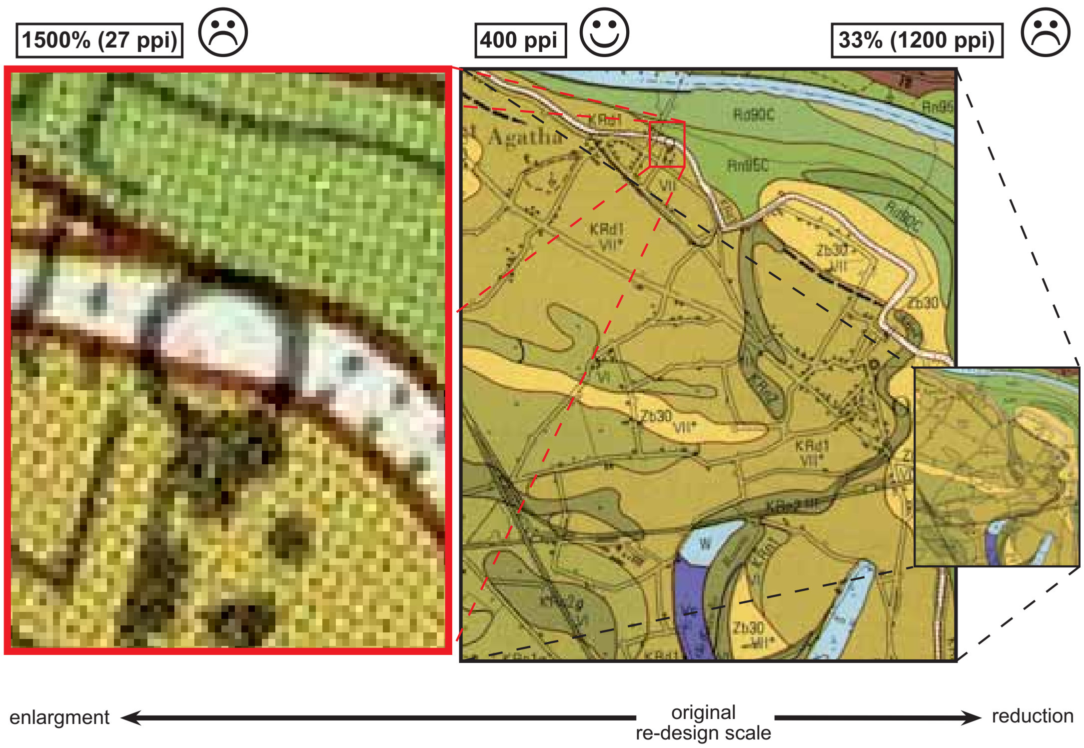


Figure 4.4: Example of a scanned map for publication

higher resolution, in fact too high for the printer or screen, so a waste of (digital) space. It is also illegible. Further, the result in the final PDF document will not include this 1200 ppi, but will resample to the resolution indicated for output.

The centre figure is printed at the original scan resolution: it is legible and smooth (not pixelated at this print resolution). This is what is wanted.

In conclusion:

- **always scan such that the resolution corresponds with the output settings**; a higher resolution is wasted.
- Always include the scan in the document **without rescaling** it; doing so changes the resolution, and you already have the desired resolution from the scan.

4.1.4 Resolution of printed output

This subsection presents more information on how printed output is produced. Understanding this may help you prepare proper graphics for printing.

All monochrome printers, including the laser printers used for produce ITC theses, are incapable of showing grey-tones; they print either 100% or 0% ink (toner or dye), i.e. one bit (0 or 1) of information. These printers produce grey tones by printing the image in dots of various sizes; see Figure 4.5.

Figure 4.5: The 1-bit print capabilities of a printer, 0 = white, 1 = black

A digital printer has a fixed resolution. For example a "1200 dpi printer" can rasterize the image to be printed to a resolution of 1200 dpi. As seen in Figure 4.6, the continuous image must be converted to dots of a certain size, which are placed as a fixed array in a standard cell called a **halftone**. The halftone cell size is a known property of the printer.

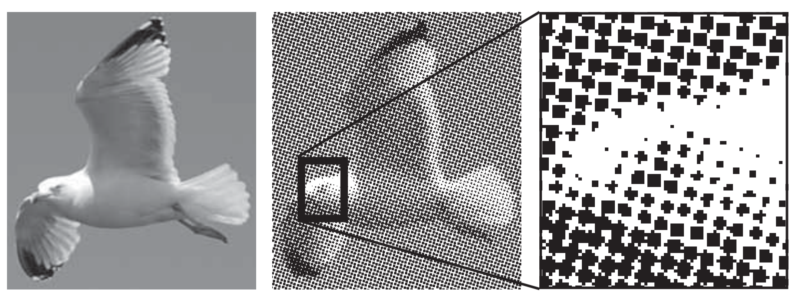


Figure 4.6: A continuous-tone image, as printed by a laser printer

A typical halftone cell is a 5x5 array of printer dots, implying this cell can produce 26 different tones (including full black and white), as shown in Figure 4.7. The larger the halftone cellsize, the more grey values can be produced, resulting in better output quality. Dividing the printer resolution, e.g. 1200 dpi by the halftone cell size, e.g. 5, gives 240, which is the number of halftone cells per inch. This is the printer (halftone) **screen ruling**, also called (halftone) **screen frequency**[9]. A larger halftone cellsize at a given printer resolution has more printer dots per cell, which reduces the screen frequency, thus giving a coarser output (but with a finer scale of grey tones). So, a large halftone cell size (to give many grey tones) requires a high resolution printer for high quality output.

[9] The term "screen" here refers to a halftone print screen, not a computer display

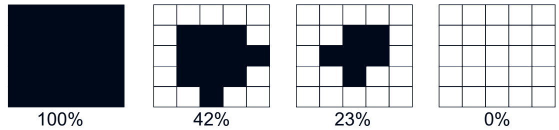


Figure 4.7: A continuous-tone image, as printed by a laser printer

The number of grey values that human eye can distinguish is about 100, thus the screen frequency of high-resolution printers are adjusted so that this about many levels can be shown. For high resolution printing (1200 dpi or higher) the screen frequency is typically 133 or 166 lines per inch, implying a halftone cell of 9x9 or 8x8.

Image resolution should be matched to the screen frequency of the specific printer. Too low resolution gives a fuzzy image; too high resolution is a waste of size. As a rule of thumb, an image resolution of 1.5 to 2 times the screen frequency should be used. For example, for a printer with a 133 line screen, the image should be somewhere between 200 and 266 ppi.

If your work will be printed on an offset press (used for magazines or books), images should be adjusted to a screen frequency of 150 lpi to 225 lpi (lines per inch), which requires an image resolution of 225 to 300 ppi. Higher quality offset printing use a screen frequency of 175 to 225 lpi.

4.1.5 Computer display ouput

For screen output the resolution of the image is the same as the monitor resolution, e.g. 72 or 90 pixels per inch[10] (ppi). For example, on a 1024x768 pixel display on a nominal 17" diagonal screen, the horizontal and vertical resolution is about 72 pixels per inch. So a figure that is displayed over half the width of the screen has a resolution of 512 pixels horizontally. Bitmaps with higher resolution are resampled at the lower (screen) resolution by the computer graphics card. For example, the 300 pixel per inch map fragment of Figure 4.3 is about 4.3 or 3.3 times too fine for screen display.

However, the viewer is able to enlarge (zoom in on) the figure on-screen. In this case the higher resolution will be reached at some magnification (in this example, approximately 3x or 4x). If it is anticipated that the viewer will zoom in on the screen version of the document, the image resolution should be correspondingly higher. The example in Figure 4.3 simulates the zoom effect for 2x, 4x, and 12x. In this case 4x looks good

screen

[10] 1" = 25.4 mm exactly

at normal screen resolution, since it corresponds to the original 300 DPI of the bitmap.

Thus if it is not anticipated that the viewer will enlarge the image, it can be prepared already with the appropriate magnification.

A monitor's ppi resolution depends on the physical size of the pixel and the display settings; there is no way to know the resolution at which the reader may be viewing the output. Two examples: A typical laptop[11] has a monitor with a diagonal of 15.4" and a resolution of 1680 x 1050 pixels. This gives a resolution of 129 ppi and a dotpitch (distance between dots of the same colour) of 0.1974 mm. A large monitor for image processing may have a diagonal of 24" and a resolution of 1920 x 1080 pixels. This gives a resolution of 92 ppi and a dot pich of 0.2767 mm.

4.1.6 Use of colour

There are two reasons to use colour, rather than grayscale, in graphics:

- Colour is a **visual variable** and thus can communicate information in addition to what is presented by other visual variables such as size, shape, font, and pattern;
- Colour is an **aesthetic element** which can make information more attractive, even if the colour itself conveys no extra information.

Colour can add meaning, and if so it should be used. Note that **colour is often not required** or even desireable. When comparing a grayscale and colour figure, ask if the colour is merely **decorative** or truly **connotative**. Colour can even confuse the reader if used improperly – so it is best avoided unless necessary.

When using colour as an aesthetic element only, remember that "beauty is in the eye of the beholder". That is, whether the colour is "attractive" is subjective. When colour is used to add meaning, as a necessary element of the graphic, it is easier to agree objectively that the colour adds value; for purely aesthetic reasons opinions may well differ.

! → **Correct use of colour is not easy**; users are urged to consult a handbook such as ITC's [4]. Colour conventions for applications should also be consulted; for example Robinson *et al.* [29, §21] discusses conventions and perceptual issues for map design.

Colour figures may also be rendered on gray-scale printers. Attractive and connotative colours will in general not look good, or even be distinguishable, in gray scale. There are colour schemes that are designed to also reproduce well in grayscale. A counter-example is the scheme

[11] late 2009

of Figure 4.8; note how all the different colours translate to almost the same grey.

Figure 4.8: A continuous-tone image, as printed by a laser printer

Another solution is to have two versions of the document, or have the colour versions of figures as on-line or on-CD supplementary information.

Figures should be designed for a specific output device. For example, if an illustration is to be printed in greytones, it makes no sense to create it in colour. Further, different colours convert to similar greytones, so a poor combination of colours (that are perceived separately in colour) may not be perceived separatedly in greytone. (see figure 1) Further, any connotation from the colours will most likely be lost. Instead, a connotative greytone legend, probably with patterns, should be designed.

Colour space

Brown & Feringa [4, §3] explain how colours are described in various **colour spaces**, and how these relate to human perception. Here we give a brief summary.

A colour space is "[A] three-dimensional geometric representation of the colours that can be seen and/or generated using a certain colour model"[12]. The concept of "colour" is hard to define or even describe. The various systems that have been developed to describe "colour" agree on using a coordinate system in a 3D (x, y , z) space to describe colour; this is the **colour model**. The x, y and z values stand for **Hue**, **Lightness** (Intensity, brightness, value) and **Saturation** (vividness, value, purity, chroma). A colour space has no influence on how we experience the colour but is a system to specify, create and visualise colour. A colour space describes the maximum amount of colours that can be used or visualised for that specific space.

For monitors the **RGB colour space** (Figure 4.9) is generally used; for colour printers the **CMYk colour space** (Figure 4.10).

Every device has its own specific maximum range (gamut) which is smaller than the colour space in which it appears. Colour printers are not able to print all colours which humans can perceive; their gamut depends on

[12] http://www.iec.ch/

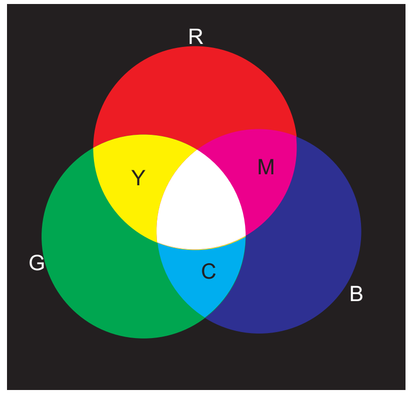
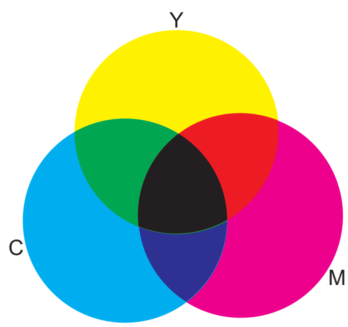


Figure 4.9: RGB colour mixing. RGB are the primary colours, the results of mixing CMY are the printing primaries. Black (absence of light) is the starting point, light (colour) is added. Therefore this method is called “additive” colour mixing



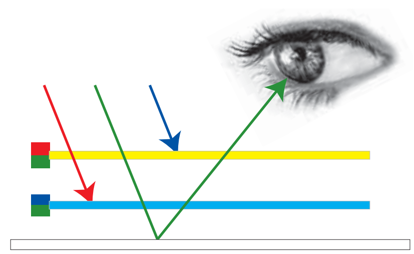


Figure 4.10: CMYk colour mixing. The inks are translucent, and act as a filter. A filter allows its own colour to pass and absorbs the others. Colour is subtracted from white (maximum mixture of all colours) and this method is called “subtractive” colour mixing

the printing technique (offset, inkjet, laser); the specific type of pigments used to create the ink, dye or toner; the quality of the paper; the callibration of the printer, etc. Likewise, monitors vary in their technology and pigments, and so also have different gamuts.

There is no direct translation between colour spaces; a **Colour Management System** is required. RGB is created by mixing light (starting point is black, we add colour). On a monitor tiny Red, Green and Blue lamps send out light of different intensities (in steps from 0 - 255) and in that way can create more than 16 million different combinations (figure 2 shows how full red-green-blue mix).

When printing, the starting point is the white of the paper, colour is subtracted from the white paper by the ink. The translucent ink acts as a filter, which allows its own colour of light to pass and absorbs others. For example, yellow is a combination of Red and Green and so the yellow ink allows Red light and Green light to pass, Blue will be absorbed. See figure 3 for a further explanation.

When creating illustrations for the thesis, always keep in mind that output on different systems give different result. Never blame the monitor or printer, they are not good or bad, they just have their limitations.

Colour composites use combinations of colour to show combinations of factors. Figure 4.11 is an example. Note how the gradations in three colours can be immediately interpreted. Note however that the colours that display well on a monitor, to give a desired visual impression, may not print well in grayscale.

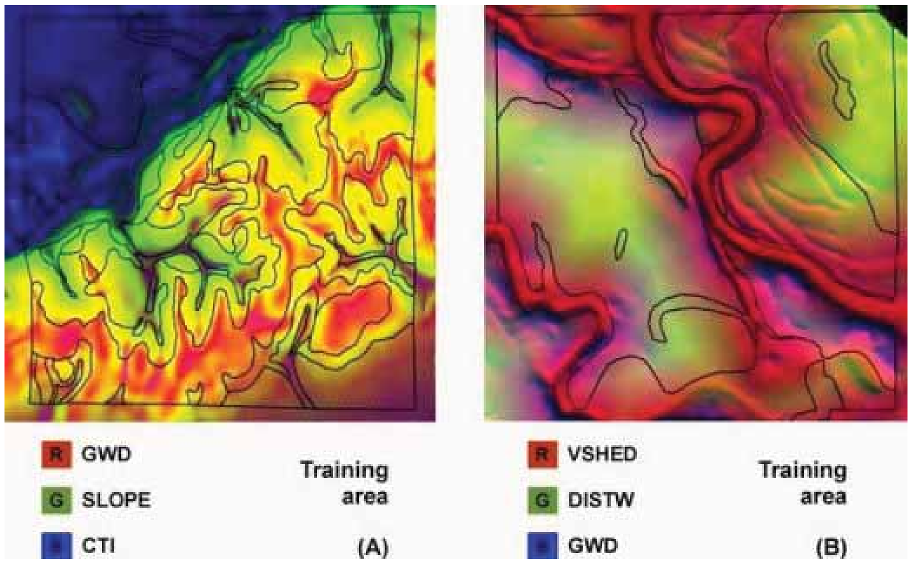


Figure 4.11: Colour composites of three terrain parameters, Baranja, HR; Source: Hengl [12]

Colour ramps

Colour ramps are sequences of colour that correspond to human perception of the desired sequence. Some are well-established, for example in topographic mapping; an example is shown in Figure 4.12.

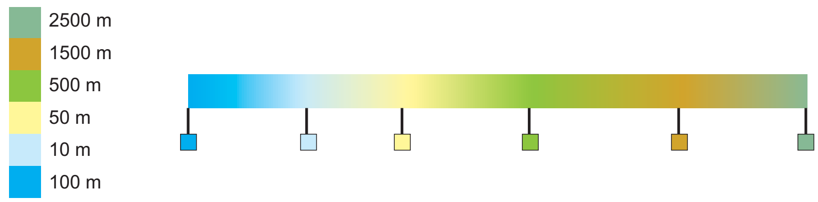


Figure 4.12: A colour ramp created from an "altitude" range, creating gradual tints for the intermediate altitudes

Other ramps suggest a physical phenomenon such as heat or a psychological phenomenon such as risk (Figure 4.13).

