Madagascar - a powerful software package for multidimensional data analysis and reproducible computational experiments

Adrian Smith*, Sergey Fomel, Robert Ferguson

adrsmith@ucalgary.ca

Abstract

Reproducibility of published scientific findings is critical toward exposure of ideas and results to indpendent testing and replication by other scientists. Computational experiments are made readily reproducible in theory due to systematic characteristics of computer programs, but this proves more difficult in practice. Madagascar is a Unix-based open source software package that provides an environment for computational data analysis in geophysical and related fields. It incorporates functionality from pre-existing geophysical analysis libraries, and it allows the end user to completely package publications in a reproducible format using SCons and LaTeX. We present two simple computational examples illustrating the functionality of Madagascar. A local reconstruction of several figures from a published paper is given to higlight the power of Madagascar as a vehicle for generating reproducible research. Existing programs developed within CREWES can be incorporated into Madagascar's library. The installation of Madagascar on CREWES servers is highly recommended.

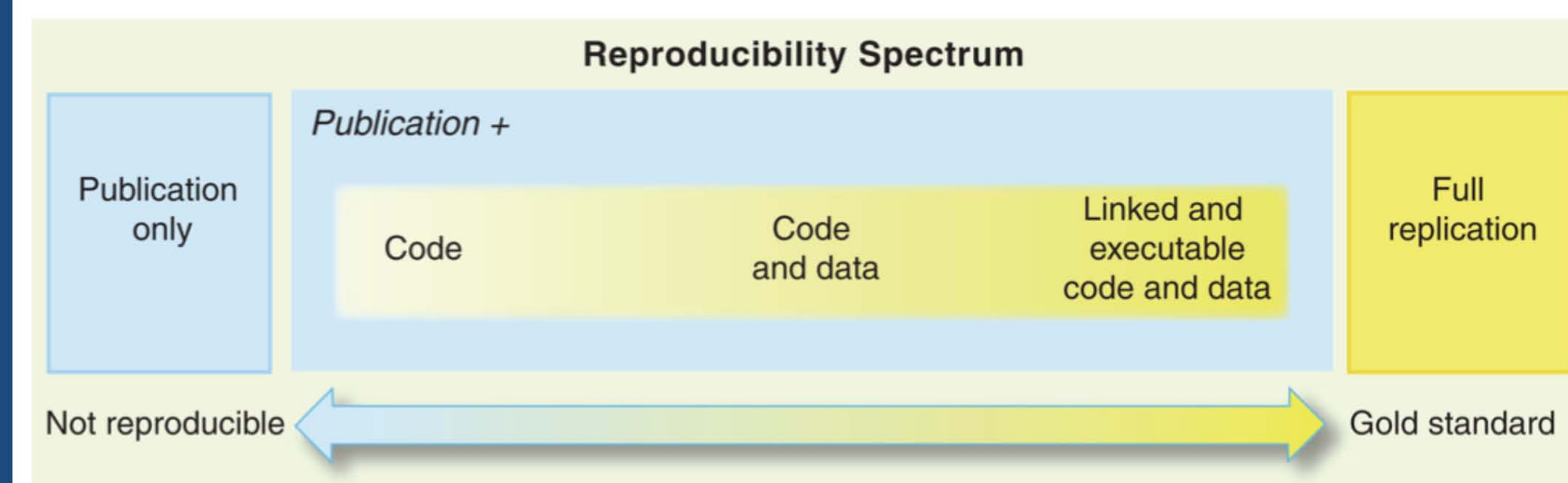## Reproducibility in Computational Science

