Recommendations for Modeling Land-Surface Processes: Report of the Community Surface Dynamics Modeling System (CSDMS) Terrestrial Working Group Winter 2009 Meeting

Report prepared on behalf of the Terrestrial Working Group[1] by Greg Tucker (University of Colorado), Jon Pelletier (University of Arizona), Joe Wheaton (Idaho State University), and Taylor Perron (MIT)

July, 2009

1. Introduction

Developing integrated models of earth-surface dynamics across a wide range of scales represents an exciting challenge and opportunity for the research community. This document summarizes the deliberations and recommendations of the second annual meeting of the CSDMS Terrestrial Working Group (TWG). The meeting was held in Boulder, Colorado, in February 2009, and was attended by about two dozen participants. One of the main goals of the meeting was to develop a set of guidelines and recommendations in three areas:

1) Reviewing current state of the art with respect modeling terrestrial environments, and highlighting knowledge gaps and research needs. This includes compiling an inventory of basic knowledge, existing computer models, and knowledge/model gaps, as well as identifying essential components of a first-generation model.
2) Developing criteria for proof-of-concept applications, identifying specific applications that are of high priority to the community, and analyzing key requirements for model-data comparison.
3) Identifying issues, needs, risks, and opportunities pertaining to technical aspects of modeling and the development of a comprehensive model-component repository.

This report is accordingly divided into three sections that cover each of the above items. A summary of recommendations can be found in the final section of this report.

2. Knowledge Gaps and Research Needs

2.1 The Challenge

Solving major problems in Earth surface processes requires understanding coupled systems. This section begins to address the question of the state of the discipline: how close are we to realizing this goal? The fact that CSDMS has been envisaged not as a single "super-model" but rather as a framework means that there is flexibility in how we represent different processes, and at different scales. However, it also requires the community to make decisions about how to prioritize efforts and to identify key knowledge gaps.

---

[1] A list of members of the Terrestrial Working Group can be found on the CSDMS web site.

2.2 A framework for identifying necessary processes and evaluating our ability to model them ("scoping")

An Earth system can be thought of as consisting of a set of "boxes" that represent different subsystems (such as climate, ecosystems, and tectonics), with fluxes of quantities such as mass and energy between them (Figure 1). Other geoscience fields have "exploded" their boxes, assessing and organizing the community's knowledge of constituent processes. Examples include the Computational Infrastructure for Geodynamics (CIG) and the Community Climate System Model (CCSM). The surface-process community (geomorphology, sedimentology, and related sub-disciplines) is now in the process of "exploding the box" and examining the state of the contents. Some key elements in the domain of CSDMS Terrestrial Working Group include:

- Pathways of mass (solid or solute) from source, via transport, to sink.
- Continental focus.
- Source = bedrock weathering and erosion.
- Sink = delivery to a reservoir where storage occurs for a time long with respect to the timescale for system evolution, such as: continental shelf/ocean, sedimentary basins, and continental water bodies.
- Transport = any intermediate process that causes mass flux
- Note that sign of net transport (i.e., flux divergence) can create local source or sink.

Thus, the "domain" of terrestrial processes includes all major processes responsible for mass transport across the earth's land surface.

The source-to-sink path that one draws, and the list of processes that must be included, depends on the question being posed. As an example, consider the transport path that would apply to one of the CSDMS Grand Challenges: "tracking surface dynamics through glacial cycles." A more specific science question within this theme is: How does a fluvial system respond to changes in sediment supply, sea level, and other factors during rapid glacial terminations like those that occurred in the Pleistocene? A schematic illustration of the system (Figure 1) helps to identify the necessary components of such a model. The diagram illustrates, in one dimension, a transport pathway for sediments and solutes from generation in uplands (upper left) to a coastline (lower right). Processes include weathering (in the sense of both chemical reactions and rock disintegration by mechanical and chemical processes), hillslope and fluvial transport, grain-size distribution and evolution, hydrologic forcing and feedbacks, tectonic forcing, and biological influences.

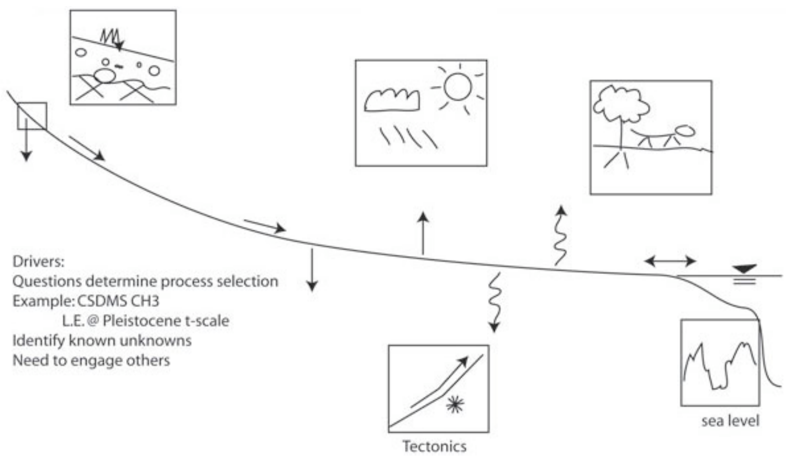


Figure 1: Schematic illustration of processes that must be considered in a model of surface dynamics across glacial-interglacial cycles.

These represent the only classes of mechanisms we need to worry about in this particular "box." Limiting the problem to a subset of processes is not an attempt to trivialize or oversimplify the problem, but rather represents the top layer of a top-down approach. There have of course been many quantitative efforts to understand the details of some of the arrows and sub-arrows in Figure 1. This exercise is a critical part of solving the problem: which terms can we neglect, which can we parameterize? Sample dilemmas include: How deeply do we need to understand weathering? Do we need to model microbial metabolism, or is a soil production rate constant sufficient? Are chemical mass fluxes sufficiently small that they can be neglected? Addressing these questions also helps us identify the gaps in our knowledge: terms that cannot be crossed out, but for which we lack well-developed theory or methods.

To fully evaluate the state of the art for a given process or set of processes, one would "explode an arrow" (or a box), and address the following:

- List the processes that fall within this category
- Evaluate "readiness" of each process component, or at least a subset that spans the full range of our level of understanding
    - Theory
    - Field or experimental validation
- Determine whether there are established methods for modeling the process
- Inventory available codes

Table 1 provides a preliminary identification and ranking of a set of important geomorphic erosion/transport processes. The table is organized around several criteria for each process or phenomenon: the degree to which a quantitative framework has been discovered, the extent of calibration or validation efforts, the human effort being devoted to it, and the degree to which the process has been expressed numerically via computer code. For each of these attributes, processes may be classified on a five-point scale, ranging from "in the dark" to "enlightenment" (Table 1). Examples of processes or phenomena for which the state of knowledge is "in the dark" or "faint flame" include the dynamics of bedrock-dominated landscapes, controls on grain-size production and evolution, incision and land sculpture by debris flows, ice erosion, deep-seated landsliding, and chemical denudation. By contrast, processes and phenomena in the "sunshine" to "enlightenment" categories include catchment-scale groundwater flow, small-scale Darcy flow, free-surface and open-channel flow, suspended-sediment transport (when the bed texture is known), annual to decadal ice dynamics, and lithosphere flexure. In all of these cases, one can find existing codes that solve a generally agreed-upon set of equations. A great many processes, however, lie between these extremes. There may be, for example, multiple competing erosion/transport laws (as in the case of bedrock river erosion), limited but growing data sets, and a significant ongoing research effort.

Table 1 illustrates the tremendous breadth of terrestrial processes, as well as some of the significant challenges ahead. In order to meet these challenges, it is essential that members of the community share their expertise and contribute their understanding and models.

	In the dark	**Faint flame**	**Lighthouse**	**Sunshine**	**Enlightenment**
Quantitative framework	None	A few straw-man expressions based on intuition	Multiple competing hypotheses based on observations and measurements	Widely accepted, mechanistic theory has emerged	Solved problem. Universally accepted physical principles
Calibration/validation efforts	None	Initial efforts to calibrate expressions are underway, but no real tests have been performed.	Several calibration exercises have been performed. Initial efforts to test predictions against field or laboratory data are underway.	Parameters have been calibrated for many scenarios. Predictions have been tested against multiple laboratory and field measurements by independent groups.	Moot, except for efforts to measure parameter values for specific sites
Human effort	We know it's important, but almost nobody is working on it	A handful of groups are working on it	Every other group is working on it	A few groups are working to refine the details	No need to work on it. Everyone uses it.
Existing code	None	A few in-house efforts	Many different in-house versions, a few longer-term development efforts, some distributed packages	Community models, widely available commercial packages	Shipped with textbooks
Examples [and names of existing codes/developers, if applicable]	> hillslope grain size production & comminution > large-scale development of bedrock landscapes	> debris flow incision and routing > landscape-scale glacial erosion > long-term overland flow erosion > deep-seated landsliding > chemical denudation > long-term ice sheet dynamics	> bedload sediment transport [Parker, Wilcock, Cui] > bedrock river incision > structural development of orogens > soil production > local (cm to m-scale) glacial erosion > river meandering [Tucker, Lancaster, others] > hydraulic geometry: fluvial channel width and depth	> Catchment-scale groundwater flow [MODFLOW] > free-surface/open-channel flow [Delft3D, MD-SWMS] > suspended sediment transport > short-term (years) ice dynamics	> Lithospheric flexure > small-scale (meters) Darcy flow

> shallow
landsliding
[SHALSTAB]
> debris flow
dynamics
> hillslope
sediment transport
> fluvial sorting
and patch
dynamics
> delta formation

3. Applications of Terrestrial CSDMS models

The processes and feedbacks acting on Earth's surface are richly varied and depend on complex initial conditions and forcing mechanisms. A key goal of CSDMS is to facilitate the development of coupled models that allow previously uncoupled process domains to be linked so that the complex, nonlinear behavior of the Earth surface system can be better understood and predicted. The goal of this section is two-fold: to develop criteria for proof-of-concept problems that illustrate the power of coupled model development and to identify specific proof-of-concept applications that are of high priority to the terrestrial geomorphic community. Our focus is primarily on what scientific questions can be tackled in the short term, which we consider to be a 6-12 month timeframe, that will both advance our understanding of a portion of the earth surface system, demonstrate that CSDMS can produce results, and provide a template for future development of the CSDMS effort.

Any plan must include a strategy for engaging as large a portion of the terrestrial geomorphic community as possible. To this end, we believe it is important to identify problems that draw from a range of existing component models rather than choosing one or two particular component models. Model intercomparison can be a useful means of engaging the broader geomorphic modeling community in these proof-of-concept application efforts. The primary goal of model intercomparison is not necessarily to provide a definitive answer to a scientific question, but, rather, to focus many developers on a focused scientific problem or set of problems in order to explore what techniques work best, solve the inevitable technical challenges that will arise when coupling component models, and facilitate collaboration, especially among scientists who focus on different process domains. Model intercomparison is also essential for validating individual component models.

The proof-of-concept problems and associated coupled models developed by the terrestrial working group within CSDMS should, at a minimum, meet the following criteria:

- The model should integrate at least two separate process domains (components) of the Earth surface system.
- The model should address an issue of widespread interest within the CSDMS community and society as a whole.
- The problem should be well-posed from the standpoint of initial and boundary conditions, and should have a wide range of accessible data with which to verify model results.

In addition, the ideal problems will possess natural lines of future inquiry, some of which may not be feasible at present, but that hold significant promise for new insight with additional data and/or model development. Problems of this sort generally fall into one of two categories – those problems that represent a process-domain transition within the Earth surface community (e.g., hillslope-channel coupling) and those that reach across traditional disciplinary boundaries with other communities. Examples of each are given below, but, in general, the working group discussions highlighted the fact that many of the exciting problems in our discipline lay at the interface between Earth's surface and atmospheric dynamics, lithospheric deformation, and ecosystem behavior. We suggest that the CDSMS Terrestrial working group should focus on two proof-of-concept efforts over the next 6 to 12 months. One of these would ideally focus on relatively short-term (annual to decadal) interactions between process domains, while the other should focus on landscape evolution over long (geologic) timescales and involve the coupling of a landform evolution model with a state-of-the-art atmospheric, lithospheric, or ecological model.

