How to organize and structure a paper

Ronald E. McRoberts
University of Minnesota

Scientific method –

- Context of ***Discovery***

- Context of ***Justification***

Hans Reichenbach 1938

Discovery

- the generation, development, or “discovery” of ideas that lead to hypotheses and models
- as much “art” as “science”
- “Ideas are coin of the realm in the intellectual world.”

  Bradley Efron
  AmStat News

- “Scientific research is the competition of ideas.”

  U.S. Forest Service

Discovery

There is no logic!

There are no rules!

Anything goes!

Justification

- The purpose of *Justification* is an inference
- Inference: to accept from evidence or premises

Justification

Very much rule-based!

Rigorous and mathematical!

Anything DOES NOT go!

A scientific paper focuses on Justification

Introduction

- The context (1-2 paragraphs)
  - Everything in the Introduction should focus on the objectives
  - Convince the reader that the general context is important
  - Examples:
    - Mapping and estimation of AGB are crucial for climate change studies
    - More accurate prediction techniques produce more precise estimates which complies with the IPCC guideline to reduce uncertainty

Introduction

- The literature review (2-4 paragraphs)
  - Review AND summarize the relevant literature
  - Identify the knowledge gaps
  - The literature review should lead naturally and logically to the objectives
  - For a methods paper, persuade the reader that the current methods are inadequate or need refinement

Introduction

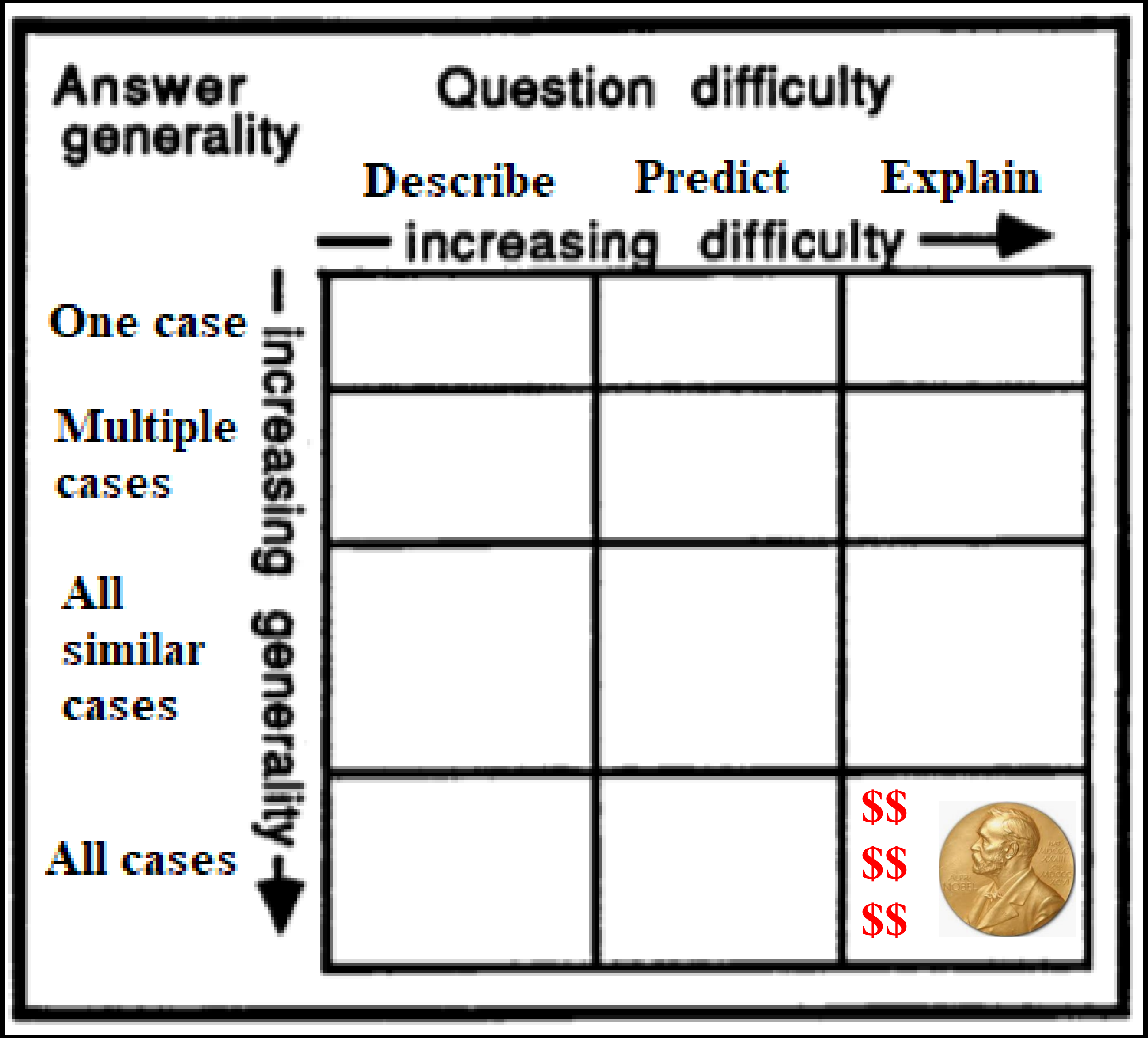
- The literature review (2-4 paragraphs)
  - Avoid long lists of references, 2-3 are usually sufficient
  - Be sure at least some of the references are to recent publications. Too many old references suggest unfamiliarity with the literature.
  - In remote sensing, 10 years ago can be ancient history.
  - Classic or foundational references are acceptable, regardless of age

