GPU Ocean:
Heterogeneous computing of drift in the ocean

1 Relevance relative to the call for proposals

Problem statement: Numerical ocean models are used for oil spill tracking, search and rescue, and in costly offshore operations involving large floating structures. The natural variability in the ocean currents is large, and a major problem is the lack of ocean observations that can be used to constrain the ocean model. Ongoing efforts focus on exploiting advanced data assimilation techniques and devising efficient observation sampling strategies. Unfortunately, traditional ocean modeling systems are too computationally demanding for the full range of forecasting uncertainties to be explored.

Our approach: We propose to *complement* the traditional ocean modeling systems with a massive ensemble of simplified ocean and drift models. These simplified models will be combined with fully non-linear data assimilation and the codes will be executed on massively parallel architectures. We aim to use simplified physical and mathematical models that are valid for the short predictive time scales necessary for operational oceanography, which is typically 1-3 days. These simplified models will be initialized from the traditional modeling system and hence provide incremental updates to the full description of the physical state of the ocean. The result is accurate ocean current predictions with detailed uncertainties.

Relevance: Predicting ocean currents is a daunting task, and efficient use of ocean observations and assessment of forecasting uncertainties is of critical importance. Our approach requires a highly interdisciplinary effort, covering statistics, oceanography, computational mathematics, and informatics. We do not know of any existing efforts that are similar to the novel approach taken herein. If successful, our methodology can pave the way for solving other similar problems, such as transport of volcanic ash in the atmosphere.

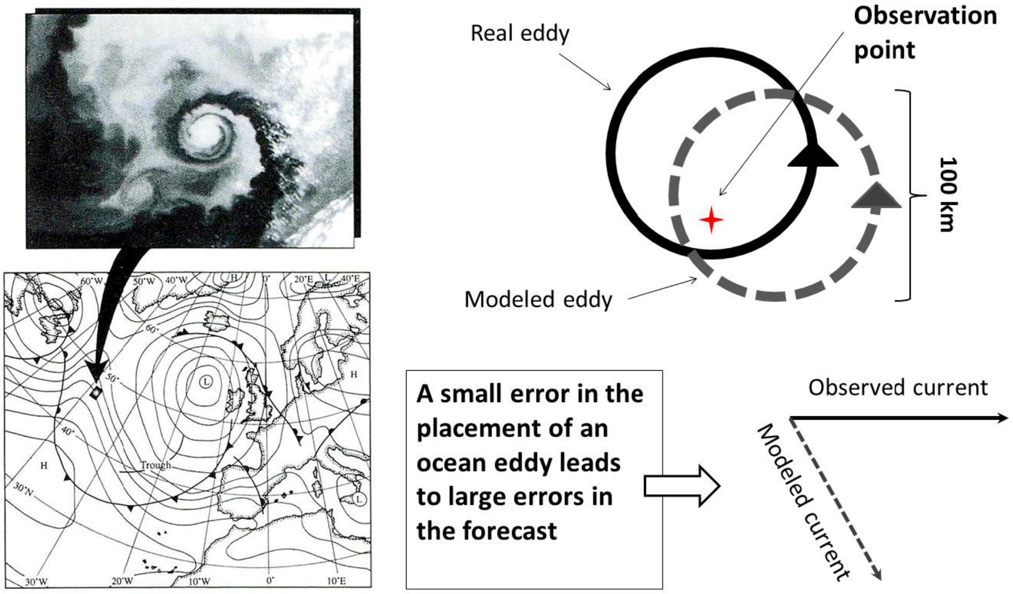

Figure 1: Although the dynamics is similar, there is a huge difference in scale between an atmospheric low pressure system (bottom left) and an oceanic low pressure system (top left), or ocean eddy, as these features are usually called. The modelled position of an atmospheric "eddy" may be wrong by 10-30 km without having any real consequence with regard to the accuracy of the weather forecast. As shown on the right, an error of similar size in the placement of an ocean eddy has a large negative impact on the ocean forecast.