3.1 Short-time-scale problem

The landscape response to intense wildfires poses a societally important, data-rich proof-of-concept problem for a CSDMS model. Post-fire erosion involves complex changes to the hydrological and erosional properties of hillslopes. The hillslope response often delivers a pulse of sediment to downstream reaches that leads to a fill-and-cut cycle and which has strong negative impact on riparian ecosystems and human infrastructure. Modeling this landscape response would require linkage of modules that include:

- Climatic forcing from actual or synthetic data sets
- Runoff/infiltration modeling reflecting the evolving state of regolith and vegetation.
- Regolith detachment by runoff reflecting the evolving state of regolith and vegetation – amount and grain size distribution.
- A model of vegetation recovery (perhaps empirical, but likely highly parameterized).
- Coupling of erosion and vegetation history to regolith state, including armoring, bioturbation, changes in critical shear stress, etc., as appropriate.
- Modules for extreme events, mass movements like landslides and debris flows.
- Routing of sediment through the channel system, including multiple grain sizes, fan and terrace development and incision, and timeline of delivery of sediment to reservoirs (e.g. Gabet, 2003).

The strong perturbation, rapid evolution, and highly coupled nature of the hillslope-channel system offer both opportunity and challenge to model development. Initial models would not necessarily involve full coupling of models of vegetation growth and recovery with physical process models, but the opportunity is there for future development of this sort. An important aspect of such a model will be to explore the sensitivity of the response time of the system as a whole to the disturbance imposed by such an event, and how this depends upon the response times of each of its components. At some level (to be determined by the time and length scale of the model) this model requires that the geomorphic modeling community reach out to the ecological modeling community.

We suggest that data from fires and their aftermath in Southern California over the last 50 years may be employed. This includes sedimentation rate in manmade sediment basins, debris flow generation, etc. (e.g. Lave and Burbank, 2004). Large fires in forested or chaparral landscapes result in significant short-term hazards from increased sediment yield, more frequent and larger discharges from a given precipitation event, and hillslope instability as landslides/debris flows. In some settings, such events may comprise the majority of sediment input to the fluvial system and dominate the landscape morphology. Large-scale, post-event data collections efforts in places such as the San Gabriel Mountains in California and after the 1996 and 2002 forest fires in the Colorado Front Range form reference databases for model development and validation. Data may include:

- Vegetation changes through time
- Documentation of hillslope channeling and soil loss
- Analysis of regolith properties in affected and unaffected comparative sites
- Movement of the sediment wave through the fluvial system, including volumes and grain size distribution
- Deposition in reservoirs, including grain size distribution
- High-resolution topography from aerial photography, LiDAR, or TLS, preferably as a time-lapse.
- Timelines of sediment yield, vegetation recovery, routing of the sediment pulse
- Precipitation history from local or nearby weather stations

3.2 Long-time-scale problem

Some of the most exciting problems in geomorphology involve the history of large-scale landscapes over millions to tens of millions of years. Tectonic geomorphology is informing our understanding of the tectonic history of mountain belts, and there is a growing appreciation that mountain belts can incite complex feedbacks between uplift, erosion, and climate change at a wide range of spatial and temporal scales. To better understand these feedbacks, it is essential to develop models that include the geophysical and atmospheric processes involved in the evolution of landscapes. At the scale of a mountain range, it is individual faults whose slip generates a rock-uplift pattern on which the geomorphic processes act. Faults redistribute mass in the upper crust and therefore incite flexural deformation of the lithosphere. In addition, on long time scales the growth of significant topography affects the flow of the atmosphere and the resulting distribution and phase of precipitation. Such models therefore require linkage of modules that include:

- Elastic dislocation along prescribed faults
- Flexural accommodation of changing load
- Viscous deformation of the substrate that sets a time scale for flexural-isostatic adjustment
- Orographic precipitation
- Hydrology
- Hillslope processes
- Fluvial bedrock incision

The development of coupled landscape-lithosphere and/or landscape-atmosphere models encourages a connection to ongoing efforts to develop such models outside of CSDMS, including the Computational Infrastructure for Geodynamics (CIG) efforts. Specific field areas where these models could be focused include the Sierra Nevada and areas of the Basin and Range or the Himalayas. In these areas, a wealth of data currently exists to calibrate and validate models including:

- Topography
- Seismically constrained stratigraphy
- Thermochronology, radioisotope and cosmogenic isotope dating
- Viscous times scale from deformation of Bonneville shorelines (e.g. Bills et al., 1994)
- Closed basin in the Great Basin allows closure of sediment budget
- Knowledge of the range of climates in the Quaternary (from Last Glacial Maximum versus present water budgets in pluvial lakes, glaciers)

3.3 Model Calibration, Validation, and Uncertainty Estimation

The engagement of the terrestrial community in model intercomparison projects provides a useful opportunity to advance the techniques we use for calibrating and validating process-based numerical models in geomorphology. Currently, model calibration and validation is, to a large extent, a process of trial and error that does not take into account uncertainty in the input data and, hence, does not quantify uncertainties in model outputs. Ideally, the proof-of-principle applications that the terrestrial working group of CSDMS focuses on will involve the testing of new techniques that have been developed in the Earth science modeling communities for improved model calibration, validation, and uncertainty estimate. To take one example of these new techniques, Markov Chain Monte Carlo (MCMC) algorithms for parameter optimization and uncertainty estimation are “adaptive search” algorithms that mimic the processes of biological evolution (random mutations and fitness selection) in order to determine optimal parameter sets for complex, nonlinear systems with multiple types of output (e.g. discrete and continuous, point-based and spatially-distributed) (e.g. Vrugt et al., 2003). The uncertainties of “known” parameters are propagated through these algorithms to provide quantitative estimates of the uncertainty in the “unknown” or inferred parameters. The hydrology community has been successful in using of MCMC algorithms for inferring spatially and temporally distributed input data (e.g. hydrologic conductivity, rainfall intensity) given a hydrological model and observational data (e.g. station hydrographs). Now is an ideal time to improve the model inference protocols currently used by the geomorphic community and CSDMS is ideally suited to lead that effort. The newly acquired supercomputer at the CSDMS Integration Facility will be essential to this effort.

3.4 Linkage with Educational and Knowledge Transfer Working Group

It is important that the proof-of-principle projects developed in CSDMS be quickly disseminated in the form of animations, interactive Java-based simulations, and curriculum materials. The EKT working group has expertise in the development of these educational and outreach activities. Wei Luo, for example, has developed a Java-based landform evolution model, WILSIM, that teaches students about river incision and drainage basin self-organization. Wei’s model, and the materials that support it, have reached hundreds of thousands of students worldwide. Proposals are currently in review to expand these activities, and the results of prototype CSDMS models should be distilled into animations, interactive inquiry-based learning modules, and curricular materials for use by undergraduate educators.

3.5 Longer-term Research Topics

Our recommendation for criteria for proof-of-concept problems and associated coupled models can be carried over to longer-term goals (i.e. not prioritized for the next 6-12 months, but for 2010-2012). The first criterion that models should integrate at least two separate process components of the Earth surface system, provides a number of key coupling problems that would yield new insights:

- glacial-fluvial transition (example: melt on the Greenland Ice Sheet propagated into the fluvial systems or retreating glaciers over a glacial cycle and its impact on the local valley morphology).
- ecology – land surface processes (for example, the role of vegetation interacting with weathering and erosion processes during landform evolution in temperate and humid landscapes)
- terrestrial-coastal transition (for example, the role of sea level changes on landform evolution in coastal regions)
- morphological transitions: can we combine meandering river models and braided river models to dynamically transition when controlling conditions change.

The second criterion is that the model should address an issue with widespread interest within the CSDMS community and society as a whole. A number of key challenges were posed:

- predictions of changes in earth surface dynamics looking forward to 2050 and 2100 by coupling to CCM efforts
- coupling of terrestrial-hydrological earth surface models to policy tools (e.g. floodplain risk mapping)

4. Computational Challenges and Needs

4.1 Mitigating Risk

One way to support the success of a complex effort like CSDMS is to envisage potential "failure modes": risks to the project that can be avoided if proper steps are taken. The Terrestrial Working Group identified a set of twelve potential risks, and highlighted ways to avoid them.

Risk 1: Poor Quality Control – Currently, the CSDMS model repository has a low threshold for participation: a developer only needs to request that his or her model be listed. CSDMS is not currently evaluating the quality of contributed models. This runs the risk that users may be disappointed in not finding the information they require about model suitability, past performance, etc. At one stage in the development, a quality-control "pyramid" had been envisioned, from donated, caveat-emptor software at the base to fully tested/validated code at the top. The implication at the time was that working groups would play an evaluative role, but at the meeting the practical limitations of such an approach were noted.

Several measures were recommended to address quality-control issues. Providing information on the frequency or number of downloads would give some indication of popularity to prospective users; while popularity does not necessarily correlate with quality, it does provide an indication that a particular code is in active use. Model contributors should also be encouraged to provide analytical test cases and/or unit tests (which are also useful for checking compiler dependency). These do not test the applicability of the code to any particular natural phenomenon or target problem, but they do demonstrate robustness of the numerical algorithms with regard to the underlying equation set. In addition, it is recommended that the web site include standard flags/symbols/icons that would indicate the degree to which a particular code is CSDMS-compliant, provides standard test cases, or provides actual field data for testing. It is also recommended that developers/contributors be encouraged to provide references to literature in which the model is applied, described, or tested. Finally, a user discussion forum for models is recommended, as this would promote sharing of information about a model as well as problems encountered, frequently asked questions, etc.

Risk 2: Poor Documentation – Any model can be used badly. The probably that a particular model, tool, or component will be used inappropriately is greatest when the documentation supporting it is weak or nonexistent. Currently, the standard model-submission form does ask for documentation. In addition, the OpenMI interface provides methods for exchanging meta-data among components, so to some extent this problem will be addressed by developers who implement OpenMI interfaces. Aside from these steps, the Working Group recommended adopting a “wait and see” approach to this risk.

Risk 3: The Complexity of CCA will Discourage Users and Developers – The Working Group acknowledged that the CCA tool-chain is indeed rather complex – challenging for computationally oriented geoscientists, and possibly daunting for others. The recommended solution is to shield users from the full CCA system by allowing most users to rely on a suitably modified version of the CCaffeine graphical user interface, which is much simpler to work with than the complete CCA system. The full CCA tool-chain would still be available, but would only be necessary for certain high-level operations.

Risk 4: It will be difficult to transition from simply listing models to hosting/encouraging simulations – At present the CSDMS Models web site is primarily a listing service. The supercomputer is seen as something that will attract users. In addition, the possibility of a “build server” for download configurations is suggested.

Risk 5: CSDMS may make it difficult to do “offline” modeling work – To the extent that the key tools and models are centralized on computers hosted by the CSDMS Integration Facility, it may become more difficult for users to work “offline” on their own platforms. However, even if this were to prove true, it was not seen as necessarily a bad thing. Climate modeling has, for example, brought about a slight change in the mode in which climate scientists operate.

Risk 6: CSDMS models and tools will be mis-applied – This is particularly a risk when a “modeling environment” like Ccaffeine makes it possible for naïve users to connect incompatible models, leading to “garbage in, garbage out.” In fact, this is a risk in science in general. The Working Group noted that there is little that can be done about this risk, apart from offering training and ensuring that documentation and bibliographies are readily available (as noted above).

