Performance Study of Parallel Algorithms in pWASH123D

Jing-Ru C. Cheng (ruth.c.cheng@erdc.usace.army.mil)
Major Shared Resource Center, Information Technology Laboratory, U.S. Army Engineer Research and Development Center, Vicksburg, MS 39180, USA

Hwai-Ping Cheng (hwai-ping.cheng@erdc.usace.army.mil) and
Hsin-Chi Lin (hsin-chi.j.lin@erdc.usace.army.mil)
Coastal and Hydraulics Laboratory, U.S. Army Engineer Research and Development Center, Vicksburg, MS 39180, USA

Robert M. Hunter (robert.m.hunter@erdc.usace.army.mil)
Major Shared Resource Center, Information Technology Laboratory, U.S. Army Engineer Research and Development Center, Vicksburg, MS 39180, USA

David R. Richards (david.r.richards@erdc.usace.army.mil)
Information Technology Laboratory, U.S. Army Engineer Research and Development Center, Vicksburg, MS 39180, USA

Earl V. Edris (earl.v.edris@erdc.usace.army.mil)
Coastal and Hydraulics Laboratory, U.S. Army Engineer Research and Development Center, Vicksburg, MS 39180, USA

Abstract. Since January 2003, the U.S. Army Engineer Research and Development Center has been developing a parallel watershed code, pWASH123D, and its graphical user interface in the Department of Defense Groundwater Modeling System in collaboration with Brigham Young University. To date, pWASH123D has a complete suite of water flow simulations in a watershed system that can be conceptualized as a combination of one-dimensional (1-D) channel network, 2-D overland regimes, and 3-D subsurface media. This paper presents the outcome of a performance study on the parallel algorithms currently employed in pWASH123D. The experimental area includes a 570-square-mile domain covering most of the land south of the Tamiami Train in South Florida and north of the Gulf of Mexico, Florida Bay, and Biscayne Bay. This area is discretized to three mesh resolutions—coarse, medium, and fine meshes—to identify the performance bottleneck. Problems are designed to investigate the parallel strategy for the 1-D component. Major findings from the study are presented in this paper.

Keywords: watershed modeling, parallel computing, scientific computing, coupled application, performance study

1. Introduction

A parallel watershed code, pWASH123D, based on a first-principle, physics-based numerical model, WASH123D (Yeh et al., 2006), has been developed in the Environmental Quality Modeling and Simulation (EQM)-related computational technology area to simulate large watershed problems on scalable computing systems. It is also being coupled with the ADCIRC (Advanced Circulation)(Luettich et al., 2006) nearshore ocean model (Cheng et al., 2006b) by using the Earth System Modeling Framework (ESMF Team, 2006) in the Battlespace Environments Institute (BEI), one of the six selected High Performance Computing Software Application Institutes sponsored by the Department of Defense High Performance Computing Modernization Program Office.

In WASH123D, the 1-D channel flow is computed by solving the cross-sectional area-averaged diffusive wave equation with the semi-Lagrangian finite element method (FEM), while the 2-D overland flow is computed by solving the depth-averaged diffusive wave equation with the semi-Lagrangian FEM, and the 3-D subsurface flow is computed by solving the Richards' equation for variably saturated porous media with the Galerkin FEM. The continuity of flux and state variables (e.g., water head) is enforced on the interface of two media (Yeh et al., 2006). Rainfall, evapotranspiration, rule-controlled flow, and injection/withdrawal are sources/sinks in respective media. The computational domain is discretized with unstructured meshes, where each element can be assigned with a different material type to account for heterogeneity, and each material may have its own set of physical model parameters.

WASH123D employs different time intervals for computations in different dimensions to resolve various flow processes in a watershed that can be conceptualized as a coupled system of 1-D channel networks, 2-D overland regimes, and 3-D subsurface media. For instance, a watershed system with strong surface and subsurface interactions through infiltration and seepage would have time intervals of tens of minutes to hours, tens of seconds to minutes, and seconds for flow computations of 3-D subsurface, 2-D overland, and 1-D channel, respectively. The multitemporal scale not only resolves the physics correctly but also makes computation affordable. In this manner, each 3-D time interval would contain more 2-D time intervals, and each 2-D, many 1-D. Figure 1 describes the time and coupling loop structure of WASH123D (Cheng et al., 2006c).