2 Aspects relating to the research project

Background and status of knowledge

Ocean current and drift trajectory modeling: In general, the ocean is stably stratified from the surface down to the bottom. The upper 10-100 meters is usually well mixed due to wind and waves, and this mixed layer lies on top of increasingly dense layers of water with increasing depth. On long time scales, the overall ocean circulation is dominated by effects due to changes in temperature and salinity (thermohaline circulation), while on short time scales the circulation is dominated by wind, waves and tides. The traditional ocean forecast model includes all these mechanisms.

The ocean general circulation model used by MET Norway has a high level of complexity and is specifically tailored to run on the Norwegian national supercomputers. Model runs are executed twice per day and for each of these runs it consumes computing power approximately equivalent to 200 hours on a powerful desktop PC. The strength of this "traditional" model is the ability to give a coherent physical description of the overall state of the ocean. The model is unable, however, to capture the exact positioning of ocean eddies, which are the ocean equivalents of high and low pressure systems in the atmosphere. Because ocean eddies are much smaller than their atmospheric counterparts and there are typically very few direct observations of the ocean, they are very difficult to to position correctly in time and space. This has a profound impact on the accuracy of our day-to-day ocean current forecasts (see Figure 1). When accurate predictions are required, such as for important marine operations or an oil spill, tailored observations are sometimes acquired by deploying buoys, for example. Detailed observations are of paramount importance to capture the correct dynamical balance in the ocean models, but such observations will unfortunately remain a scarce luxury for the foreseeable future.

In addition to eddy placement, the most important factors for surface currents are wind stress and waves, tidal phase, and correct bathymetry. All of these parameters have uncertainties associated with them, and will affect the trajectory of any floating object. Given that a surface current prediction exists, it is possible to simulate and predict drift trajectories for floating objects. Such transport or advection equations are well studied and can either be modeled as passive advection, independent of the underlying velocity field, or as a coupled system. The passive transport can be viewed as representative for plankton or other components of the water which do not affect the currents themselves. On the other hand, an oil spill may change surface rheology or bulk properties, and thereby also change the flow [e.g., 5]. Such systems therefore need to be tightly coupled with the current model.

Uncertainty quantification and nonlinear data assimilation: The size of forecasting errors can be estimated using so-called ensemble prediction systems (EPS) [6]. Ensemble methods have been used for over 20 years to improve the accuracy in operational weather forecasting [9], and have been especially important to avoid missing extreme weather developments. EPS methods are based on running many different instances (ensemble members) of the forecast model. For example, if the wind forcing of each ensemble member is varied slightly, the results will also vary. This difference between members is related to the model error, which again can be different from one geographical location to another, thus giving an indication of where the prediction is less reliable.

The ensemble can also be combined with observational data, known as ensemble based data assimilation, to produce the statistically most likely estimate of the ocean currents. Present-day data-assimilation techniques assume Gaussian distributed errors in the model and in observations, and solve the nonlinear data-assimilation problem iteratively through a set of linear problems. For highly nonlinear problems with a limited number of observations, like the present problem, such methods cannot guarantee convergence. Nonlinear methods are available and most promising for the present application is the fully nonlinear particle filter. In the particle filter method, each ensemble member is referred to as a particle and the filter refers to the comparison with observations and subsequent handling of the particles. A particle filter provides a description of the full probability density function, so best estimates, their errors, and probabilistic information can be extracted [14].

The present application is ideally suited for a particle filter. The model is realistic, highly nonlinear, but the number of uncertain input parameters (eddy placement, wind stress, tidal phase, bathymetry, etc.) creates an enormous demand on the number of ensemble members. An essential part is therefore to reduce the number of particles required [15]. By exploring the freedom in the so-called proposal density, it is possible to dramatically increase the effectiveness of the particles. One can force all particles to end up relatively close to the observations, whilst simultaneously ensuring that the correct full nonlinear data-assimilation problem is solved. This allows for statistically sound estimates of e.g., the ensemble mean and its error covariance.