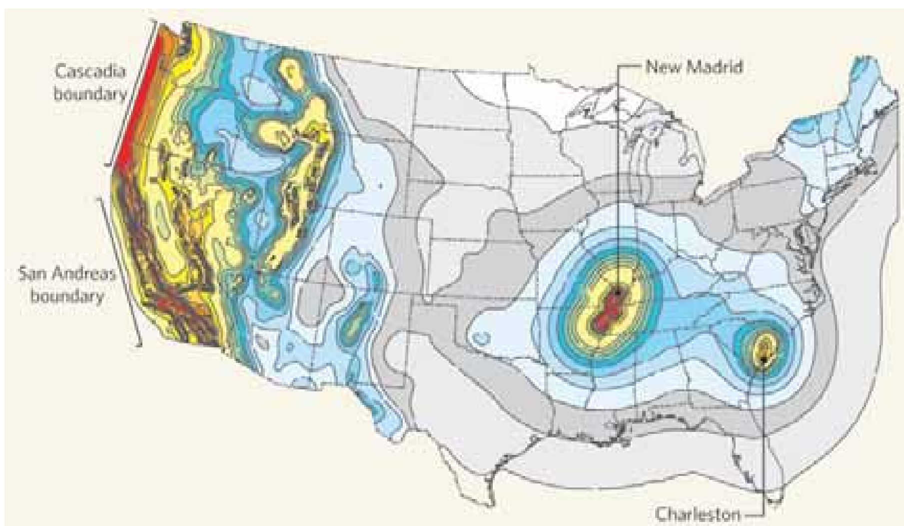


Figure 4.13: Probabilistic assessment of seismic hazard in the United States; Source: [11]

Q85 : *How does the colour ramp of Figure 4.13 convey the idea of earthquake risk? What is missing in this figure?* *Jump to A85* •

Colour ramps are implemented in many graphics systems; Figure 4.14 shows an example from ArcMap. The `grDevices` package of the R environment for statistical computing and graphics [30] provides several palettes, some of which are colour ramps[13].

[13] e.g.`heat.colors` and `topo.colors`

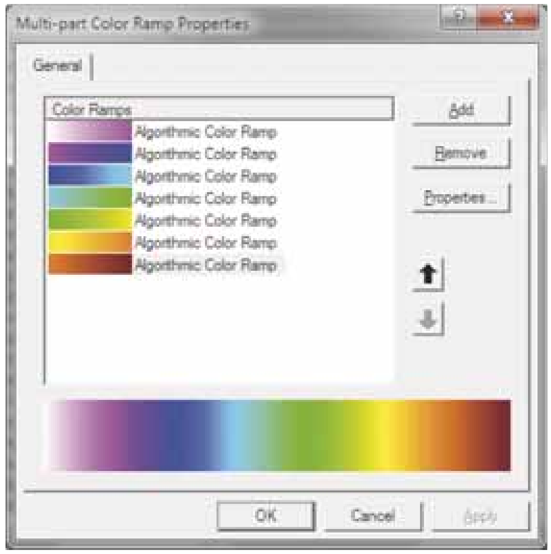


Figure 4.14: A colour ramp as available in ArcMap. It is a constructed ramp, created from seven separate ramps

Use of contrasting colours

Contrasting colours are those that are clearly separated on the colour circle (Figure 4.15). They are used to distinguish different things. Non-contrasting or **similar colours** to show gradations or sub-classes of similar things (Figure 4.16). It is tricky to choose a good set of similar colours; in the figure colours are taken from one quadrant of the colour circle

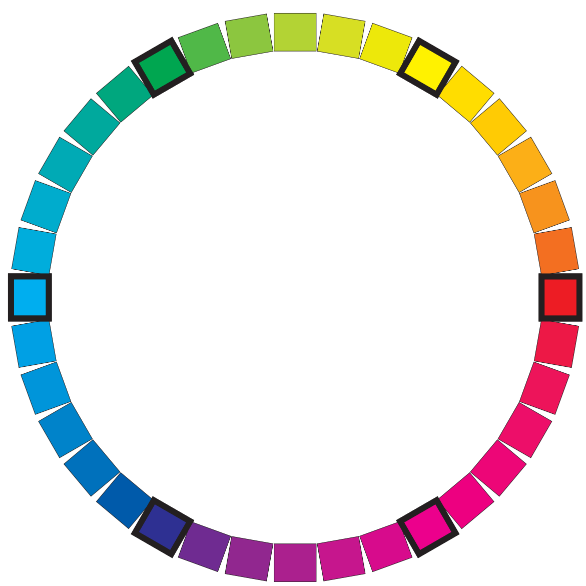


Figure 4.15: Colour circle (hues) with several clearly-seprated contrasting colours

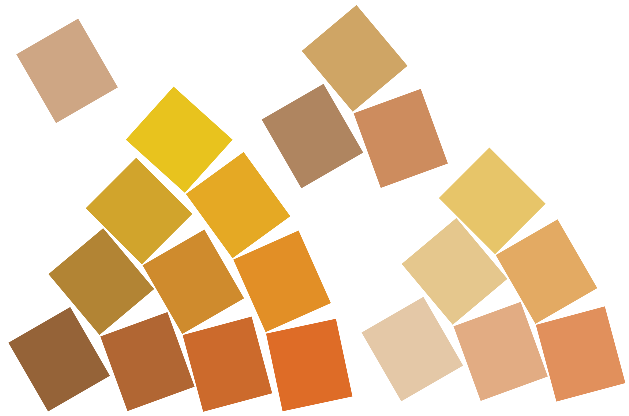


Figure 4.16: An example of similar colours

When features in a map show nominal differences (e.g. in a landuse map: forest, agriculture, desert, built-up areas) contrasting colours have to be used. Another example is a political map – similar colours would imply political linkage (so, perhaps could be used to distinguish autonomous regions within a country).

For **ordinal** differences, **tints** (percentages) of a base colour have to be used, to show that one element is "more" or "less" than another element (darker is more, lighter is less). Examples are population density, temperature, and altitude. Figure 4.17 shows a well-chosen tint range. The darkest green should be used for the "densist", "largest", "heaviest" etc. end of the range, and progressively lighter tints for "more dilute", "smaller", "lighter" etc.

Figure 4.17: A well-chosen tint range

For elements that fall within the main class, i.e. sub-classes, similar colours are required (Figure 4.16), but they should not show gradual differences. Often patterns are used for a better distinction. Examples are the subclasses in soil, geomorphological or geological maps. However, similar colours rely on differences in lightness or value, which can lead to mis-interpretation.

Colour connotations

Brown & Feringa [4, §7] discuss perceptual and psychological factors which must be taken into account when choosing colour.

Colours can be **connotative**, suggesting something about the class. These can be conventions established over the years. An example from soil mapping is purple for organic soils, blue for clayey soils, browns for loamy soils, and yellows for sandy soils.

But, colour is often **culture-specific**, so if a colour is supposed to represent a concept, it may not have the intended effect. For example, bright red generally signals "danger" or "hazard" in the USA, but "good fortune" in China; white is the colour of "purity" in most northern European cultures but represents "death" in China. The traditional colours showing topographic regions, with green for low elevations, grading to browns and finally white, will not be connotative to areas where lowlands are arid (pale brown in nature) and mid-slopes are moister (green in nature).

4.1.7 Typography

Typography is an essential part of the graphic design of a thesis. It has a large effect on its readability, and thus makes the content more or less accessible. Type was in ancient days a pictographic system, which has evolved to different systems, including the Latin alphabet which is used in English-language documents such as an ITC thesis. Before development of printing, text was duplicated by means of handwriting, mostly the person that reproduced the text was responsible for the choice of type its design.

With the development of printing, techniques were also developed to standardize the appearance of characters and symbols; a consistent appearance is known as a **typeface**. The first classification of typefaces dates from 1964, into eleven distinct styles (e.g. serif, sans-serif, script, roman and handwriting). A single typeface can contain a wide variation of styles (e.g. normal, bold, italic, condensed, expanded, black). The variations of a single typeface design taken together form a **font family**.

The elements of a character all have their specific names, see Figure 4.18. In this figure, all characters are of different typefaces, but all are specified with the same size. This example makes clear that the size of typefaces depend on the design. A 24 point size for one type is not equal to that of another one.

For each printed work the designer must choose a typeface, and further a **font**, which is a complete character set of a single size and style of a particular typeface. For example 'Arial' is a typeface, 'Arial 12 points bold' is a font.

Font **size** is specified in points; using one of three different systems:

- DTP-point (mainly used nowadays): 1 pnt = 1/72 Inch = 0.353 mm

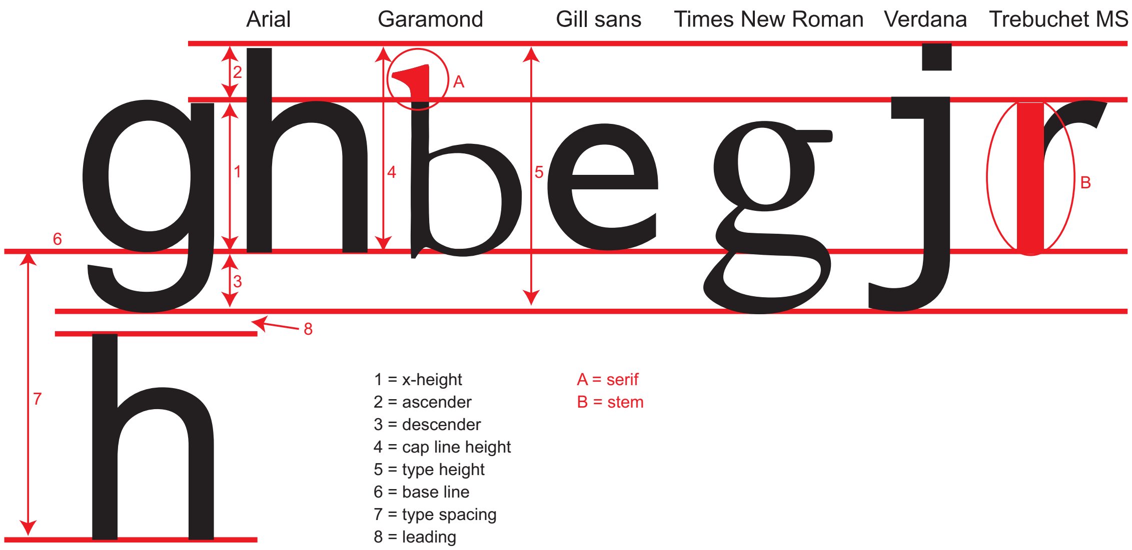


Figure 4.18: Some fonts and font terminology

- Pica-point: 1 pnt = 0.351 mm (12 pnt = 1 Pica, 6 Pica = 1 Inch)
- Didot-point: 1 pnt = 0.375 mm (12 pnt = 1 Cicero)

Unfortunately different programs use different rules (not always made explicit) for type size definition, for example “the size of a font, measured in points, from the lowest descender to the highest ascender” or “the height of a character together with the amount of space between lines of text”. Different fonts vary a lot in their x-heights as well as in caps heights. For example, 12 point Times New Roman Regular seems to be smaller than 12 point Verdana Regular (Figure 4.19).

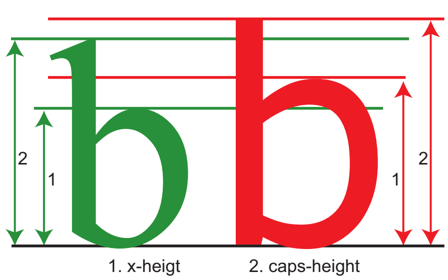


Figure 4.19: Font measurements

Screen versus printed output

A typeface is designed for a certain purpose, either for use on paper or for use on screen. Most people however use them wherever they please, without taking note of the characteristics of print on paper or showing on screen. The main issue is resolution or detail: for a paper print this is much higher than for screen. Screens have a fixed resolution, expressed in dpi. For a screen (or monitor) this is somewhere between 72 and 110, on average 96 dpi. The resolution of a print on paper depends on the resolution of the printer, which is nowadays 1200 dpi, i.e, 12.5 times better than screen resolution. Type for use on the monitor thus has to fulfill other requirements than type for use on paper. Figure 4.20 shows the characteristics of a good font for use on Web pages.

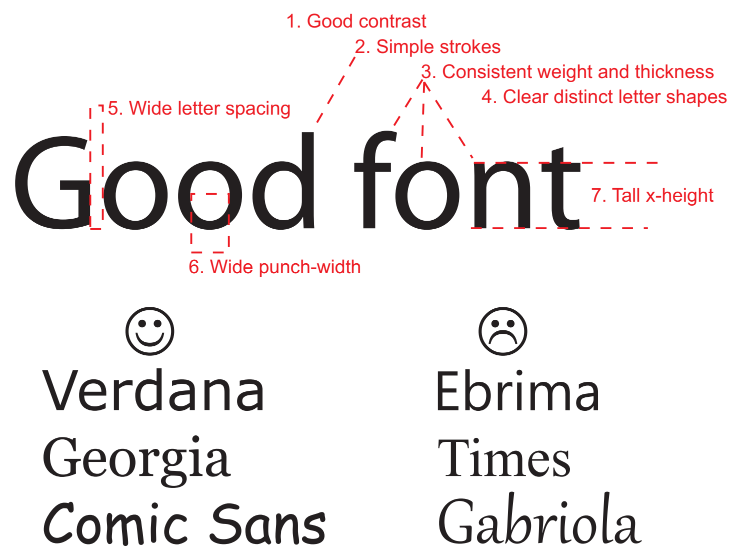


Figure 4.20: Characteristics of a good web font

For the ITC thesis, a font designed for printed output must be used. The ITC LaTeX thesis class[14] uses New Century Schoolbook for running text and Computer Math for mathematical typesetting. ITC also has a license to use the Lucida Bright fonts[15]. You could choose to use these in figures, except that the running text fonts are serif.

Use of fonts

It is important not to use too many different type faces or font families in one document; in general a maximum of two, where one family can be used for running text and one for headings. A general rule of thumb for print is to use a a sans-serif typeface (Arial, Helvetica or similar) for the headings and a serif typeface for the running text. As a rule[16] use a maximum of 60 characters per line and around 40 lines per page.

[14] itcmsc
[15] These are used in the present document
[16] http://desktoppub.about.com/od/typography/Typography.htm

Lengthy text should not be set smaller than 9 points and no larger than 11 points. Line space should be type height + 2 point.

Good typography in illustrations

- Use the same font family and the same base-sizes throughout the whole document.
- Create all illustrations on the correct scale (as used in the document). Re-scaling the illustrations will change the fontsize in the final output.
- Use a sans serif font (e.g. Arial, Helvetica, Verdana).
- Use a maximum font size of 11 points and a minimum font size of 7 points, at the output scale.
- Avoid text overlapping/crossing other graphic elements (like lines or patterns) of the same colour.
- Maintain a good contrast between the text and the background.
- Place text in the illustration near the element it is referring to.
- Respect the correct reading direction for the application, e.g. contour values facing uphill.

4.1.8 Answers

A*85* : *Red colours imply "danger" (in all cultures?); blue colours imply "cool" or "relaxed", i.e. little danger; white implies "background" or "not relevant", i.e. the concept of danger does not apply in those zones. Intermediate colours (grey next to white; shades of blue from light [low intensity] to dark [high intensity]; blue grading into green and then yellow on the way to red. Note that it is impossible to have non-adjacent colours adjacent in the map.*

There is no scale of measurement attached to the colours. Red is of course highest probability, but is this a return period? If so, what?

There is no scale; but since many other maps of the USA can be consulted, and the map is meant to give an overview, this may not be so serious. There is no North arrow, so the reader assumes that the map is North oriented. There is no coordinate reference system information. If the map is provided digitally, this must be known. *Return to Q85* •

4.2 Graphical elements in a document

Key points

1. Five types of graphical elements are described: **tables**, **graphs**, **illustrations**, **maps**, and **images**.

This section discusses five kinds of graphical elements:

1. Tables (§4.2.1)
2. Graphs (§4.2.2)
3. Illustrations (§4.2.3)
4. Maps (§4.2.4)
5. Images (§4.2.5)

Each has its own use and construction issues.

4.2.1 Tables

Key points

1. Tables are used to present **numerical data** in a **compact** form.
2. Tables must be **self-explanatory**.
3. Tables must be **referred to** in the text, and contribute to the **argument**.
4. Use **spaces** to separate elements; **horizontal lines** just to separate headings from contents.

Purpose

Large amounts of numerical or descriptive data are poorly-served by expository text. They can generally be much better understood as summary tables. For example, the text:

☹

> "Of the visitors to the web portal during the trial period, 76% were from Europe (32% Netherlands, 22% Germany, 9% Belgium, 5% France, 5% United Kingdom, 3% other European) and 24% from rest of the world (10% Americas, 12% Asia/Pacific, 2% Africa)."

can much better be presented as a table and commented on in the text:

"Visitors to the web portal were predominantly from Europe, especially the Netherlands and Germany (Table X)."

☺

Europe						**Other**		
total 76						*total 24*		
NL	D	B	F	UK	other	America	Asia/Pacific	Africa
32	22	9	5	5	3	10	12	2

Table X: Web site visitors, location (%)

Conversely, a few numbers should be presented in the text, and do not require a table. For example, the following table:

☹

Yes	**No**
20	10

Table Y: Wetness in soil profiles

can much better be presented as:

☺ "Twenty of the thirty soil profiles showed signs of subsoil wetness"

Appearance and types

Here are some guidelines for table design.

- Tables should be understandable **without reference to the text**; the text should comment on and interpret a table, but should not be needed to understand its contents.
- Each table must have a **caption**, which is understandable without reference to the text. This is repeated in the **List of Tables**.
- Each column and row must have a **heading** which very briefly identifies it.
- A table should be as **simple** as possible, given the information it must present. The position of data items in the table should be as intuitive as possible.
- All rows, columns, sub-rows or columns etc. must be clearly **labelled** so the reader can tell to what a particular item (cell) refers.
- The **same data** should generally not be presented in the text, table, and figures; if they do, there should be a clear reason. In particular, avoid phrases like

☹ "According to Table *X*, 35.3% of the respondents preferred the first option . . . "

The table already gives the exact number, so the text does not need to repeat it; instead, the text comments on it:

☺

> "A substantial minority of the respondents preferred the first option . . . (Table *X*)."

Here the author interprets ("substantial minority") the exact number (35.3%) shown in the table, and relates it to the overall argument.