Currently in the parallel watershed model, pWASH123D, a parallel data management toolkit, DBuilder (Hunter and Cheng, 2005), developed within the ERDC, is used in both 2- and 3-D computation to balance computational load on each processor. On the other hand, each processor reads complete 1-D channel information and executes 1-D computation without partitioning to avoid excessive run time overhead from data exchange among processors. Figure 2 shows the hierarchical data structures of pWASH123D (Cheng et al., 2006c). In this study,

```
Foreach 3-D flow time step (△t_3DF) do
    Foreach 3-D coupling/nonlinear iteration do
        Foreach 2-D flow time step (△t_2DF) do
            Incorporate infiltration/seepage for 2-D/3-D coupling
            Foreach 2-D coupling/nonlinear iteration do
                Foreach 1-D flow time step (△t_1DF) do
                    Incorporate infiltration/seppage for 1-D/3-D coupling
                    Incorporate infiltration/seepage for 1-D/2-D coupling
                    Foreach 1-D coupling iteration loop do
                        Solve linearized 1-D flow equation
                    Endfor
                    Incorporate infiltration/seepage for 1-D/2-D coupling
                    Solve linearized 2-D flow equation
                Endfor
            Endfor
        Endfor
        Incorporate infiltration/seepage for 1-D/3-D coupling
        Incorporate infiltration/seepage for 2-D/3-D coupling
        Solve linearized 3-D flow equation
    Endfor
Endfor
```

Figure 1. 1-D/2-D/3-D coupling algorithm in pWASH123D

the performance of pWASH123D was examined to investigate the parallelization strategy mentioned above, which could be affected by mesh resolution and by 1-D computational nodes.

2. Experimental Examples

The test example employed in this study considered an area (∼570 square miles) south of the Tamiami Trail in South Florida, which is bounded by Tamiami and C-4 canals in the north and coastal shores of the Gulf of Mexico, Florida Bay, and Biscayne Bay in the south. The model boundary was determined based on the available surface and groundwater gauge data so that adequate boundary conditions can be applied. Figure 3 shows the topographical color contours in and around the computational domain adopted.

2.1. Material setup in each component

The simulation model was composed of 1-D canal networks, 2-D overland regimes, and 3-D subsurface media. The canals included in the

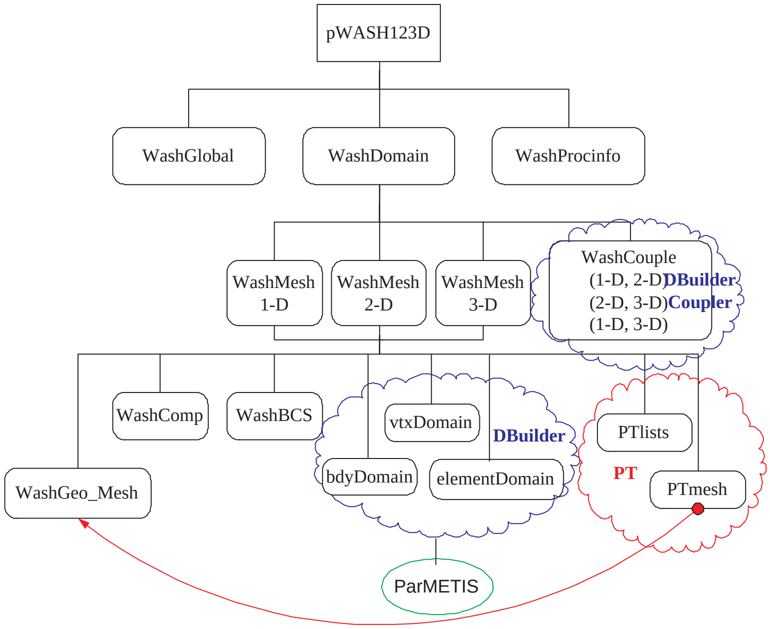

Figure 2. Hierarchical data structures designed in pWASH123D

1-D canal network are L-31N, C-111, C-103, C-102, C-1, C-3, and L-31E (Figure 4). Each canal reach was assigned a Manning's roughness coefficient (i.e., n_1) to characterize the canal flow in this reach, where