A difficulty with ensemble simulation is the extensive resources required. Each ensemble member requires the same computing time as a single run, permitting only a limited number of members (e.g., the atmospheric EPS at MET Norway has 21). One way of allowing more ensemble members is to decrease the model resolution, but this is often counterproductive for our target application areas since important small scale features are lost.

Current ensemble simulations typically use the increasing computational power offered by new hardware for higher grid resolution or more physically correct simulation models. For short term forecasts an alternative is to reduce the model complexity by neglecting effects that are only important for long term predictions. Such simplified ocean models were used extensively a few decades ago when available computing resources were just a fraction of what we have today. The models are able to capture the most important features of the ocean circulation, but do not contain enough physical effects to be suitable for general purpose ocean monitoring and forecasting. They can therefore be used to give accurate short term predictions of the ocean currents if properly initialized.

In ensemble simulations, classical Monte Carlo methods have a relatively poor convergence rate, which can be greatly improved by moving to multi-level Monte Carlo, for example. In multi-level Monte Carlo, the samples are drawn from different grid resolutions, which yields a much better convergence rate. Looking at accuracy versus work performed, multi-level Monte Carlo is on the same order of magnitude as a single solve of the original deterministic problem [8]. Multi-level Monte Carlo methods have successfully been applied to mathematical models that closely resemble the simplified ocean circulation models targeted in this project [12].

Massively parallel simulation: Over the last decade, we have seen a dramatic change in computer architecture. Single-core CPUs have been displaced by multi-core CPUs and many-core processors such as graphics processing units (GPUs), capable of performing thousands of operations in parallel, are today present in everything from laptops to the most powerful supercomputers. In particular, many-core processors have shown to give orders of magnitude speedup over CPUs for a wide range of problems [1]. However, such processors are challenging to program, as they have complex memory hierarchies and multiple levels of parallelism that a programmer explicitly has to take into consideration. High-level programming languages, such as C++ AMP [7] or OpenACC [10], can hide these complexities from a programmer, but they often have an associated performance impact. There are also programming languages, such as OpenCL [13], that operate at a lower level and aim to be hardware independent. However, portable code does not equate portable performance, and one often has to resort to manual optimization for each target architecture, or auto-tuning strategies.

Exploiting massively parallel architectures is still a significant challenge, in particular since many scientific algorithms have been, and still are, developed without regard to concurrency. However, we have previously shown that mathematical simulations resembling the simplified ocean models considered herein map very well to these architectures [2, 3, 4]. Furthermore, GPUs are ideal for Monte-Carlo simulations and data-assimilation methods such as the particle filter, due to the extreme degree of parallelization. In general, current hardware standards allow for an ensemble with $10^3 - 10^4$ members for the simplified ocean circulation models we target. This high number of ensemble members enables rapid exploration of a huge number of different scenarios, contributing to a detailed quantification of uncertainties.

Although the GPU will be our architecture of choice, the algorithms and code developed in this project will also be transferable to other similar massively parallel architectures.

Approaches, hypotheses and choice of method

This project will develop mathematical and physical models and efficient software implementations of these. Our research can be broken down into separate parts or components, and our work with each component will follow the same general strategy. Starting with existing software components and mathematical models, we will implement first prototypes and perform verification to check that our implementation correctly solves the mathematical problem. We will then continue by coupling the different components to make the first full system prototype. This prototype can then be used to check the validity of our approach for simple problems, and will further developed and refined throughout the project to solve the end goal: accurate drift trajectories with uncertainty quantification. A common challenge with research projects such as this is the lack of good test cases to check the validity of the approach. MET Norway has an extensive inventory of data sets that are relevant for this project (e.g., the drifter data collected in the RCN project BIOWAVE), and integration with the ongoing RCN project RETROSPECT (Petromaks2) will give access to more field data. These data sets will serve as benchmarks and test cases during the research in this project.

Our approach requires a highly interdisciplinary effort, covering statistics, oceanography, computational mathematics, and informatics. These disciplines are covered by the participating institutions, which have extensive experience from working interdisciplinary in previous projects. Throughout the project, we will have a focus on reproducible research, and our software and results will be made available online under an open source software license.

