Parallelization of a Fully-Distributed Hydrologic Model using Sub-basin Partitioning

Enrique R. Vivoni¹, S. Mniszewski², P. Fasel², E.S. Springer², V.Y. Ivanov³ and R.L. Bras³

¹. Department of Earth and Environmental Science, New Mexico Tech, Socorro, NM. (vivoni@nmt.edu)
². Computational and Environmental Sciences, Los Alamos National Laboratory, Los Alamos, NM.
³. Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Abstract

A primary obstacle towards advances in watershed simulation has been the parallelization of hydrologic models. Distributed hydrologic models typically run sequentially on a single computer, preventing their application over spatial and temporal extents of interest.

Problem Statement

Computational limitations prevent distributed hydrologic models from being used over spatial and temporal extents of interest:
- Regional to continental-scale basins (possibly to 10 km2).
- Hydroclimatic studies (seasonal to multi-year periods).
- Scales comparable to field and remote-sensing resolutions.

We seek to eliminate the existing computational limits in distributed hydrologic modeling to enable:
- Efficient simulations with high-resolution datasets.
- Ability to use complex topologies (geometric, topographic).
- Ensemble forecasting and data assimilation.

Hydrological Model

TIN-based Real-time Integrated Basin Simulator (IRIBS)

IRIBS Model Features:
- Continuous simulation over storms and interstorm periods.
- Spatially-distributed forcing parameters, boundary conditions.
- Complex terrain representation from hillslope to basin scales.
- Complete water-energy cycle (surface-subsurface coupling).
- Captures hydrologic capacities (flowing, storage, evaporation).

Additional details: Iverson et al. [2004]; Vivoni et al. [2009]

Overview and Approach

Distributed hydrologic models typically function sequentially on a single computer processor (Singh and Woolhiser, 2002). In groundwater and atmospheric models, parallelization into rectangular domains is common (e.g., Pa et al. 2002; Zhang et al. 2005). Here, we propose a new strategy for parallelization of watershed models based on:
1. Terrain discretization using irregular meshes.
2. Domain decomposition via sub-basin partitioning.
3. Hydrologic exchanges for lateral surface-subsurface fluxes.

A sub-basin as an organizational unit provides an excellent framework for parallelization. Physical constraints on hydrological fluxes between sub-basins can be exploited to reduce communication time and provide a modular approach.

Parallellization Strategy

1. Terrain Discretization based on Triangular Irregular Networks

Irregular discretizations allow terrain representations with low node numbers without sacrificing accuracy (Vivoni et al. 2004). Here, a 30 x 60 DEM (70,960 nodes) is sampled into different TINs (A = 3,201; B = 1,515), with low RMSE (1.1-2.4 mm for A and B).

2. Domain Decomposition using Sub-basin Partitioning

Sub-basin definition allows partitioning the domain into smaller units of variable size connected via the channel network. Here, a single basin is decomposed into many sub-basins (A = 203; B = 43), using a cost-area threshold (0.225 and 0.9 km²).

3. Hydrologic Exchanges in Channel Network and Subsurface

Parallelization is achieved by exchanging data among reaches and sub-basins, each with a TIN with minimal reaches. Hydrologic data is exchanged downstream along reaches (1 and 2 to 3) and in subsurface between sub-basins along flow direction (1-2, 2-3, 1-3).

Implementation

Objective-oriented programming allows modular design of parallelization algorithms with flexibility in terms of processor partitioning:
(A) Simple parallelization of IRIBS model
(B) Diagram of class structures for pIRIBS
(C) Options for partitioning sub-basins (N subbasins) into available processors (M total processors).

Software Development using Object-Oriented Approach

- Common Input/Output Functionality
- Initialize MPI
- Simulation Control
- Mesh Construction
- Common Input/Output Functionality

A simple example

Example basin used for parallelized model development and testing versus single processor version.
- Defined shower event that allows for a 2 x speedup is achievable for the example basin, but it is uncertain how long this performance will continue.
- Scales comparable to field and remote-sensing resolutions.
- Complete water/energy cycle (surface-subsurface coupling).
- Multiple capabilities (forecasting, estimation, synthesis).

Results

Hydrological Performance of pIRIBS

Comparison of the single processor IRIBS and the pIRIBS model for the SMALLBASIN (7 processors: full partition) demonstrate the parallel version is adequately constructed. For a synthetic event (by Iverson), both the output hydrograph (a lumped measure) and the basin saturated area (a distributed measure) are identical (over 10 hr simulation time).

Computational Performance of pIRIBS

A 2 x speedup is achievable for the example basin, but a significant leveling off after 4 processors is observed. This is due to the small basin and model time spent on solving (initial EOF mesh generation). For larger problems, the computational efficiency is anticipated to be higher. Parallelization gains achieved as individual processors balance computational effort.

Future Efforts

Given the confidence in the parallelization of the distributed model, next steps include further validations and tests:
- Speed-up tests for larger basins (10 km² to 100 km²).
- Analysis of effect of balanced sub-basin partitioning.
- Analysis of effect of EOF resolution on parallelization.
- Additional performance metrics from a hydrologic perspective.

A few studies have used a watershed-based parallelization taking advantage of network organization. One notable exception is the work of Apostolopoulos and Georgakakos [1997]. We believe this effort has significant potential to advance the state of modeling science to hydrology.

References