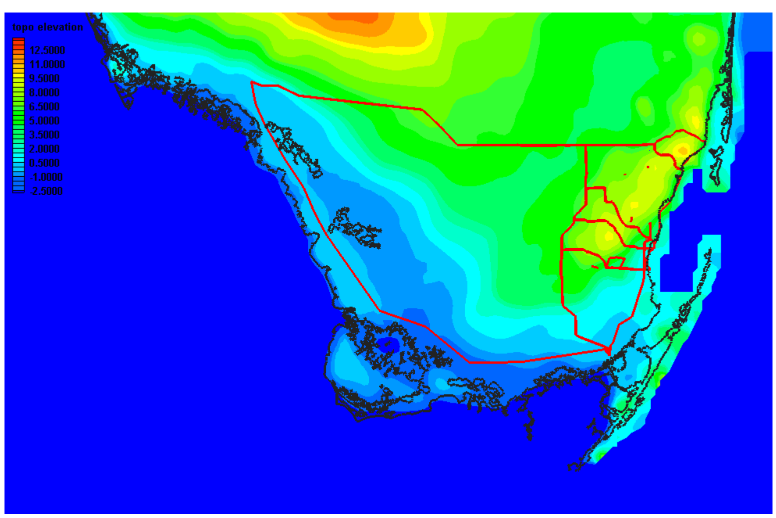

Figure 3. Topography in and around domain of test example (domain boundary and canals taken into account highlighted in red)

the ends of a canal reach may be an upstream boundary, a downstream boundary, a dead end, a canal junction, the headwater of a canal structure, or the tailwater of a canal structure. Because of the similarity of the canals, only one Manning's roughness coefficient value was used for all canal reaches. The 2-D overland regimes were divided into five subregions (Figure 4) mainly based on the land-use type, where each subregion was associated with a Manning's roughness coefficient (i.e., n_2) representing the overland flow characteristics in the subregion. The 3-D subsurface media included 17 materials with various hydraulic conductivities (i.e., $\mathbf{K}$) representing the flow characteristics through these materials.

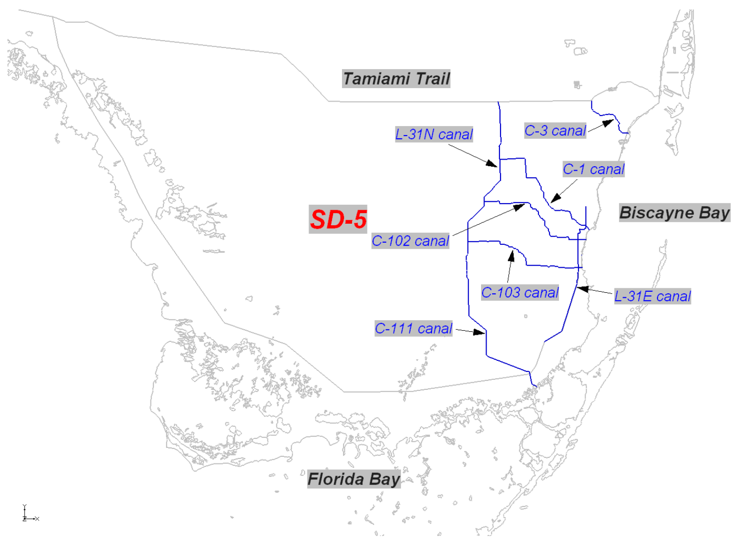

Figure 4. Canal network considered in computational model

Model parameters described above for simulation runs were determined based on the calibration and validation results from former studies in both the Biscayne Bay Coastal Wetlands (Cheng et al., 2005; Cheng et al., 2006a) and the C-111 Spreader Canal projects (Lin et al., 2004). Tables I and II list the values of flow parameters (i.e., n_1, n_2, and $\mathbf{K}$) used for all simulation runs considered in this study. The magnitude of interaction of canal water and groundwater through infiltration/seepage is controlled by the hydraulic conductivity of the "Canal (3-D)" and the "L-31E (3-D)" materials, as shown in Table II.