Risk 7: CSDMS will become dependent on the success of outside initiatives and organizations – CSDMS is already adopting products and methods from projects such as OpenMI, CCA, and even Java, and in some cases this has slowed development (for example, waiting for critical bug fixes in CCA). This represents a tradeoff between risk and efficiency: the risk could be avoided by creating a new set of model interaction protocols and tools, but that in turn would bring its own risk in the form of increased development time. The recommended solution is to choose to rely only on projects that are well established and have solid support (like CCA and OpenMI) and to try as far as possible to become engaged with their personnel.

Risk 8: Coupling models and components will lead to problems in conservation of mass, momentum, and/or energy, especially with dissimilar grids – Interpolation methods can lead to information loss, particularly when interpolation is repeated. Several complementary solutions are recommended: (1) work on good re-mapping tools (or incorporate existing ones), (2) encourage use of models with similar grid structures, (3) incorporate parallel remapping tools (e.g., ESMF, MCT).

Risk 9: Overwhelming data volumes – There is potential for increased volumes of computer-generated data. There are a number of potential issues that can arise. When grids become very large, performance of some numerical algorithms can become very poor. In addition, large data sets involve increasingly large volumes of memory that must be addressed and accessed, and may involve reduced performance even beyond the limitations of a numerical algorithm simply due to frequent I/O operations. There are several potential solutions. Recent work in the GIS and computer science communities has led to I/O-efficient GIS algorithms, and similar approaches could be applied to common numerical algorithms. Partnering with these groups is recommended. In addition, the Working Group noted that there are special funding opportunities for petascale computing.

Risk 10: The community may be reluctant to transition to HPC – High-performance computing requires expertise that many geoscientists do not possess, and there is a risk that few in the community will feel motivated to make the necessary investment. Potential solutions include coding camps, demonstrations of "success stories," and strong technical support.

Risk 11: Traditional supercomputer-based HPC will be superceded by new technologies like cloud and GPU-computing – It is possible that a significant investment in traditional HPC will strike a dead end as scientific computing shifts toward cloud-based (e.g., Amazon, EC2, Eucalyptus) and graphical-processor-based computing. The Working Group noted that it is notoriously difficult to predict the direction that technological innovations will take. At present, supercomputer and cluster-based computing have a solid foundation in the sciences. The best way to mitigate this risk therefore is simply to keep an eye on developments and be prepared to adapt.

Risk 12: The CSDMS/CCA communication layer is a black box that will be difficult for users to manage – Some commercial attempts at common, modular development environments have failed (e.g., Mozilla's object model; ESRI). Recommendations to avoid problems arising from a complicated "black box" include: requesting different degrees of diagnostic output (for debugging/development), encouraging use of transparent code (e.g., option to write netCDF output), working top-down (e.g., initially wrapping an entire model and only later breaking into components), and inter-model comparison testing. It was noted that the target audience is reasonably sophisticated in terms of scientific model development. One recommendation is to provide support in the form of a support@csdms.edu email hotline.

4.2 General Recommendations on Modeling and Software Development Issues

Several additional ideas and recommendations emerged from the Working Group's deliberations on the topic of modeling and software development:

- A pragmatic approach is to work top-down: begin by wrapping ("RTF'ing") whole models, and split as/when needed.
- Be on the lookout for code duplication
- Encourage swapping of modules as cross-checks of models (a form of inter-model comparison)
- Provide "support@csdms.edu"
- In prioritizing models, begin with an overview of main processes; give modules usable names like "1d flow," "turbulence closure," etc.
- Provide coding camps
- Include "stick figure" cartoons in manuals/guides

4.3 Recommended Components and Toolkits

The Working Group noted that, in addition to process codes, there are many types of utility software that will be critical for some aspects of CSDMS. These include terrain-modeling tools that are common in GIS packages, for performing operations such as computation of slope, aspect, curvature, and calculation of watershed and drainage pathways. They also include a wide range of hydrologic and hydrodynamic modeling codes (for example, 2D shallow-water equation solvers). Ultimately such components should be HPC compatible. **Figure 2** illustrates some fundamental components.

CSDMS should transition over the next few years from being primarily focused on coupling models to providing a framework supporting model construction from the ground up, with basic pre-existing software components to handle common tasks such as terrain representation and stratigraphy.

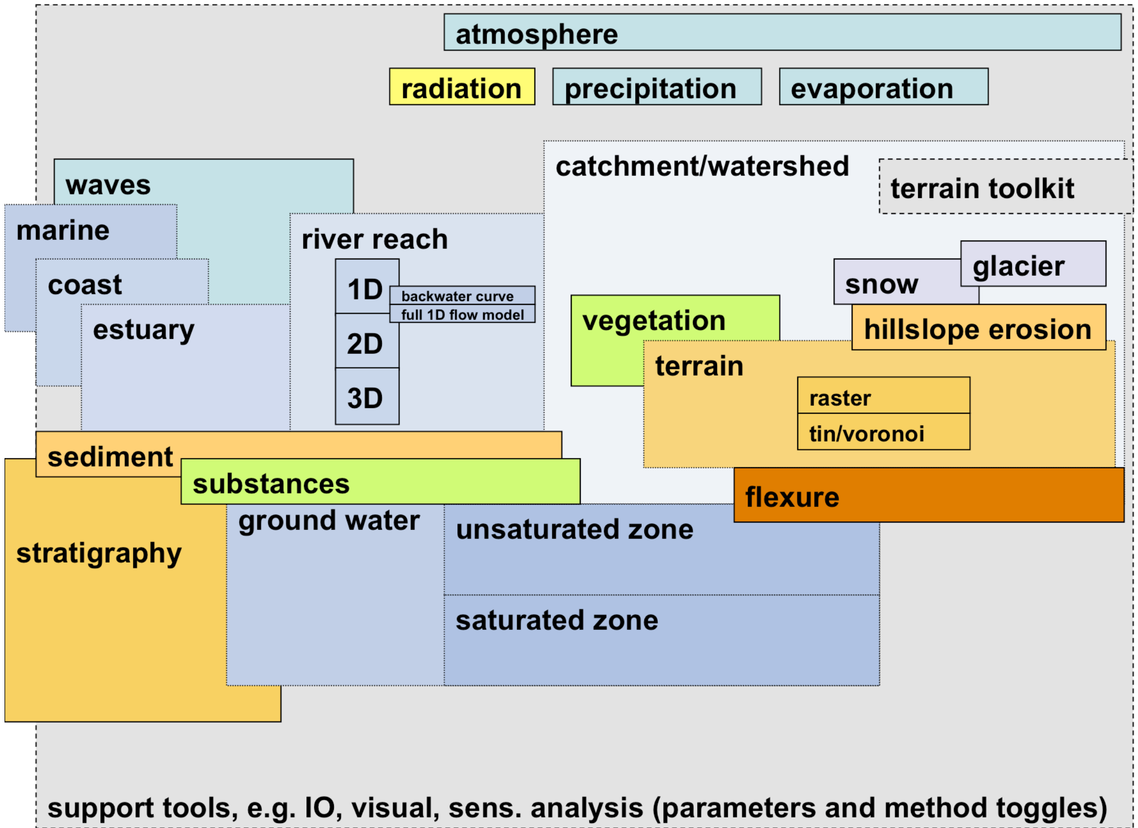


Figure 2: Schematic illustration of some common components.

4.4 Categories of Modules

Modeling Needs

The plug & play components that ultimately will make up the CSDMS library will be a mix of complete models, tools and individual algorithms. There are many ways to categorize this hierarchal and overlapping mix of components, but the system advocated here reflects the typical categories anticipated for terrestrial modeling needs. In order to highlight model development needs, a break-out group from the Terrestrial Working Group February 2009 Meeting highlighted that a more thorough inventory of what available open-source components already exist is needed. To effectively build that inventory, it was suggested that a more comprehensive list of open-source tools could be established. One of the key oversights of the existing 'model lists' are that the meta-data necessary to determine OpenMI , CCA and CSDMS compliance (or potential for compliance) is not explicitly collected. **From this inventory, the sub-group hopes to be able to identify where effort should be invested over the next year to build a basic core library of components for building terrestrial models.**

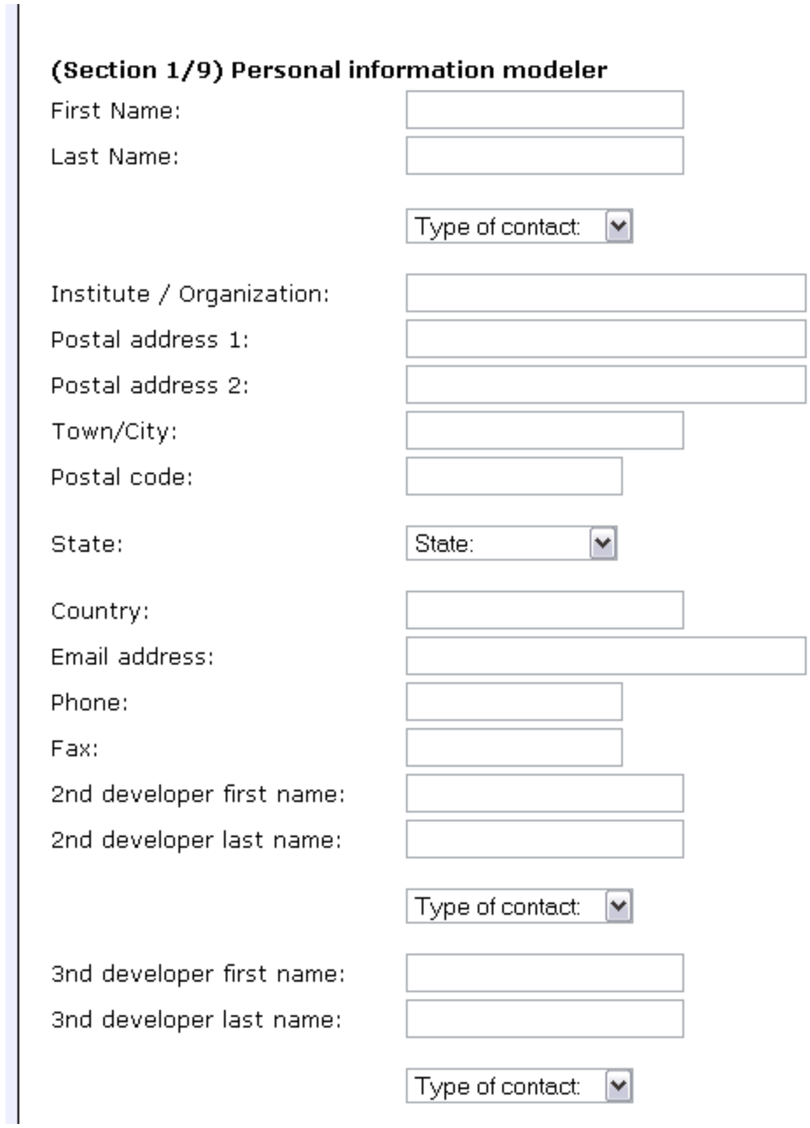
List of Categories for Modules

There is a list of basic attributes for all modules that should establish basic meta-data for the code and ultimately be used to determine compatibility and interoperability. Many of these attributes are already well defined through the CSDMS 'Model Questionnaire': http://csdms.colorado.edu/wiki/index.php/Models_questionnaire. The attribute categories include: 1) Personal Information on Modeler, 2) Model Identity, 3) Technical Information, 4) Input/ Output Description, 5) Process Description, 6) Model Testing, 7) User Groups, 8) Documentation, 9) Additional Comments. This questionnaire is currently targeted at just the authors of the code. Despite the richness of information solicited, the lists are currently only queried by a single field (model domain). **Here we would like to extend that attribute list to (a) serve as a more comprehensive survey and inventory of what is available and exists (not necessarily just entered by the developer); and (b) allow users and CSDMS integration personnel to be able to perform more sophisticated queries of a more comprehensive list to aid in model integration and helping highlight most pressing community needs.** The questionnaire can easily be modified to include these additional attributes in the module database (goal a). However, the Wiki may not be the best web-tool available for allowing users to query the database.