- Use of **symbols, units and terms** in tables should agree with the text. For example, a column heading "water bodies" in the table should not correspond to a category called "lakes" in the text.
- If categories given in rows and columns are naturally **ordered**, this should be respected in the table. For example, educational levels "none", "primary", "secondary", "tertiary" should follow this order in rows or columns.
- Tables must be **numbered** sequentially, either by chapter or for the whole document. They must be **referenced** in the text, either in parentheses or as a noun (subject or object):
  - *Table 3 shows the model parameters.*
  - *The model parameters are shown in Table 3.*
  - *Initial model parameters were set according to literature (Table 3).*
- The combination of human eye and human brain is optimized to detect patterns. Thus if the table has a good structure, **most horizontal and vertical lines are superfluous**, and in fact tire the reader (by breaking the reading flow) and can make it more difficult to see related items. Font variations (e.g. bold, italic) can be used to provide contrast of table headers from body. Compare Table 4.1 with Table 4.2. Table 4.1 is too "busy" and difficult to read.

☹

Europe						Other		
total 76						total 24		
NL	D	B	F	UK	other	America	Asia/Pacific	Africa
32	22	9	5	5	3	10	12	2

Table X: Web site visitors, location (%)

Table 4.1: A table design with too many lines

☺

By contrast, Table 4.2 is clear and easy to read. Note also that the numbers are centred under the headings in this design.

Europe						Other		
total 76						*total 24*		
NL	D	B	F	UK	other	America	Asia/Pacific	Africa
32	22	9	5	5	3	10	12	2

Table X: Web site visitors, location (%)

Table 4.2: A table with spaces

Examples

☺ All the following examples here have good design.

The structure of a typical table is shown in Table 4.3.

Table xx This is the table caption

	Spanner head		Spanner head	
Stub head	*Col. head*	*Col. head*	*Col. head*	*Col. head*†
Row head				
Row subhead	xxx	xxx	xxx	xxx
Row subhead	xxx	xxx	xxx	xxx
Row subhead	xxx	xxx	xxx	xxx
Row head				
Row subhead	xxx	xxx	xxx	xxx
Row subhead	xxx	xxx	xxx	xxx
Row subhead	xxx	xxx	xxx	xxx
Total	yyy	yyy	yyy	yyy

Source: …

†: an explanatory footnote

Table 4.3: A typical table – structure

Cells with numbers should be aligned on the decimal point or comma, as shown in a worked example, Table 4.4.

Table 4.5 shows the use of a table to introduce a large amount of notation that will be subsequently used in the text, figures or other tables.

Table 4.6 shows a table with results, in this case a confusion matrix. The data is from a paper by Congalton & Mead [8], also used as a teaching example by Skidmore [34, Table 12.2]). This table has a clean design; however it can be enhanced, see Figure 4.21. Here grey tints at the back instead of lines enhance readability; leading is changed between main classes; the difference horizontally between classes is 4 points extra.

Table xx Fatty acid composition of selected dietary fats and oils (wt %)

	Saturated		Unsaturated	
Fat,oil	*Palmitic*	*Stearic*	*Oleic*	*Linoleic*†
Animal fat				
Lard	29.8	12.7	47.8	8.7
Butter	25.2	9.2	29.5	10.8
Beef	29.2	21.0	41.1	5.3
Vegetable oil				
Corn	8.1	2.5	30.1	59.2
Peanut	6.3	4.9	21.8	‡
Soybean	9.8	2.4	28.9	57.7

Source: adapted from [5, Fig. 31-1]

†: and miscelleanous

‡: not reported

Table 4.4: A typical table – filled example

‘

Symbol	Meaning	Computation
$\mathbf{X}$	confusion matrix	
	→ rows $[1 \ldots r]$ are *classified*	
	→ columns $[1 \ldots r]$ are *reference*	
r	number of rows and columns of $\mathbf{X}$	
x_{ij}	number of observations in row i, column j, i.e. in reference class j but mapped as class i	as observed
x_{i+}	marginal sum of row i	$\sum_{j=1}^{r} x_{ij}$
x_{+j}	marginal sum of column j	$\sum_{i=1}^{r} x_{ij}$
n	total number of observations	$\sum_{i=1}^{r} \sum_{j=1}^{r} x_{ij}$
	… or	$\sum_{i=1}^{r} x_{i+}$
	… or	$\sum_{j=1}^{r} x_{+j}$
$\mathbf{P}$	proportions matrix	$\mathbf{X}\ ./n$
p_{ij}	proportion of observations in row i, column j	x_{ij}/n
p_{i+}	proportion of mapped data in row$i$	x_{i+}/n
	… or	$\sum_{j=1}^{r} p_{ij}$
p_{+j}	proportion of reference data in column j	x_{+j}/n
	… or	$\sum_{i=1}^{r} p_{ij}$

Table 4.5: Table summarizing notation (here, for confusion matrices)

Construction

Tables are automatically formatted by common “office” programs such as MS-Word; however, without any attention to sound design principles.

		Reference Class				Total
		A	B	C	D	p_{i+}
	A	35	14	11	1	61
		0.2147	*0.0859*	*0.0675*	*0.0061*	*0.3742*
	B	4	11	3	0	18
Mapped		*0.0245*	*0.0675*	*0.0184*	*0.0000*	*0.1104*
Class	C	12	9	38	4	63
		0.0736	*0.0552*	*0.2331*	*0.0245*	*0.3865*
	D	2	5	12	2	21
		0.0123	*0.0307*	*0.0736*	*0.0123*	*0.1288*
	Total	53	39	64	7	163
	p_{+j}	*0.3252*	*0.2393*	*0.3926*	*0.0429*	*1.0000*

Roman type: counts; *Italic type: frequencies*

Table 4.6: Table summarizing results; for enhancement see Figure 4.21

			Reference Class			**Total**
		A	***B***	***C***	***D***	***pi+***
Mapped Class	***A***	35 *0.2147*	14 *0.0859*	11 *0.0675*	1 *0.0061*	61 *0.3742*
	B	4 *0.0245*	11 *0.0675*	3 *0.0184*	0 *0.0000*	18 *0.1104*
	C	12 *0.0736*	9 *0.0552*	38 *0.2331*	4 *0.0245*	63 *0.3865*
	D	2 *0.0123*	5 *0.0307*	12 *0.0736*	2 *0.0123*	21 *0.1288*
Total	***p+j***	53 *0.3252*	39 *0.2393*	64 *0.3926*	7 *0.0429*	163 *1.0000*

Figure 4.21: Table summarizing results (enhanced from Table 4.6)

LaTeXprovides a `tablular` environment for fine control of tables, and a `table` environment for floating tables within running text. As usual for LaTeX, it can be time-consuming to format a table, but once done, it does not change. See a good LaTeX reference [e.g. 19, 20] for details. Enhancements such as Table 4.21 require the use of specialized packages, e.g. `tikz`.

4.2.2 Graphs

Key points

1. Graphs are used to report the results of **exploratory data analysis** or **statistical analysis**.
2. The visual form of the graph must be interpretable by the reader, and convey the result **honestly** and **efficiently**.
3. Graphs must be labelled, captioned, and referred to in the text, and contribute to the **argument**.
4. There are a wide variety of **graph types**, the selection depends on the **form of the data** and the **message** to be conveyed to the reader

Graphs are most commonly used to report the results of exploratory data analysis or statistical analysis. Statisticians have identified effective and honest visualisation techniques, as well as confusing and misleading ones. Two of the best-known works are by Tufte [39, 40]; these are fascinating and full of ideas for effective graphical presentation. Two semi-humorous books [14, 18] titled *How to lie with ...* (*statistics* and *charts*, respectively) give amusing examples.

Cleveland [6, 7] developed novel visualization techniques especially for multivariate and conditioned data; these techniques have been implemented as "Trellis" graphics in S+ and as the `lattice` package [31, 32] in the R environment for statistical computing [30]. Geographers have long had effective methods for presenting geographical statistics, for example Dickinson [10].

As with all graphical elements, graphs must be labelled, captioned, and referred to in the text. The graph must contribute to the argument. It is common to present a graph and then comment on what it is supposed to show about the result.

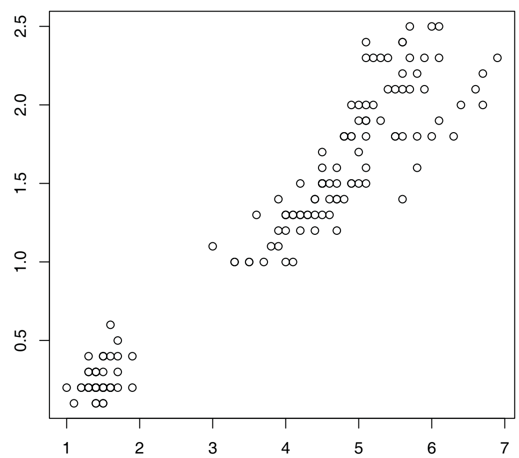
Composition and labelling

Graphs consist of (usually two) **axes**, possibly with a second scale at the right, with information in the centre. They should be as simple ("clean") as possible, while presenting the relevant information. By convention, a **bivariate** graph presents the "independent" variable on the horizontal axes, and the "dependent" variable on the vertical; **trivariate** graphs present the dependent variable on the vertical axis, with two independent variables spanning a plane that appears to come out of the 2D page.

Axes of graphs must be labelled, also with units of measure. Divisions of the axis should be simple numbers (e.g. integers) unless there is a special reason (for example, a threshold or a mean value).

Information content of graphs Before discussing specific graph types, we illustrate the concept that a graph must be **interpretable** by itself, i.e. it must have sufficient information for the reader.

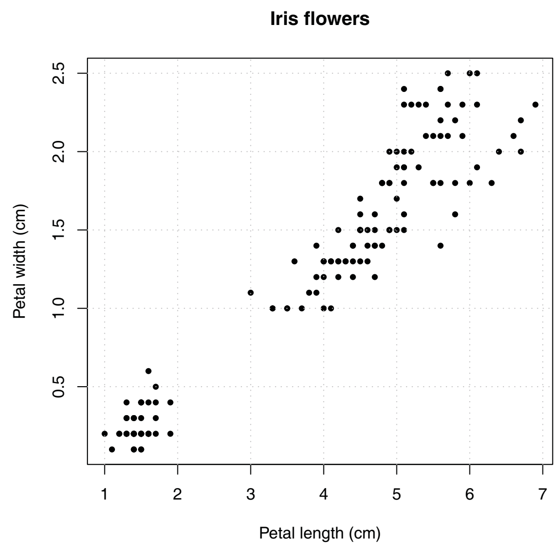
Consider this scatterplot of the petal width vs. petal length of the well-known Anderson dataset[17] of 150 iris flowers, 50 from each of three species [2]:



Q86 : *What information is missing in this graph?* *Jump to A86* •

Here is an attempt at modification:

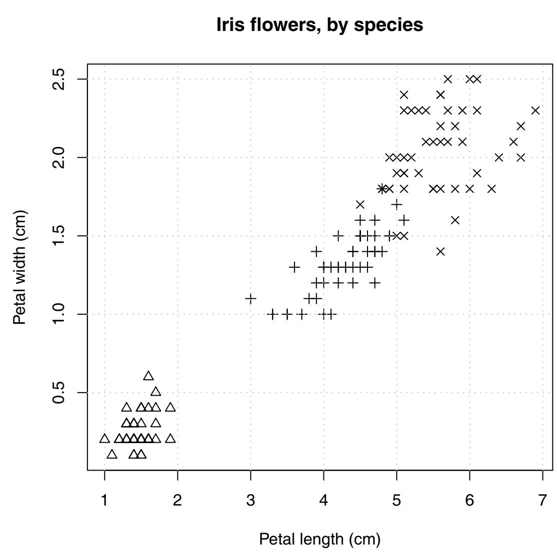
[17] Available as R dataset `iris`



Q87 : *What information has been added? What is still missing?* *Jump to A87* •

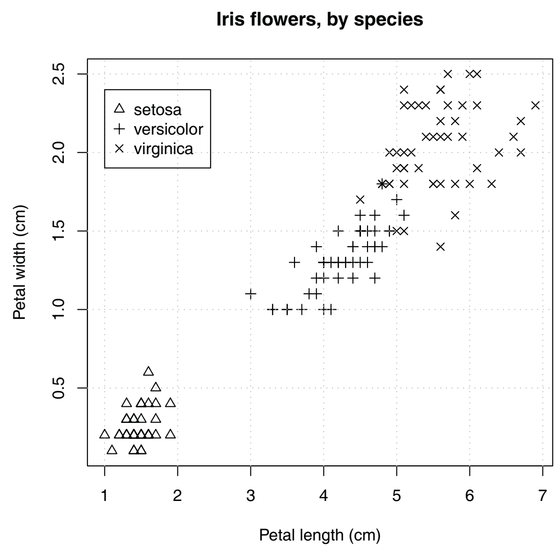
This modification has a grid to help the viewer estimate values of each point. The points are now clearer and a bit smaller.

This scatterplot combines observations from three different species; these were not shown in the previous plots, so we add that information now, using different symbols to indicate the species:

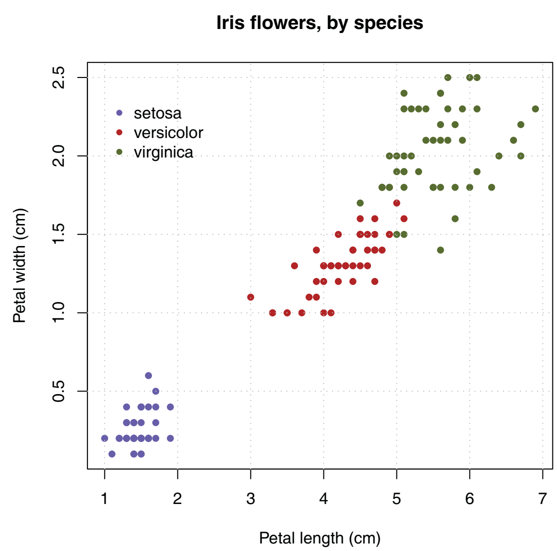


Q88 : *What information has been added? What is still missing?* *Jump to A88* •

The final version, now with a legend, is:



This same information could have been added in colour:



Q89 : *What could be added to these last two plots to enhance the viewer's understanding of the data?* *Jump to A89* •

An example of an interpreted graph is shown in Figure 4.22.

Graph types

Various types of graphs have been developed to visually present data and analytical results. Graphs appeal directly to the user's visual perception, and as such can greatly influence the viewer's interpretation. It is the responsibility of the designer to prepare graphs that are:

- **honest**: the graph must not mis-represent information (e.g. by omitting elements);
- **informative**: maximum information should be conveyed;
- **clear**: it must be possible to interpret the graph and its caption directly, without reference to the text;
- **useful**: the graph must add to the viewer's understanding.

Following we show some common types; these are by no means the only ones; many others are presented by Cleveland [6, 7], Schmid [33], and two introductions by Jacoby, one on univariate and bivariate data [16] and the other on multivariate data [17].

Histograms

Histograms are used to show the univariate distribution of some variable, either as a **frequency** (count) or **density** (proportion)

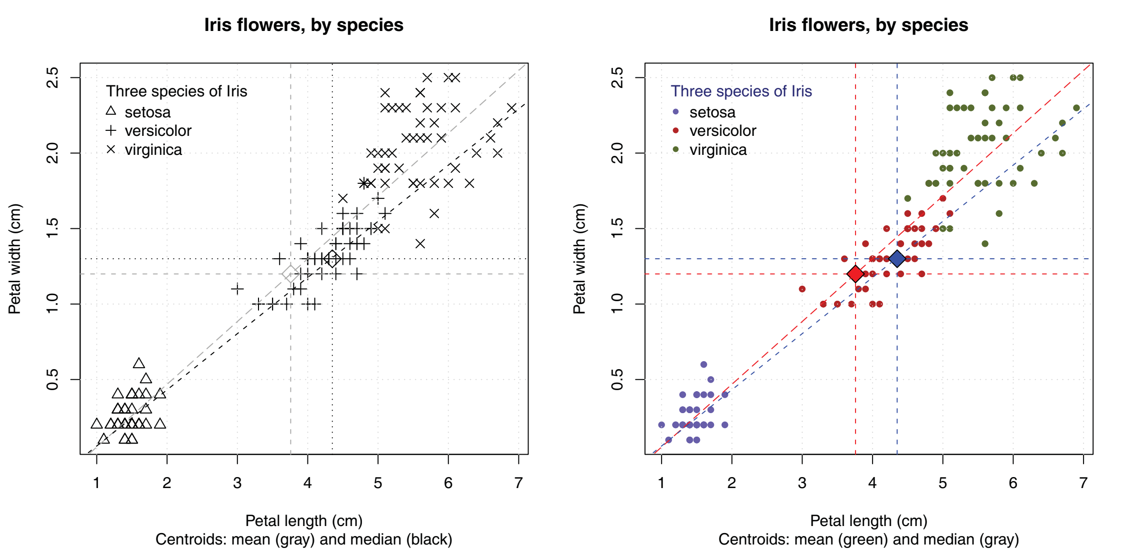


Figure 4.22: Example of interepreted graphs

Figure 4.23 (left) is a typical histogram; the same figure (right) shows the same histogram, enhanced by showing the actual counts, precise values with a **rug** plot under the x-axis, and a **grayscale ramp** colours to indicate the relative intensity of the histogram bins.

