8-5-2021

Collaborative Research: Deciphering the role of extreme rainstorms and hydroclimatic regime on arid escarpment retreat and sub-cliff slope evolution

Joel Pederson
Utah State University, joel.pederson@usu.edu

Data Management Plan

Roles and Responsibilities

The data management activities during the project period will be performed in adherence with NSF policies and the policies of the partner universities. The investigators will coordinate on the execution of this data management plan.

Data Types and Definition

Types of project data will include: digital topography and various derivatives; geochronology (including cosmogenic nuclide analyses and optically stimulated luminescence measurements); rainfall measurements from gages and from rainfall; data from field infiltration experiments; and outputs from computational models of rainfall and landform evolution. Project software is also considered to be data, as are the various program scripts used to process data, configure numerical model runs, and post-process model output.

Period of Data Retention

Project data will be released either before or in conjunction with publication. The data will be publicly available for at least a three-year period, in adherence with NSF Policies.

Data Sharing

The general approach is as follows. A web-hosted, version-controlled document repository will be created for the project, and will serve both as a collaboration platform for the team and as a data sharing platform for the wider community and the public. Because there are limits to the size and format of data files that can be used on such a system, we will use this platform primary for text-based data files that are smaller than about 1 Gb in size. This will include the source code for project software, including various data-processing scripts along with numerical models. Geochronologic and field data will be included, as will point-based (i.e., gage) rainfall data.

Where well-established community repositories are available for a particular type of data, we will provide data to those repositories with appropriate metadata. For example, the HydroShare platform provides a natural distribution point for hydrologic data. The Community Surface Dynamics Modeling System (CSDMS) Model Repository will be used to share and archive numerical modeling software, and the GeoChron repository will be used for geochronologic data.

Large data sets, such as those associated with high-resolution digital topography and its derivatives, present a challenge for storage and dissemination. Our strategy will be to use OpenTopography for the primary elevation data. Analyses of these primary data will be scripted, and the scripts openly shared, such that the various derivatives of the primary data can be reproduced. For radar rainfall data, we will rely on a similar strategy: the primary data will remain with their original point of origin, while documented scripts for the derivation of secondary products, statistics, and similar information, will be provided in the open project repository. Digital copies of the full set of project data, including all primary and derived data, will be maintained on storage media at each of the partner universities.

Model I/O Standards

For output from grid-based numerical models, the NetCDF format will be used as a model input and output standard. NetCDF is a widely used file format that has been specifically designed for storing scientific data. NetCDF is a flexible, self-describing format that bundles data together with any relevant metadata. It can be used to store virtually any kind of data including spatial grids, time-indexed grid stacks, time series, time-indexed profiles and time indexed "cube series". When appropriate, data can be stored efficiently and compactly in binary form with transparent byte order conversion.