Figure : 1. Illustration of the concept of reproducibilty in computational research. The ultimate goal should be to make any research land as far to the right-hand-side of the spectrum as possible.
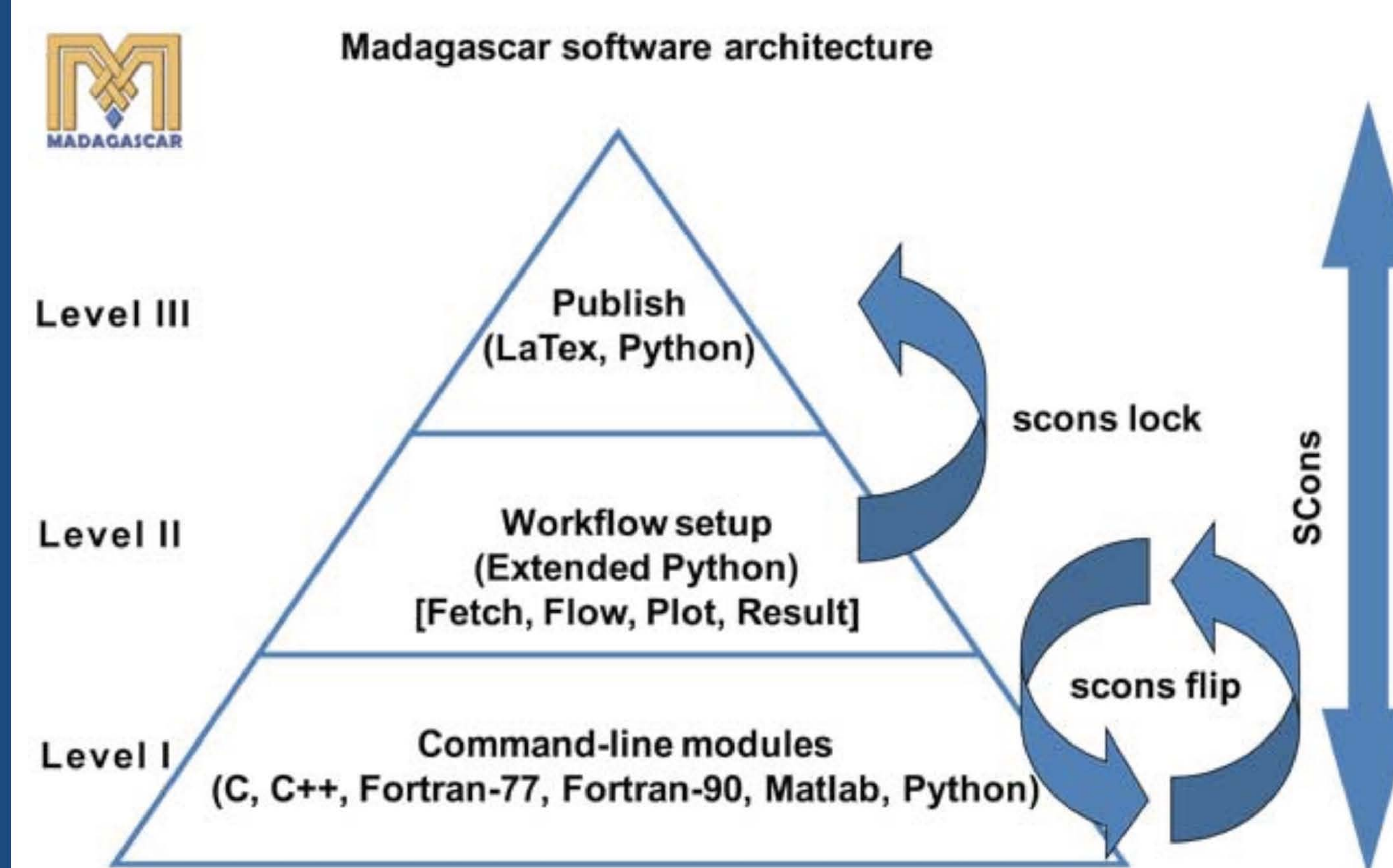

Figure : 2. Illustration of the architecture of the Madagascar software package. The three levels are: (I) - Implementation of new computational algorithms for data analysis, involving writing low-level programs (II) - Testing of new algorithms or workflows on data by assembling workflows from existing command-line modules and tuning their parameters (III) - Documentation level. Results (figures) get referenced in the output publication.