Simplified ocean current model: We will initially construct a simplified ocean model based on a description of the ocean as two separate layers of constant density fluid, with no heat or mass exchange across the interface [11]. Provided proper initialization, the model is able to sustain typical features of the ocean circulation for a limited period. The traditional ocean forecast model at MET Norway will be used to generate a "qualified first guess" for these initial fields and boundary conditions; hence thermohaline effects are implicitly included. The atmospheric forecast model will be used to generate wind forcing.

Utilizing Python and existing software components, we will ensure a rapid research and development phase in which the first two-layer prototype will be quickly up and running to demonstrate the basic concepts of our approach. We will continue by extending our prototype with OpenCL capabilities to utilize the GPU, thereby making it suitable for ensemble simulations. This is key for the efficient evaluation of a large number of ensemble members.

Drift trajectory model: We will extend the simplified ocean current model to include a decoupled passive drift trajectory model in the first prototype. This model will be refined during the project to include the uncertainties of inputs, and to support different physical source terms such as air and water side drag coefficients and Coriolis force, which is relevant for e.g., icebergs. Further developments include explicitly modeling horizontal dispersion, which is relevant for e.g., oil spill drift.

Ensemble generation: Once we have the prototype drift trajectory model in place, we will develop and prototype methods for perturbing the initial state to generate ensemble members. A major undertaking in the project will here be to generate an ensemble that spans the probability density function of our uncertain input parameters, and this constitutes one of the key research challenges. Random perturbation (i.e., uniform Monte Carlo sampling), has shown to be quite inefficient, as randomly perturbing the initial state does not necessarily give a good sampling of the probability distribution function of the input. Multi-level Monte Carlo methods will therefore be explored.

In operational weather forecasting, the initialization process is often performed using empirical orthogonal functions to create perturbed initial conditions that sample the fastest growing modes. We will adapt these experiences to our simplified models. However, since we target a large number of ensemble members, we also require that the initialization of ensemble members is highly efficient and can be executed rapidly. One approach to achieve this is to parametrically add ocean eddies in the model domain by perturbing the vorticity field. The tidal phase and/or bottom topography will equivalently be perturbed parametrically to create a representative set of input initial conditions.

Non-linear data assimilation: With an ensemble simulation framework in place, we can develop particle filters for data assimilation. Using particle filters, ensemble members are removed and reinitialized if they have a poor match with observations. In addition to this, we will explore new developments in particle filtering that will allow us to greatly reduce the number of particles needed [15]. The particle filter allows one to find statistically sound estimates of e.g. the ensemble mean and its error covariance, essential for operational applications. An example of the use of the error covariance is that it contains information on which direction the uncertainty is largest, which is useful for decision support, search and rescue operations, and when planning deployment of tracking buoys.

3 The project plan, project management, organisation and cooperation

We have composed a cross-disciplinary research team consisting of members that together represent the expertise required to successfully complete the project. Members of the research team have previously worked together on several papers, projects, and project proposals. The key personnel are

Dr. Kai Christensen (PI) (Senior scientist, Norwegian Meteorological Institute) is an expert in oceanography and ocean transport, and has worked extensively with ocean currents and drift trajectories.

Dr. Øyvind Sætra (Senior scientist, Norwegian Meteorological Institute) is an expert on wave and drift modeling and has extensive experience with ensemble prediction systems.

Dr. Martin L. Sætra (Scientist, Norwegian Meteorological Institute) has worked with GPU simulation of shallow water systems, which are highly related to simplified ocean circulation models.

Prof. Knut-Andreas Lie (Chief scientist, SINTEF ICT, Professor, NTNU) has long experience with the mathematical and numerical modeling of physical problems similar to those at hand.

Dr. André R. Brodtkorb (Scientist, SINTEF ICT) has worked extensively with numerical simulation on massively parallel architectures.

