Karlsruhe Institute of Technology

Data Analysis in Geoscience Remote Sensing Projects

Introduction and working in geoscientific projects

Dr. Hendrik Andersen | November 15, 2023

Schedule and overview

The exercises will be in person, on Wednesdays 8 am in the “großer Hörsaal” in building 10.50

- Remote lecture: Introduction, aspects of working in geoscientific projects
- Remote lecture: Time series analysis, change point detection
- Remote lecture: Linear and logistic regression
- Remote lecture: Machine learning fundamental concepts
- Remote lecture: Machine learning techniques and explainability
- Jan. 10: Exercise, python: Introduction to data visualization and descriptive statistics
- Jan. 17: Exercise, change point detection
- Jan. 24: Exercise, regression analysis and sensitivity estimation
- Jan. 31: Exercise, machine learning workflow and regression
- Feb. 7: Exercise, machine learning for classification
- Feb. 14: Exercise, independent data analysis

A note for the exercises

Before first exercise starts, make sure to install python in an Anaconda environment. The tutorial on how to do this can be found in ILIAS. During the exercises there will be no support for installing software - if you encounter difficulties installing the environment, contact me before January 5, 2024 via email (hendrik.andersen@kit.edu).

The title of this course

Data Analysis in Geoscience Remote Sensing Projects

The title of this course

Data Analysis in Geoscience Remote Sensing Projects

Today, we will discuss

- Data
- The scientific method and working in scientific projects
- Geoscience as a system science
- Remote sensing applications

What are data, information?

What are data, information?

Data

- "Individual facts"
- Numbers/impressions gained from observations/models

Information

- Organization and interpretation of data
- Putting data into context
- Combination of data
- Analysis of data

What are data, information?

Data

- “Individual facts”
- Numbers/impressions gained from observations/models

Information

- Organization and interpretation of data
- Putting data into context
- Combination of data
- Analysis of data

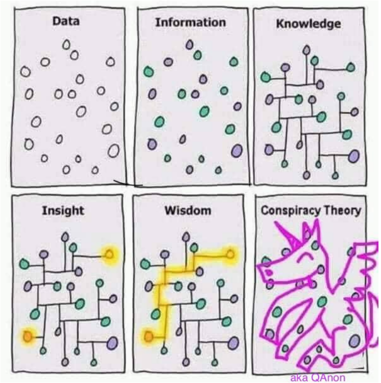


Figure: `https://flowingdata.com/2020/08/18/data-information-conspiracy/`

From data to information/insights

1. Measurement
2. Collection in a data set
3. Correction / quality control
4. Analysis and visualization of data

From data to information/insights

1. Measurement
2. Collection in a data set
3. Correction / quality control
4. Analysis and visualization of data

$\rightarrow$ Is this already science?

The scientific method

Figure: Ibn al Haytham [wikipedia]

6 steps of the scientific method

1. Observe/ask questions
2. Research
3. Formulate a hypothesis
4. Test hypothesis: experiment/data analysis
5. Conclude
6. Share results

The scientific method 1: Observe and ask questions

Converts your sense of wonder into a focused research question.

Examples:

- Under what conditions do clouds form?
- How long is the ash of a volcano eruption in the atmosphere?
- ...

$\rightarrow$ With your initial question, you move on to step 2

The scientific method 2: Research

Find out if others have asked this or a similar question (i.e. check the state of the art on the topic).

Typically involves:

- Searching in online data bases
- Reading text books/papers

$\rightarrow$ After your literature research, you know a lot more on the topic. You are likely to be able to specify a more focused research question and have an assumption on the answer...

The scientific method 3: Formulate a hypothesis

A hypothesis is an idea or proposition that can be tested by observations, experiments or analyses.

Typical hypotheses could be:

- Different cloud types form in different conditions.
- The lifetime of atmospheric ash particles can be up to years.

$\rightarrow$ With your hypothesis you move on to step 4...

The scientific method 4: Test hypothesis

*A hypothesis can be tested with an **appropriate** experiment and/or data analysis.*

To answer your research question and test your hypothesis you could

- Sample the conditions in which different cloud types form, perform statistical analysis
- Track volcanic aerosols in satellite imagery for a specific volcano eruption

Take care: In this step, it is critical that the data and methods that you use are appropriate to adequately test your hypothesis (e.g. you may not have data of e.g. ambient atmospheric conditions of clouds (or these are only available for some clouds but not for others), or when the ash dispurses satellites may not be able to detect them any more.).