Madagascar Package Overview

- Madagascar is Unix-based. It breaks the data analysis chain into steps by using short programs that perform simple tasks.
- The primary language of Madagascar is C, but interfaces to other languages (C++, Fortan 90, Python, MATLAB) are available.
- SCons is a Python-based software construction tool used by Madagascar to check dependencies and package research in a reproducible format.
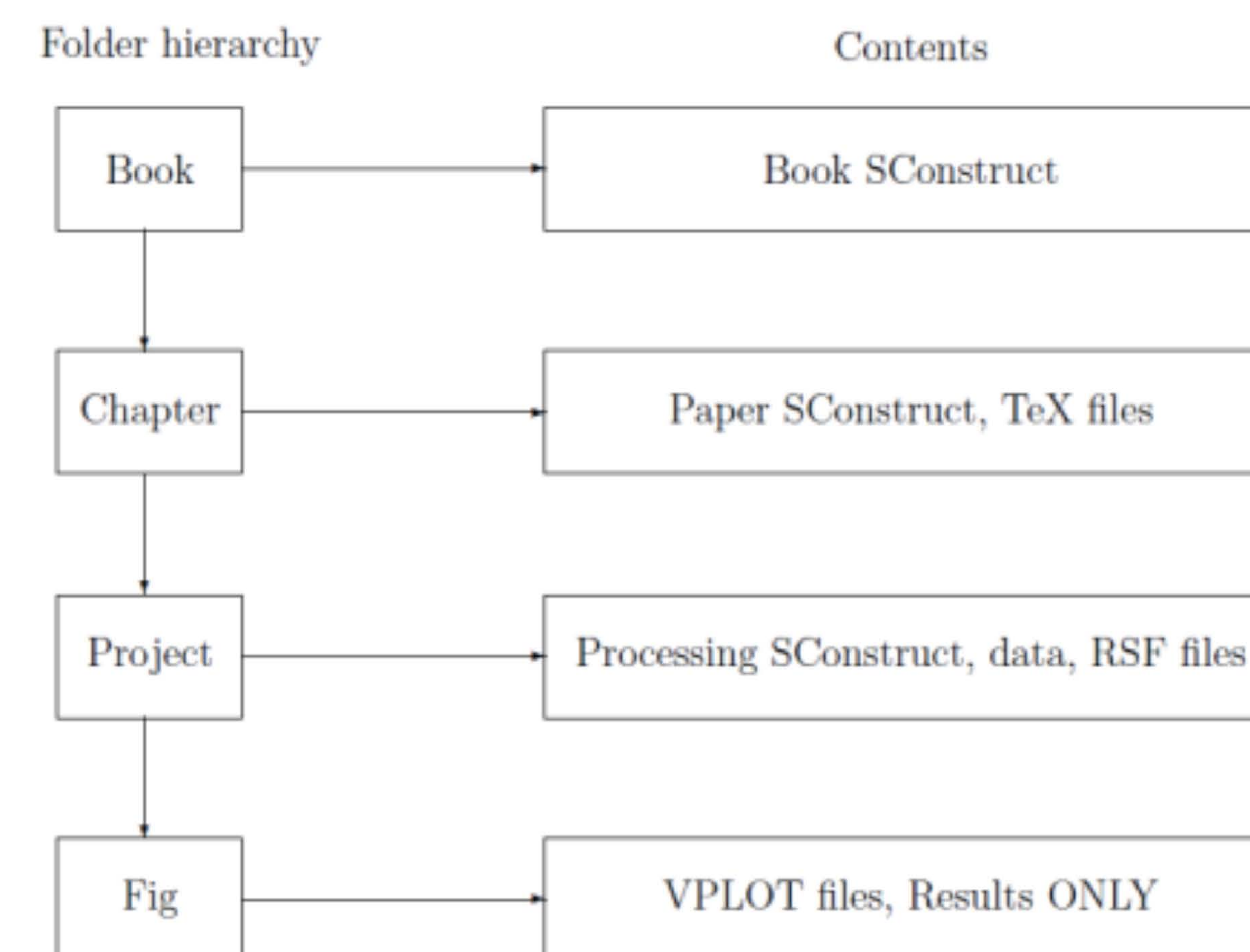