As a matter of semantics, the existing questionnaire allows users to 'contribute their model, tools or algorithms', but solicits these contributions under the banner of 'model'. It might be clearer and more accurate to call this the module or component questionnaire, wherein a complete model, tool or algorithm can still be contributed. Also, there are many tools and libraries potentially available that may be appropriate for use in CSDMS terrestrial models, but that are not authored by members of the CSDMS community. As such, the questionnaire needs to allow members of the CSDMS community to upload components to the database that are authored by someone else (provided they are open-source). So that the process of filling out the questionnaire is educational, we strongly suggest hyperlinking any possible terms or names in the survey (with pop-up windows) to a website with more information on that term (e.g. clicking on "CCA" could provide a link to the Common Component Architecture website).

As the attribute categories already defined in the 'Model Questionnaire' are already a logical starting point, this document and suggested changes and additions are organized around the nine existing categories. The subsections that follow make specific recommendations for the modification and expansion of these categories. It is intended that this will provide a clear workflow for modifying the form, database and how that information is accessed via the CSDMS website. For each category, a screen shot of the existing 'Model Questionnaire' is provided as a starting point. The key recommendations are bolded in each section.

Personal Information on Model


(Section 1/9) Personal information modeler

First Name:

Last Name:

Type of contact:

Institute / Organization:

Postal address 1:

Postal address 2:

Town/City:

Postal code:

State: State:

Country:

Email address:

Phone:

Fax:

2nd developer first name:

2nd developer last name:

Type of contact:

3nd developer first name:

3nd developer last name:

Type of contact:

Figure 1 – Existing Section 1 Fields

Section 1 is very logical for components being submitted by the code author or member of the development team. However, for those components not actually submitted by the author, the form makes no provision. **We suggest adding field(s) to list the name and contact details of the actual individual submitting the information.** For user friendliness, a check box can be provided, which 'Use Same Information as Primary Model Contact' to allow author contributors to avoid entering information twice (when checked it grays out the 'Submitted by' fields. Sub-headings should clearly distinguish between the information solicited that describes the personal information of those involved in the module itself, from someone simply recommending a module. Given these changes, it might make sense to change the section heading from 'Personal Information Modeler' to simply 'Contact Details'.

Model Identity

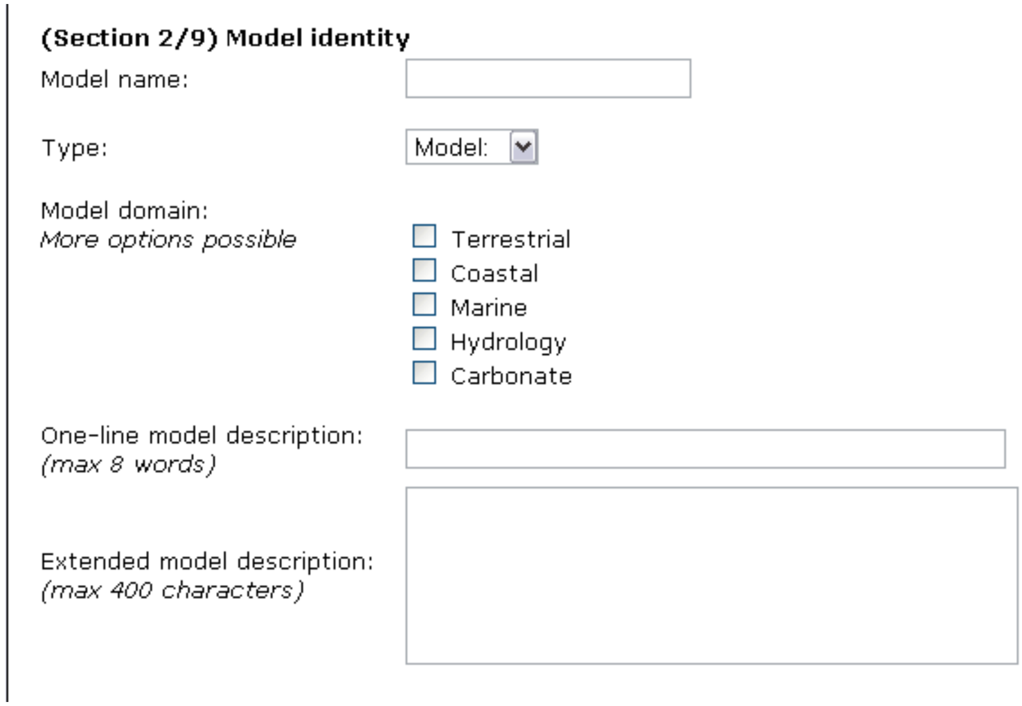


Figure 2 – Existing Section 2 Fields

The model identity is where we propose the most comprehensive additions. These additions are primarily specific to the Terrestrial and Hydrology model domains, but may serve as a template for similar changes by the other working groups. First, the **following changes are suggested to the existing fields**:

- Change ‘Model Identity’ section title to ‘Module Identity’
- Change ‘Model Name’ to ‘Module Name’
- Change ‘Type’ options from a) model, b) tool, c) single, and d) modular (not clear what these all are) to: a) Model (stand-alone), b) Pre-Processing Software, c) Post-Processing Software, d) Project Management tools, d) Visualization Software, e) Analysis & Generic Algorithm Tools, f) Process Subroutine/Function, g) other.
- Include some notes as to what the types are defined by
- Change ‘model description’ to ‘module description’.

Based on the user selections for module domain and type, a different set of metadata categories should be provided.

Generic Fields to Be Added

A variety of generic fields could be added to this section to help distinguish between module contributions of any type. These include spatial scale, spatial dimensions and temporal scale. These are described in the following sub-sections.

Spatial Scale & dimensions

Under spatial scale, spatial extent, resolution and dimensions need to be defined.

For a check-box field ‘Spatial Extent’ of module, the following options might apply (admittedly fluvio-centric) (note: multiple categories allowed):

- € Global
- € Continental (order 1,000 km)
- € Regional-Scale (order 100 km)
- € Landscape-Scale (order 10 km)
- € Watershed-Scale (order 1 km)
- € Reach-Scale (order 100m)
- € Patch-Scale (order 1-10m)
- € Grain-Scale (order 0.00001 to 1m)
- € Point-Based

Check box field of ‘spatial dimensions’ of module:

- € 1D (e.g. profiles)
- € 1.5D (e.g. 2D projections extracted from 1D profiles)
- € 2D (e.g. a DEM grid with one value of z for every x-y location)
- € 3D (e.g. multiple z values possible for every x-y location)

Check box field of ‘spatial resolution’ of module, or fill-in text box with “typical computational element size.”

Temporal Scale

The following temporal fields need to be defined:

- A checkbox filed for ‘Temporally Dynamic Model?’, with options a) Steady-State, b) Dynamic, c) Time evolving
- If ‘Temporal Dynamic Model’ is true,
  - a text field for ‘Temporal Resolution (i.e. range of possible time-steps)
  - a radio button for ‘Time Step type’ – Fixed, or Variable
  - a text field for Temporal Extent (i.e. range of possible simulated model durations)

Fields to be added based on module domain

Based on whether the user filling out the questionnaire selects terrestrial, coastal, marine, hydrology or carbonate, a different selection of ‘process algorithms’ may be appropriate to display. For this document, we only address those that might be appropriate to Terrestrial or Hydrology (there is likely to be some overlap). The other working groups may wish to undertake a similar exercise.

Fields to be added based on module type

This is one of the most critical fields collected for determining the utility of the submitted module. The next eight subsections, describe fields that should be added to discriminate common attributes of those specific tools.

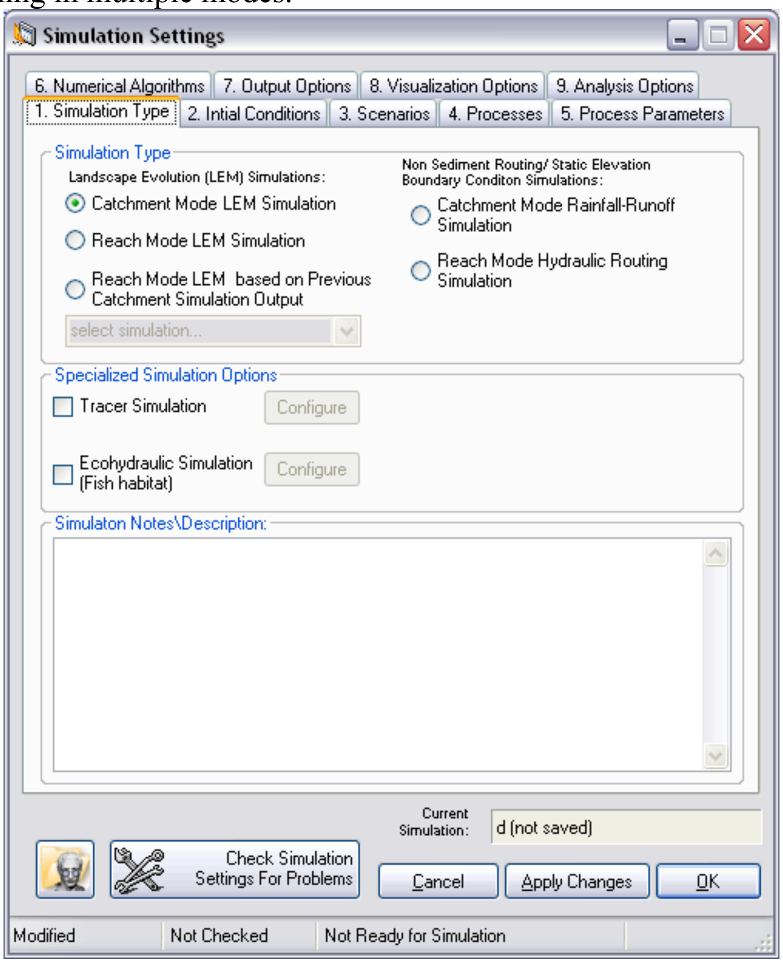
Model (stand-alone)

From a terrestrial perspective, a model types field should be added with a check-box selection provided from the following primary fields:

- € Landscape Evolution Model
- € River channel morphology model

€ Morphodynamic Model
€ Soil-erosion Model
€ Eolian Model
€ Hillslope process model
€ Hydrologic Model
€ Hydraulic Model
€ Groundwater Model
€ Other _________