Prof. Peter Jan van Leeuwen (Professor and head of the Data Assimilation Research Centre, University of Reading) is an expert in data assimilation, and has worked extensively with non-linear models and geophysical fluid dynamics.

The majority of the project will be carried out at MET Norway and SINTEF in Oslo, which are physically located in close proximity to each other. This facilitates rapid interaction between project participants through face-to-face project meetings. The project team will in addition include one Ph.D. candidate employed at SINTEF, and one post-doc employed at MET. In addition to the work performed in Oslo, the post-doc and/or the Ph.D. candidate will each have a 3-6 month stay at the University of Reading under the supervision of Prof. van Leeuwen. The focus of the stay will be on non-linear data assimilation techniques, for which Prof. van Leeuwen is an internationally leading expert.

MET Norway has access to supercomputing resources on the Vilje supercomputer. SINTEF ICT is a CUDA research center, and has good access to machines with several GPUs that are suitable for use in this project. Most of the day-to-day development will be undertaken on standard computers equipped with a standard, yet powerful, GPU. In addition, we will apply for access to resources in the national eInfrastructure, and in particular the Abel supercomputer which has several GPU nodes relevant to this project[1].

The project can be broken down into four work packages, each with a specified set of milestones. The project plan is graphically visualized in the Gantt diagram in Figure 2. There will be a significant interaction between the different work packages, as we already now foresee that the specialized requirements from the different milestones will affect the work performed in different work packages.

Work package 1: Simplified ocean current model This work package consists of developing an efficient two-layer ocean current model capable of utilizing GPU resources. The work package can again be broken down into several milestones:

Milestone A: The first milestone will create a simulator (simplified ocean model) based on existing one-layer codes, but extended to two layers.

Milestone B: The second milestone will verify and validate the simulator against know test cases and existing codes to ensure validity of results.

Milestone C: Milestone C will extend the simulator to utilize the GPU efficiently through OpenCL.

---

[1] Projects financed by the research council of Norway are eligible to apply for computing time from the national eInfrastructure.

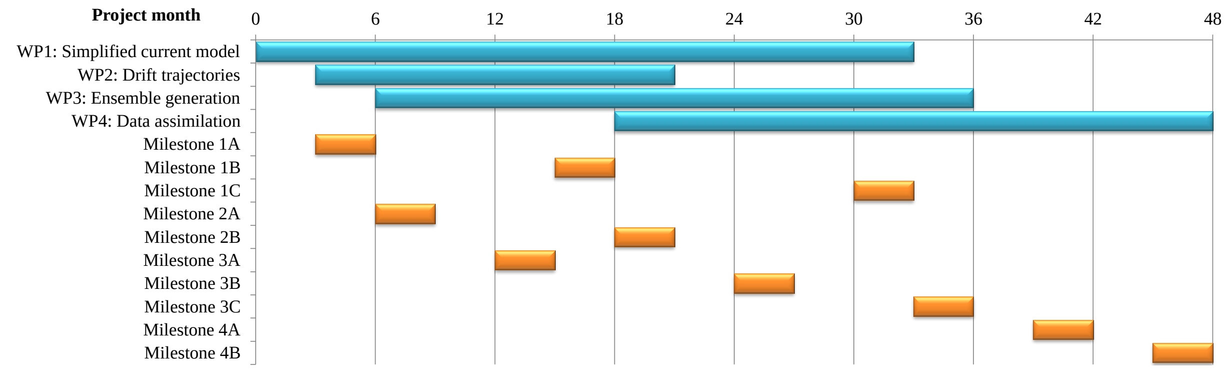

Figure 2: Gantt chart of project plan. The project plan is to develop the initial prototypes and integrate these in the first "wave". In the second "wave", these prototypes are iteratively refined to include more realistic features. In the third "wave" we will focus on performance and GPU acceleration, before we finally focus on the non-linear data assimilation.

Work package 2: Drift trajectory model The second work package will focus on extending and the ocean circulation model from work package 1 to include a drift component. Our initial version will be a decoupled passive drift component, which will be refined over the project.