2.2. Flow boundary and initial conditions

Observed daily data of groundwater head and surface water stages were imposed on the domain boundary to set up the boundary conditions

Table I. Model parameters for 1- and 2-D components

Material ID	n_1 or n_2 (dimensionless)
Canal (1-D)	0.05
Canal, Wetland, Urban (2-D)	0.05
Cropland, Rangeland (2-D)	0.1

Table II. Model parameters for 3-D component

Material ID	**K** (ft/hr)	
	Horizontal	Vertical
Lower3, Lower2, Lower1 (3-D)	100	10
Middle4, Middle3, Middle2, Middle1 (3-D)	1000	100
FS, SL, SIL, MUCK, UWB, GRV-L, MARL, URBAN (3-D)	0.1	0.01
CANAL (3-D)	250	25
L31E (3-D)	0.001	0.0001

for simulation runs, where linear interpolation was employed to calculate surface water stages and groundwater heads between two adjacent observation locations. At canal junctions, the continuity equations of both flow and stages were enforced. As the 17 canal structures were taken into account here, the derived daily-average flow data were used to specify the flux-type boundary conditions for their immediate upstream and downstream canal reaches. A zero-velocity condition was applied at the canal nodes representing dead ends. The canal levees were considered as drainage divides for the overland flow, and a zero-depth boundary condition was applied at the corresponding overland nodes. A total head Dirichlet boundary condition was applied to the side boundary faces on 3-D boundaries. The bottom face of the 3-D domain was assumed impermeable. On the top face of the 3-D domain, i.e., the ground surface, the interaction of surface and subsurface water was accounted for, and an adequate boundary condition, either the head type or the flux type, was applied at a 3-D top boundary node to satisfy the continuity of flux and head.

To start each simulation with a reasonable and stable initial condition, the initial canal stages were first calculated through an interpolation process based on the given observed stage information. The calculated initial canal stages were then applied at the canal-corresponding subsurface nodes as boundary conditions for steady-state subsurface flow computation, which ensures the continuity of state variables (i.e., canal stage equals groundwater head) on the canal-subsurface interface. The steady-state subsurface flow solutions were then used as the initial condition for the subsequent transient simulations. Based on rainfall, evapotranspiration, initial canal stage, and initial subsurface total head along the domain boundary, the entire overland domain was assumed initially dry for all model runs. This assumption was validated by the pressure head solution of the steady-state flow computation, i.e., all the overland corresponding subsurface nodes had negative pressure heads, representing a dry ground surface.

2.3. Computational meshes

Three meshes of different resolutions were generated to investigate the pWASH123D performance on the U.S. Army Engineer Research and Development Center Major Shared Resource Center (ERDC MSRC) high performance computing (HPC) machine—Cray XT3 named Sapphire containing 4,176 processors, with each containing a 2.6-GHz Opteron 64-bit processor and dedicated memory. Table III lists the specification of the three meshes generated for this study (i.e., Coarse, Medium, and Fine). Figures 5 and 6 depict the 2- and 3-D meshes, respectively, generated by the Groundwater Modeling System (GMS) 6.0 (GMS Team, 2006) for the Coarse mesh, where each color represents a material in these two figures, and the magnification in the vertical direction is 300 in Figure 6. It is noted that the vertical spacing ranges from 2 feet for the top layers through several tens of feet for the middle and bottom layers depending on the thickness of the hydrogeologic units (e.g., aquifers, aquitars, and aquicludes) taken into account.

3. Simulation Runs

The collected information to set up for simulation runs includes the following:

- Computing 1-D channel flow: cross-sectional geometry for each canal reach, canal bottom elevation, headwater stage, tailwater stage, flow at each gate structure in the canal network, rainfall, and evapotranspiration

- Computing 2-D overland flow: topography, land use, rainfall, and evapotranspiration

- Computing 3-D subsurface flow: hydrogeology, groundwater pumping, and groundwater head

The time-step sizes are 0.5 hour for 3-D, 5 seconds for 2-D, and 0.5 second for 1-D component. A total of 12 jobs are submitted for execution in this study.

