Parallel computing efficiency of SWAN 40.91

Christo Rautenbach1,2,3,4, Julia C. Mullarney4, and Karin R. Bryan4

1Institute for Coastal and Marine Research, Nelson Mandela University, Port Elizabeth, South Africa
2Department of Oceanography and Marine Research Institute, University of Cape Town, Cape Town, South Africa
3Research and development, MetOcean (a division of the MetService), Raglan, New Zealand
4Environmental Research Institute, University of Waikato, Hamilton, New Zealand

Correspondence: Christo Rautenbach (rautenbachchristo@gmail.com)

Received: 18 September 2020 – Discussion started: 2 December 2020
Revised: 18 May 2021 – Accepted: 10 June 2021 – Published: 6 July 2021

Abstract. Effective and accurate ocean and coastal wave predictions are necessary for engineering, safety and recreational purposes. Refining predictive capabilities is increasingly critical to reduce the uncertainties faced with a changing global wave climatology. Simulating WAVes in the Nearshore (SWAN) is a widely used spectral wave modelling tool employed by coastal engineers and scientists, including for operational wave forecasting purposes. Fore- and hindcasts can span hours to decades, and a detailed understanding of the computational efficiencies is required to design optimized operational protocols and hindcast scenarios. To date, there exists limited knowledge on the relationship between the size of a SWAN computational domain and the optimal amount of parallel computational threads/cores required to execute a simulation effectively. To test the scalability, a hindcast cluster of 28 computational threads/cores (1 node) was used to determine the computation efficiencies of a SWAN model configuration for southern Africa. The model extent and resolution emulate the current operational wave forecasting configuration developed by the South African Weather Service (SAWS). We implemented and compared both OpenMP and the Message Passing Interface (MPI) distributing memory architectures. Three sequential simulations (corresponding to typical grid cell numbers) were compared to various permutations of parallel computations using the speed-up ratio, time-saving ratio and efficiency tests. Generally, a computational node configuration of six threads/cores produced the most effective computational set-up based on wave hindcasts of 1-week duration. The use of more than 20 threads/cores resulted in a decrease in speed-up ratio for the smallest computation domain, owing to the increased sub-domain communication times for limited domain sizes.

1 Introduction

The computational efficiency of metocean (meteorology and oceanography) modelling has been the topic of ongoing deliberation for decades. The applications range from long-term atmospheric and ocean hindcast simulations to the fast-responding simulations related to operational forecasting. Long-duration simulations are usually associated with climate-change-related research, with simulation periods of at least 30 years across multiple spatial and temporal resolutions needed to capture key oscillations (Babatunde et al., 2013). Such hindcasts are frequently used by coastal and offshore engineering consultancies for purposes such as those related to infrastructure design (Kamphuis, 2020) or environmental impact assessments (Frihy, 2001; Liu et al., 2013).

Operational (or forecasting) agencies are usually concerned with achieving simulation speeds that would allow them to accurately forewarn their stakeholders of immediate, imminent and upcoming metocean hazards. The main stakeholders are usually other governmental agencies (e.g. disaster response or environmental affairs departments), commercial entities, and the public. Both atmospheric and marine forecasts share similar numerical schemes that solve the governing equations and thus share a similar need in computational efficiency. Fast simulation times are also required for other forecasting fields such as hydrological dam-break models (e.g. Zhang et al., 2014). Significant advancement in operational forecasting can be made by examining the way in which the code interfaces with the computation nodes, and how results are stored during simulation.