Introduction

- Objectives/aims (1 paragraph)
  - State the 2-5 objectives clearly and concisely
  - Be sure the objectives align with the knowledge gaps
  - State the objectives using terms that help the reader determine if the objectives were achieved
    - Poor choices: *to explore, to investigate, to discuss, to examine*
    - Good choices: *to estimate, to compare, to illustrate*

Introduction

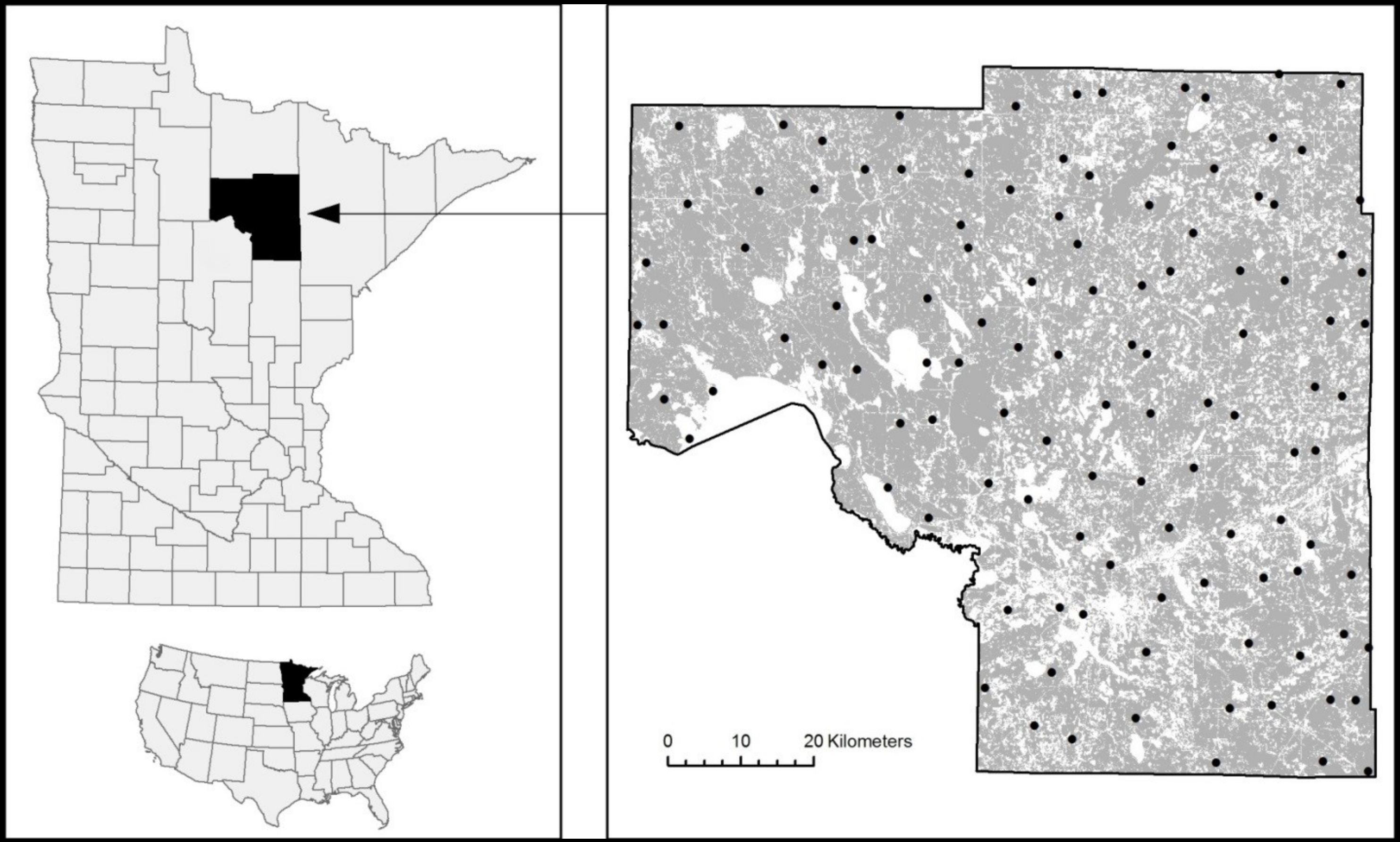
- Objectives/aims (1 paragraph)
  - As scientific disciplines mature, they transition from *descriptive* studies, to *predictive* (mapping) studies, to *inferential* studies (confidence intervals, tests of hypotheses)
  - For forest inventories, models and maps are only intermediate products on the path to an inference in the form of a confidence interval for a population parameter such as mean AGB/ha.
  - Focus on the higher level, more mature objectives



Rolfe A. Leary: Some characteristics of high quality research

Data

- Describe the study area and include a map of where it is located.



Data

- Carefully describe the sampling design.

- The term “random” is not sufficient because simple random, stratified random, systematic with a random starting point, etc all have random components.

Methods

- Overview (1 paragraph)
  - If the methods are complex, consider including an overview

  - State the primary technical components and how they contribute to the overall objectives

- Subsections with headings
  - Describe each methodological component with references

  - Describe the methods in order of application

Methods

- Put long, complex derivations in appendices

- Consider an “Analyses” sub-section at the end of the Methods section to describe how and why the methods are applied

Results and Discussion

- Consider combining Results and Discussion sections
  - Easier to discuss results immediately after reporting results
  - Do not have to recycle back through topics twice

- To the degree possible:
  - Order the subsections similar to the ordering of the objectives and methods

- Consider a summary subsection that discusses the overall significance of the results in the broader context as stated in the Introduction

Results and Discussion

- Limit the numbers of tables and figures to no more than 3-5 of each.
- Include only enough tables and figures to support the conclusions. "Nice-to-know" tables and figures can be placed in Supplementary Material
- Remember, a journal article is not a PhD dissertation with unlimited length.