Table III. Specification of three basic meshes with different numbers of 1-D nodes

Mesh ID	Coarse Meshes		Medium Meshes		Fine Meshes	
NP[a]	16		64		128	
Component	ND[b]	NE[c]	ND[b]	NE[c]	ND[b]	NE[c]
2-D	8,487	16,583	42,941	84,996	101,148	200,935
3-D	59,409	99,498	558,233	1,019,952	2,124,108	4,018,700
1-D	ND[b]	NR[d]	ND[b]	NR[d]	ND[b]	NR[d]
Case 1	89	1	206	1	316	1
Case 2	127	2	298	2	492	2
Case 3	200	4	463	4	710	4
Case 4[e]	214	5	495	5	760	5

[a] Number of processors
[b] Number of nodes
[c] Number of elements
[d] Number of reaches
[e] 3-D mesh is modified to 1,314,924 nodes and 2,411,220 elements

4. Experimental Results

Figures 7 through 9 depict the wall-clock time in hours and in percent versus number of 1-D nodes with 16 processors for the Coarse, 64 for Medium, and 128 for modified Fine meshes, respectively. The average wall-clock time spent on the computation of a 1-D node, a 2-D node, and a 3-D node is given in Tables IV through VI for the four cases of 1-D nodes, i.e., one, two, four, and five canal reaches, with the Coarse, the Medium, and the modified Fine meshes, respectively.

Figures 7 through 9 show that the wall-clock time for 1-D computation is approximately proportional to the number of 1-D nodes included in the simulation, which results in the increase of both total wall-clock time and the wall-clock time percent for 1-D computation. It also explains why the wall-clock time percent for 2- and 3-D computations decreases with the number of 1-D nodes computed even though the wall-clock time for 2- and 3-D computations basically does not vary with the number of 1-D nodes. With the current parallelization strategy in pWASH123D, it is obvious that the fewer the 1-D nodes are considered for computation, the less time is spent for 1-D computation. Moreover, it is also observed, from these three figures, that the time spent in couplers is negligible when compared with that spent in 1-, 2-, or 3-D computation.

By examining the average wall-clock time spent on each node in various dimensions (Tables IV through VI), the following is observed.

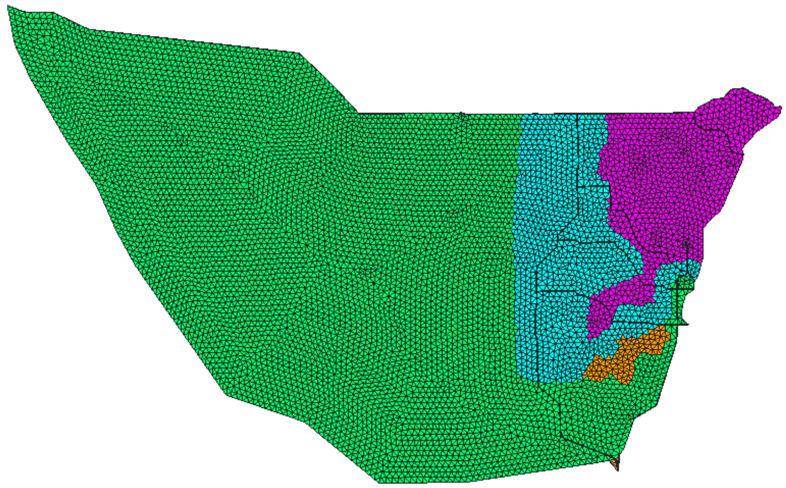

Figure 5. 2-D mesh generated from GMS 6.0 for Coarse mesh

First, the simulation setup in this study required the most time to achieve computation on each 1-D node. Second, the least time required is on each 3-D node. The above observation is related to the time-step sizes used: 0.5 hour for 3-D, 5 seconds for 2-D, and 0.5 second for 1-D. Thus, increasing the 1-D time-step size without sacrificing accuracy may reduce the wall-clock time for 1-D computation significantly and benefit the use of pWASH123D with the current parallelization strategy. It is noted that concerning both numerical stability and accuracy, the time-step size used for computation is also closely related to the element size that can be translated in a way with the number of nodes, especially in the case of nonlinear systems.