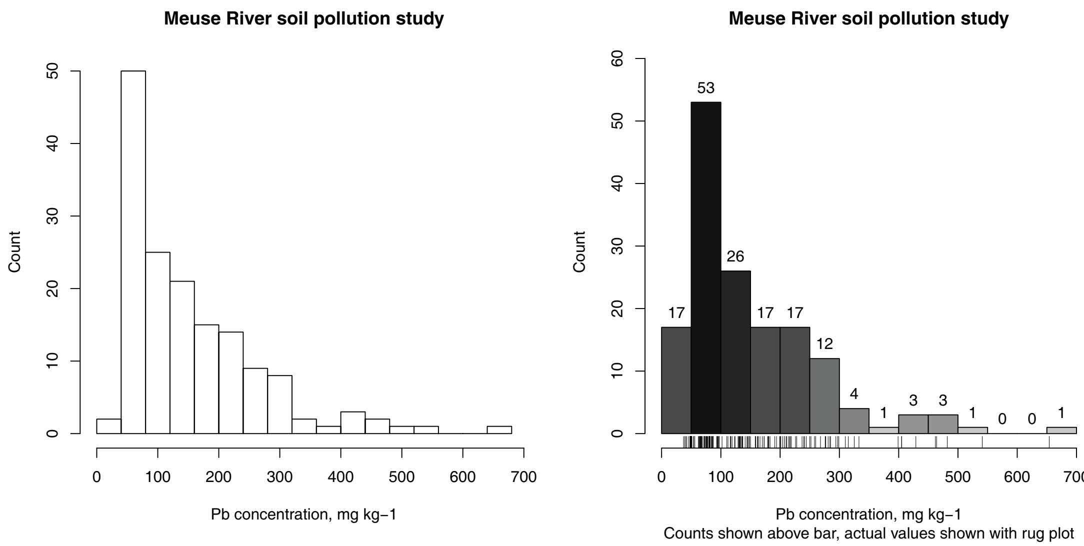


Figure 4.23: Plain (left) and Fancy (right) histograms

Figure 4.24 shows density histograms for six metals in surface soils in the same figure. This allows a direct comparaison of the shape of their distributions. The scales on both axes are different in each graph, because each metal has a different distribution, and the importance is in showing these as density distributions. These histograms are further enhanced with the **kernal density** estimation of the continuous frequency, as estimated from the histogram.

Boxplots

A boxplot, also called "box-and-whiskers" plots, gives another view of the univariate distribution of a variable. The precise definition varies, but always half of the observations are inside the box, defined by the 25% and 75% percentiles (also called the **quartiles**), with the **median** (50% percentile) also shown, and the range of observations outside the box until **fences**, generally defined as 1.5 times the **inter-quartile range** (IQR). Any more extreme observations are shown as **boxplot outliers**. The boxplot is only meaningful for **uni-modal** distributions.

Figure 4.25 is a typical boxplot, here with the observations **grouped** by a

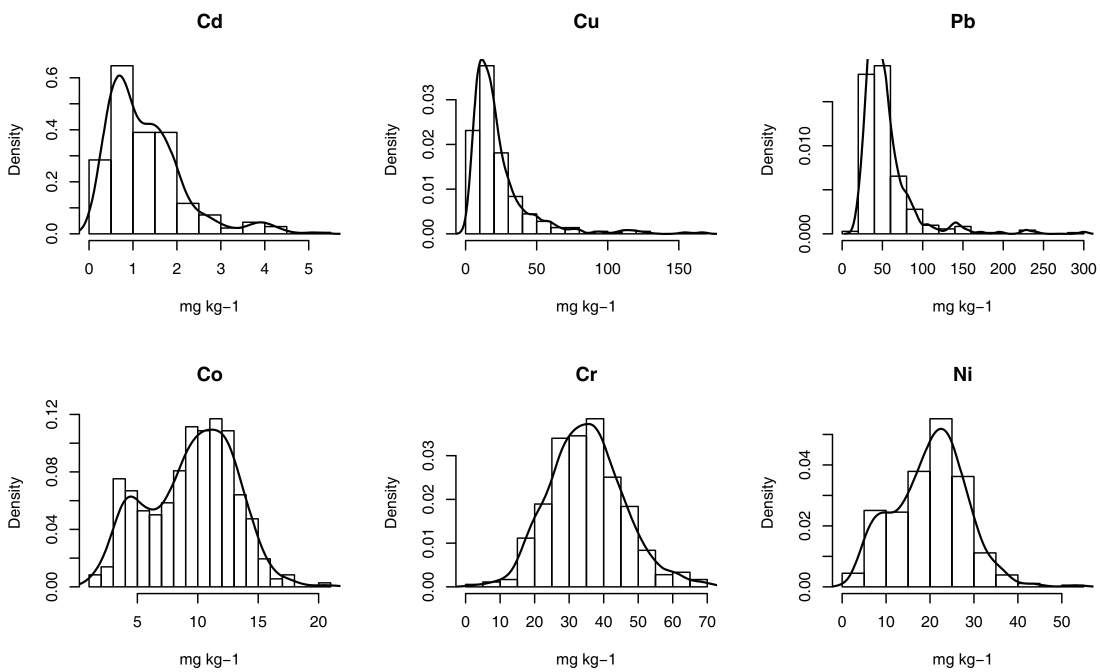


Figure 4.24: Example of multiple density histograms in one figure

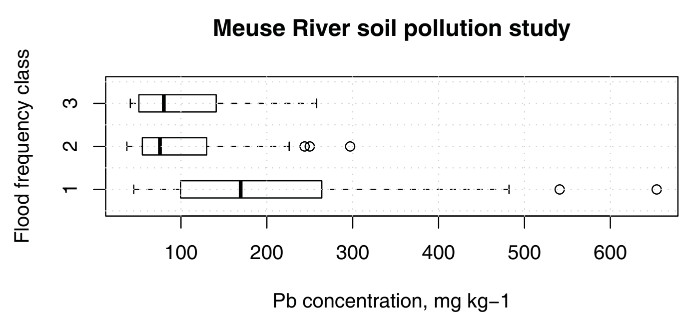


Figure 4.25: Example grouped (classified) boxplot

classifying factor. The skewed distribution for all classes is clear, as well as the boxplot outliers in classes 2 and 3.

Scatterplots

Bivariate scatterplots show the relation of two variables. The left graph in Figure 4.26 is an example. The two plots in the figure show the relation of a single **dependent** variable (here, volume of trees) to two different **independent** variables, here tree girth and height. From a scatterplot the reader can judge the nature of the relation (positive? negative? linear? ...).

Scatterplots can be enhanced with empirical or fitted **functions**. This projects the author's interepretation of the scatterplot on the viewer; the right graph in Figure 4.26 is an example. The presence of the fits makes it difficult for the viewer to imagine other patterns.

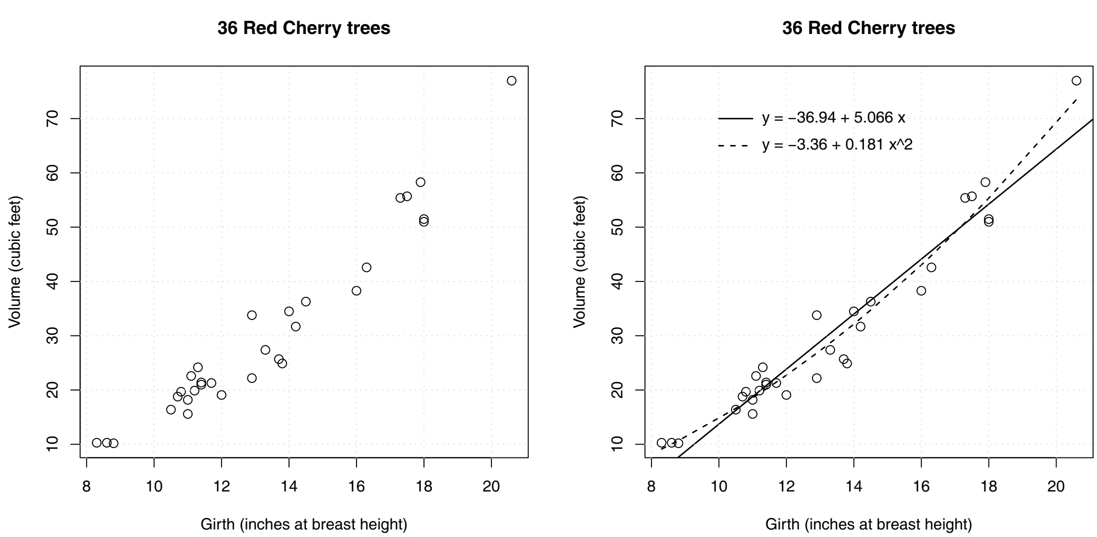


Figure 4.26: Example scatterplot,without (left) and with (right) fitted functions

Scatterplots with multiple series

If several data series have the same two variables, it is possible to show them all in the same scatterplot, as long as the points and/or functional lines are clearly differentiated. This allows direct comparison of related series.

Figure 4.27 shows how symbols (left) and colour (right) can be used to differentiate series in a plot, here four different sub-samples for which variograms were computed. Notice that the colours of lines for the functions can match the symbols for the observations on which they are based, but in the monochrome plot of the function the legend must show both the point and line symbols.

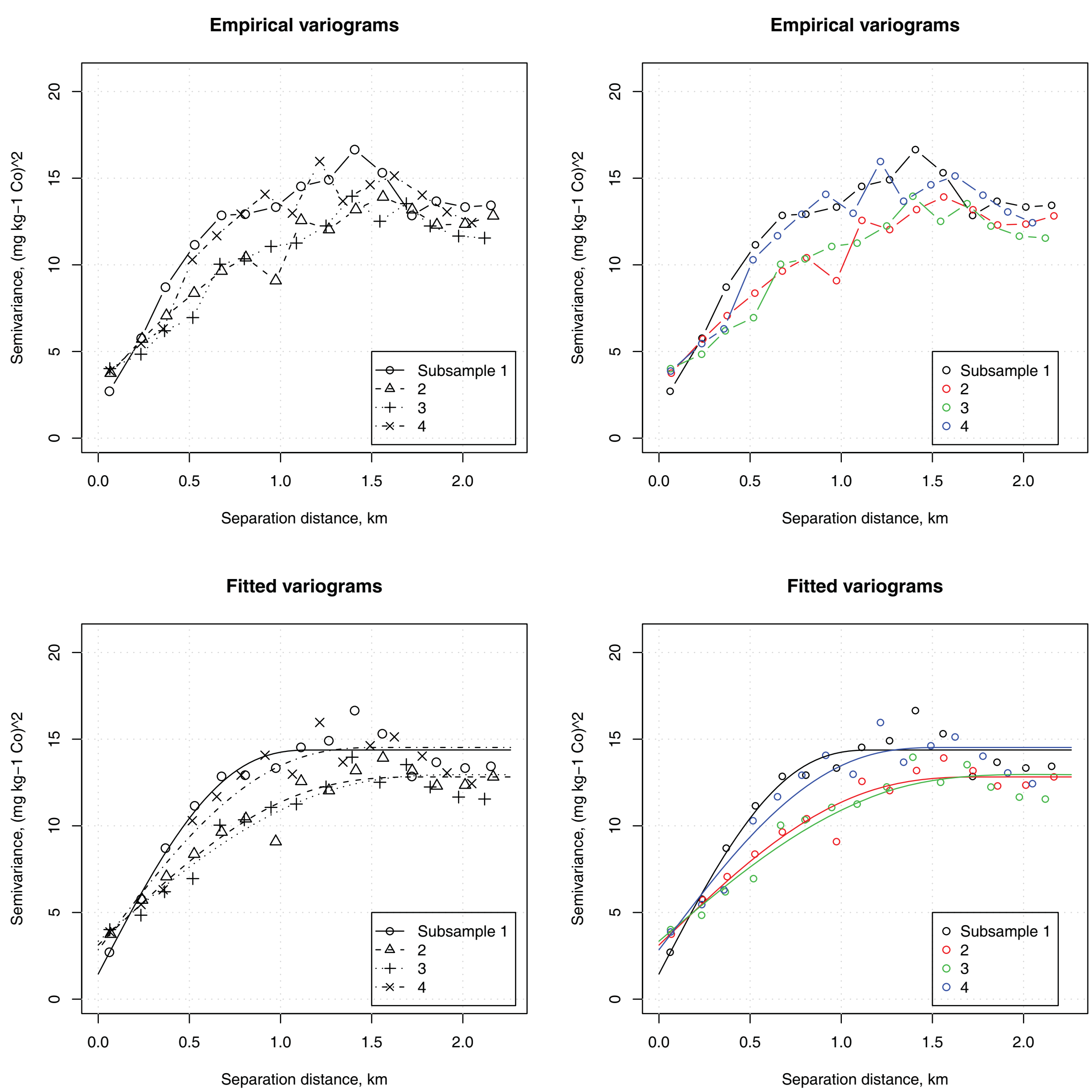


Figure 4.27: Variograms of Co concentration in Jura surface soils, four random samples; empirical (top) and fitted (bottom); in black-and-white (left) and colour (right)

Postplots

A postplot shows the **location** of an observation in **geographic space** (as shown by the two coördinate axes) with the of some **feature-space** variable shown by the size. For continuous variables, the value is generally shown by symbol size; for classified variables by a symbol showing the class membership.

Figure 4.28 is an example continuous postpolot. Note the difference in visual perception when the symbol radius is proportional to variable (left) or to its square root (right). This illustrates that the graph designer can easily (de-)emphasize contrasts.

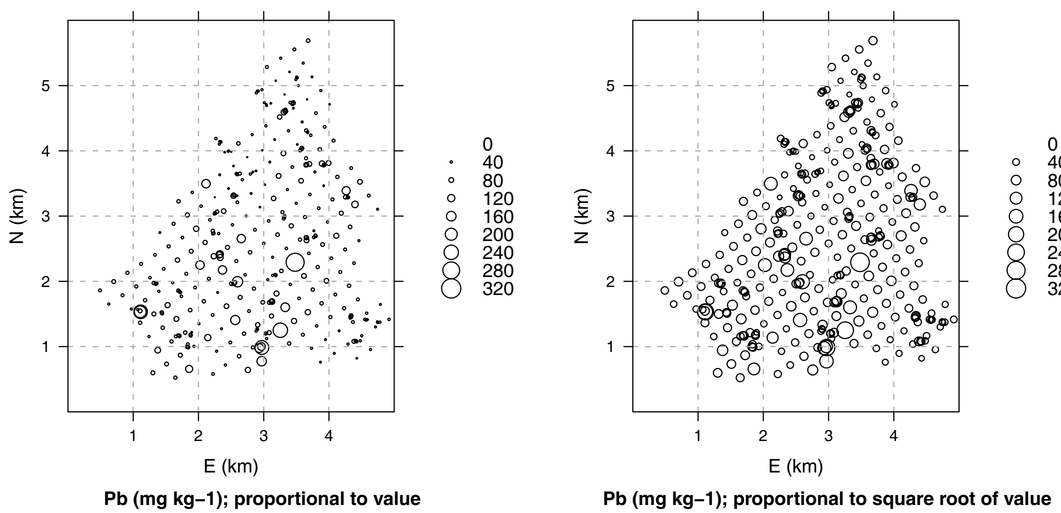


Figure 4.28: Example postplots

Related graphs

In a series of graphs that are logically related, they should be presented so that the viewer getws the correct impression of the similarities and differences. The **same symbols and terms** should be used in all of them. In general they should have the same **scale** to allow direct comparison, although this may not be possible if the magnitudes differ greatly (e.g. time-series of precipitation by month).

An example is Figure 4.29, which compares the same thing (here, a kriging prediction variance) computed by two different methods (here, ordinary and block kriging). Notice that the scales are the same, and the colour ramp is also the same, so that the reader gets a correct visual impression of the difference (here, that block kriging substantially reduced the prediction variance).

This example also shows the use of a colour ramp[18] that also prints well in monochrome; thus the PDF can be viewed on-screen in colour or printed in monochrome.

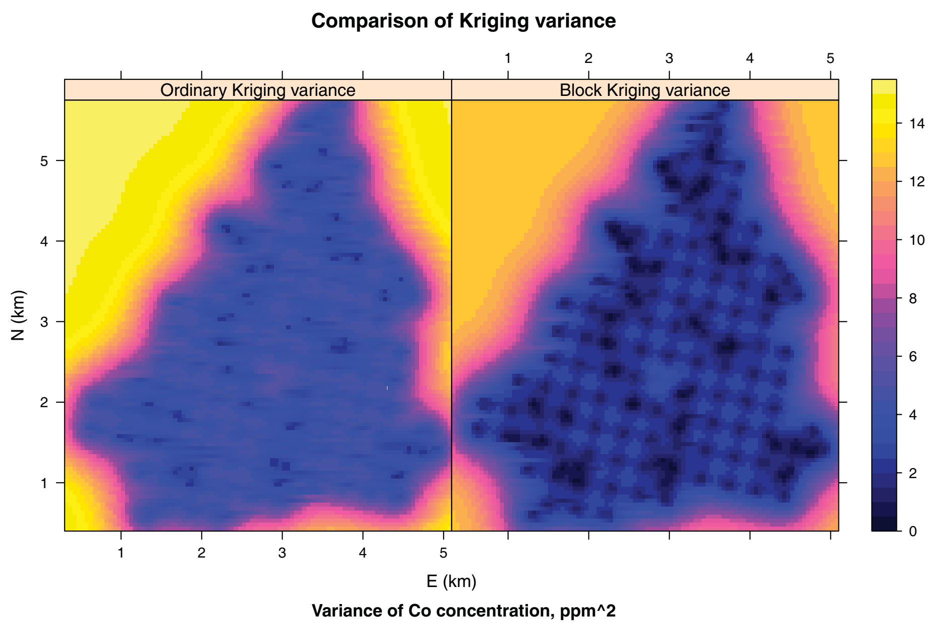


Figure 4.29: Example of paired graphs with the same visual scale

Choosing the correct graph type

The graph form must match the logic of the data being presented. For example, a scatterplot naturally relates two independent variables; a boxplot or histogram shows the distribution of a single variable.

Within the graph, there must be logic in the elements presented. For example, points in a graph should only be connected if there is a logical connection between them. Consider Figure 4.30, which shows a single variable: the ages of a group of 34 ITC students[19].