Figure : 3. Chart illustrating the organization of file\folder locations used to generate a reproducible document. Image from: *www.ahay.org/wiki/Guide_to_RSF_file_format*
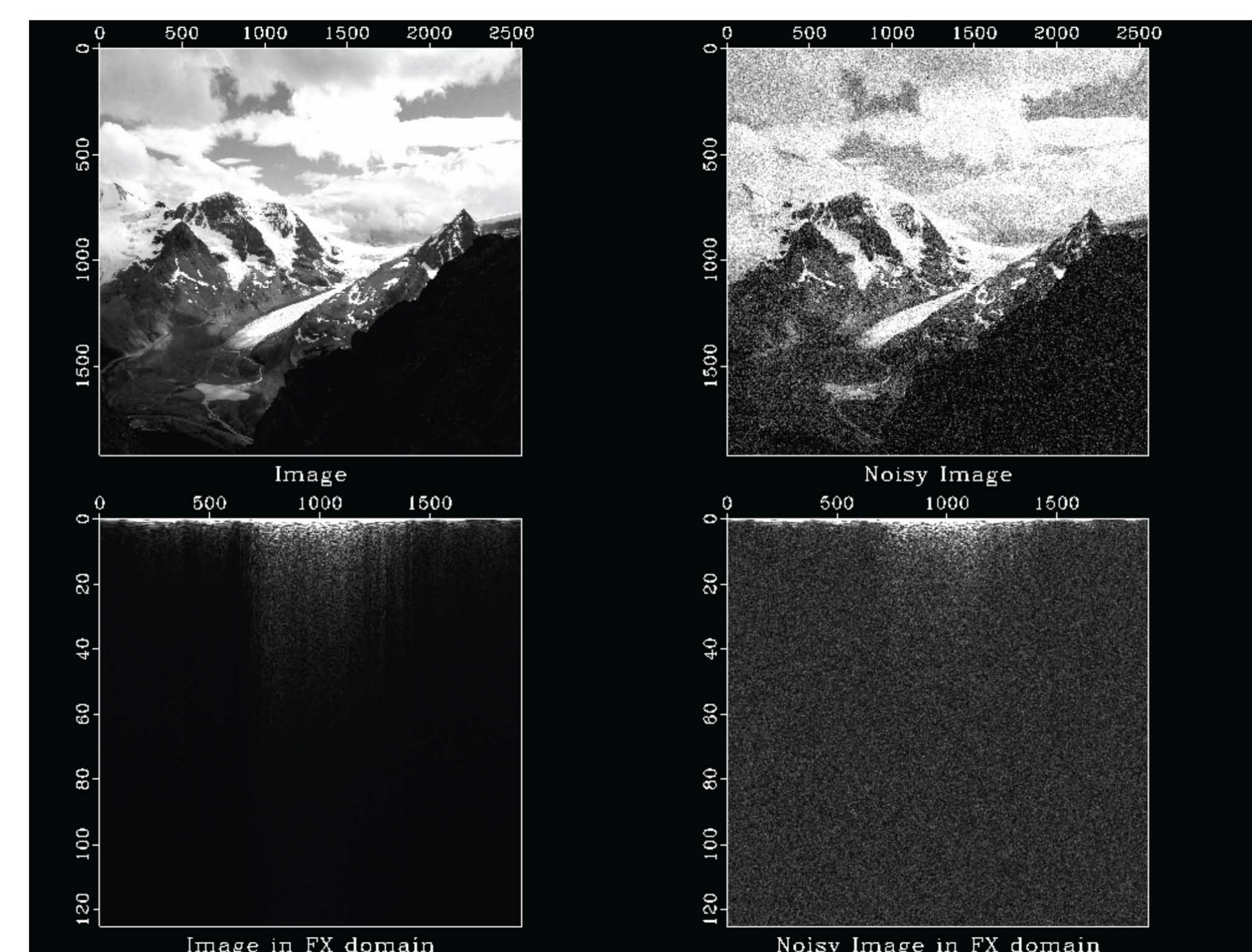
Madagascar Examples