To account for the situation when there are many 1-D nodes considered for computation, and the time-step size for 1-D computation has to be small, it is necessary to develop another parallelization strategy to efficiently compute the coupled 1-/2-/3-D system. To do this, investigating time-space parallelism on the lower dimensional domains is recommended.

5. Summary and Future Plans

Because of the current computational strategy for the 1-D component, the following conclusion can be made. The 1-D computation signif-

Table IV. Average wall-clock time in coupled 1-/2-/3-D simulation with Coarse mesh and 16 processors

Case No.	Mesh Component	No. of nodes (A)	Wall-clock time (hr) (C)	Average wall-clock time (sec) (D=3600*C*B/A)[a]
1	1-D	89	0.3414	13.8107
	2-D	8,487	0.8112	5.5058
	3-D	59,409	1.4034	1.3607
2	1-D	127	0.8143	23.0820
	2-D	8,487	0.8420	5.7144
	3-D	59,409	1.4883	1.4430
3	1-D	200	1.2045	21.6815
	2-D	8,487	0.8754	5.9409
	3-D	59,409	1.5403	1.4934
4	1-D	214	1.2490	21.0112
	2-D	8,487	0.8918	6.0527
	3-D	59,409	1.5548	1.5074

[a] B = 1 for 1-D and B = 16 for 2-D and 3-D

Table V. Average wall-clock time in coupled 1-/2-/3-D simulation with Medium mesh and 64 processors

Case No.	Mesh Component	No. of nodes (A)	Wall-clock time (hr) (C)	Average wall-clock time (sec) (D=3600*C*B/A)[a]
1	1-D	206	1.8866	32.9699
	2-D	42,941	1.2730	6.8304
	3-D	558,233	6.8156	2.8130
2	1-D	298	2.7714	33.4797
	2-D	42,941	1.2182	6.5362
	3-D	558,233	6.8865	2.8423
3	1-D	463	4.0121	31.1954
	2-D	42,941	1.2745	6.8384
	3-D	558,233	6.8857	2.8419
4	1-D	495	4.4664	32.4832
	2-D	42,941	1.2147	6.5177
	3-D	558,233	7.9639	3.2870