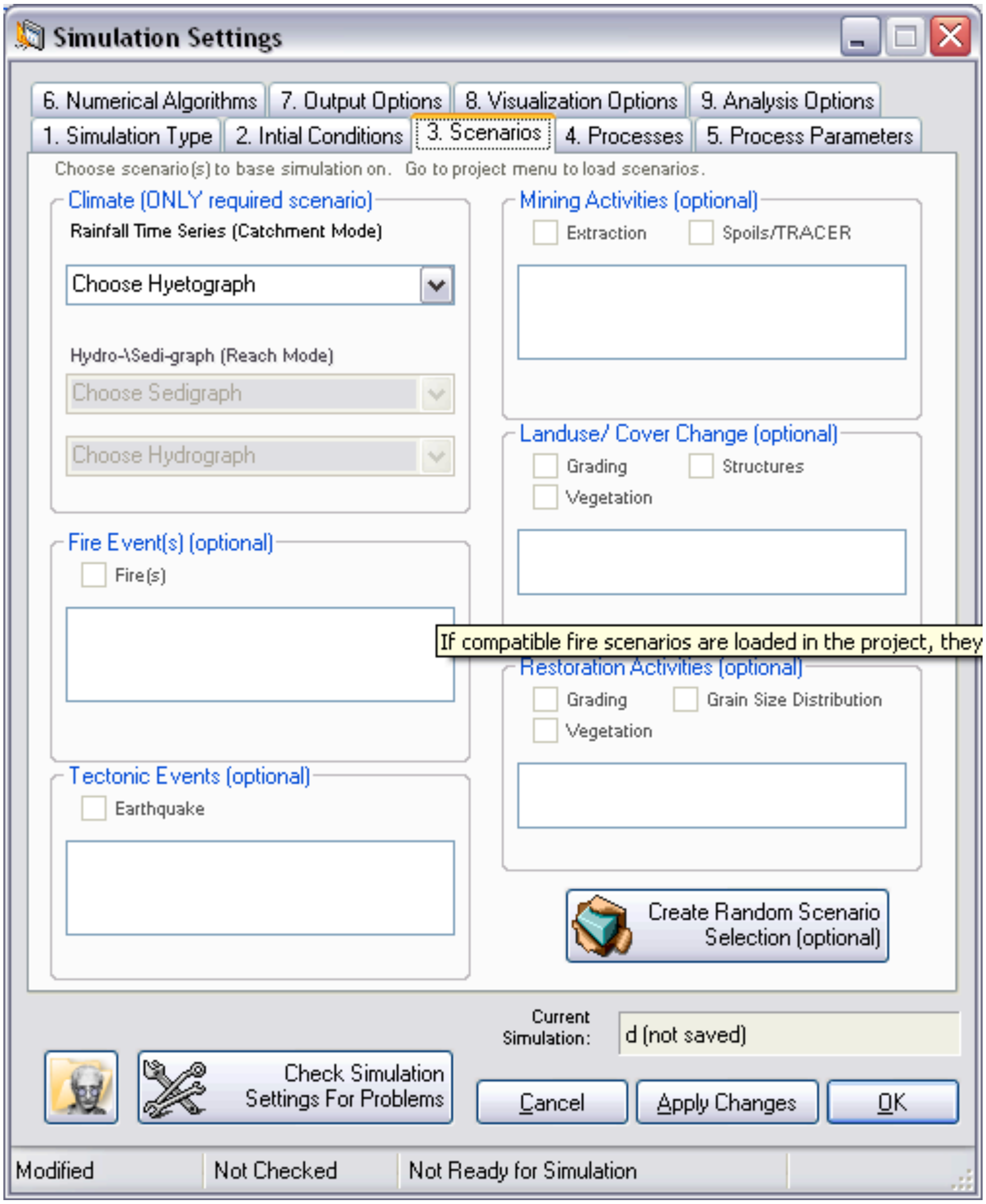
It should be a checkbox selection, as some models (e.g. below) are capable of running in multiple modes.



There might be an additional checkbox field called 'model style' with choices of:
€ Spatially-Distributed (Raster-Based)
€ Spatially-Distributed (TIN-Based)
€ Spatially-Distributed (Unstructured Grid)
€ Spatially-Distributed (Agent-Based)
€ Lumped
€ Schematic

Pre-Processing Software

Most dynamic terrestrial models have some sort of scenario drivers. For hydrologic models, this is a hyetograph, for hydraulic models, a hydrograph. Some of these scenarios may be based on continuous time-series data (e.g. rainfall and streamflow), whereas others may be based on discrete events (e.g. earthquakes, fires, etc.). For post-diction modeling simulations these may be based off actual data, but for many post-diction and prediction simulations these may be entirely synthetically produced scenarios (e.g. IPCC climate change scenarios). There are generic tools for preparing such scenarios, which may be usefully submitted as modules in CSDMS. A screen shot from ooCAESAR below illustrates these concepts.



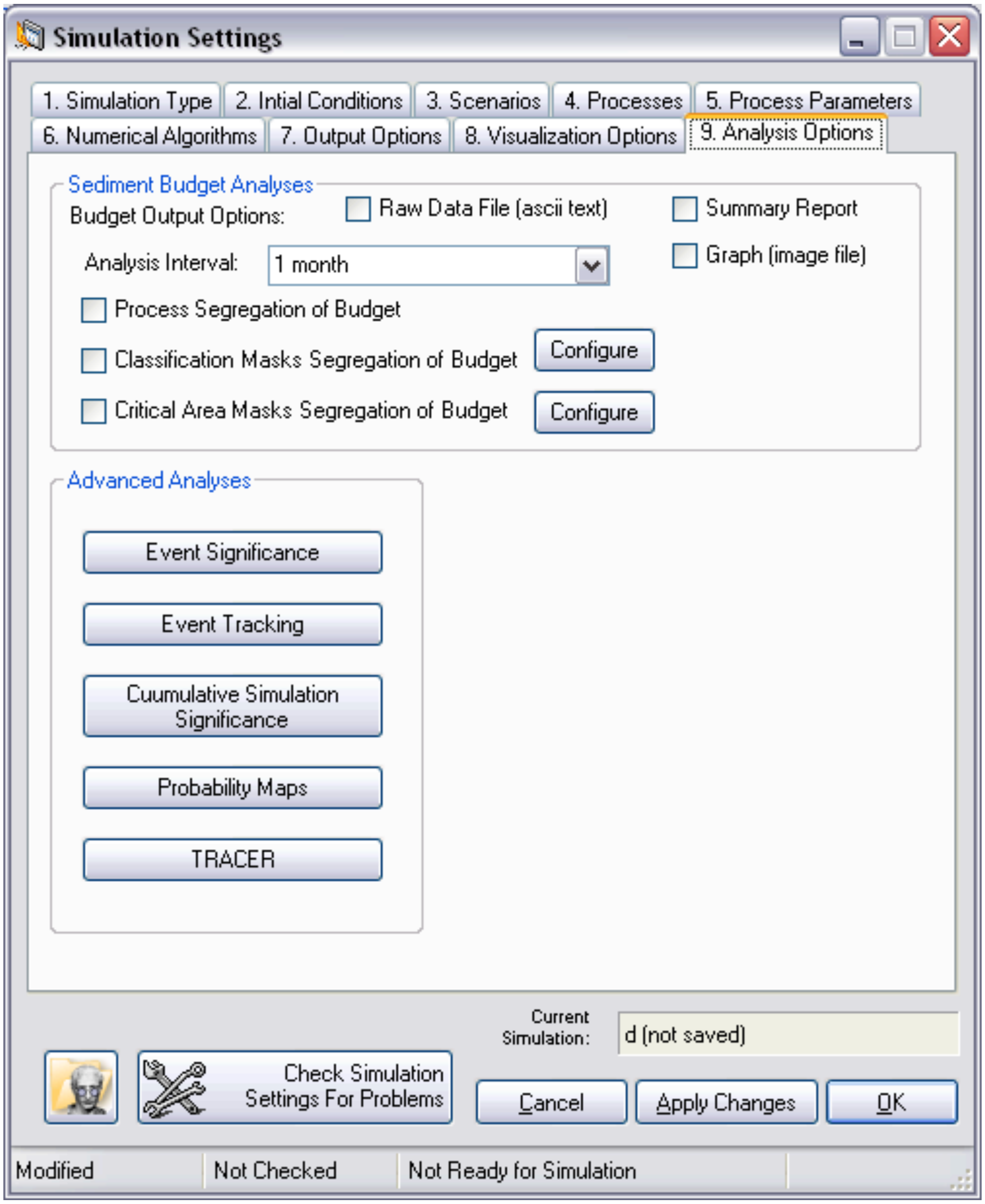
It may make sense to define the following fields for Pre-Processing Software for scenarios (allowing "both" as a possibility):

Scenario-Type: a) discrete (events), b) continuous

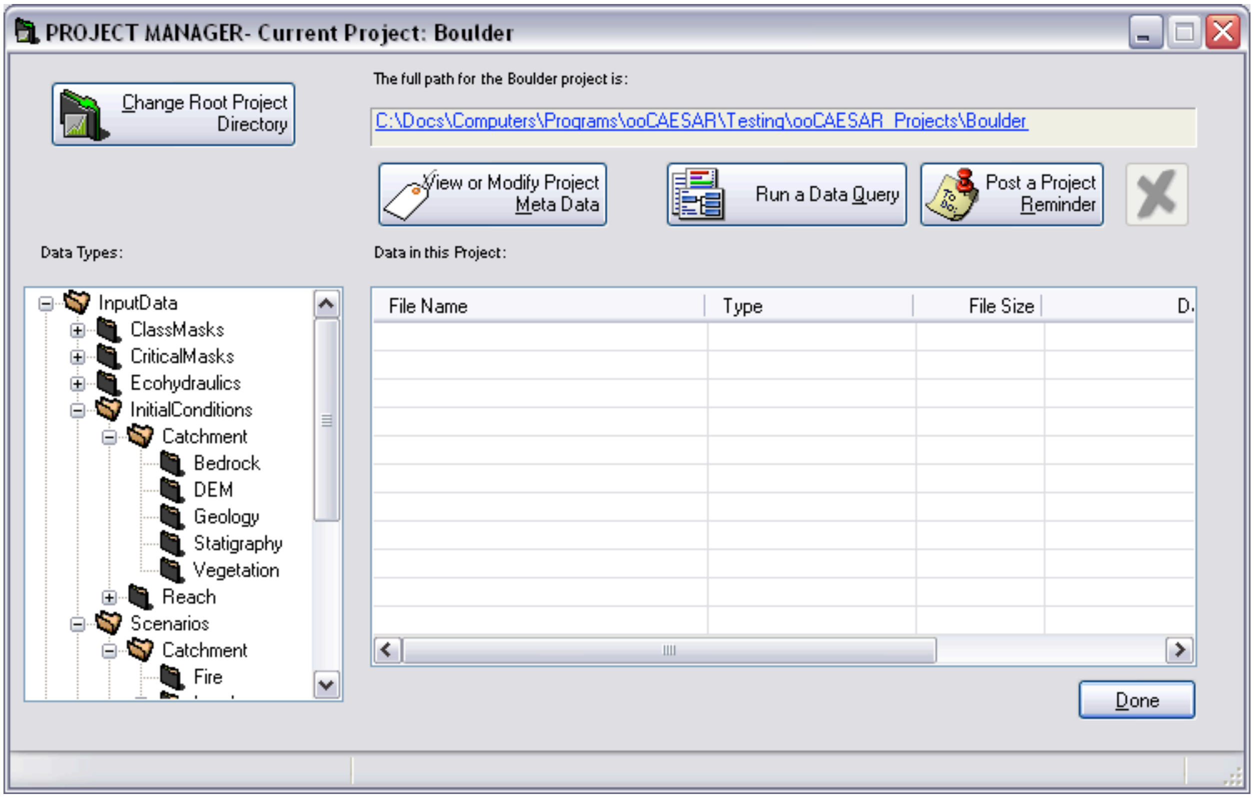
Other Pre-Processing Software fields might include:

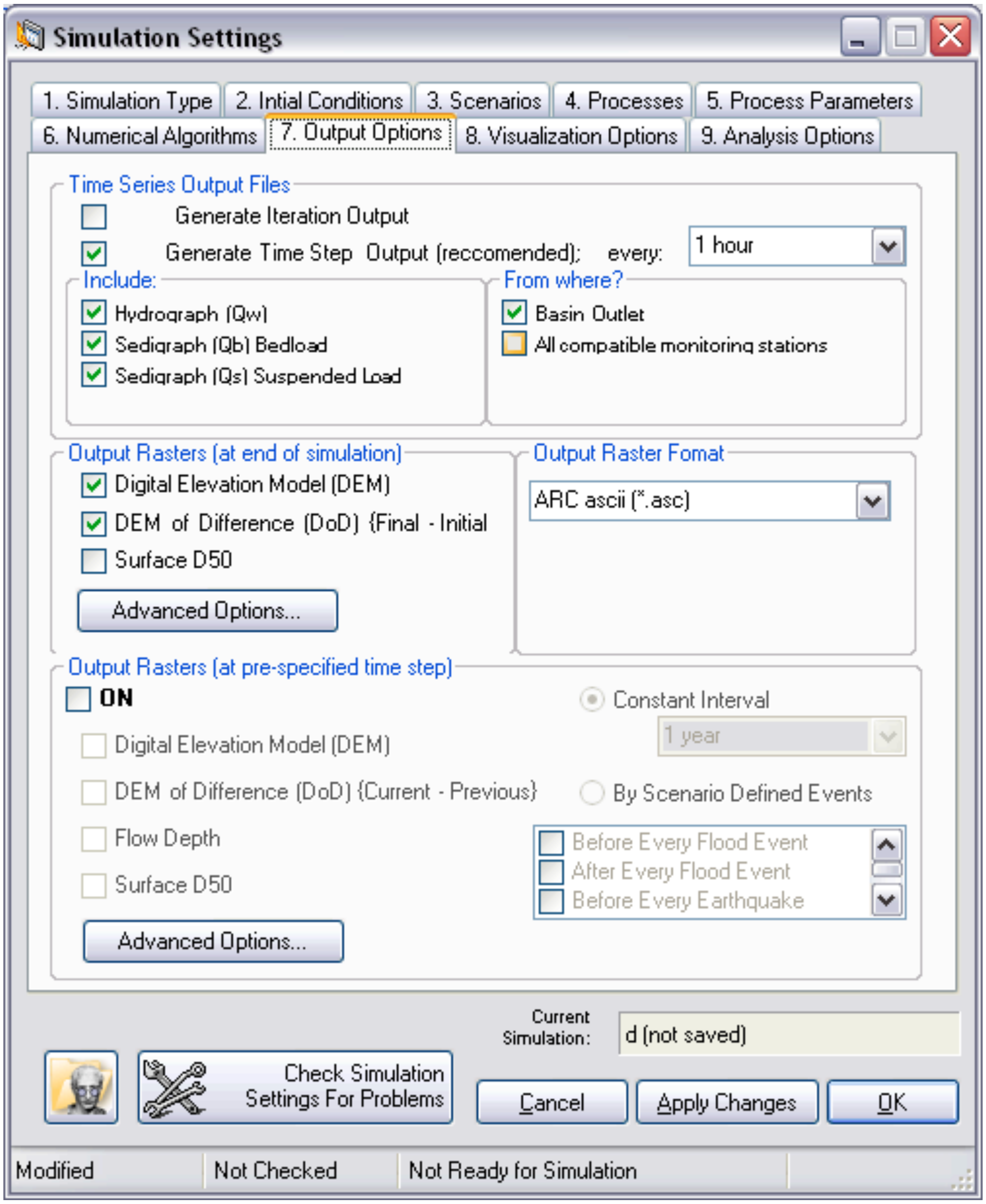
- Pre-Processor Type: a) Scenario Preparation, b) Data Conversion, c) Parameter estimation, d) Grid Construction, e) Boundary Condition

Post-Processing Software

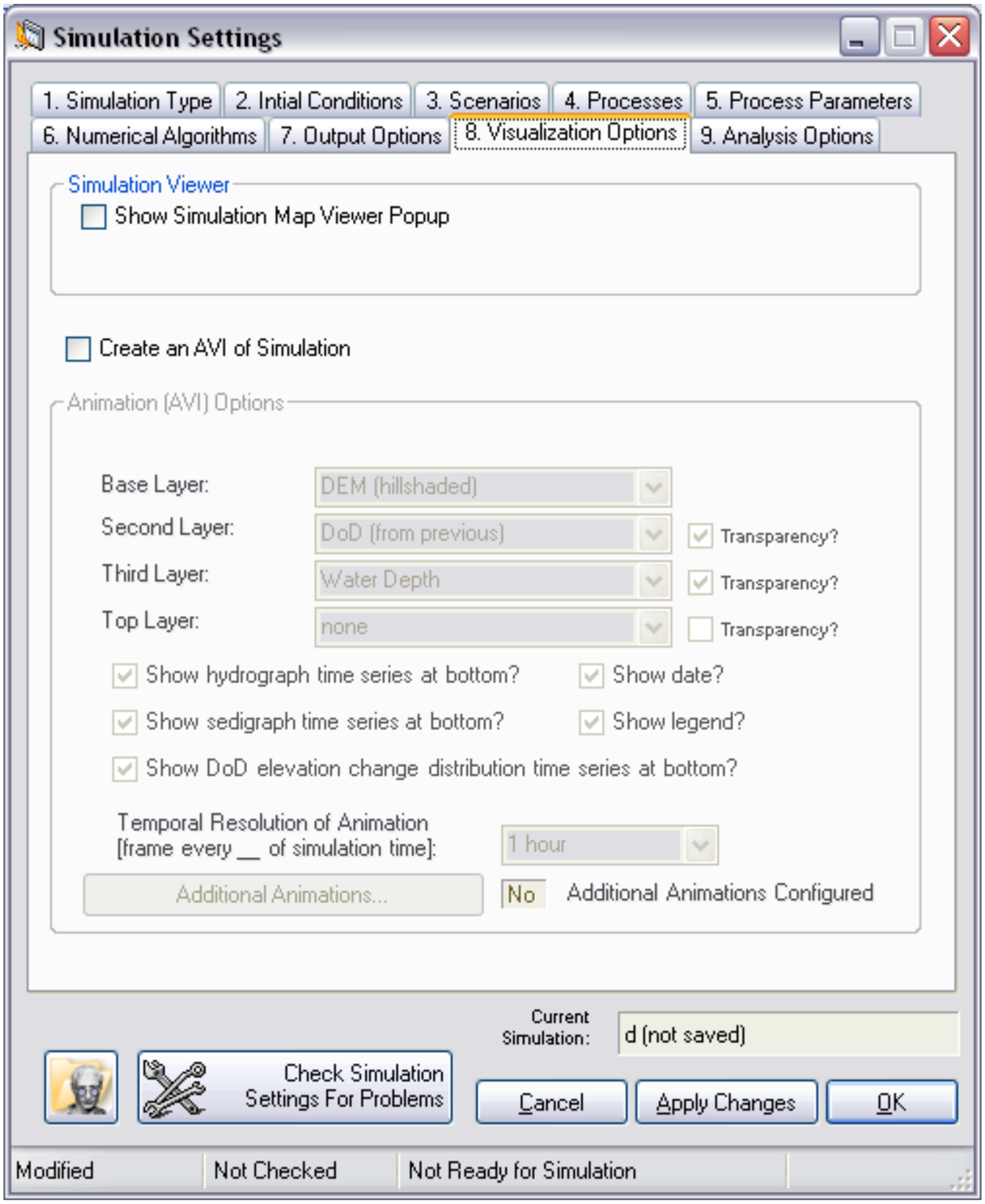


Project Management tools


PROJECT MANAGER- Current Project: Boulder
Change Root Project Directory
The full path for the Boulder project is:
C:\Docs\Computers\Programs\ooCAESAR\Testing\ooCAESAR_Projects\Boulder
View or Modify Project Meta Data
Run a Data Query
Post a Project Reminder
Data Types:
Data in this Project:
InputData
ClassMasks
CriticalMasks
Ecohydraulics
InitialConditions
Catchment
Bedrock
DEM
Geology
Statigraphy
Vegetation
Reach
Scenarios
Catchment
Fire
File Name
Type
File Size
Done


Simulation Settings
1. Simulation Type
2. Intial Conditions
3. Scenarios
4. Processes
5. Process Parameters
6. Numerical Algorithms
7. Output Options
8. Visualization Options
9. Analysis Options
Time Series Output Files
Generate Iteration Output
Generate Time Step Output (reccomended); every:
1 hour
Include:
Hydrograph (Qw)
Sedigraph (Qb) Bedload
Sedigraph (Qs) Suspended Load
From where?
Basin Outlet
All compatible monitoring stations
Output Rasters (at end of simulation)
Digital Elevation Model (DEM)
DEM of Difference (DoD) {Final - Initial
Surface D50
Advanced Options...
Output Raster Fomat
ARC ascii (*.asc)
Output Rasters (at pre-specified time step)
ON
Constant Interval
1 year
Digital Elevation Model (DEM)
DEM of Difference (DoD) {Current - Previous}
By Scenario Defined Events
Flow Depth
Before Every Flood Event
After Every Flood Event
Before Every Earthquake
Surface D50
Advanced Options...
Current Simulation:
d (not saved)
Check Simulation Settings For Problems
Cancel
Apply Changes
OK
Modified
Not Checked
Not Ready for Simulation

Visualization Software



Analysis & Generic Algorithm Tools

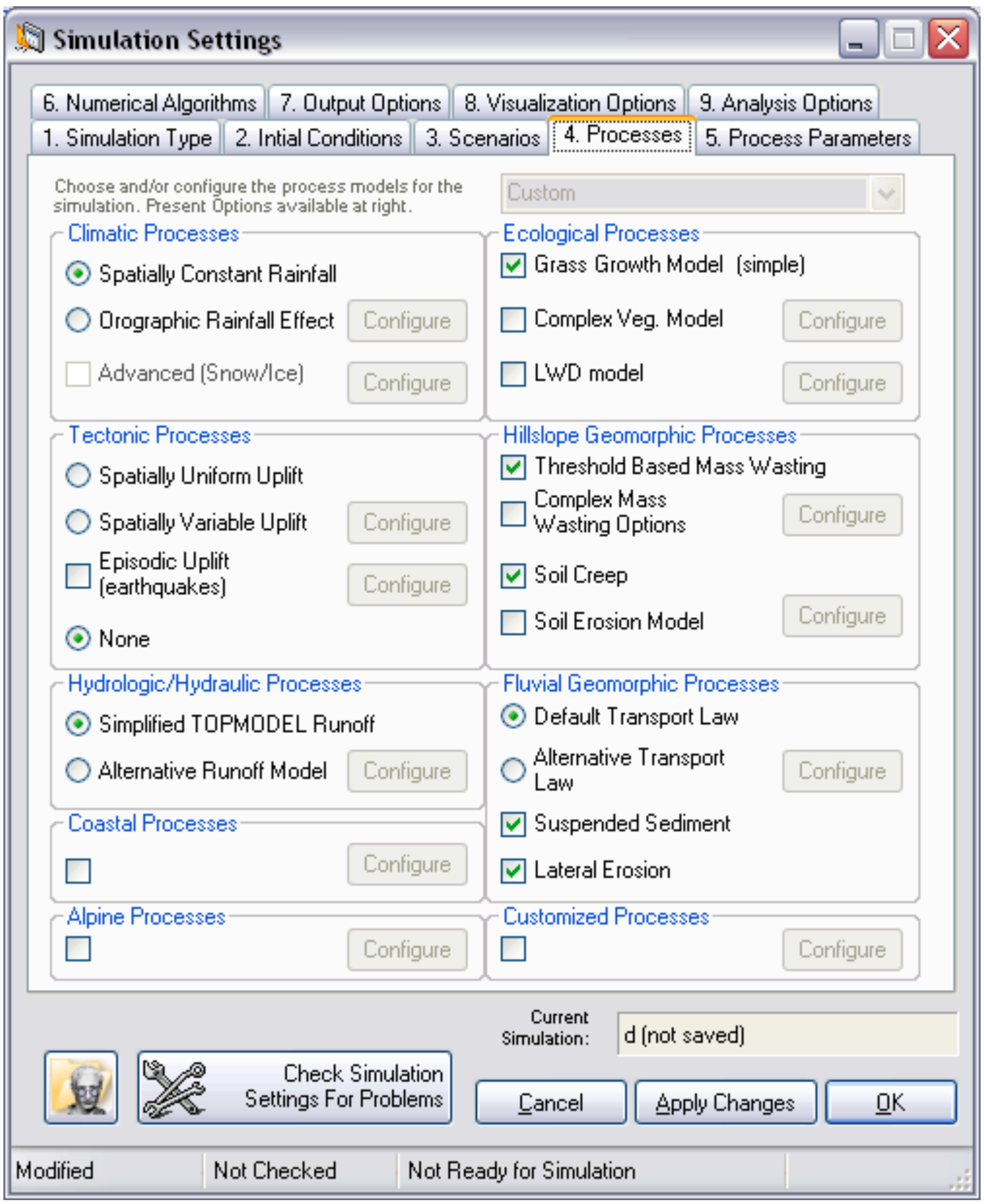
Raster Based Analyses
Vector Based Analyses
Numerical Solvers