Numerous operational agencies (both private and public) make use of Simulating Waves in the Nearshore (SWAN) to predict nearshore wave dynamics (refer to Genseberger
and Donners, 2020, for details regarding the SWAN numerical code and solution schemes). These agencies include the South African Weather Service (e.g. Rautenbach et al., 2020), MetOcean Solutions (a division of the MetService of New Zealand) (e.g. de Souza et al., 2020), the United Kingdom Met Office (e.g. O’Neill et al., 2016) and the Norwegian Meteorological Institute (e.g. Jeuring et al., 2019). In general, these agencies have substantial computational facilities but nonetheless still face the challenge of optimizing the use of their computational clusters between various models (being executed simultaneously). These models may include atmospheric models (e.g. the Weather Research and Forecasting (WRF) model), Hydrodynamic models (e.g. Regional Ocean Modeling System (ROMS)) and the Semi-implicit Cross-scale Hydroscience Integrated System Model (SCHISM) and spectral wave models (e.g. Wave Watch III (WW3) and SWAN; Holthuijsen, 2007; The SWAN team, 2006, 2019b, a). Holthuijsen (2007) presents a theoretical background to the spectral wave equations, wave measurement techniques and statistics as well as a concluding chapter on the theoretical analysis of the SWAN numerical model. There must also be a balance between hindcast and forecast priorities and client needs. Some of these agencies use a regular grid (instead of irregular grids, e.g. Zhang et al., 2016), with nested domains in many of their operational and hindcast projects. Here, we focus only on the computational performance of a structured regular grid (typically implemented for spectral wave models). Kerr et al. (2013) performed an inter-model comparison of computational efficiencies by comparing SWAN, coupled with ADCIRC, and the NOAA official storm surge forecasting model, Sea, Lake, and Overland Surges from Hurricanes (SLOSH); however, they did not investigate the optimal thread usage of a single model. Other examples of coupled wave and storm surge model computational benchmarking experiments include Tanaka et al. (2011) and Dietrich et al. (2012), who used a unstructured meshes to simulate waves during Hurricanes Katrina, Rita, Gustav and Ike in the Gulf of Mexico. Results from these models were presented on a log–log scale, and their experimental design tested computational thread numbers not easily obtainable by smaller agencies and companies. The latter rather require sequential versus paralleled computational efficiencies using smaller-scale efficiency metrics. Genseberger and Donners (2015) explored the scalability of SWAN using a case study focused on the Wadden Sea in the Netherlands. By investigating the efficiency of both the OpenMP (OMP) and MPI version of the then current SWAN, they found that the OpenMP was more efficient on a single node. They also proposed a hybrid version of SWAN, to combine the strengths of both implementations of SWAN; using OpenMP to more optimally share memory and MPI to distribute memory over the computational nodes.

Here we build on the case study of Genseberger and Donners using results produced in the present study for southern Africa, to answer the following research questions:

1. When using SWAN, is it always better to have as many threads/cores as possible available to solve the problem at hand?
2. What is the speed-up relationship between number of threads/cores and computational grid size?
3. At what point (number of threads/cores) do the domain sub-communications start to make the whole computation less effective?
4. What is the scalability of a rectangular computational grid in a typical SWAN set-up?

2 Methodology and background

Details of the model configuration can be found in Rautenbach et al. (2020a, b). The computational domain (refer to Fig. 1) and physics used here were the same as presented in those studies.

All computations were performed on Intel Xeon Gold E5-2670 with 2.3 GHz computational nodes. A total of 28 threads/cores each with 96 GB RAM was used with 1 Gbyte inter-thread communication speed. Given that the present study was performed using a single computational node, inter-node communication speeds are not considered. Thus, given a computational node with similar processing speed, the present study should be reproducible. In general, these node specifications are reasonably standard, and therefore the present study is an acceptable representation of the SWAN scalability parameters.

SWAN 40.91 was implemented with the Van der Westhuysen whitecapping formulation (van der Westhuysen et al., 2007) and Collins bottom friction correlation (Collins, 1972) with a coefficient value of 0.015. Fully spectral wave boundary conditions were extracted from a global Wave Watch III model at 0.5 geographical degree resolution.

Here, the main aim was not the validation of the model but rather to quantify the relative computational scalabilities, as described at the end of the previous section. However, it should be noted that no nested domains were employed during the present study. Only the parent domain was used as a measure for scalability. The computational extent given in Rautenbach et al. (2020a, b) contains numerous non-wet grid cells that are not included in the computational expense of the current study. In Table 1, the size of the computational domain and resolution together with the labelling convention is given. For clarity, we define the resolutions as low, medium and high, denoted L, M and H, respectively, in the present study (noting that given the domain size, these resolutions would be classified as intermediate to high regional resolution for operational purposes).