[a] B = 1 for 1-D and B = 64 for 2-D and 3-D

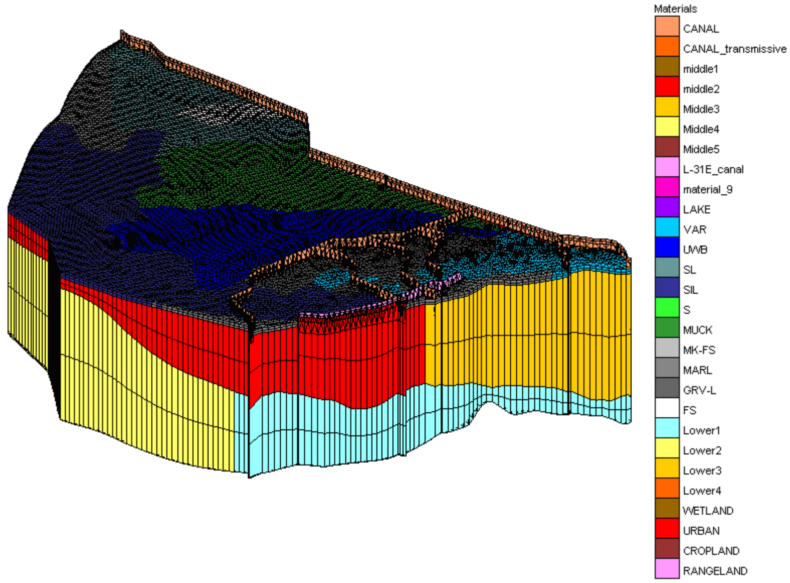

Figure 6. 3-D mesh generated from GMS 6.0 for Coarse mesh

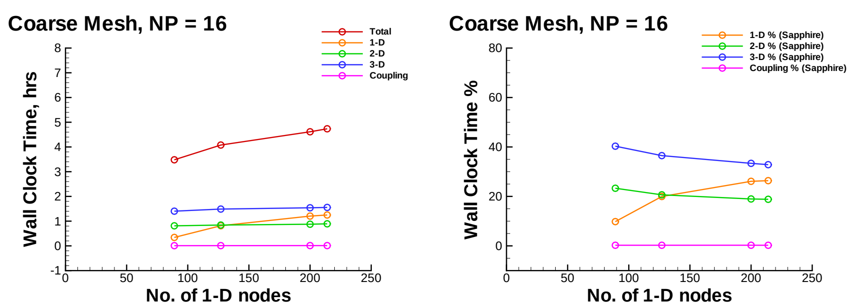

Figure 7. Comparison of wall-clock time and its percentage among simulations with various numbers of 1-D nodes with Coarse mesh

icantly takes up a portion of overall wall-clock time of simulations when using a larger number of processors. From the results, one can also conclude that, from the comparison of the averaged wall-clock time spent on a nodal computation in each component, computation requires the most time on a 1-D node, and the least time on a 3-D node. This result is directly related to the implemented algorithm, which has different time-step sizes in each component, i.e., 0.5 hour for 3-D, 5 seconds for 2-D, and 0.5 second for 1-D, used for these problem setups. Thus, an autonomous approach, which can guarantee convergence of the nonlinear system using larger time-step sizes, will highly benefit such a parallel watershed model. It is worthwhile to investigate time-

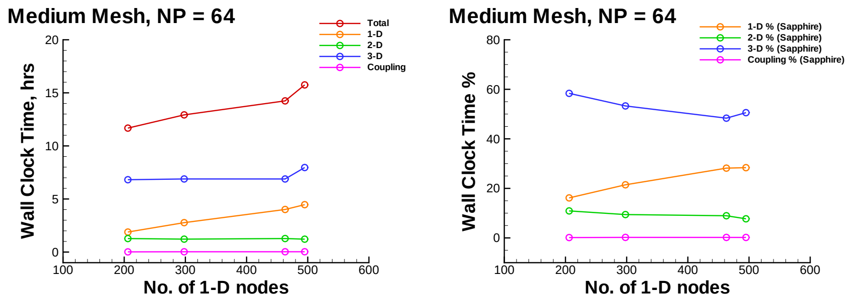

Figure 8. Comparison of wall-clock time and its percentage among simulations with various numbers of 1-D nodes with Medium mesh

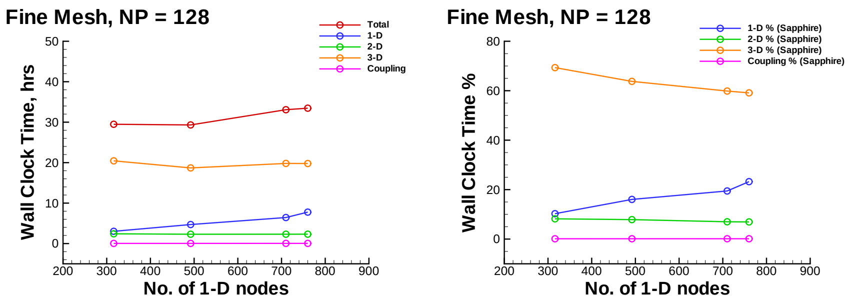

Figure 9. Comparison of wall-clock time and its percentage among simulations with various numbers of 1-D nodes with Fine mesh

Table VI. Average wall-clock time in coupled 1-/2-/3-D simulation with Fine mesh and 128 processors

Case No.	Mesh Component	No. of nodes (A)	Wall-clock time (hr) (C)	Average wall-clock time (sec) (D=3600*C*B/A)[a]
1	1-D	316	3.0270	34.4844
	2-D	101,148	2.4129	10.9922
	3-D	1,314,924	20.4437	7.1643
2	1-D	492	4.7016	34.4023
	2-D	101,148	2.3067	10.5088
	3-D	1,314,924	18.6967	6.5521
3	1-D	710	6.4354	32.6302
	2-D	101,148	2.3139	10.5416
	3-D	1,314,924	19.8040	6.9401
4	1-D	760	7.7713	36.8115
	2-D	101,148	2.3201	10.5697
	3-D	1,314,924	19.7824	6.9325

[a] B = 1 for 1-D and B = 128 for 2-D and 3-D

space parallelism on the lower dimensional domains. To sum up, several research topics mentioned have been established to improve the parallel watershed software, pWASH123D, further.