$\rightarrow$ You now have conducted an analysis to test the hypothesis, it is time to...

The scientific method 5: Conclude

Conclude from the results of your experiment/data analysis if or to what extent your hypothesis was right/wrong.

To draw the correct conclusions this step typically involves

- Statistical tests/methods
- Visualizing the results in figures
- Discussing uncertainties in the data/method

→ When you have drawn your conclusions it is time to...

The scientific method 6: Share results

If it is not published [or communicated] it never happened

Typical ways of communicating scientific results are via:

- Research paper
- Conference presentation (talk/poster)
- Outreach

$\rightarrow$ After you have shared/communicated your results, they will update the state-of-the-art on that topic, which other researchers can then use to formulate hypothesis and design their analyses.

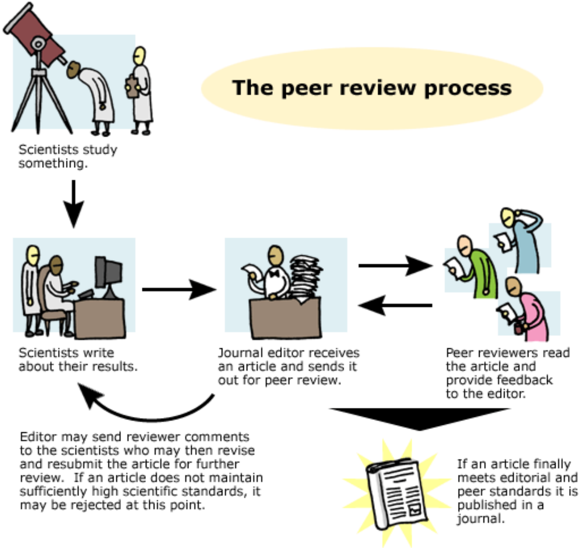


Figure: Understanding Science, berkeley.edu 2012

How do I find literature

Karlsruhe Institute of Technology

A literature search always starts with a question.

Suggestion: write down the question and pin it to a location where you can see it during your search.

Approaches to literature search:

- Search in databases
- Snowballing
- Inverse snowballing
- Literature administration

Search in databases

There are many different databases for searching and finding scientific literature. Some examples:

- Web of Science, accessible from all university computers (`http://isiknowledge.com/`)
- Scopus (`http://scopus.com/`)
- Google scholar (`http://scholar.google.com/`)

Which of these databases yields the best results for you may depend on your research question; so it's best to try more than one.

Snowballing

ECOSPHERE
AN ESA OPEN ACCESS JOURNAL

Article | Open Access

Fog drip maintains dry season ecological function in a California coastal pine forest

Douglas T. Fischer, Christopher J. Still, Colin M. Ebert, Sara A. Baguskas, A. Park Williams

First published: 28 June 2016 | https://doi.org/10.1002/ecs2.1364 | Cited by: 6

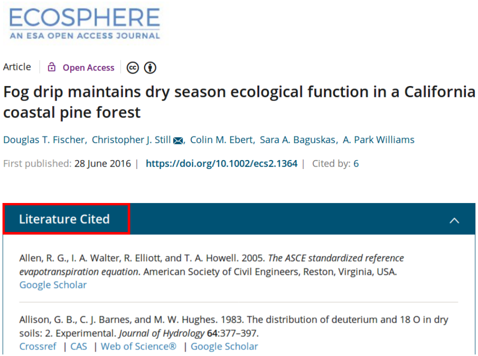
Literature Cited

Allen, R. G., I. A. Walter, R. Elliott, and T. A. Howell. 2005. *The ASCE standardized reference evapotranspiration equation*. American Society of Civil Engineers, Reston, Virginia, USA.
Google Scholar

Allison, G. B., C. J. Barnes, and M. W. Hughes. 1983. The distribution of deuterium and 18 O in dry soils: 2. Experimental. *Journal of Hydrology* **64**:377–397.
Crossref | CAS | Web of Science® | Google Scholar

Pros and cons

Snowballing



ECOSPHERE
AN ESA OPEN ACCESS JOURNAL

Article | Open Access

Fog drip maintains dry season ecological function in a California coastal pine forest