The test for scalability ability of a model used here was the ability to respond to an increased number of computations with an increasing amount of resources. In the present study these resources are computational threads/cores. Computations were performed for an arbitrary week to assess model performance. Model spin-up was done via a single stationary computation. The rest of the computation was per-
formed using a non-stationary computation using an hourly time step, which implied wind-wave generation within the model occurred on the timescale of the wind forcing resolution. The grid resolutions used in the present study corresponded to 0.1, 0.0625 and 0.05 geographical degrees. Local bathymetric features were typically resolved through down-scaled, rotated, rectangular grids, following the methodology employed by Rautenbach et al. (2020a). A nested resolution increase of more than 5 times is also not recommended (given that the regional model is nested in the global Wave Watch III output at 0.5 geographical degree resolution, (Rautenbach et al., 2020a). Given these constraints, these resolutions represent a realistic and typical SWAN model set-up, for both operational and hindcast scenarios.

The three main metrics for estimating computational efficiency are the speed-up, time-saving and efficiency ratios. A fourth parameter, and arguably the most important, is the scalability and is estimated using the other three parameters as metrics.

The speed-up ratio is given as

$$S_p = \frac{T_1}{T_p},$$

(1)

where $T_1$ is the time in seconds it takes for a sequential computation on one thread and $T_p$ is the time a simulation takes with $p$ computational threads/cores (Zhang et al., 2014).

The time-saving ratio is given by

$$T_1 S_p = \frac{(T_1 - T_p)}{T_1},$$

(2)

and the efficiency ratio is defined as

$$E_p = \frac{S_p}{p}.$$  

(3)

The scalability of SWAN was tested based on the speed-up ratios for the grid resolutions in Table 1.

Zafari et al. (2019) recently presented some of the first results investigating the effect of different compilers on the scalability of a shallow water equation solver. Their experiments compared a model compiled with GNU Compiler Collection (gcc) 7.2.0 and linked with OpenMPI and Intel C++ compilers with Intel MPI for relatively small computational problems. Their numerical computation considered models with 600, 300 and 150 K grid cell sizes (what they called matrix size). These computational grid sizes were deemed “small”, but they still acknowledged the significant computational resources required to execute geographical models of this size due to the large number of time steps undertaken to solve these problems.

From a practical point of view, regular SWAN grids will rarely be used in dimensions exceeding the resolutions presented in the previous section. The reason for this statement is twofold: (1) to downscale a spectral wave model from a global resolution to a regional resolution should not exceed a 5-fold refinement factor and (2) when reasonably high resolutions are required in the nearshore (to take complex bathymetric features into account), nested domains are preferred. The reasoning will be different for an unstructured grid approach (Dietrich et al., 2012). Given these limitations with the widely used structured SWAN grid approach, SWAN grids will almost exclusively be deemed as a low spatial computational demand model. Small tasks create a sharp drop in
performance via the Intel C++ compiler due to the “work stealing” algorithm, aimed at balancing out the computational load between threads/cores (Zafari et al., 2019). In this scenario, the threads/cores compete against each other resulting in an unproductive simulation. In our experiments, each task performed via Intel was approximately 13 times faster but the overall performance was 16 times slower than the equivalent gcc version of the compiled shallow water model presented by Zafari et al. (2019).

3 Results

In Fig. 2, the computational scalability of SWAN is given as a function of number of computational threads/cores. Figure 2a shows the computational time in seconds and here the model resolutions grouped together with not much differentiation between them. These results also highlight the need for performance metrics, as described in the previous section. From Fig. 2b the MPI version of SWAN is more efficient for all the computational domain sizes. There is also a clear grouping between OMP and MPI. Figure 2c presents the speed-up ratios and clearly indicates that the MPI version of SWAN outperforms the OMP version. The closer the results are to the 1 : 1 line, the better the scalability.

Near-linear speed-up is observed for a small number of computational threads/cores. This result agrees with those reported by Zafari et al. (2019). In Fig. 2d the results are expressed via the time-saving ratio. In this case, the curves start to asymptote with thread counts larger than approximately 6.