Q90 : *Which of these four graphs does not show a univariate distribu-*

[18] `bpy.colors` in the `sp` package of the R environment
[19] AES Block 2, 2009

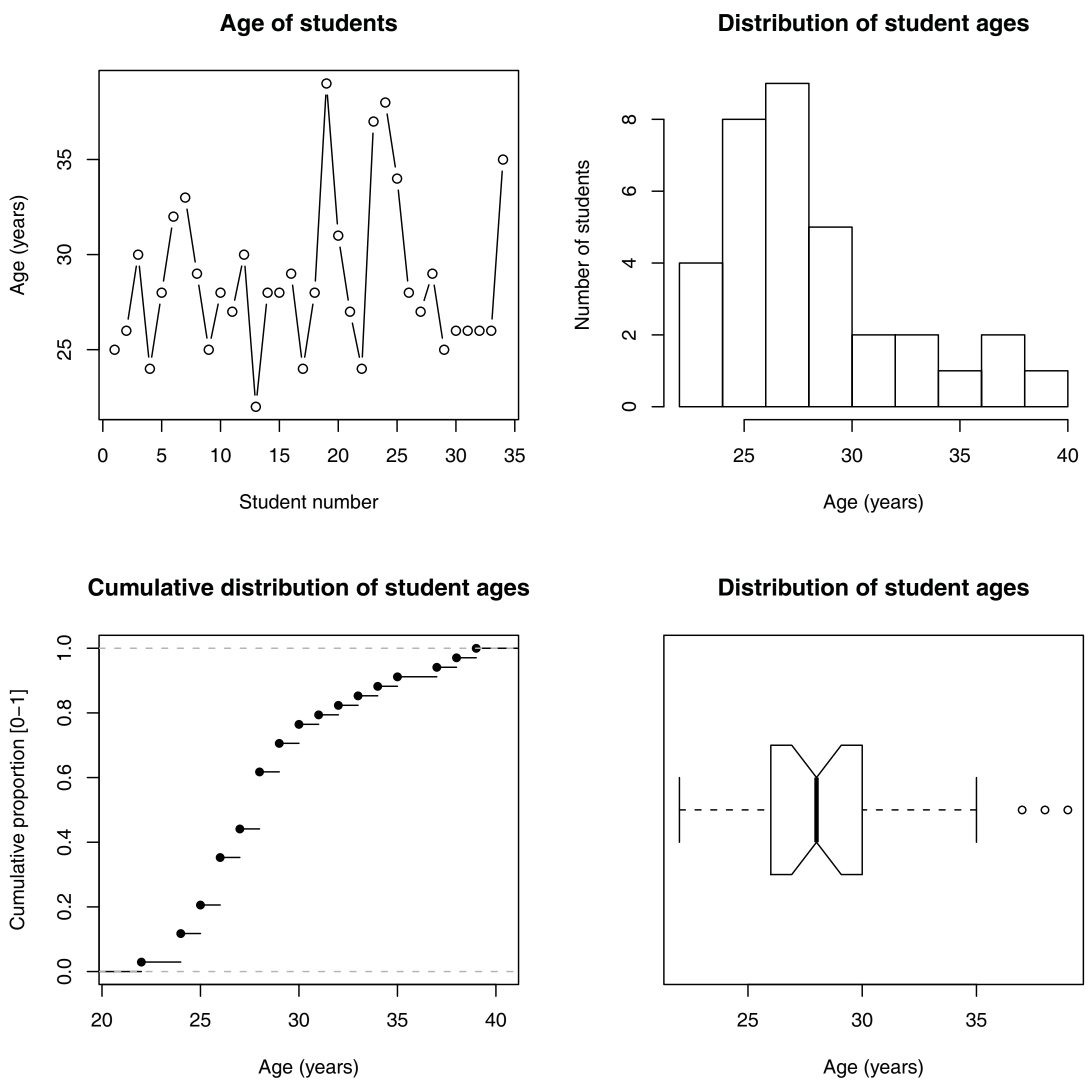


Figure 4.30: Distribution of student ages: (a) connected line graph; (b) histogram; (c) empirical cumulative distribution function; (d) boxplot

tion? *Jump to A90*
•

Figure 4.31 shows four ways to present a time series, here of well levels over time.

Q91 : *What do the two axes represent? Is their a logical connection between adjacent points? Which of the four graphs shows this? Which of these graphs best shows the fluctuation in water table over time?* *Jump to A91* •

Graphs can be manipulated to exaggerate or minimize features. Compare the three versions of the time series in

---

Q92 : *What is the visual effect of varying the aspect ratio between the ground water level (y-axis) and date (x-axis)?* *Jump to A92* •

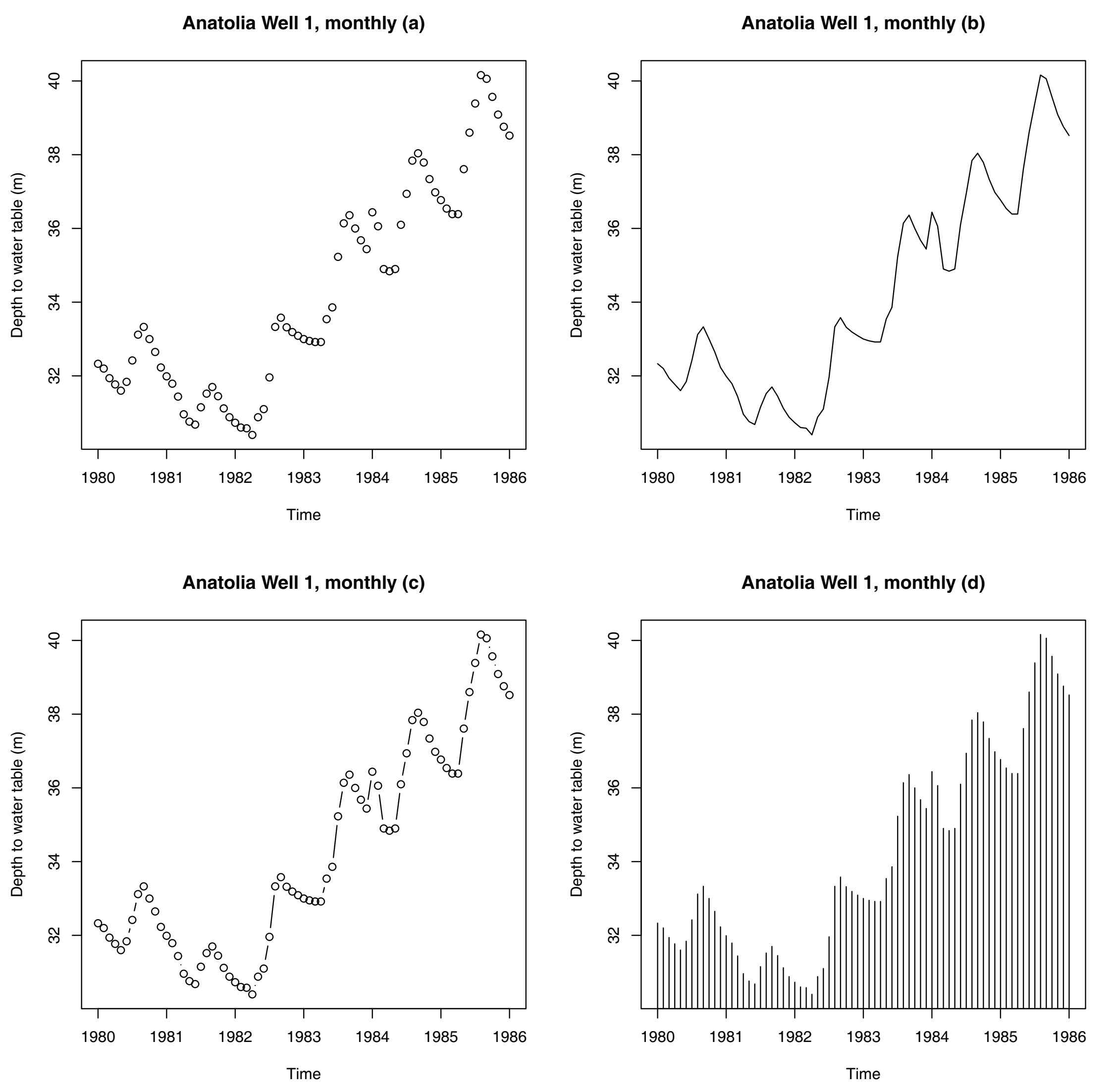


Figure 4.31: Four ways to show a time series: (a) points only; (b) line only; (c) points and line; (d) bars

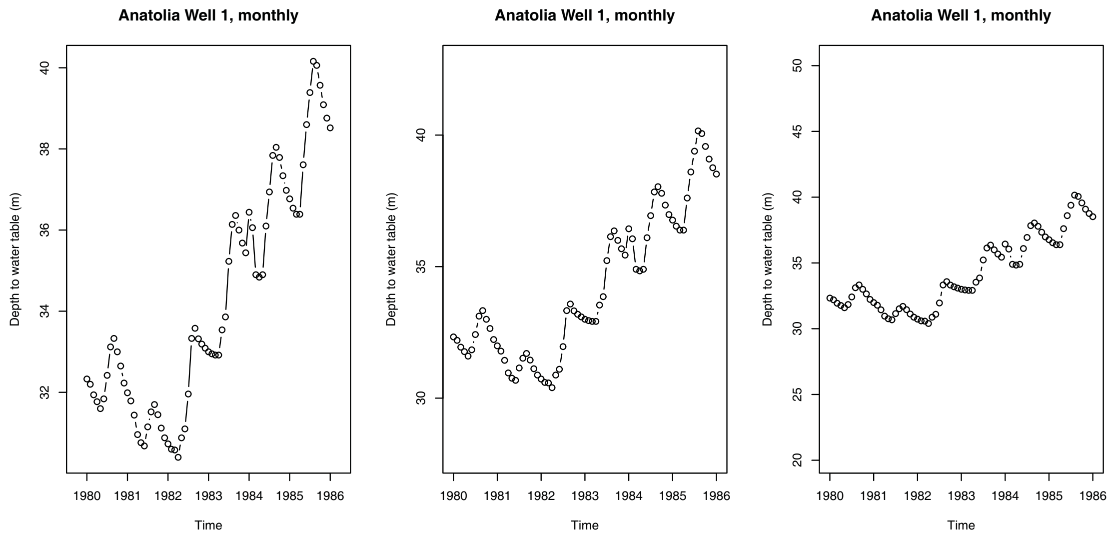


Figure 4.32: The same time series, three aspect ratios

4.2.3 Illustrations

Line drawings are original artwork that illustrate some aparatus (e.g. schematic diagrams) or procedure (e.g. flow diagrams). They are usually monochrome (either black-white or with some grey-scale shading) but may have colour elements if these add to the reader's understanding.

As with all figures, line drawings should be **legible** and as **simple** as possible, while showing the required information. They must be labelled, captioned, and referred to in the text.

We now discuss some kinds of illustrations.

Flow diagrams

A flow diagram shows how information, materials, procedures or concepts "flow" (either in reality or concept). They are often used for logical frameworks. Figure 4.33 shows a typical flow diagram.

It may be necessary to split a complex figure into parts, with labelled **connectors** between them.

Schematic diagrams

Diagrams show some real object, such as an instrument, sensor, landscape, or geomorphic feature, as a line drawing which is structured to

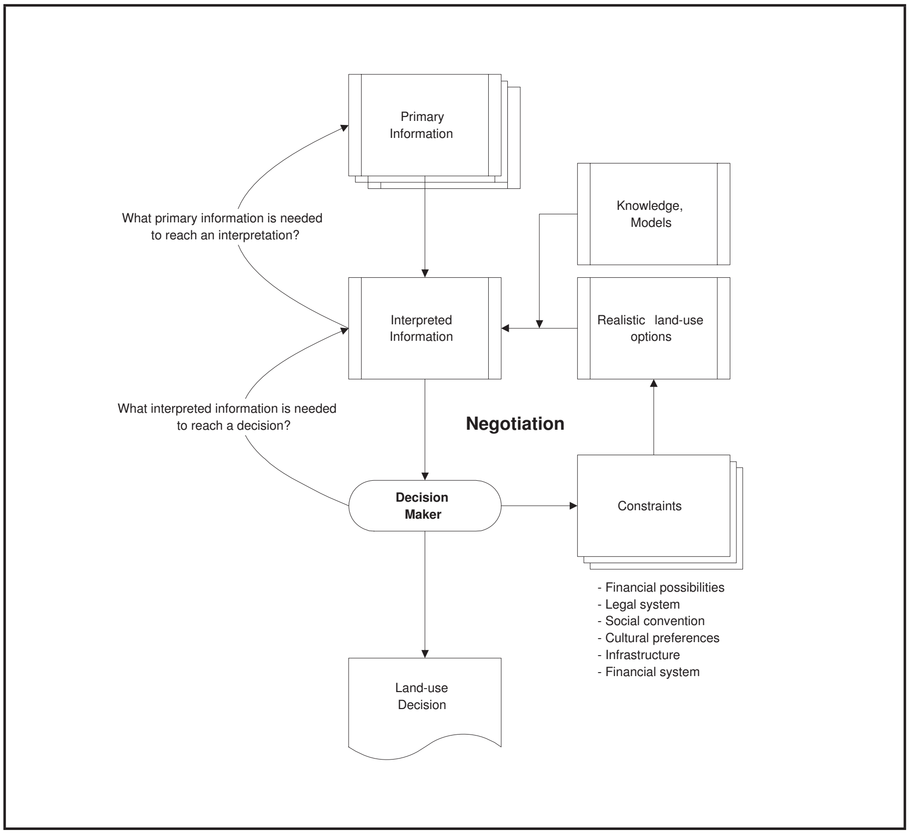


Figure 4.33: Example flow diagram; Source: Bacic [3, Fig. 1]

bring out the main features clearly to the reader. Figure 4.1, already used as an example of a floating figure, shows a typical schematic diagram of an instrument.

Sequences (animation in stills)

Animations can be embedded in digital documents, and used to show the **dynamic** behaviour of a system, impossible to show statically. An example is the animation of how the MODIS sensor covers the earth, both its track and its side-to-side scan[20].

Of course, it is impossible to show a full animation in a printed document; however, several frames from the animation can be shown in sequence (Figure 4.34).

[20] http://terra.nasa.gov/About/MODIS/modis_swath.php, accessed 28-Nov-2009

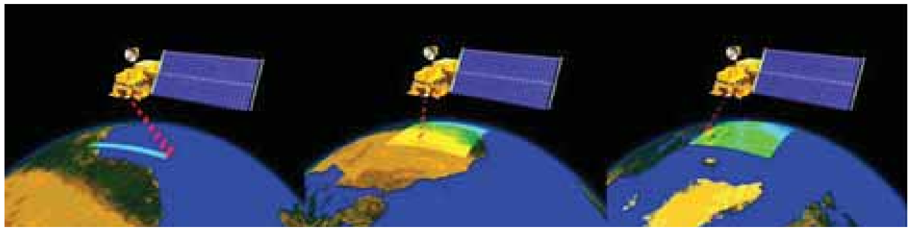


Figure 4.34: Example stills from an animation; Source: [25]

4.2.4 Maps

Key points

1. Maps are a special class of figure that present geographical information.
2. Maps must include enough information to determine location, scale and orientation.
3. Most maps must be georeferenced, at least implicitly; an exception are topological maps.
4. Colour in a map must not be confusing and, if possible, should be connotative.

Geographic information is naturally presented as **maps**: figures that represent some part of the Earth's surface schematically, with sufficient geographic reference for readers to identify the locations shown and, if desired visit them.

Note: There are also maps of other celestial bodies (the Moon, Mars) and of non-georefernced spaces such as the inside of buildings.

> "Not only is it easy to lie with maps, it's essential. To portray meaningful relationships for a complex, three-dimensional world on a flat sheet of paper or a video screen, a map must distort reality." – Monmonier [23, p. 1]

Cartographic design is a specialised subject which these notes can not cover in detail; refer to standard cartography text such as Robinson *et al.* [29] or Slocum *et al.* [35]. Here we only deal with some graphical issues.

Map types

Maps are made for a particular reason, each map has a specific purpose. There are three main map types, each with a design related to its purpose:

1. **Topographic** maps with an accurate and complete visualisation of the area as whole.

   Note that accuracy and completeness are related to the map scale; there may be deliberate distortions to be able to show the required information.

2. **Thematic** map showing distribution, kind and/or quantity of attributes, often with a (partial) topographic background for georeference.

   These maps must show the relations between the different attributes or their distribution. Thematic maps can be classified according to how this visualisation is effected:

   2.1. **dot maps** shows the geographic distribution of a phenomenon as "points" (the symbol of course will have a certain size);

   2.2. **choropleth maps** use the visual variable "grey value" (or a colour ramp) for the visualisation of relative quantities, interval or ordinal data related to predefined areas. These are often administrative areas but could be soil map units, where the visual variable is related to the relative concentration of some soil constituent in the unit;

   2.3. **chorochromatic maps** are used to show nominal data: distribution, kind and location of a phenomenon related to areas, for example kind of administrative unit (city, province, autonomous region . . .) where there is no natural ordering;

   2.4. **proportional point symbol** maps are used to show absolute quantitative (ratio) data related to points or areas, for example concentration of some chemical in well water, at each well location;

   2.5. **diagram maps** make use of (pie- or bar) charts as symbols instead of colours of patterns. For example, the breakdown of population in each administrative unit by age group.

3. **Charts** for (aero)nautical navigation; the main information is the route network and navigational aids.