Milestone A: The first milestone will use a passive Lagrangian particle model to represent e.g., a floating object. This prototype will be verified and validated against existing simulations to ensure validity of results.

Milestone B: The second milestone will extend the first prototype to also support transport of icebergs, including air and water side drag and Coriolis force. This will be performed by extending the ocean circulation model with a new component.

Work package 3: Ensemble generation Work package 3 will be a major undertaking in the project, and concerns the initialization of ensemble members that sample the probability density function. We can break this task down into several milestones:

Milestone A: The first milestone in this work package will use classical Monte Carlo sampling to generate a set of ensemble members. This prototype will then represent the first fully integrated simulation system with uncertainty quantification.

Milestone B: The second milestone will move from Monte Carlo to multi-level Monte Carlo to sample the probability density function of our uncertain input parameters.

Work package 4: Non-linear data assimilation Work package 4 concerns non-linear data assimilation, whereby observations are used to correct the simulation. The key objective is here to make optimal use of the few observations we are expected to have.

Milestone A: The first milestone will implement a particle filter without any special handling with respect to the probability density function. It is therefore expected that the ensemble will gradually collapse into only a few uncorrelated samples.

Milestone B: The second milestone in this work package will take special care to avoid the problem of a collapsing ensemble when reinitializing ensemble members.

National research network: The core project group is physically located in Oslo (SINTEF, MET Norway), which facilitates strong collaboration through face-to-face meetings, and will serve to strengthen the existing collaborations and facilitate network-building between the participating organizations and individuals.

Impact: The project will develop mathematical methods and software that addresses an important societal challenge. The methodology will further support a large number of similar problems within geosciences, and thereby have a significant impact for challenges important to our society.

Teaching environment and international collaboration: The key personnel from MET Norway and SINTEF have a track record for supervision of master and Ph.D. students, and this experience will be leveraged in this project. Collaboration with the University of Reading will include a 3-6 month research stay of the Ph.D. candidate under the supervision of Prof. van Leeuwen, who is the head of the Data Assimilation Research Centre at the University of Reading, which is at the front of data assimilation research. The centre has a large group of Ph.D.-level students, has long experience with visiting students, and offers a very good research environment for a Ph.D. candidate.

4 Key perspectives and compliance with strategic documents

Compliance with strategic documents: The proposed project is clearly in line with the strategic policy of MET Norway, which mission is to provide the best possible forecasts in order to protect life and property. The researchers at MET Norway are constantly working to improve the forecast models and investigate new ways of using observations and infrastructure.

SINTEF ICT has invested a significant amount of strategic research funds into developing heterogeneous computing using multi-core CPUs and many-core GPU accelerator units as explained above, and sees a research alliance with MET Norway as an important strategic step for future developments. MET Norway also views research into heterogeneous computing and accelerator technologies as strategically important, and both the project and the research alliance with SINTEF ICT would be in line with this strategy.

Relevance and benefit to society: Better ocean current forecasts are useful for the planning of marine operations and for protection of offshore installations. For example, there is increasing offshore activity in the north, e.g., in the Barents Sea, related to oil and gas exploration. In this area there is a risk of collision with icebergs, and icebergs drifting close to offshore installations are closely monitored (e.g., by GPS trackers). If necessary, an iceberg can be towed away, but this is of course a very costly operation. The proposed forecasting system would be ideal for predicting the drift trajectory of icebergs and hence be a valuable tool in the decision making process. This use would not be limited to icebergs, but include cargo containers, oil spills, etc. Search and rescue services would also be potential users of the proposed forecasting system.

Environmental impact: The environmental impacts of this project are positive: Ocean current forecasts are used for oil spill mitigation and tracking of other pollutants, and more accurate predictions will have a positive effect in such an event.