4 Discussion

The behaviour noted in the results is similar to the dam-breaking computational results reported by Zhang et al. (2014). Genseberger and Donners (2020) present the latest finding on the scalability and benchmarking of SWAN. However, their focus was quantifying the performance of their new hybrid version of SWAN. In their benchmarking experiments (for the Wadden Sea, in the Netherlands), they obtained different results to those presented in Fig. 2a, with OMP running faster than MPI for small thread/core counts. For larger thread/core counts MPI performs better in the present study. This difference in performance is probably related to the particular hardware configuration (Genseberger and Donners, 2015). It must still be noted that with an increased number of nodes, and thus threads/cores, the total computational time should continue to decrease up until the point where the internal domain decomposition, communication efficiencies, starts to outweigh the gaining of computational power. Based on results of Genseberger and Donners (2020), we can estimate that, for our node configuration and region of interest, the communication inefficiencies will become dominant at approximately 16 nodes (16 × 24 threads/cores). One of the possible explanations for the non-perfect speed-up observed in Fig. 2c is related to the computational domain partition methods used, and the wet and dry (or active and inactive) point definitions in the model. In the present study the dry points were the bathymetry or topography values above mean sea level (MSL). The employed partition method is currently stripwise because of the underlying parallel technique, namely the wavefront method (Genseberger and Donners, 2015; Zijlema, 2005). The stripwise partition is thus potentially not the most effective method to optimize speed-up.

In the present study this partition leads to an optimal point around six threads/cores without losing too much parallel efficiency. In general, increasing the number of threads/cores will still produce results faster, but in an increasingly inefficient manner. This trend is clear from Fig. 2c and d where the total computational time (speed-up ratio, time-saving ratio) does not scale linearly with the increasing number of threads/cores. The ideal scenario (linear scalability) would be if the computational results followed the 1 : 1 line in Fig. 2c. In Fig. 2d the non-linear flattening of the time-saving ratio is also evident, although the ratio still slightly increases beyond six threads/cores. This result implies the total computational time will marginally decrease with an increasing number of threads/cores. This marginal decrease in computational time could, however, still make significant differences in the total simulation times when extensive simulation periods are considered.
Figure 2. Model performance as a function of the number of computational threads/cores. (a) Computing time in seconds, (b) efficiency (Eq. 3), (c) speed-up ratio (Eq. 1) and (d) the time-saving ratio (Eq. 2).

5 Conclusions

The present study investigated the scalability of SWAN, a widely used spectral wave model. Three typical wave model resolutions were used for these purposes. Both the OpenMP (OMP) and the Message Passing Interface (MPI) implementations of SWAN were tested. The scalability is presented via three performance metrics: the efficiency, speed-up ratio and the timesaving ratio. The MPI version of SWAN outperformed the OMP version based on all three metrics. The MPI version of SWAN performed best with the largest computational domain resolution, resulting in the highest speed-up ratios. The time-saving ratio indicated a decrease after approximately six computational threads/cores. This result suggests that six threads/cores are the most effective configuration for executing SWAN. The largest increases in speed-up and efficiency were observed with small thread counts. According to Genseberger and Donners (2020), computational times decrease up to ~16 nodes (16 × 24 threads/cores), indicating the wall-clock optimal computational time for their case study. This result suggests that multiple nodes will be required to reach the optimal wall-clock computational time – even though this turn point might not be the most efficient computational configuration. Ultimately, the efficiencies recommended here can improve operational performance substantially, particularly when implemented over the range of modelling software needed to produce useful metocean forecasts. Future studies might consider investigating the scalability employing a gcc compiler.

Code and data availability. The open-source version of SWAN was run for the purposes of the present study. SWAN may be downloaded from http://swanmodel.sourceforge.net/ (SWAN, 2020). To ensure a compatible version of SWAN remains available, the current, latest version of SWAN is permanently archived at https://hdl.handle.net/10289/14269 (SWAN, 2021). The bathymetry used for the present study may be downloaded from https://www.gebco.net/ (GEBCO, 2020), and the wind forcing may be found from https://climatedataguide.ucar.edu/climate-data/climate-forecast-system-reanalysis-cfsr (National Center for Atmospheric Research Staff, 2017).