Figure : 4. Various images of the Athabasca glacier generated using a basic processing flow in an SConstruct file. The upper-left image is the original image, with its FX spectrum in the lower-left. The upper right is the original image with random noise added and the corresponding FX spectrum in the bottom-right image. The horizontal and vertical axes on the upper images represent pixel numbers. The horizontal axes on the lower images represent pixel numbers, with vertical axes of frequency in Hz.
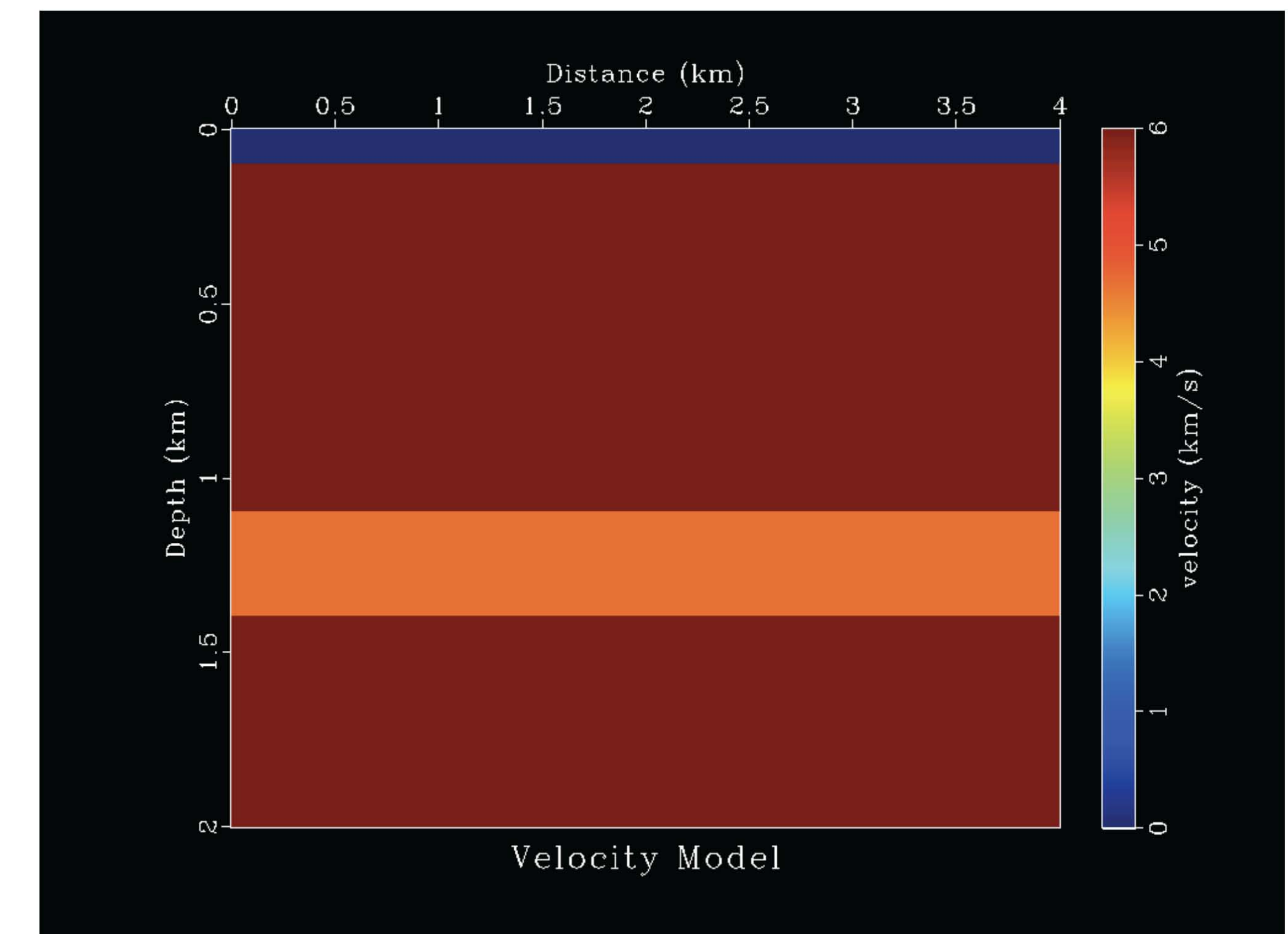
Examples Continued