Douglas T. Fischer, Christopher J. Still, Colin M. Ebert, Sara A. Baguskas, A. Park Williams

First published: 28 June 2016 | https://doi.org/10.1002/ecs2.1364 | Cited by: 6

Literature Cited

Allen, R. G., I. A. Walter, R. Elliott, and T. A. Howell. 2005. *The ASCE standardized reference evapotranspiration equation*. American Society of Civil Engineers, Reston, Virginia, USA.
Google Scholar

Allison, G. B., C. J. Barnes, and M. W. Hughes. 1983. The distribution of deuterium and 18 O in dry soils: 2. Experimental. *Journal of Hydrology* **64**:377–397.
Crossref | CAS | Web of Science® | Google Scholar

Pros and cons

- pro: *quick* overview of many similar/connected sources
- con: risk of too *narrow focus*
- con: all referenced sources are by definition *older*

Inverse snowballing

ECOSPHERE
AN ESA OPEN ACCESS JOURNAL

Article | Open Access

Fog drip maintains dry season ecological function in a California coastal pine forest

Douglas T. Fischer, Christopher J. Still, Colin M. Ebert, Sara A. Baguskas, A. Park Williams

First published: 28 June 2016 | https://doi.org/10.1002/ecs2.1364 | Cited by: 6

Literature Cited

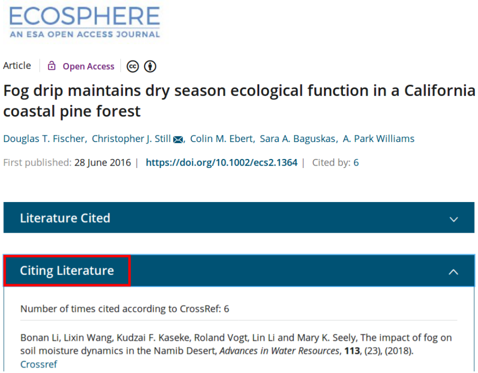
Citing Literature

Number of times cited according to CrossRef: 6

Bonan Li, Lixin Wang, Kudzai F. Kaseke, Roland Vogt, Lin Li and Mary K. Seely, The impact of fog on soil moisture dynamics in the Namib Desert, *Advances in Water Resources*, **113**, (23), (2018).
Crossref

pros and cons

Inverse snowballing


ECOSPHERE
AN ESA OPEN ACCESS JOURNAL

Article | Open Access

Fog drip maintains dry season ecological function in a California coastal pine forest

Douglas T. Fischer, Christopher J. Still, Colin M. Ebert, Sara A. Baguskas, A. Park Williams

First published: 28 June 2016 | https://doi.org/10.1002/ecs2.1364 | Cited by: 6

Literature Cited

Citing Literature

Number of times cited according to CrossRef: 6

Bonan Li, Lixin Wang, Kudzai F. Kaseke, Roland Vogt, Lin Li and Mary K. Seely, The impact of fog on soil moisture dynamics in the Namib Desert, *Advances in Water Resources*, **113**, (23), (2018).
Crossref

pros and cons

- pro: *quick* overview of many similar/connected sources
- con: risk of too *narrow focus*
- pro: all referenced sources are by definition *newer*

Literature administration

Over the course of your studies, and especially when working on term papers or a thesis, you will collect a lot of references. To stay on top of things, keeping a literature database is a good idea. There are many software programs for this purpose. Some examples:

- Mendeley (`https://www.mendeley.com/`)
- Jabref (`https://www.jabref.org/`)
- Zotero (`https://www.zotero.org`)

How do I read a scientific paper?

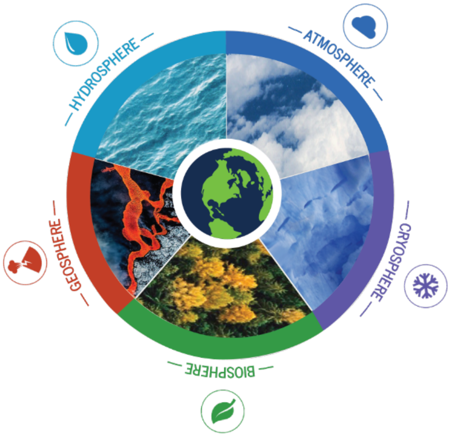
A scientific paper may be look very complicated at first, but you can exploit the standardized structure of scientific publications to get a quick overview.