Author contributions. CR conceptualized the study, executed the experiments and wrote the paper. He also secured the publication funding. JCM and KRB reviewed the paper.
Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This research was funded by the National Research Foundation of South Africa (grant number 116359).

Disclaimer. Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Financial support. This research was funded by the National Research Foundation of South Africa (grant number 116359).

Review statement. This paper was edited by Julia Hargreaves and reviewed by two anonymous referees.

References

Babatunde, A., Pascale, B., Sin, C. C., William, C., Peter, C., Fatima, D., Seita, E., Veronika, E., Forest, C., Peter, G., Eric, G., Christian, J., Vladimir, K., Chris, R., and Markku, R.: Evaluation of Climate Models, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner, G. K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, UK, 2013.

Collins, J. I.: Prediction of shallow-water spectra, J. Geophys. Res., 77, 2693–2707, https://doi.org/10.1029/JC077i015p02693, 1972.

de Souza, J. M. A. C., Couto, P., Soutelino, R., and Roughan, M.: Evaluation of four global ocean reanalysis products for New Zealand waters – A guide for regional ocean modelling, New Zeal. J. Mar. Fresh., 55, 132–155, https://doi.org/10.1080/00288330.2020.1713179, 2020.

Dietrich, J. C., Tanaka, S., Westerink, J. J., Dawson, C. N., Luetttich, R. A., Zijlema, M., Holthuijsen, L. H., Smith, J. M., Westerink, L. G., and Westerink, H. J.: Performance of the Unstructured-Mesh, SWAN+ADCIRC Model in Computing Hurricane Waves and Surge, J. Sci. Comput., 52, 468–497, https://doi.org/10.1007/s10915-011-9555-6, 2012.

Frihy, O. E.: The necessity of environmental impact assessment (EIA) in implementing coastal projects: lessons learned from the Egyptian Mediterranean Coast, Ocean Coast. Manage., 44, 489–516, https://doi.org/10.1016/S0964-5691(01)00062-X, 2001.

GEBCO: General Bathymetric Chart of the Oceans, available at: https://www.gebco.net, last access: 2 February 2020.

Genseberger, M. and Donners, J.: A Hybrid SWAN Version for Fast and Efficient Practical Wave Modelling, Procedia Comput. Sci., 51, 1524–1533, https://doi.org/10.1016/j.procs.2015.05.342, 2015.

Genseberger, M. and Donners, J.: Hybrid SWAN for Fast and Efficient Practical Wave Modelling – Part 2, edited by: Krzhizhanovskaya, V. V., Závodský, G., Lees, M. H., Dongarra, J. J., Sloat, P. M. A., Brissos, S., and Teixeira, J., Springer International Publishing, Cham, vol. 12139, 87–100, 2020.

Holtluujsen, L. H.: Waves in Oceanic and Coastal Waters, Cambridge University Press, Technische Universiteit Delft, The Netherlands, https://doi.org/10.1017/CBO9780511618536, 2007.

Jeuring, J., Knol-kauffman, M., and Sivle, A.: Toward valuable weather and sea-ice services for the marine Arctic: exploring user – producer interfaces of the Norwegian Meteorological, Polar Geogr., 43, 139–159, https://doi.org/10.1080/1088937X.2019.1679270, 2019.

Kamphuis, J. W.: Introduction to coastal engineering and management – Advanced series on ocean engineering, World scientific publishing Co. Pte. Ltd., Singapore, Vol. 48, 2020.

Kerr, P. C., Donahue, A. S., Westerink, J. J., Luetttich, R. A., Zheng, L. Y., Weisberg, R. H., Huang, Y., Wang, H. V., Teng, Y., Forrest, D. R., Roland, A., Haase, A. T., Kramer, A. W., Taylor, A. A., Rome, J. R., Feyen, J. C., Signell, R. P., Hanson, J. L., Hope, M. E., Estes, R. M., Dominguez, R. A., Dunbar, R. P., Semeraro, L. N., Westerink, H. J., Kennedy, A. B., Smith, J. M., Powell, M. D., Cardone, V. J., and Cox, A. T.: U.S. IOOS coastal and ocean modeling testbed: Inter-model evaluation of tides, waves, and hurricane surge in the Gulf of Mexico, J. Geophys. Res.-Oceans, 118, 5129–5172, https://doi.org/10.1029/jc010j6.2013.