There is another group of "maps" which do not preserve scale or all details. Instead their purpose is to give the user a specific kind of information, e.g. the route from 'A' to 'B'. This map type is referred to as a **topological** map, and is more illustrative than showing correct geographical information. Perhaps the most famous example is London

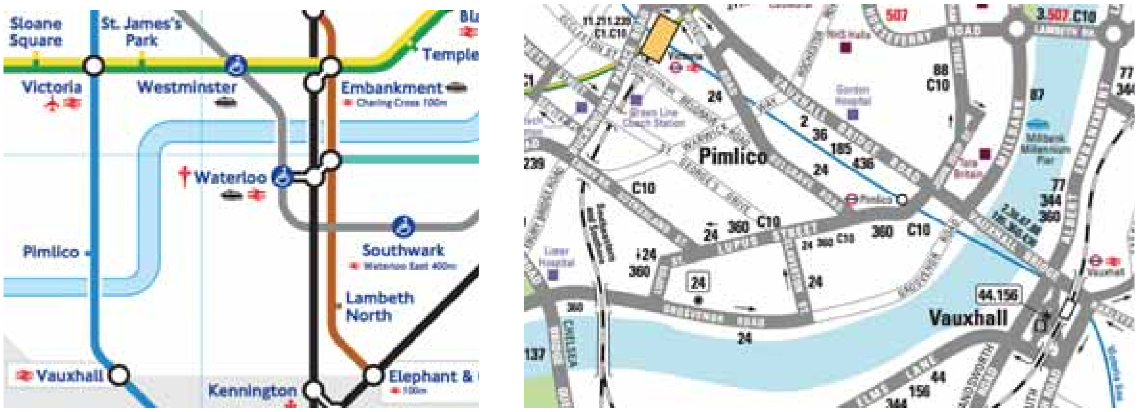


Figure 4.35: Topological vs. topographic maps; Source: [38]

Underground's "Tube Map" (Figure 4.35, left).

Q93 : *Compare the two maps in Figure 4.35. Both give information on transport in London. Which is topological and which is geometrically-correct? Why would a user prefer one or the other of these examples, for the purpose of navigating in London?* *Jump to A93* •

Scale and orientation

A map must contain sufficient information for the reader to be able to orient the map (e.g. to North) and approximately measure distances; otherwise the "map" is merely a sketch.

> **Note:** Note that the sketch may convey useful information; as in the London Underground line "map", as explained just above (§4.2.4).

One way to orient the user is with an overprinted **grid** (metric coordinates) or **graticule** (geographic coordinates), with the coordinates printed in the map margin, along with a description of the **Coordinate Reference System** (CRS), i.e. datum, projection, and coordinate system [15]. Figure 4.36(a) is an example of this method of orientation and scale. The map is overprinted with the UK Ordnance Survey 1 km grid (in blue), so the reader knows (1) the vertical lines are oriented to grid north; (2) the blue squares are 1x1 km.

Another way is with a **North** arrow or **compass rose** for orientation and a **visual scale** for distance. Note that the nominal map scale (e.g. 1:10 000) will in general not be preserved in a digital document or non-precision

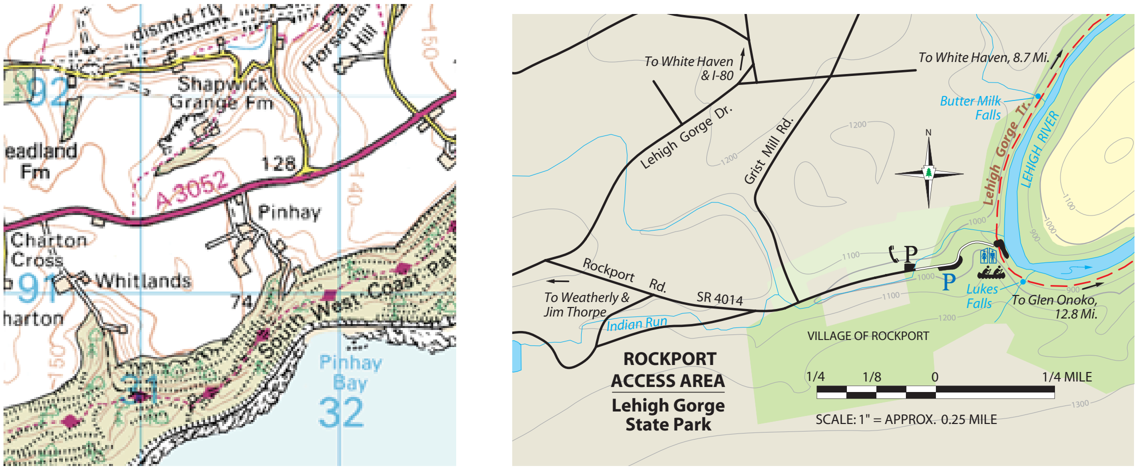


Figure 4.36: (a): Ordnance Survey (UK); portion of Explorer (1:25 000) series; Lyme Regis & Bridport sheet; (b): Lehigh Valley State Park (PA), Rockport access; Source: [27]

paper printing, so it is incorrect to state the scale as printed. The scale of the source map, however, may be reported. Figure 4.36(b) is an example of this method of orientation and scale.

Georeferencing

The reader should be able locate the area on the Earth's surface. This can be:

- by means of absolute coordinates in a specified CRS;
- with a hierarchical location map;
- with overprinted cultural features (e.g. roads);
- with overprinted natural features (e.g. contour lines, streams).

Figure 4.36(a) is an example of the use of absolute metric coordinates. The map is overprinted with the UK Ordnance Survey 1 km grid; all that is needed for unambiguous georeference is the name of the system and, if applicable, the zone. In this case the system is the National Grid, which is well-documented [e.g. 26]. The same figure also shows both natural and cultural features, thus allowing readers to identify the locatation.

Map contents

A map is in effect an illustration, which is to communicate with the user. Therefore the same rule holds (§4.3.3): the content has to be to the main focus of the map. The map should not have any visual “noise”, i.e. useless information that only distracts the map user from the main message.

When making a map, always ask: “what is the minimum information required to understand the content, to know where it all takes place and to understand the theme?”. As with all illustrations, the map must be readable: every included element or object should be clearly distinguishable. Design of the map for this reason always has to be related to the way of publication. A monitor is of lower resolution, so offers less details to be included in the map. Screen however has the advantage of quickly changing the view scale, offering possibilities to zoom-in and so showing a smaller area in more detail. The on-line London Transport maps are excellent examples of this[21]. Paper can hold detailed information (depending on the printer) but has no interactive options. If for a certain area more detail has to be shown, an inset map has to be included showing the area in a larger scale.

Use of colour in map design

The use of colour in graphics was discussed in §4.1.6; here we discuss some issues unique to maps.

Contrasting vs. similar colours These were discussed in §4.1.6. In the context of map design, contrasting colours are those that are easy to differentiate at a glance. They should be used to distinguish different map components. **Similar** colours are close to each other in colour space; they differ slightly in hue, saturation or value. They should be used show gradations of an attribute or sub-classes.

Figure 4.37 shows a typical but incorrect use of contrasting colours in a class-area map. ArcGIS has assigned colours seemingly at random to a map of soil types, with no regard to the properties or similarity of the different classes. In this map, several soil types with quite similar properties [41] have strongly contrasting colours, visually implying a large difference, when in fact there is very little. Of course, ArcGIS can not be expected to know which soils are similar; the designer must choose colours for this purpose. One approach is to add a field to the data table with the appropriate colour.

> For example, legend categories “MaA” (shown in brown) and “MaB” (shown in dark green) are the same soil type (Marlboro loamy sand), only differing in their (adjacent) slope class (0-3% vs. 3-8%). These soils have very similar properties and management, yet the colours imply that they are very different.

[21] http://tfl.gov.uk

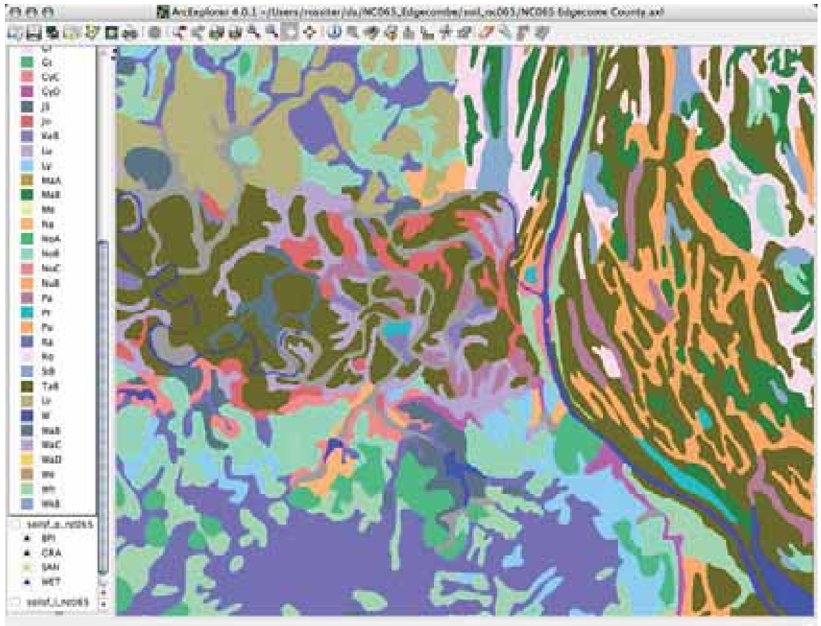


Figure 4.37: Soil map units, Old Sparta, Edgecombe County, NC (USA); Source: [41, 42]

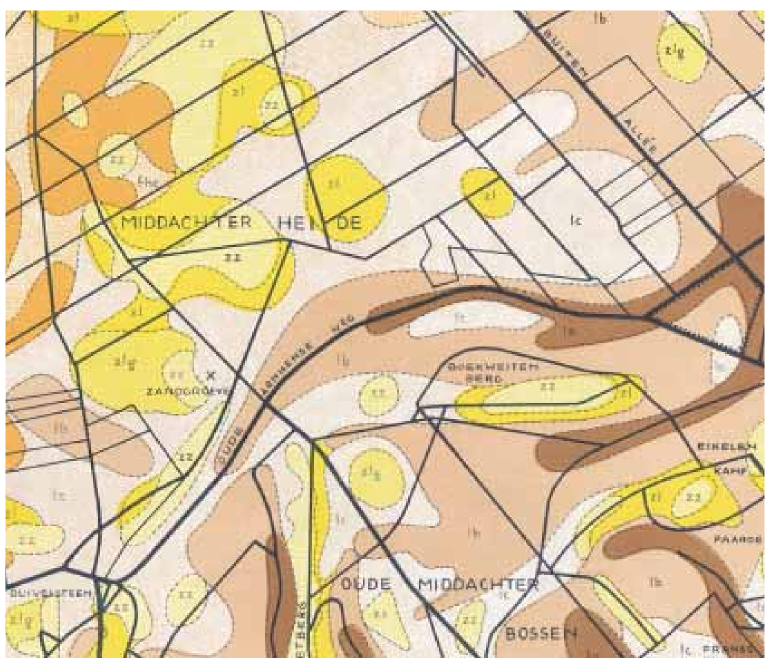


Figure 4.38: Soil map units, Middachten estate, Dieren (NL); Source: [43]

Connotative colours Colours may be connotative, i.e. suggesting something about the class from the visual perception. These can be conventions established over the years. An example from soil mapping is purple for organic soils, blue for clayey soils, browns for loamy soils, and yellows for sandy soils.

Figure 4.38 is an example where colours have been used to show soil properties controlling land use. Bright yellows show coarse loamy sands, lighter shades of yellow show progressively coarser soils (more coarse sand, less fine sand and silt), dark browns show the wettest soils where water collects in dales, pinkish brown shows the most fertile soils, with a high proportion of wind-blown silt and very find sand[22] [43]. For all the colours, a lower value (darker) indicates higher water-holding capacity of the soil. Thus the map helps the user visualize land use.

Topographic maps have well-established uses of connotative colours. For example Figure 4.36(b) shows the typical use of blue for water, green for forest and black for cultural features.

[22] "löß", English "loess"

4.2.5 Images

Key points

1. Images are **bitmapped graphics**, they should only be used when an actual image is needed, not as a rasterized drawing.
2. Images that can be presented are screendumps, photographs, scans of paper maps, and remote sensing imagery.
3. **Screendumps** illustrate specific processing steps, they should not be used for program output.
4. **Photographs** should help the reader visualize a complex object that can not be adequately described in words.
5. Pay careful attention to **resolution**, which depends on the type and capabilities of the output device.

This subsection discusses several kinds of images (bitmaps) that may be included in the thesis.

Screendumps

Illustrating a specific result of actions shown on screen as a result of settings/actions with a software are often difficult to reproduce and have to be captured at the spot at that specific moment. A so-called **screendump** gives the solution for capturing the temporary image from screen, resulting in a raster file. Figure 4.14 is an example: the author wants to show how ArcMap can be used to prepare multi-part colour ramps.

A screendump should *not* be used for saving a map or figure that is created and which could be saved from within the application.

A screendump is always low in resolution compared to figures prepared by an application. The resolution is controlled by the monitor, which is typcially between 72 and 96 dpi. For print a higher resolution image is required, somewhere near 300 dpi for a laser print and 600 dpi for an offset print. So, how to get the best possible result in the printed document?

1. Decide **which part** of the view on screen is required for the illustration in the thesis;
2. **Maximise** the view, spread it over the total space your monitor offers; this maximizes the resolution;
3. Use a special "**screen-grabbing**" software[23], save the image;

[23] e.g. freeware `MWSnap`; `Grab` for Mac OS X

4. **Edit** the image to make it ready for inclusion in the thesis (convert from RGB to CMY, make annotations, add text etc.). When saving, do not compress the image, keep maximum quality (which is already much lower than the printer).

Photographs

Photographs help the reader visualize a complex object that can not be adequately described in words, for example a landscape or an apparatus. They can also help the reader appreciate how a procedure was carried out. They should be used when "a picture is worth a thousand words", i.e. when visualization can explain more than text description.

These should show the object of interest as clearly as possible, without distractions. The subject mentioned in the text should be obvious. For example:

> "Figure 4.39 shows the typical landscape elements at a retreating glacial margin: the glacial front in a U-shaped valley, terminal and lateral moraines, and a pro-glacial lake."

Figure 4.39: Glacial geomorphic features, Kviárjökull, Iceland

> "Dune stabilization begins with small patches of grass (Figure 4.40), which stabilization in turns promote more grass in a positive feedback loop."

A photo can include its own annotation:

> "Where the the mid-Atlantic ridge between tectonic plates reaches the surface in the Rekyanes peninsula, sulfur accumulates in

Figure 4.40: Small dunes, Kootwijkerzand, Barneveld, NL

surface soils due to geothermal alteration (Figure 4.41)."

Figure 4.41: Sulfur accumulation on soil surface, Rekyanes, Iceland

Sometimes both an **overview** and **detail** photo should be included side-by-side:

> "Weathering can proceeed rapidly even in cool climates. Figure 4.42 shows a memorial stone engraved in 1921. The detail (right) shows clearly the pitting of the granite, with the resistant quartz grains standing out prominantly, and the accumulation of moss on the surface."

Figure 4.42: Weathering of granite memorial stone, Judithoeve, Losser, NL

If the size of an object in the photograph is important, it should include some sort of **scale**: either an object of obvious size, a scale placed into the scene as the photograph was taken, or a scale drawn on the finished photograph. A typical example is the meter stick in the photo of a soil profile, which also often has the horizon labels, as shown in Figure 4.43.

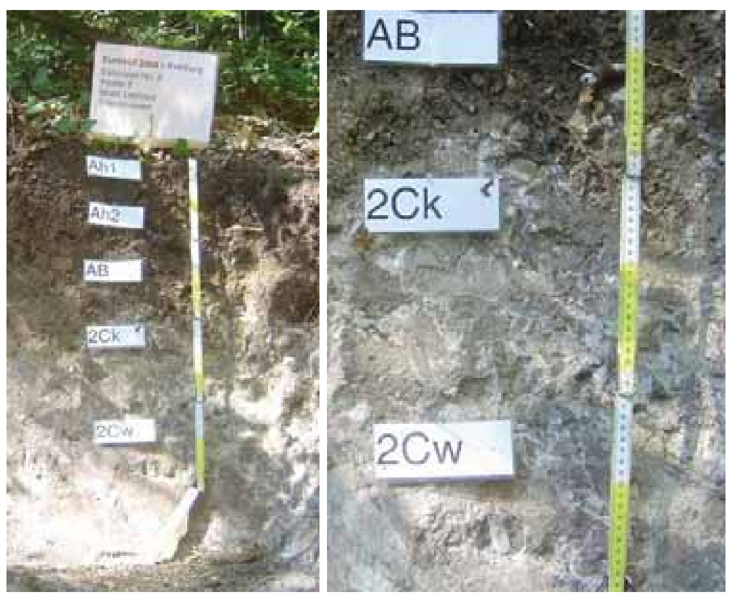


Figure 4.43: Soil profile (left); detail of AB and 2Ck (right); Kaiserstuhl, Baden-Württemberg (D)

Scans of paper maps or aerial photographs

Images printed on paper, especially maps and airphotos, are often used in an information system and may be presented as a thesis illustration. These must be made digital by scanning[24]. Section 4.1.3 explains how to choose a scan resolution to match the printed or screen output.

Remote sensing imagery

Many ITC theses include remote sensing imagery, often optical satellite images but also thermal, gamma-ray, radar etc. By their nature these are rasters; they were captured, manipulated, and delivered as rasters.

The main issue here is the resolution of the imagery vs. the output device. These do not have to match: it may be desireable to show the pixellated nature of the image, in which case the print resolution will exceed the image resolution (the image is expanded).

4.2.6 Formats

The vector or raster formats that can be included in a thesis depend on the document preparation program.

- **Microsoft Word**: wmf (Windows Metafile), jpeg, gif, png, bitmap, pict, tiff, eps. Among the (many) disadvantages of this program is that it embeds the image in the file itself and does not maintain a link to the actual file, so that if the original is changed the file needs to be included again.
- **PDFLaTeX**: pdf, jpg, png, eps, ps, tiff, pcx, pict, bmp. A package such as `graphics` or `graphicx` must be added to the document preamble. Among its (many) advantages, LaTeX incorporates the graphics file into the final document (PDF) via commands such as `\includegraphics`, therefore if the graphic is changed, simply re-compiling the LaTeX document will automatically include the updated version of the graphic.
- **InDesign** (a desktop publishing software): Ai, eps, pdf, tiff, psd, dcs, bmp, jpeg, gif.