How do I read a scientific paper?

A scientific paper may be look very complicated at first, but you can exploit the standardized structure of scientific publications to get a quick overview. One possible approach:

1. **Read the abstract.** An abstract is a special kind of summary focusing on the central aspects of the text; it contains the essence of the article.
2. **Read the introduction.** The introductory section will outline the underlying problem and often includes an outline of the article structure.
3. **Look at the figures.** In a good scientific article, all core points are communicated in a figure, and each figure is fully explained in its caption.
4. **Read the last chapter.** The last chapter often has a title like “discussion”, “conclusions”, “outlook” or similar. Here, the article’s most important findings are discussed in a wider scientific context.
5. **Read everything else.** If the article seems really relevant after you have gone through the previous steps, read the other chapters (or a selection thereof).

Geoscience: The Earth as a system



Atmosphere	The thin layer made up of a mixture of gases and particles suspended in the air that surround the Earth (predominantly N_2, O_2, Ar, CO_2, and H_2O)
Hydrosphere	A sphere that includes the liquid ocean, inland water bodies, and groundwater
Cryosphere	A subset of the Hydrosphere that consists of frozen water
Geosphere	A sphere that includes the solid Earth; the core, mantle, crust, and soil layers
Biosphere	A sphere that Includes all of Earth's organisms, including humans, and matter that has not yet decomposed

Figure: `https://mynasadata.larc.nasa.gov/basic-page/about-earth-system-background-information`

Typical questions in geoscience research, applications of remote sensing

Typical geoscientific research questions

- What is the current state of the Earth system?
- How does the Earth system change?
- How are different components of the Earth system related?

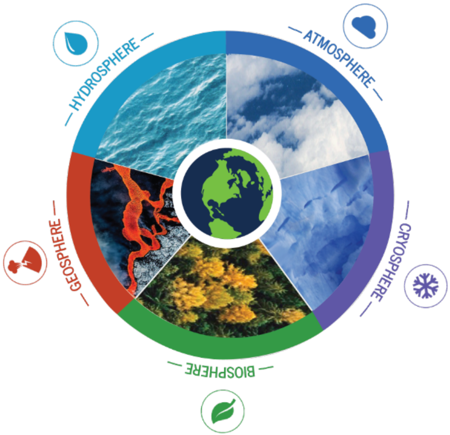
Applications of remote sensing

- Observing the current state of the Earth system and recent changes
- Weather observations and short-term forecasting
- Analyzing relationships of observables within the Earth system to improve system understanding
- Combine with earth system models or use observations to evaluate them

Typical questions in geoscience research, applications of remote sensing

https://www.ted.com/talks/kristin_poinar_what_s_hidden_under_the_greenland_ice_sheet?language=en#t-534403

Geoscience: The Earth as a system



- **Atmosphere**: The thin layer made up of a mixture of gases and particles suspended in the air that surround the Earth (predominantly N_2, O_2, Ar, CO_2, and H_2O)
- **Hydrosphere**: A sphere that includes the liquid ocean, inland water bodies, and groundwater
- **Cryosphere**: A subset of the Hydrosphere that consists of frozen water
- **Geosphere**: A sphere that includes the solid Earth; the core, mantle, crust, and soil layers
- **Biosphere**: A sphere that Includes all of Earth's organisms, including humans, and matter that has not yet decomposed

Figure: `https://mynasadata.larc.nasa.gov/basic-page/about-earth-system-background-information`

Overview

Today's contents

- From data to information
- The scientific method
- Scientific literature
- Geoscience as a system science
- Remote sensing applications in geosciences

Further reading

- Chapter 1: Jacobson, Michael C.. Earth System Science: From Biogeochemical Cycles to Global Changes. Academic Press, 2000. [ebook in KIT library]
- Chapter 5: Smerdon, Jason. Climate change: the science of global warming and our energy future. Columbia University Press, 2018. [ebook in KIT library]
- Chapter 3: Finn, John A. Getting a PhD: An action plan to help manage your research, your supervisor and your project. Psychology Press, 2005. [ebook, KIT library]