Liu, T. K., Sheu, H. Y., and Tseng, C. N.: Environmental impact assessment of seawater desalination plant under the framework of integrated coastal management, Desalination, 326, 10–18, https://doi.org/10.1016/j.desal.2013.07.003, 2013.

National Center for Atmospheric Research Staff (Eds.): The Climate Data Guide: Climate Forecast System Reanalysis (CFSR), available at: https://climatedataguide.ucar.edu/climate-data/cclimate-forecast-system-reanalysis-cfsr (last access: 1 June 2020), 2017.

O’Neill, C., Saulter, A., Williams, J., and Horsburgh, K.: NEMO-surge: Application of atmospheric forcing and surge evaluation, Technical report 619, available at: http://www.metoffice.gov.uk/binaries/content/assets/mohippo/pdf/library/frtr_619_2016p.pdf (last access: 1 February 2020), 2016.

Rautenbach, C., Daniels, T., de Vos, M., and Barnes, M. A.: A coupled wave, tide and storm surge operational forecasting system for South Africa: validation and physical description, Nat. Hazards, 103, 1407–1439, https://doi.org/10.1007/s11069-020-04042-4, 2020a.

Rautenbach, C., Barnes, M. A., Wang, D. W. and Dykes, J.: Southern African wave model sensitivities and accuracies, J. Mar. Sci. Eng., 8, 77, https://doi.org/10.3390/jmse8100773, 2020b.

SWAN (Simulating Waves Nearshore): a third generation wave model, Delft University of Technology, available at: http://swanmodel.sourceforge.net, last access: 18 September 2020.

SWAN (Simulating Waves Nearshore): a third generation wave model, Delft University of Technology, available at: https://hdl.handle.net/10289/14269, last access: 1 April 2021.

Tanaka, S., Bunya, S., Westerink, J. J., Dawson, C., and Luetttich, R. A.: Scalability of an unstructured grid continuous Galerkin ocean modelling testbed: Inter-model evaluation of tides, waves, and hurricane surge in the Gulf of Mexico, J. Geophys. Res.-Oceans. 118, 5129–5172, https://doi.org/10.1029/jc010j6.2013.

The SWAN Team: USER MANUAL SWAN Cycle III version 40.51, Cycle, available at: http://falk.ucsd.edu/modeling/swanuse.pdf (last access: 1 March 2021), 2006.
The SWAN team: Implementation Manual Swan Cycle III version 41.31, available at: http://swanmodel.sourceforge.net/online_doc/swanimp/swanimp.html (last access: 1 March 2021), 2019a.
The SWAN team: User Manual Swan Cycle III version 41.3, available at: http://swanmodel.sourceforge.net/download/zip/swanimp.pdf (last access: 2 July 2021), 2019b.

van der Westhuysen, A. J., Zijlema, M., and Battjes, J. A.: Nonlinear saturation-based whitecapping dissipation in SWAN for deep and shallow water, Coast. Eng., 54, 151–170, https://doi.org/10.1016/j.coastaleng.2006.08.006, 2007.

Zafari, A., Larsson, E., and Tillenius, M.: DuctTeip: An efficient programming model for distributed task-based parallel computing, Parallel Comput., 90, 102582, https://doi.org/10.1016/j.parco.2019.102582, 2019.

Zhang, S., Xia, Z., Yuan, R., and Jiang, X.: Parallel computation of a dam-break flow model using OpenMP on a multi-core computer, J. Hydrol., 512, 126–133, https://doi.org/10.1016/j.jhydrol.2014.02.035, 2014.

Zhang, Y. J., Ye, F., Stanev, E. V., and Grashorn, S.: Seamless cross-scale modeling with SCHISM, Ocean Model., 102, 64–81, https://doi.org/10.1016/j.ocemod.2016.05.002, 2016.

Zijlema, M.: Parallelization of a nearshore wind wave model for distributed memory architectures, in: Parallel Computational Fluid Dynamics 2004, Elsevier, 207–214, 2005.