Figure : 5. Simple four layer velocity model generated in an SConstruct file. A low velocity layer is located between 1.1 and 1.4 km depth.
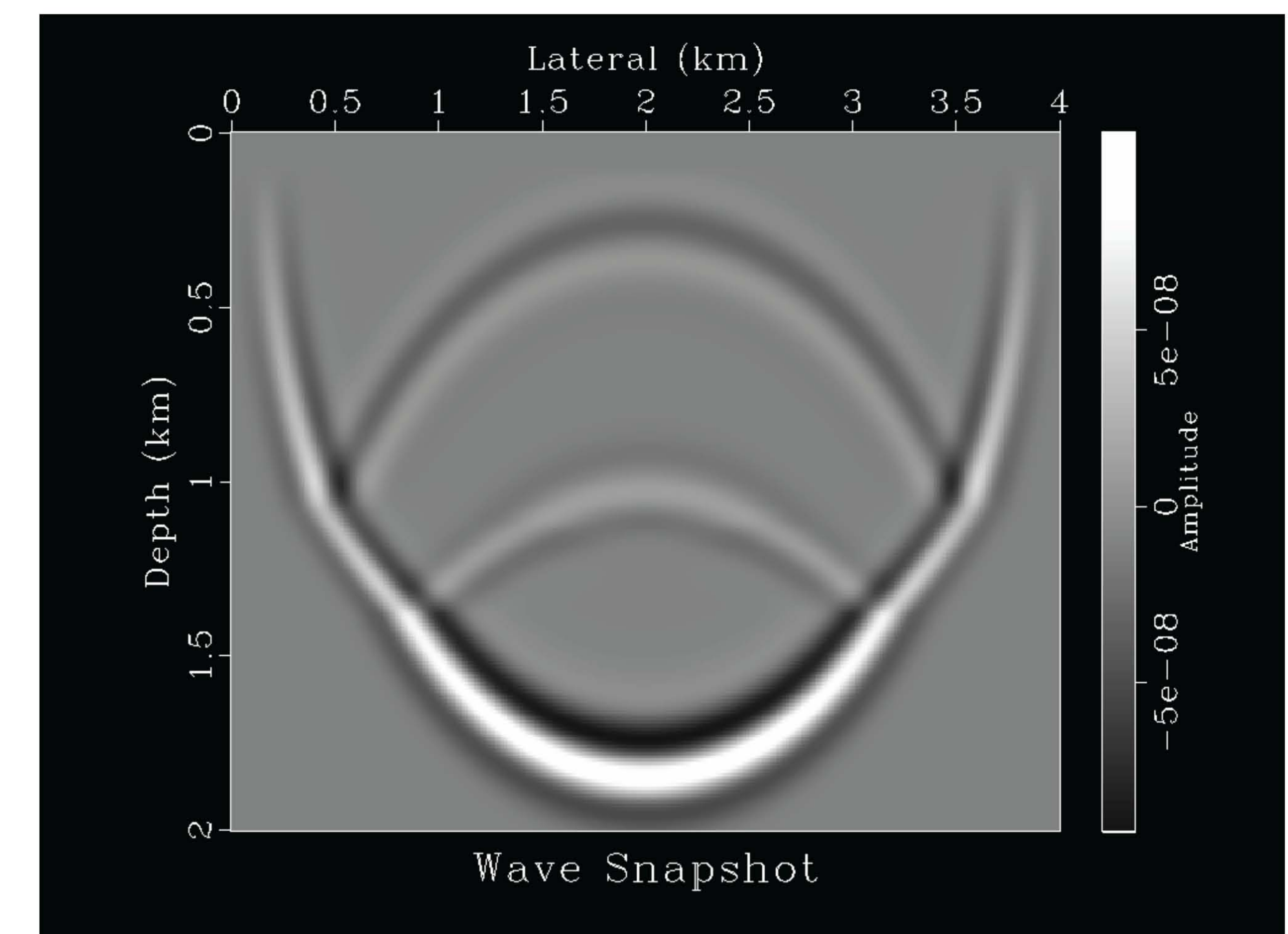

Figure : 6. Snapshot of a wavefield modelled using 2D FD acoustic forward modelling code made available in Madagascar. The velocity model used is shown in Figure 5, with the source location at 2km lateral distance and 0km depth. The two reflections associated with the top and bottom of the low-velocity layer are quite visible.

Conclusions

It is recommended that CREWES install Madagascar on our servers for the following reasons:

- It will allow colleagues within CREWES to more efficiently follow and use workflows that have been previously developed.
- Complete, reproducible packaging of CREWES publications could be made available for internal use and for sponsors. Previous work could also be archived more efficiently.
- Exiting software developed at CREWES can be incorporated within Madagscar.

NSERC CRSNG · UNIVERSITY OF CALGARY