Process Representation

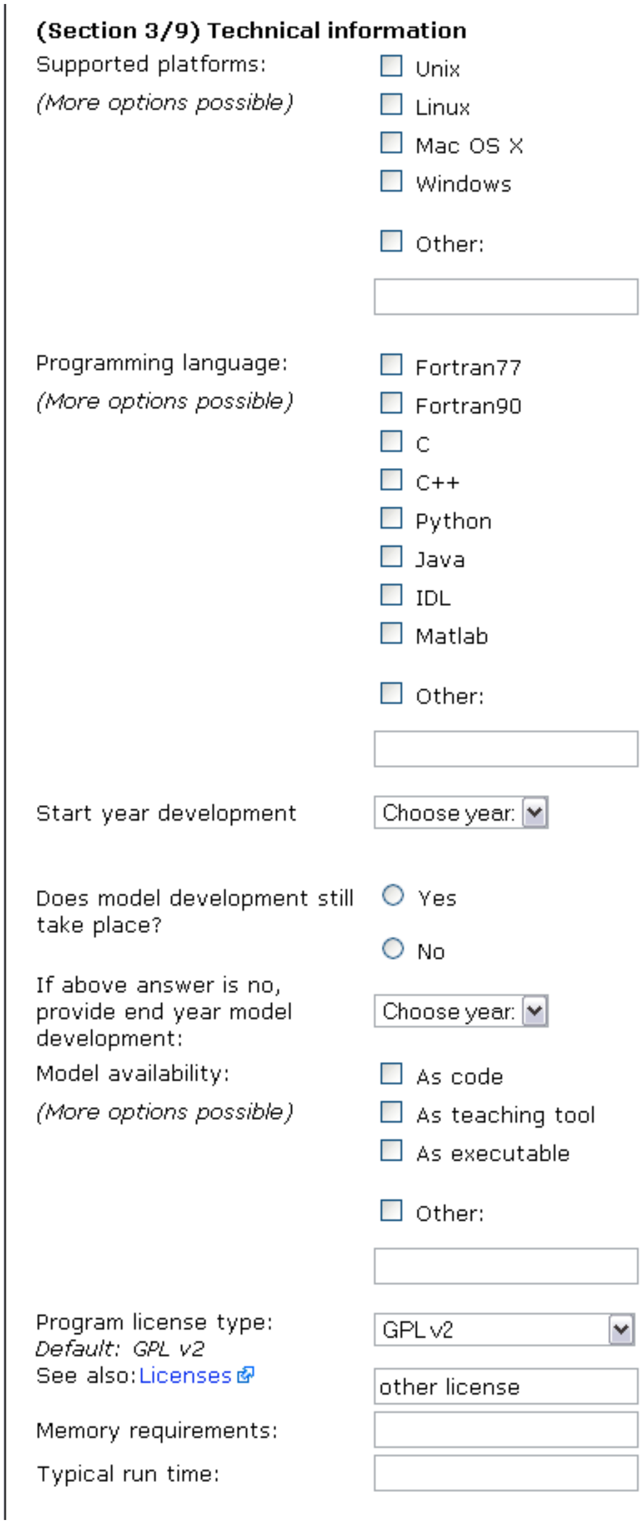
A field for Process Representation Type(s):

€ Aeolian
€ Hillslope
€ Hydrologic
€ Rainfall

€ Runoff
€ Fluvial
€ Landsliding
€ Soil creep
€ Other hillslope processes
€ Tectonics
€ Ecological
€ Climatic
€ Glacial
€ Soil production / rock weathering
€ Geochemistry / solute flux
€ Dissolution / karst



Technical Information



(Section 3/9) Technical information

Supported platforms: *(More options possible)*
- ☐ Unix
- ☐ Linux
- ☐ Mac OS X
- ☐ Windows
- ☐ Other:

Programming language: *(More options possible)*
- ☐ Fortran77
- ☐ Fortran90
- ☐ C
- ☐ C++
- ☐ Python
- ☐ Java
- ☐ IDL
- ☐ Matlab
- ☐ Other:

Start year development: Choose year:

Does model development still take place?
- ○ Yes
- ○ No

If above answer is no, provide end year model development: Choose year:

Model availability: *(More options possible)*
- ☐ As code
- ☐ As teaching tool
- ☐ As executable
- ☐ Other:

Program license type: GPL v2
Default: GPL v2
See also: Licenses
other license

Memory requirements:

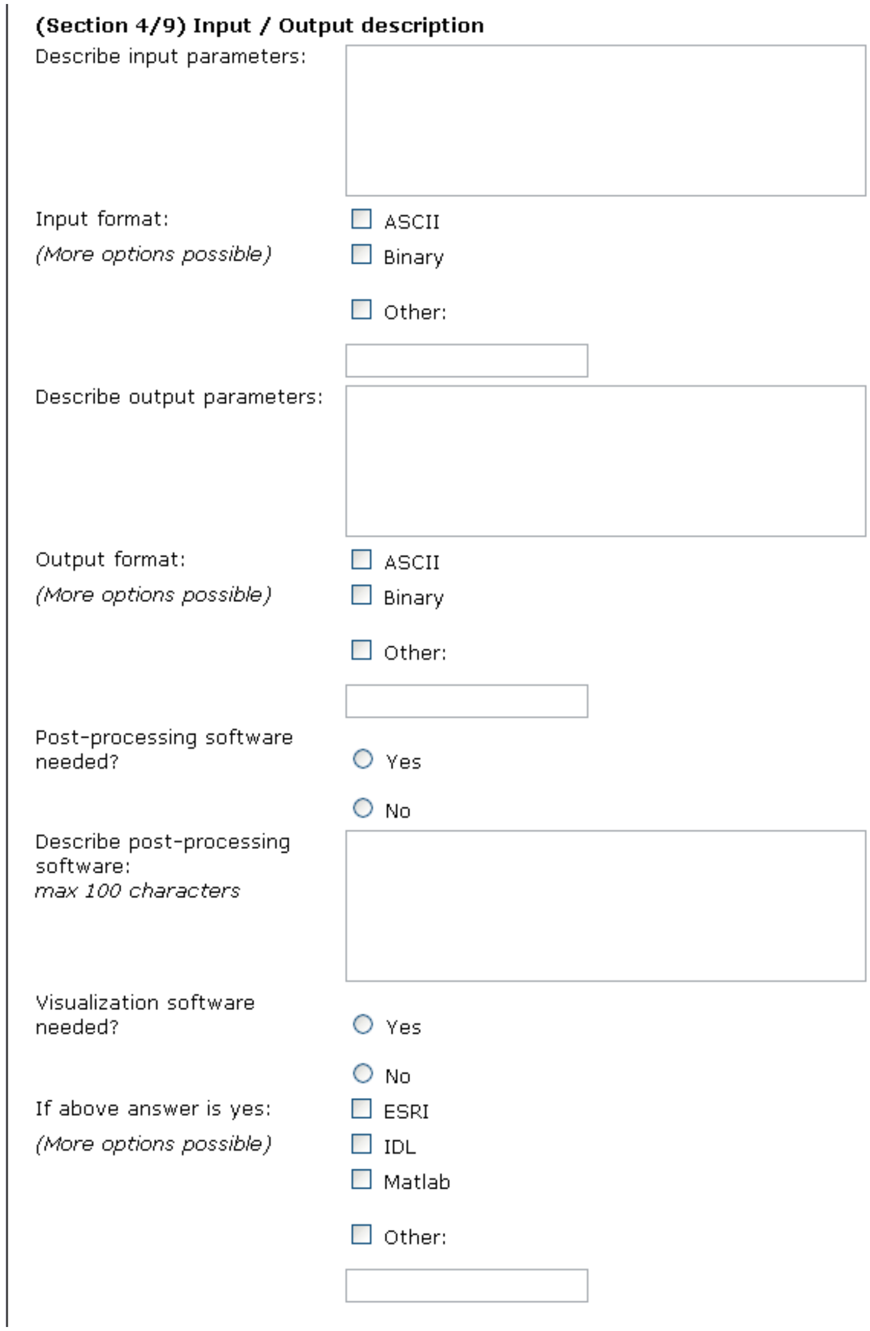
Typical run time:

Figure 3 – Existing Section 3 Fields

This section is generally applicable to all modules. We suggest:

- Adding a checkbox field for ‘Compiled Code is Distributed as’: a) GUI, b) Web-Application, c) Command-Prompt Application, d) Library (e.g. DLL), e) script, f) other ________
- Adding Checkbox field for: ‘Code optimized for:’ a) Single Processor, b) Parallel Computing, c) High-throughput computing, d) High-Performance Computing
- Adding radio button field and text field (for notes) for ‘OpenMI compliant’ with options: a) Yes, b) No, but Planned, c) No, but possible, d) No, not possible
- Adding radio button field and text field (for notes) for ‘CCA compliant’ with options: a) Yes, b) No, but Planned, c) No, but possible, d) No, not possible
- Adding radio button field and text field (for notes) for ‘Fully CSDMS compliant’ with options: a) Yes, b) No, but Planned, c) No, but possible, d) No, not possible (somewhere the web site should explain what “Fully CSDMS compliant” entails)
- Adding radio button field and text field (for notes) for ‘Is code already in “IRF” interface?’ with options: a) Yes, b) No, but Planned, c) No, but possible, d) No, not possible
- Modifying the ‘Typical Run Time’ tool to include another fields for ‘On what type of system’ and ‘For what type of tasks’.

Input/ Output Description



(Section 4/9) Input / Output description

Describe input parameters:

Input format: *(More options possible)*

- ☐ ASCII
- ☐ Binary
- ☐ Other:

Describe output parameters:

Output format: *(More options possible)*

- ☐ ASCII
- ☐ Binary
- ☐ Other:

Post-processing software needed?

- ○ Yes
- ○ No

Describe post-processing software: *max 100 characters*

Visualization software needed?