Resolution

For high-resolution output, all included files must also have a high resolution. All vector formats (scalable graphics) are suitable for high-resolution printers; of course the original graphic must be good. Text

[24] For use in a GIS they must also be georeferenced

should be included in scalable graphics as text, i.e. a font (not as a drawing), since scalable fonts are also vector graphic. Raster files should be high-resolution (large dpi) and uncompresed. Many compression algorithms reduce the quality of the raster file, even when keeping high resolution. A special problem is text in a raster file: even with a proper font it is not scalable and compression can only reduce its sharpness.

! → **Use vector formats whenever possible.** Only if there is no other option (a screendump, photograph, satellite image, ...) use a raster format.

Colour

If your document (or certain pages) will be printed on a monochrome ("black and white") printer, convert colour to greyscale. If the document or parts will be printed in colour, make sure that the colourspace is according the printers requirements; always ask before delivering the final files.

For the ITC thesis, it is best to decide which pages should be printed in colour and collect these on color plates. They can be interspersed with text pages, but must be on their own pages. This can be ensured with appropriate pagebreaks[25].

4.2.7 Answers

A*86* : *There is no **caption** to indicate what the graph represents, and there are no **axis labels**, with **units of measurement**, to show what the numbers mean.* *Return to Q86* •

A*87* : *Now there is a caption and axis labels with units of measurement. A grid has been added to help the viewer estimate the vvalue of each observation. However, there is no way to tell which species (class) is represented by each observation.* *Return to Q87* •

A*88* : *Now each observation is coded with a symbol indicating its species, but there is no **legend** to link symbol with species name.* *Return to Q88* •

A*89* : *These plots are purely **descriptive**; they show the relation but with no interpretation. The analyst may want to add lines for means and medians, either overall or per group, and lines or curves of the presumed positive relation between the variables. These are **interpretations**.* *Return to Q89* •

---

[25] e.g. `\newpage` or `\cleardoublepage` in LaTeX

A90 : Graph (a) does not show the distribution; it connects "adjacent" students, but the ages were not collected in any sequence, so such a connection has no meaning. Graphs (b), (c), and (d) are different ways to visualize the distribution. *Return to Q90*
•

A91 : The x-axis (abscissa) represents time, the year from 1980–1986. The y-axis (ordinate) represents the depth to ground water at a particular date, in meters from the surface. Since time is continuous, there is a logical connection between adjacent points; they are also adjacent in time. Graphs (b) and (c) show this. Graph (a) is difficult to read: no connection between points and some have big gaps on the y-axis. Graph (d) implies the connection, although it doesn't show it. To my eyes, graph (c) is best, because it shows the fluctuations but also the exact times when a reading was made. *Return to Q91* •

A92 : A larger aspect ratio emphasizes the fluctuations; a lower ratio emphasizes the long-term trend. *Return to Q92*
•

A93 : The toplogical map is on the left (the "Tube map"); on the right is a geometrically-correct transport map. The toplogical map is easier to read for finding Tube stations, especially interchange stations between Tube lines (open circles) and with the rail network and ferries (small symbols). However, the scale is inconsistent and grossly distored; further no streets are shown, so the map is useless for walking. The other map shows the Tube and its stations but the user has to pick them out from much other information. But this other information is very useful to the walker (true distances), bus patron (bus routes), and tourist (main points of interest, e.g. hospitals, shown in distinctive colours).

If the user's question is "How do I reach Lister Hospital[26]?" the geometrically-correct map is preferable. If the user's question is "How do I reach Pimlico station by Tube, from my present Tube station?" the topological map is preferable.

Return to Q93 •

[26] where anaesthetic was invented

4.3 Constructing effective graphics

Key points

1. A wide variety of **tools** is available to produce graphics (§4.3.1).
2. Direct output from analysis software (e.g. GIS) can be **enhanced** with a graphics tool (§4.3.2); however it must not be **distorted**.
3. **Flexible design** (§4.3.4) is the key to clever production (§4.3.5).
4. **ArcGIS** can produce reasonable maps, if a good production flow is followed (§4.3.6).

A wide variety of tools can be used to produce illustrations for your paper, each with strong and weak points. The best tool to use depends strongly on the kind of illustration to be produced. In some cases the tool is not relevant, since the research is based on the use of certain software, so that the output of these results must be presented in the thesis. Most software can generate an output file; in the rare exception a screendump bitmap must be captured; see §4.2.5 for how to achieve best quality in that case.

Figures output from the research software may not be optimal for illustrating the points to be made by the thesis. These can be manipulated (e.g. cropped, annotated, adding marginal information or symbols and text), which will require the use of another tool. Further, colour mode may need to be changed from RGB to CMYK.

4.3.1 Tools

The following is a partial list of some specialized tools for creating graphics. Note that not all are equally effective, either in their design principles or their production quality.

- **Tables**: Excel, Adobe Illustrator
- **Graphs**: Excel, R, Adobe Illustrator
- **Statistical graphs**: specialized statistics programs produce their own graphs; for example the R environment for statistical computing and graphics [30], for which many books are available [9, 21, 24, 32, e.g.].
- **Illustrations**

  - Vector: Inkscape, Powerpoint, Adobe Illustrator, Excel, Visio, Word
  - 3D: Sketch-up, Arc Scene / Globe
  - Animation: Arc Scene, Adobe Fireworks

- **Maps**: ArcMap, Adobe Illustrator, ILWIS

  ArcGis is not a real graphic software nor a dtp package. It has some options in the "Layout view" for adding title, legend, north-arrow and other loose bits and pieces and it has some options to add drop-shadows and backgrounds (§4.3.6). But it does not meet the high standards required for a professional product.

  Enhancements to the basic map output can be best made in a specialized graphics software. For raster images this can be Adobe PhotoShop, for paper output using pdf the best tool is Adobe Illustrator.

  Similarly, ILWIS can be used to make reasonable maps but not for professional publication.

- **Raster**: Paint, Gimp, Adobe Photoshop, Adobe Fireworks

4.3.2 Enhancement of raw output

As explained in the introduction of this section, many computer programs are perfect for performing analyses or calculations, but are less suitable to show the results in a good graphic presentation suitable for output. Graphs can be produced, tables generated and maps plotted, but none might meet the high graphic standards required for output. Mostly an enhancement is needed, the output file has to be opened or imported in a more graphically sophisticated tool where colour corrections, better positioning of elements and other choice of fonts and symbols (including lines) have to be done.

! → However, **never distort the ouput** in such a way that the reader is mislead. In extreme cases this is certainly fraud; but even inadvertantly the reader may get a wrong impression. Make sure that the enhancement does not obscure, and preferably highlights, the main features of the output.

4.3.3 The art of keeping it simple

The title of this paragraph does not only refer to the content (elements) of the illustration but also to the message to be passed on. Keep the message simple, use graphic elements that support the message. The purpose of a good illustration is to enhance (support) the text, not compete with it. The message conveyed by the illustration needs to be easily readable and immediately recognizable.

"Keep It Simple, Stupid!" (KISS) is a well-known engineer's maxim, which applies equally well to thesis text and graphics. "As simple as possible, but no simpler" might be a more accurate expression. As inn writing ("omit needless words"), also in graphics: omit needless elements. By simplifying, the message is emphasized.

To come to a good illustration you can follow the next steps as guide:

1. **The idea**

   Identify the need for a figure for a certain part of the text, do not illustrate because there is something available but because there is a need.

2. **What is to be communicated?**

   Decide on what the graphic should tell the reader. Ask yourself, "What do I want the reader to know after viewing this graphic?"

3. **The content**

   Specify the minimal required content of the illustration, based on the text where it should relate to or that what should be illustrated. Collect required external elements (like a satellite image, photo, tabular data . . .) that will make up the graphic.

4. **Sketch**

   Do not use the computer yet, just pencil and paper! Start by trying to figure out the simplest possible illustration for the point you want to make. Make a sketch on paper more or less to the correct scale. It's no problem if your drawing skills are poor, it even helps in forcing to keep things simple. When drawing on paper you will not focus on visualization aspects such as colour, fonts, and shadows. Instead, concentrate only on the content. Do not include elements that are redundant like cell-lines in a table. Human visual perception is well-suited to discover patterns. Lines often make it harder to read tables, because they erect visual borders (barriers) to smooth reading.

5. **Create**

   After sketching and outlining, the next step is the actual production of the figure in the appropriate software.

6. **Revise**

As with the rest of your document, the graphic must be revised several times. View it critically, the same way you read your document critically, and adjust the graphic until it is just as you want it, to communicate the message you intend.

Recall the question: "What do I want the reader to know after viewing this graphic?"; make sure it is answered as you wish.

4.3.4 Layout and design

This section presents some ideas for effective design.

1. **Design to scale, on basis of available space**

   Calculate the space you have available for the illustration, so you can create it in the correct scale. If an illustration has to be re scaled later when placing it in the article, this has great influence on the final line thickness, symbol and font size. The result will be either that all elements are too tiny and so unreadable or everything is too large and the illustration is too simple for the size. Size and content of the figure have to be related.

2. **Consistent design**

   Specify type, colour, lines, symbols and graphic elements. For good accessibility of the information all illustrations in the document (thesis) should be consistent in design. For this reason all graphic elements to be used in the illustrations need to be specified.

   - **Type**: Stick to one typeface, best for illustrations is a sans-serif such as Arial (compact and clear for print). Decide on different fonts to be used (combination of type, size and style), based on the chosen type (eg title, subtitle, names, numbers...).
   - **Colour**: Decide if colour is really required to make your point. Colour does not always make the information more accessible. If required, than make a decision on the colour palette to use. Colours can be related to the type of information (land use[27], temperature, ...), it can create a mood (sad or happy, ...) or it can be related to a certain theme of the thesis. Do not create an overkill of colours in your document, especially if it is used in graphs, flowcharts, cartoons, schematic diagrams or maps. Restrict yourself, less is better and clearer.
   - **Lines**: Variation in line styles (full, dashed, thickness) can create a good visual hierarchy. When all lines are the same, all is placed on visually the same level and makes it harder to read the illustration.

[27] blue = water, light green = pasture, dark green = forest, white = glacier ...

- **Symbols**: When using symbols in a diagram, graph or map make sure they are visually-distinct and if possible related somehow to what they represent. Do not use too many different symbols.

- **Graphic elements**: These are non-functional, but visual elements to enhance the attractiveness of an illustration. Examples are drop-shadows and shaded backgrounds. If these elements are used, they should be used throughout the document, and give the document a clearer cadence. Do not use all kind of elements on an arbitrary basis, it makes a messy unprofessional design ("Look what I can do!").

When using recurring base elements (like a map of a certain area) repeat the base legend.

3. **Figure and ground**

   An illustration has two fundamental layers of information, the main message and those elements that support the message. A map can have the base topography and the theme on top of it. The topography is required for positioning of the thematic info. The thematic info is the main info, the topography supports the theme. If these layers are not distinguished in the illustration the user has difficulty in getting the message clearly. Differentiation in the layers of information can be reached by the use of tints for the background (ground) info and full colours for the theme (figure). If the figure/-ground relation is not designed well, you create a situation where either the figure and the ground compete or the figure should be the ground and the ground should be the figure (the focus is reversed). Figure 4.44 shows poor (top) and good (bottom) separation between the figure (thematic information) and ground (supporting information).

   ☹ ☺

4. **Contrast**

   Contrast is the difference in colour and light between parts of an image or illustration. Good use of contrast sets the correct relation of figure to ground. Also, within a single layer, contrast between different elements greatly influences readability. For example, a green linear object on a blue background is hardly readable, whereas a yellow linear object on a green background gives a clear contrast and therefore very readable. So for elements that should be distinguished, use contrasting colours, i.e the colours that are opposite on the colour circle[28]. Brown & Feringa [4, §6] explains much more on colour contrast. Figure 4.45 shows poor (left) and

[28] http://www.worqx.com/

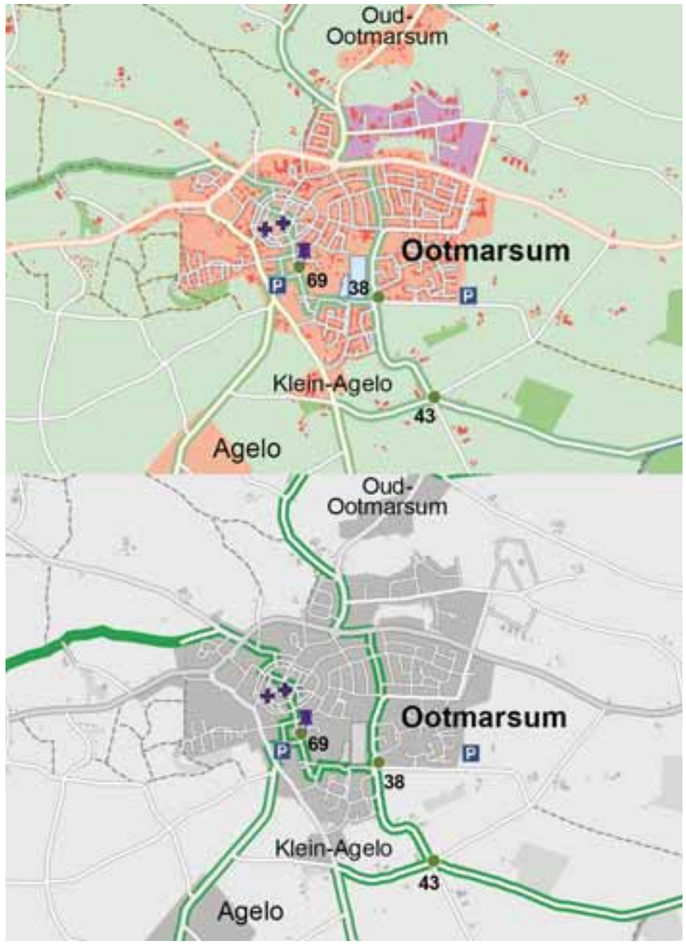


Figure 4.44: Figure and ground; poor (top) and good (bottom)

good (right) separation between the figure (thematic information) and ground (supporting information).

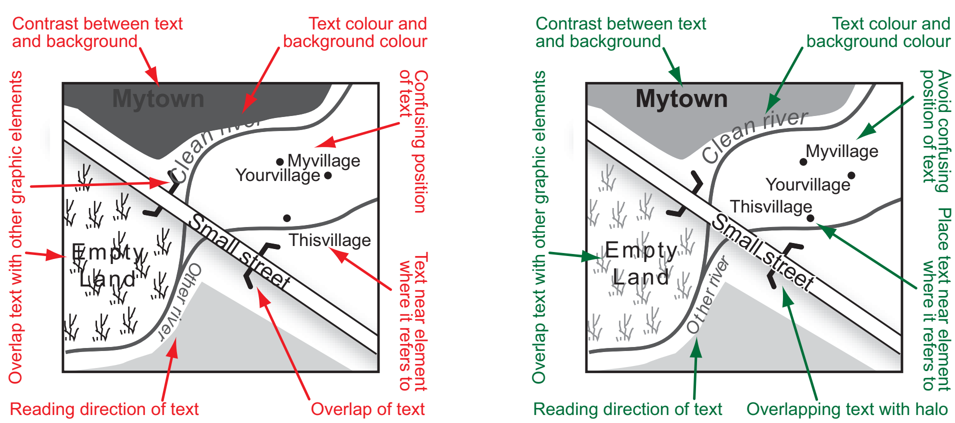


Figure 4.45: Contrast: poor (left) and good (right)

5. **Design aids**

A well thought-out organization of elements in an illustration enhances the readability. The human eye has to be guided to elements. Randomly distributed elements forces the user to study the illustration closely before understanding its message. There are several ways to organise the separate elements in an illustration (Figure 4.46):

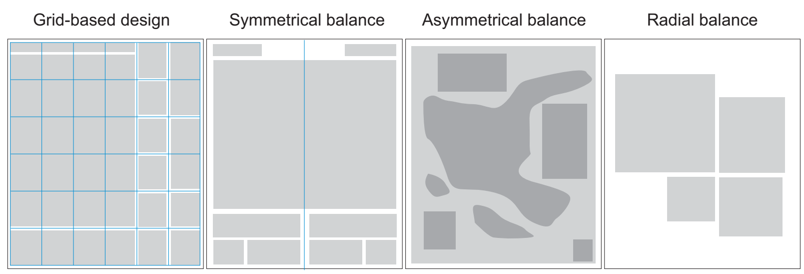


Figure 4.46: Types of balance in a graphical layout

- **Grid**: an invisible structure used to guide the placement of elements. They do not appear on the printed piece but their influence may be evident in the widths of column texts, the uniformity of space around elements, or the consistent placement of repeating elements. They are a series of guidelines that determine the margins of the piece, space between elements.
- **Symmetrical balance**: equal visual weight where a central axis divides the composition in the middle, horizontally or vertically, with the same distribution of elements on both sides (mirror reflection).
- **Radial balance**: has a central focal point in a generally square compositional format.
- **Asymmetrical balance**: off-center or created with an odd or mismatched number of disparate elements.

6. **Position of the illustration**

During the design stage it is hard to tell where in the document the figure will be placed, but if possible take into account:

- If an illustration contains an image or uses dark colours and tints (eg a satellite image with annotations) it should not be placed at the top of the page. The illustration is visually “heavy” and so should be placed at the bottom of the page.
- Place “lightweight” illustrations at the top of the page.

- Place an isolated image in the optical center of the page. The distance from the bottom to the illustration should be bigger than the distance from the top; optically this will look centred.

4.3.5 Construction

This section deals with the actual construction of the graphic, using one or more of the tools listed above (§4.3.1).

Flexible design is the key of clever production. An illustration is subject of continuous editing, it needs refinement and sometimes re positioning of elements. When making an illustration, choose for an application that offers the option to work in multiple layers. In 4.3.3 ("The art of keeping it simple") above the best procedures for design are outlined. In this same stage the overall structure of the illustration can also be set. First defining the overall layer structure helps in creating a flexible design.