Acknowledgments

This work was supported in part by an allocation of computer time from the Department of Defense High Performance Computing Modernization Program at the ERDC MSRC, Information Technology Laboratory, Vicksburg, Mississippi. The work was also supported in part by the ERDC, R&D program System-Wide Water Resources Project. The authors would like to thank Stephen England at the Philadelphia District, U.S. Army Corps of Engineers, for preparing the data set for performance evaluation.

References

Cheng, H.-P., J.-R. C. Cheng, E. V. Edris, D. R. Richards, and G. T. Yeh: June 19-22, 2006a, 'Numerical Strategies to Model Hydrologic Features for the Biscayne Bay coastal wetlands project Alternatives'. In: *Computational Methods in Water Resources XVI*. Copenhagen, Denmark, CMWR CD-ROM. paper 164.

Cheng, H.-P., J.-R. C. Cheng, D. R. Richards, T. Campbell, and R. Allard: June 26-29, 2006b, 'Coupled Watershed-Nearshore Modeling—Phase I'. In: *DoD High Performance Computing Modernization Program, 2006 Users Group Conference*. Denver, CO. poster.

Cheng, J.-R. C., R. M. Hunter, H.-P. Cheng, H.-C. J. Lin, and D. R. Richards: May 15-19, 2005, 'Parallelization of the WASH123D code—Phase II: coupled two-dimensional overland and three-dimensional subsurface flows'. In: *World Water & Environmental Resources Congress*. Anchorage, Alaska, USA, EWRI CD-ROM.

Cheng, J.-R. C., R. M. Hunter, H.-P. Cheng, D. R. Richards, and G. T. Yeh: June 19-22, 2006c, 'Parallelization of a Watershed Model—Phase III: Coupled 1-Dimensional Channel, 2-Dimensional Overland, and 3-Dimensional Subsurface Flows'. In: *Computational Methods in Water Resources XVI*. Copenhagen, Denmark, CMWR CD-ROM. paper 64.

ESMF Team: 2006, 'ESMF: Earth System Modeling Framework Web page'. http://www.esmf.ucar.edu/, National Center for Atmospheric Research.

GMS Team: 2006, 'GMS - Groundwater Modeling System Web page'. http://chl.erdc.usace.army.mil/gms/.

Hunter, R. M. and J.-R. C. Cheng: 2005, 'DBuilder: A parallel data management toolkit for scientific applications'. In: *Proceedings of the 2005 International Conference on Parallel and Distributed Processing Techniques and Applications (PDATA'05)*. Las Vegas, Nevada, USA, pp. 825–831, CSREA Press.

Lin, H.-C. J., J.-R. C. Cheng, H.-P. Cheng, E. Edris, D. R. Richards, and G. T. Yeh: June 28-July 1, 2004, 'Using the Parallel WASH123D Code to Simulate Overland-Subsurface Interactions'. In: *World Water & Environmental Resources Congress*.

Salt Lake City, Utah, USA, EWRI CD-ROM (Ground Water Issues/Ground Water Modeling).
Luettich, R., J. Westerink, R. Kolar, and C. Dawson: 2006, 'ADCIRC Coastal Circulation and Storm Surge Model Web page'. http://www.adcirc.org/.
Yeh, G.-T., G. Huang, H.-P. Cheng, F. Zhang, H.-C. Lin, E. Edris, and D. Richards: 2006, 'A First-Principle, Physics-Based Watershed Model: WASH123D'. In: V. P. Singh and D. K. Frevert (eds.): *Watershed Models*. Taylor and Francis Group, LLC, Chapt. 9, pp. 211–244.