Results and Discussion

- Table and figure captions should be limited to 1-2 sentences. If an entire paragraph is required, then the table or figure is too complex.
- Graph observations versus predictions, not the other way
  - Piñeiro et al. 2008. Ecological Modelling 216: 316–322.
- State the limitations of the study and the consequences
- **HOWEVER**, too many limitations or too much discussion of them suggests that the study has no merit or relevance.

Conclusions

- Always include a Conclusions section
  - The Conclusions section is the authors' last opportunity to secure the most important findings in the mind of the reader.

- Do not obscure the conclusions with unnecessary discussion

- State the 3-5 most important findings using 1-2 sentences per finding/conclusion

- A few sentences regarding future and/or follow-up analyses are appropriate.

Some terminology issues

- Avoid use of the term *unbiased estimate*
- How do you know an estimate is unbiased if you do not know the true value?

Some terminology issues

- Bias
  - Unbiasedness and bias are properties of an estimator (a formula, a procedure) but not an estimate, not a model, not a sample, and not a map.
  - For example: $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i$
  - If the mean of all possible estimates obtained from all possible samples of the same size obtained using the same sampling design equals the true value, then an estimator is unbiased.

Some terminology issues

- Bias
  - Any assessment of bias requires knowledge of the true value plus at least a very large number of samples (not observations).

  - Often terms such as *systematic deviation* can be used.

Some terminology issues

- Avoid terms such as *earth observation*
- Remotely sensed data do not *observe* or *measure* anything on the ground.
- Remote sensors observe/measure only frequencies and/or intensities of reflected light (e.g., Landsat) or time from emission of a pulse to receipt of the reflected pulse back at the sensor (lidar).
- However, remotely sensed data can be used to *predict* ground conditions.

Some terminology issues

- Not just a semantic issue.
- If remotely sensed data observe or measure ground conditions, then the only source of uncertainty is measurement error.
- If remotely sensed data predict ground conditions, then there are two sources of error, the original measurement error plus the error associated with the model predictions of the ground conditions from the measurements.

Some terminology issues

- Size or amount
  - The first definitions of terms such as *high*, *low*, *over*, *under*, etc refer to vertical position.

  - The first definitions of terms such as *large*, *small*, *more*, *less*, *greater than*, etc refer to size or amount.

  - In scientific papers, use the latter terms to refer to size or amount.

Some terminology issues

- Size or amount
  - When we use data with a vertical component (e.g., lidar, 3D), terms such as “high biomass” are confusing. Does it mean a large amount of biomass or does it refer to biomass near the top of the canopy or near the tops of mountains?
  - A recent native English writer used *high biomass*, *high resolution*, and *high altitude* all in the same sentence.
  - For *resolution*, use *fine* or *coarse*.

Some terminology issues

- In papers that use the term *plot* to refer to a ground sampling location, use the term *graph* to refer to the visual depiction of a relationship.

- Avoid the use of subjective terms such as *better*, *best*, *poor*, etc whose criteria are not explicitly stated.
  better with respect to what?

- Use words (one, two,…) for numbers nine and less,
  and use digits (10, 11, ….) for numbers 10 and greater.

Some terminology issues

- When uncertainty is involved, use terms such as *predict* or *estimate* rather than *derive*, *calculate*, *identify*, *quantify*, *determine, compute, extract, capture, extrapolate, etc*

- Distinguish between the term *sample units* and the term *samples* which refers to a set of multiple sample units.

- In papers with statistical content, reserve the term *significant* to refer to a formal test of significance in which case the name of the test and either the P-value of the α-level should also be reported.

Some terminology issues

- When the mathematical expression for a relationship has uncertainty, has a random residual term, or is estimated with sample data, the correct term is *model*, not *equation* or *function*.

- Avoid use of the term *prove*, because in science nothing is ever really proven.

- Avoid use of the term *believe*. The root of the English term *believe* is *faith*. As scientists, we base our conclusions on evidence, not faith.

- Avoid use of the term *feel*. As scientists we base our conclusions on evidence, not feelings.

Live long and prosper

and write great scientific papers.