- ○ Yes
- ○ No

If above answer is yes: *(More options possible)*

- ☐ ESRI
- ☐ IDL
- ☐ Matlab
- ☐ Other:

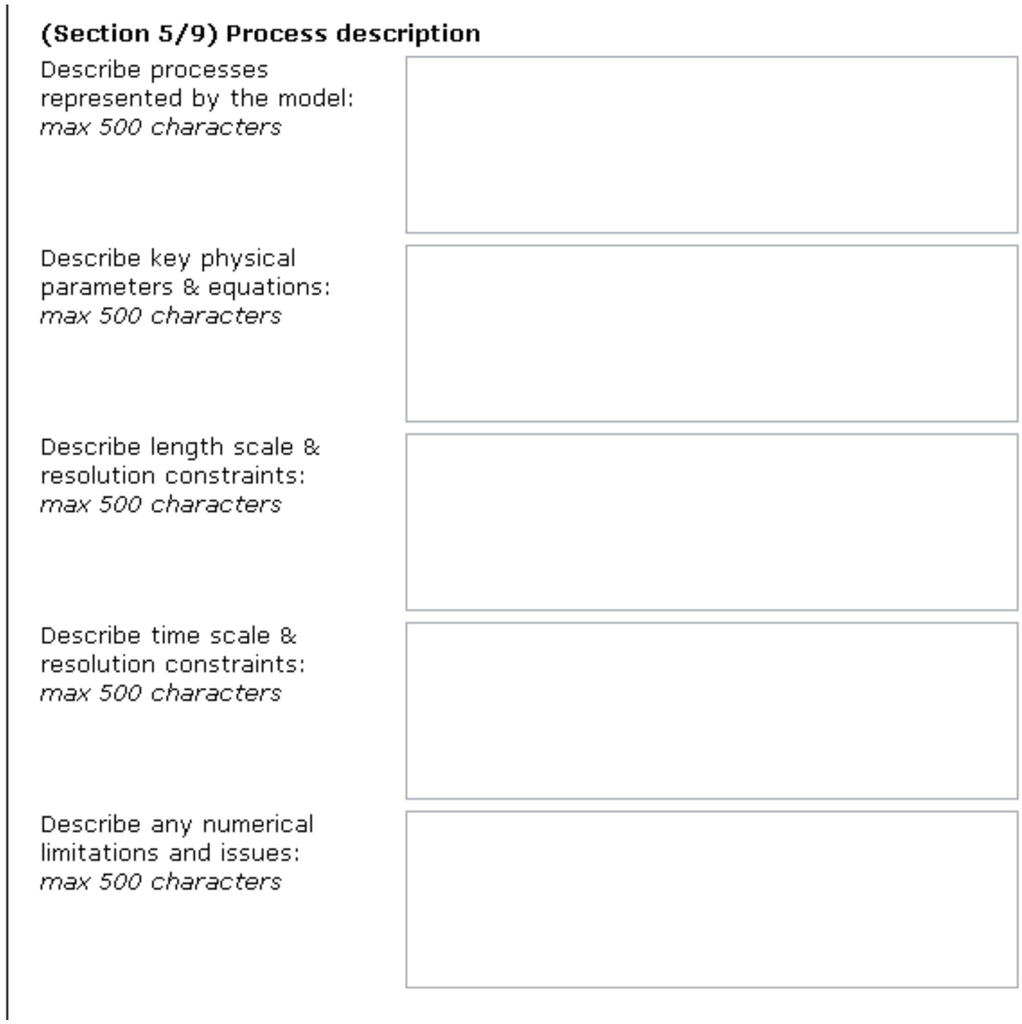
Figure 4 – Existing Section 4 Fields

This section is sufficiently generic to still apply to all 'modules'. We suggest:

- Graying out options based on user selections (i.e. if 'no' chosen, don't make available the 'if yes, questions').
- Duplicate post-processing software section with a 'pre-processing software' selection.

- Consider how this section might be extended to include other data-type standards (e.g. XML, CUAHSI-HIS, etc.)
- Adding check-box field to post-processing that enables a drop-down list of all modules of type 'Post Processing Software' already submitted to CSDMS. Repeat for pre-processing and visualization software.

Process Description



(Section 5/9) Process description

Describe processes represented by the model: *max 500 characters*

Describe key physical parameters & equations: *max 500 characters*

Describe length scale & resolution constraints: *max 500 characters*

Describe time scale & resolution constraints: *max 500 characters*

Describe any numerical limitations and issues: *max 500 characters*

Figure 5 – Existing Section 5 Fields

Under the suggestions recommended in this document, these descriptions are only applicable to those individuals submitting a module of 'type' Model or Process Representation Algorithm. Moreover, the spatial temporal fields we feel are better placed under the 'model identity' section as they are fundamental to the module's identity and eventual compatibility and interoperability.

We suggest:

- Only allowing users to fill in this section if they've selected a module type of model or process representation algorithm.
- Changing Title to 'Process Representation'
- Deleting spatio-temporal fields
- Moving the last field 'numerical limitations and issues' to 'Model Testing'
- Add a text and file field to allow the upload of an image of a flow-chart or diagram of a conceptual model the process representation is based on and a text field for its description.

Model Testing

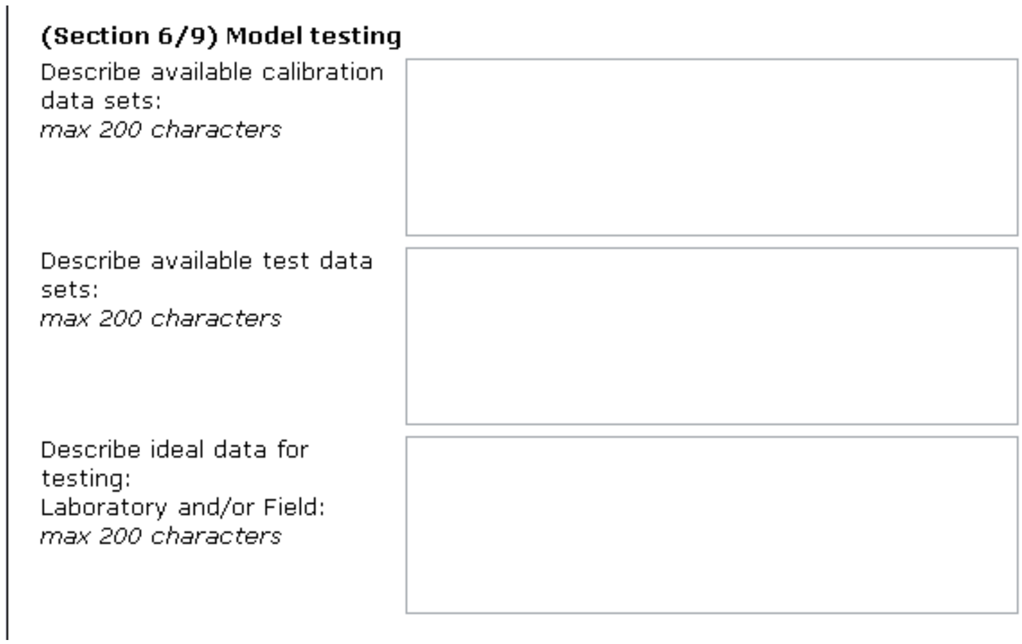


Figure 6 – Existing Section 6 Fields

Provide additional fields to provide a title and URL of places to acquire the 'available calibration data sets' described and 'test data sets'. Change name from 'Model testing' to 'Module Testing'.

User Groups

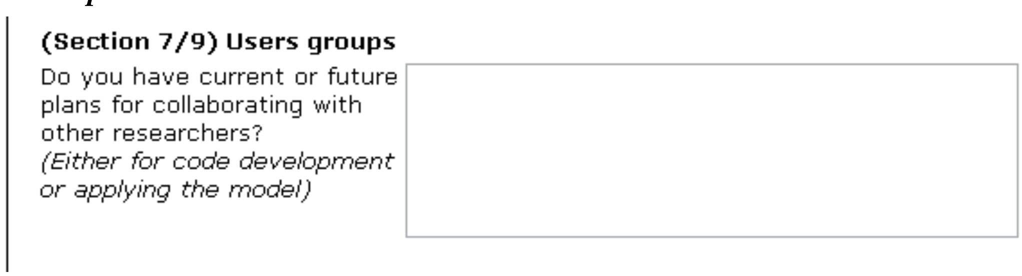


Figure 7 – Existing Section 7 Fields

This is a good start. Provide a field for adding multiple names and corresponding website URLs (if applicable) of other collaborative research groups, working groups, organizations, projects, etc.

Documentation

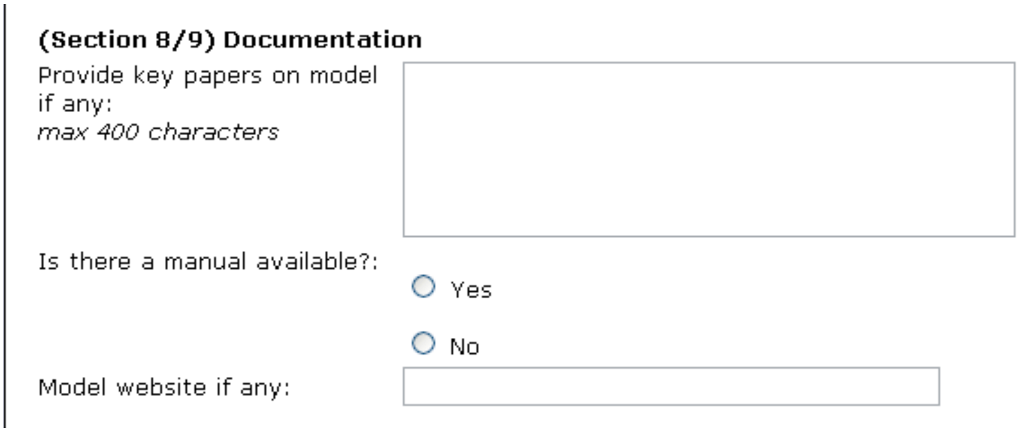


Figure 8 – Existing Section 8 Fields

For documentation, three additions are recommended.

- **Under the key papers on model field, add a button should be added to 'Add Documentation', which triggers a pop-up dialog form for entering the 'key papers' into consistent bibliographic fields.** The first prompt should be for paper type (e.g. Journal Article, Book Section, Report, etc.), which then determines which fields the user is prompted to enter. All journal entries should ask for a DOI or URL where the publication is available. This information can then be used not just to produce a consistently formatted bibliography within section 8 of the module page, but can also be used to add to complete bibliography lists which may be useful elsewhere in the CSDMS website (these could be avialble for website visitors to download as EndNote or BibTex libraries for example).

- **A button should be added under the 'manual' section to 'Add Manual' in a similar manner to the 'key papers' field above.** Again this should have the URL where this can be downloaded or a facility to upload the manual directly to CSDMS.

- **An additional field should be added for the URL to a 'Model Forum or Discussion Board (if applicable)'.**

Additional Comments


(Section 9/9) Additional comments

Comments:
max 500 characters

Figure 9 – Existing Section 9 Fields

It is always good to have a slop category for things that do not fit neatly into the eight other categories. No change is necessary here.

Summary

The way to implement the above suggestions is simple. First the 'Model Questionnaire' web form should be modified and the additional fields should be built into the database. Secondly, this break-out group will take an initial stab at populating the database based on our own knowledge of existing components. Third, we will solicit contributions from the rest of the terrestrial community and review those to make recommendations regarding model development priorities. Finally, a dynamic web-page (to be hosted on the CSDMS website) should be constructed which provides users the means to query the database in a variety of fashions needs to be built. There can still be several default lists (e.g. now there is one for each working group). The list will never be complete, but the web-form and dynamic web-page allow the list to grow indefinitely. No doubt, as the CSDMS effort matures and grows, new metadata categories may be deemed appropriate to add. As contributors to the list can always edit their entries, updating past entries will be feasible.

Secondarily, it is hoped that the above meta-data fields may become a basis for a generic CSDMS object-oriented class-structure for organizing these types of module component contributions into different libraries.

References Cited

Bills, B.G., D.R. Currey and G.A. Marshall, 1994, Viscosity estimates for the crust and upper mantle from patterns of lacustrine shoreline deformation in the Eastern Great Basin, *Journal of Geophysical Research*, *99*(B11), 22,059–22,086.

Gabet, E.J., 2003, Post-fire thin debris flows: Sediment transport and numerical. modeling, *Earth Surface Processes and Landforms*, vol. 28: 1341-1348

Lave, J., and Burbank, D., 2004, Denudation processes and rates in the Transverse Ranges, southern California: Erosional response of a transitional landscape to external and anthropogenic forcing: *Journal of Geophysical Research*, v. 109, F01006, doi: 10. 1029/2003JF000023.

Vrugt, J.A., H.V. Gupta, L.A. Bastidas, W. Bouten, and S. Sorooshian, 2003, Effective and efficient algorithm for multi-objective optimization of hydrologic models, *Water Resources Research*, 39 (8), art. no. 1214, doi:10.1029/2002WR001746.