Separate the main graphic elements as defined by your sketch design and place them in own layers, this makes it easier to hide or show individual groups or move them to better positions. Elements of an illustration can in this way be easily re-used. When all elements are on a single layer, re-ordering is difficult: it is hard to control how an element is stacked in relation to others.

4.3.6 Producing maps in ArcGIS

Producing a map in ArcGIS seems to be rather simple: import data, execute some queries, change some colours and line weights and in the "Layout view" add some text, add a legend and some more. Finally export the map to one of the available formats and that's it. This simplified procedure as described above is of course a caricature of the complex processes that lead to a map. Many maps that are published seem to work according that same caricature ignorant of some basic rules of map design.

Here is a suggested ArcGIS production flow. These steps are guidelines to reach an acceptable output; to produce a good map cartographic rules have to be applied. Such rules are beyond the scope of this text; please consult a proper cartography textbook [29, e.g.] (or better, study cartographic design).

1. Set the correct dimensions of the map in the "Page and Print Setup". Set it to the exact space reserved on the page, do not include surrounding white space.
2. In the "Layout view" make the "Data frame" fit with the "paper size" and make sure that the content fits in the data frame.

3. When creating a map for use on screen set the zoom scale to 100%, it is easier to have good control on the content of the map. For output on paper, which can show more detail, the zoom can be 250%.

4. Make sure that all details visible on screen are readable, this ensures that also the details will be readable in the final output. Set line weight, font size, colours etc. according to specifications. Especially for text make sure you use the correct fonts, do not decide this on basis of what you see on the monitor (this is especially true when you create a map for paper output).

5. Export the map in the appropriate format. ArcGIS can create different output formats: emf, eps, ai, pdf, svg, bmp jpeg, png, gif and tiff. The appropriate format depends on the output (or further workflow). For output to screen presentation most obvious formats are jpeg (if the map contains images) and gif (when making use of tints, lines and text), png also may be an option (this combines the two). For output on paper the best choice is tiff (24 bit true colour) if your map contains images and many transparencies, but pdf when working with tints, lines and text. Transparency settings in ArcGIS cause the layer that contains transparency and all lower layers to be rasterized. PDF can also show transparency, but unfortunately not the PDF version that is created by ArcGIS.

6. Check the output files for correctness (i.e. make sure the map has the right information, not just design!), after which they can be included in the final text.

4.4 Examples

In this section we've collected some examples of designs that can be greatly improved; the improved version is also shown, with comments on how it was produced.

4.4.1 Example 1: a contour map

☹

Figure 4.47 was produced within the R environment[29] [30]. Although it shows the required information, it has several problems:

- The title contains legend information (i.e. the contour interval);
- There is no legend for either the contours or the points (which are in fact monitoring wells);

[29] with the `contour` function; note that this graph could have been enhanced also from within R

- All elements are visually on the same level, i.e. same grey tone and line thickness, so it is difficult to focus on a sublevel or to read contours;
- Information is overlapping, obscuring each other, here contours, symbols, and values;
- Coordinate notation is not a cartographic standard, measurement unit not given (meters? feet?).

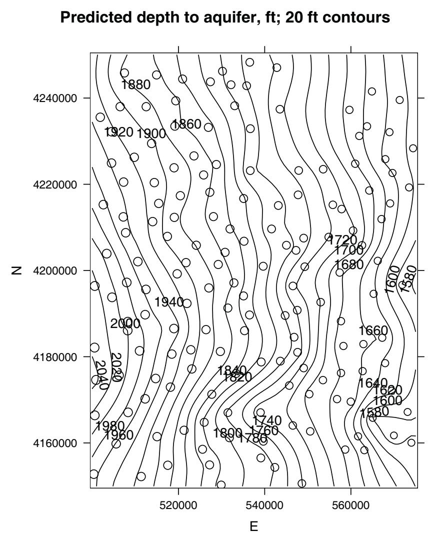


Figure 4.47: Example of scalable graphics – as output from the statistics program

Figure 4.48 is an improved version; this was also used as 4.2, above.. Note the following enhancements:

- North and East indicated at first coordinate, notation simplified and standardised;
- Contours are in two levels: index and interval;
- Contour interval is known (20 ft, given in the legend) so no need to label all contours;
- Index contours are 0.3 mm 100% black, interval contours 0.1 mm 70% black;
- Reading direction of the contour values is towards highest point;
- Contour values are placed on contour lines, which are masked out at these positions for improved readability;
- Contour values are placed at graphically “open” spaces, so that they do not interfere with other graphic elements;

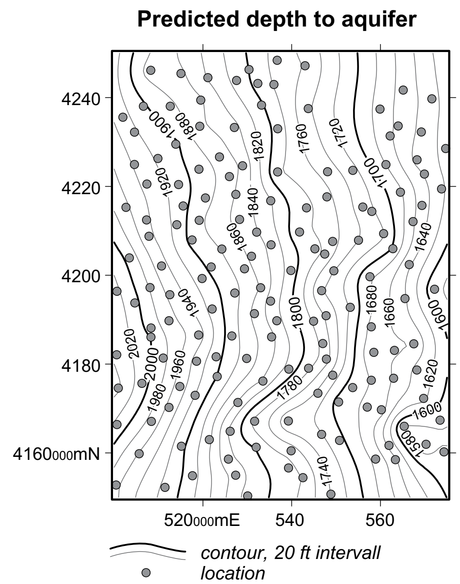


Figure 4.48: Example of scalable graphics – enhanced

- Location points (monitoring wells) are shown with 50% tint fill, for improved readability;
- All font sizes and line thicknesses are chosen for the 100% design for this document. If the illustration needs to be larger or smaller, design has to be adapted to the new size.

4.4.2 Example 2: Summary of a survey

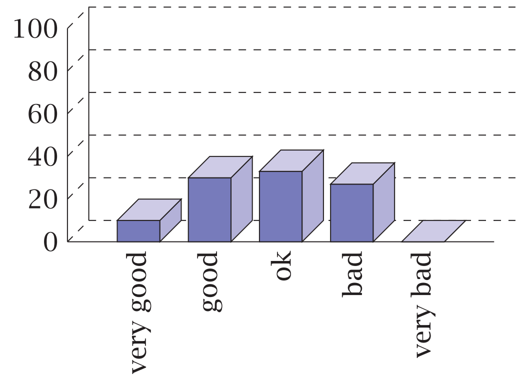
This example is adapted from Tantu [37].

At the end of a seminar a lecturer asked the participants for feedback. Of the 50 participants, 30 returned the feedback form. According to the feedback, three participants considered the seminar "very good," nine considered it "good," ten "ok," eight "bad," and no one thought that the seminar was "very bad."

This can be presented as the above text; all the information is there. But it is hard to visualize.

The survey can be summarized nicely as Table 4.7. All the information is here and accessible. It is easier to find a specific number here than in the text.

But maybe a graphic presentation will be more accessible? Here is a typical "Excel-style" attempt, a 3D bar chart.



This “looks nice” (sort of) but can the bar chart answer these questions?

- How many participants where there?
- How many participants returned the feedback form?
- What percentage of the participants returned the feedback form?
- How many participants checked “very good”?
- What percentage out of all participants checked “very good”?
- Did more than a quarter of the participants check “bad” or “very bad”?
- What percentage of the participants that returned the form checked “very good”?

In addition, the graphic has many faults:

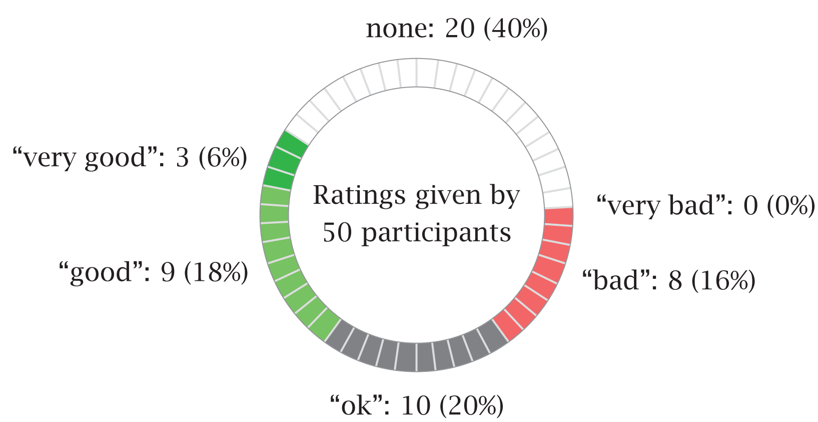
- The graphic is dominated by irritating background lines.
- It is not clear what the numbers at the left mean; presumably percentages, but it might also be the absolute number of participants.
- The labels at the bottom are rotated, making them hard to read.

Rating given	*Number who gave this rating*	*Percentage*
“very good”	3	6%
“good”	9	18%
“ok”	10	20%
“bad”	8	16%
“very bad”	0	0%
(no rating)	20	40%

Table 4.7: Satisfaction with a seminar (50 participants)

- The third dimension adds complexity to the graphic without adding information.
- The 3D setup makes it much harder to gauge the height of the bars correctly. Consider the “bad” bar. It the number this bar stands for more than 20 or less? While the front of the bar is below the 20 line, the back of the bar (actually aligned to the number) is above.
- It is impossible to tell which numbers are represented by the bars.
- What do the bar heights add up to? Is it 100% or 60%?
- Does the bar for “very bad” represent 0 or 1?
- Why are the bars blue?

Now, by thinking a bit on how to best communicates all the information, Till Tantau came up with this, which he drew in `TikZ` [37] (a LaTeX package):



Note the connotative colours, the even treatment of each class, the natural order, the clean and legible design.

4.5 References

Bibliography

[1] American Society of Agronomy; Crop Science Society of America; & Soil Science Society of America. 2004. *Publications Handbook and Style Manual*. Madison, WI: American Society of Agronomy. URL: https://www.agronomy.org/publications/style/; accessed 23-March-2009 139

[2] Anderson, E. 1935. *The irises of the Gaspé peninsula*. *Bulletin of the American Iris Society* **59**:2–5 169

[3] Bacic, I. L. Z. 2003. *Demand-driven land evaluation : with case studies in Santa Catarina, Brazil*. Thesis (phd), International Institute for Geo-Information Science and Earth Observation, Wageningen University, 2003. 183

[4] Brown, A. & Feringa, W. 2003. *Colour basics for GIS users*. Harlow, Essex: Prentice-Hall 150, 151, 157, 202

[5] CBE Style Manual Committee. 1994. *Scientific style and format : the CBE manual for authors, editors and publishers*. Cambridge: Cambridge University Press, 6th edition 139, 166

[6] Cleveland, W. S. 1993. *Visualizing data*. Murray Hill: AT&T 140, 168, 172

[7] Cleveland, W. S. 1994. *Elements of graphing data*. Murray Hill: AT&T, revised edition 140, 168, 172

[8] Congalton, R. G. & Mead, R. A. 1983. *A quantitative method to test for consistency and correctness in photointerpretation*. *Photogrammetric Engineering & Remote Sensing* **49**:69–74 165

[9] Cook, D.; Swayne, D. F.; & Buja, A. 2007. *Interactive and Dynamic Graphics for Data Analysis : with R and GGobi*. Use R! New York: Springer Verlag 198

[10] Dickinson, G. C. 1973. *Statistical mapping and the presentation of statistics*. London, New York,: Edward Arnold; Crane, Russak, 2nd edition 168

[11] Frankel, A. D.; Petersen, M. D.; Mueller, C. S.; Haller, K. M.; Wheeler, R. L.; Leyendecker, E.; Wesson, R. L.; Harmsen, S. C.; Cramer, C. H.; Perkins, D. M.; & Rukstales, K. S. 2002. *Documentation for the 2002 update of the national seismic hazard maps*. Open-File Report 02-420, US Geological Survey. http://pubs.usgs.gov/of/2002/ofr-02-420/US1hz2500v3-508.pdf 155

[12] Hengl, T. 2003. *Pedometric mapping: bridging the gaps between conventional and pedometric approaches*. ITC Dissertation Series 101. Enschede (NL): International Institute for Geo-Information Science and Earth Observation (ITC) 154

[13] Huberman, A. M. & Miles, M. B. 2002. *The qualitative researcher's companion*. Thousand Oaks, CA: Sage Publications 140

[14] Huff, D. 1993. *How to lie with statistics*. New York: Norton 168

[15] Iliffe, J. & Lott, R. 2008. *Datums and map projections for remote sensing, GIS, and surveying*. Scotland, UK; Boca Raton, FL: Whittles Pub.; CRC Press, 2nd edition 186

[16] Jacoby, W. G. 1997. *Statistical graphics for univariate and bivariate data*. Thousand Oaks, Calif.: Sage Publications 172

[17] Jacoby, W. G. 1998. *Statistical graphics for visualizing multivariate data*. Thousand Oaks: Sage Publications 172

[18] Jones, G. E. 2007. *How to lie with charts*. Santa Monica, Calif.: LaPuerta, 2nd edition 168

[19] Kopka, H. & Daly, P. W. 2004. *Guide to LaTeX*. Boston: Addison-Wesley, 4th edition 167

[20] Lamport, L. 1994. *LaTeX : a document preparation system : user's guide and reference manual*. Reading, MA: Addison-Wesley Pub. Co., 2nd edition 167

[21] Maindonald, J. & Braun, J. 2003. *Data Analysis and Graphics Using R*. Cambridge: Cambridge University Press 198

[22] Miles, M. B. & Huberman, A. M. 1994. *Qualitative data analysis : an expanded sourcebook*. Thousand Oaks, CA: Sage Publications, 2nd edition 140

[23] Monmonier, M. S. 1996. *How to lie with maps*. Chicago: University of Chicago Press, 2nd edition 184

[24] Murrell, P. 2005. *R Graphics*. Boca Raton, FL: Chapman & Hall/CRC 198

[25] National Aeronautics and Space Administration (NASA). 2011. *TERRA (EOS AM-1)*. Accessed 2011-May-03
URL http://terra.nasa.gov/ 184

[26] Ordnance Survey. 2008. *A guide to coordinate systems in Great Britain*. Southampton, England: Ordnance Survey, v1.9 edition. URL: http://www.ordnancesurvey.co.uk/oswebsite/gps/docs/A_Guide_to_Coordinate_Systems_in_Great_Britain.pdf 187

[27] Pennsylvania Department of Conservation and Natural Resources. no date. *Lehigh Gorge State Park*. URL: http://www.dcnr.state.pa.us/stateparks/parks/lehighgorge.aspx; accessed 23-April-2009 187

[28] Revesz, K. M.; Landwehr, J. M.; & Keybl, J. 2001. *Measurement of $\delta^{13}C$ and $\delta^{18}O$ isotopic ratios of $CaCO_3$ using a Thermoquest Finnigan GasBench II Delta Plus XL continuous flow isotope ratio mass spectrometer with application to Devils Hole core DH-11 calcite*. Technical Report Open-File Report 01-257, U.S. Geological Survey 142

[29] Robinson, A.; Sale, R.; Morrison, J.; & Muehche, P. 1995. *Elements of cartography*. New York: John Wiley, 6th edition 150, 184, 205

[30] Rossiter, D. G. 2009. *Introduction to the R Project for Statistical Computing for use at ITC*. Enschede (NL): International Institute for Geo-information Science & Earth Observation (ITC), 3.6 edition
URL http://www.itc.nl/personal/rossiter/teach/R/RIntro_ITC.pdf 155, 168, 198, 206

[31] Sarkar, D. 2002. *Lattice*. *R News* **2**(2):19–23 168

[32] Sarkar, D. 2008. *Lattice : multivariate data visualization with R*. Use R! New York: Springer
URL http://lmdvr.r-forge.r-project.org/ 168, 198

[33] Schmid, C. F. 1983. *Statistical graphics : design, principles and practices*. New York: Wiley & Sons 172

[34] Skidmore, A. K. 1999. *Accuracy assessment of spatial information*. In Stein, A.; Meer, F. v. d.; & Gorte, B. G. F. (eds.), *Spatial statistics for remote sensing*, pp. 197–209. Dordrecht: Kluwer Academic 165

[35] Slocum, T. A.; McMaster, R. B.; Kessler, F. C.; & Howard, H. H. 2008. *Thematic cartography and geographic visualization*. Prentice Hall series in geographic information science. Upper Saddle River, NJ: Pearson/Prentice Hall, 3rd edition 184

[36] Stichting voor Bodemkartering (STIBOKA). 1976. *Bodemkaart van Nederland : 45 Oost 's-Hertogenbosch en 46 West - 46 Oost Vierlingsbeek : schaal 1:50 000*. Wageningen: Stichting voor Bodemkartering (STIBOKA) 146

[37] Tantu, T. 2007. *TikZ and PGF, Manual for Version 2.00*. Technical report. http://sourceforge.net/projects/pgf 140, 208, 210

[38] Transport for London. 2011. *Transport for London*. Accessed 2011-May-03
URL http://www.tfl.gov.uk/ 186

[39] Tufte, E. R. 1990. *Envisioning information*. Cheshire, CT: Graphics Press 140, 168

[40] Tufte, E. R. 2001. *The visual display of quantitative information*. Cheshire, CT: Graphics Press, 2nd edition 140, 168

[41] US Department of Agriculture. 1979. *Soil survey of Edgecombe County, North Carolina*. Technical report, NRCS 188, 189

[42] US Department of Agriculture Natural Resources Conservation Service. 1998. *Soil survey geographic (SSURGO) database for Edgecombe County, North Carolina* 189

[43] Vink, A. 1948. *Bodemkundige detailkaart van "Middachten" en "Beekhuizen" (Detailed soil map of "Middachten" and "Beekhuizen")*. Wageningen: Stichting voor Bodemkartering 145, 189, 190