The use of heterogeneous computing will also have a positive environmental impact, since GPUs offer an order-of-magnitude more performance per watt than conventional CPUs. The project will therefore be a pilot study at MET Norway in so-called “green computing”. MET Norway is the primary user of national supercomputing resources, which is associated with significant energy consumption. MET Norway is therefore naturally very interested in new cost efficient and power-lean computational hardware. Furthermore, through MET Norway’s ‘MetCoOp’ project collaboration with the Swedish Meteorological and Hydrological Institute, in which a new supercomputer will be procured every other year, this experience will also impact supercomputing in Sweden.

Ethical perspectives: This project contains no foreseen issues from an ethical perspective. The applicants are aware and have considered the contents of the guide to ethics in research[2]. Should an ethical issue become apparent for a project participant, both MET Norway and SINTEF have dedicated contact points where such ethical questions can be raised and brought into consideration anonymously.

Gender issues (Recruitment of women, gender balance and gender perspectives) Gender equality is important to all members of the project team, and equal gender opportunities emphasized when employing the Ph.D. position at SINTEF.

5 Dissemination and communication of results

Dissemination plan The project will publish research findings in targeted scientific journals, and disseminate results at relevant conferences and workshops. The publications will be closely related to the milestones in the different work packages, and typically timed with these. We expect that the project will on average produce one conference and one journal paper per year, totalling to eight publications

Communication with users The Norwegian Meteorological Institute has end users in industry and government agencies for operational forecasts of ocean currents. The results of this project will be especially relevant for BarentsWatch, which is an surveillance and information system for Norwegian waters. The project will publish 1-2 news articles or blog entries every year on Yr.no (the web portal for weather forecasts by MET Norway), the Halo-project blog, and BarentsWatch to disseminate results to relevant users. Moreover, there will be yearly workshops for project partners, where preliminary findings are shared and discussed, and the project plan revised. The project will also publish key findings in popular science and news media at the end of the project. The software produced during the course of the project will be released under an open source software license and made available on Github.

---

[2] Forskningsetiske rettningslinjer for naturvitenskap og teknologi, De nasjonale forskningsetiske komiteer, 2009.

References

[1] A. R. Brodtkorb, C. Dyken, T. R. Hagen, J. M. Hjelmervik, and O. Storaasli. State-of-the-art in heterogeneous computing. *Scientific Programming*, 18(1):1 – 33, May 2010.

[2] A. R. Brodtkorb, T. R. Hagen, K.-A. Lie, and J. Natvig. Simulation and visualization of the Saint-Venant system using GPUs. *Computing and Visualization in Science*, 13(7):1–13, 2011.

[3] A. R. Brodtkorb, T. R. Hagen, and L. P. Røed. One-layer shallow water models on the gpu. Technical Report 27, Norwegian Meteorological Institute, 2013.

[4] A. R. Brodtkorb, M. L. Sætra, and M. Altinakar. Efficient shallow water simulations on GPUs: Implementation, visualization, verification, and validation. *Computers & Fuids*, 55:1–12, 2011.

[5] K. H. Christensen and E. Terrile. Drift and deformation of oil slicks due to surface waves. *Journal of Fluid Mechanics*, 620:313–332–, 2009.

[6] A. Melsom. Mesoscale activity in the north sea as seen in ensemble simulations. *Ocean Dyn.*, 2005.

[7] Microsoft. C++ Accelerated Massive Parallelism. [online], 2014. `http://msdn.microsoft.com/`.

[8] S. Mishra, C. Schwab, and J. Sukys. Multi-level monte carlo finite volume methods for uncertainty quantification in nonlinear systems of balance laws. Technical Report 2012-08 (revised), Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2012.

[9] F. Molteni, R. Buizza, T. N. Palmer, and T. Petroliagis. The ecmwf ensemble prediction system: Methodology and validation. *Quarterly Journal of the Royal Meteorological Society*, 122(529):73–119, 1996.

[10] OpenACC Members. OpenACC. [online], 2014. `http://www.openacc-standard.org/`.

[11] L. P. Røed. Modelling mesoscale features in the ocean. In *Waves and Nonlinear Processes in Hydrodynamics*. Kluwer Academic Publishers, 1996.

[12] C. Sanchez-Linares, M. de la Asuncion, M. Castro, S. Mishra, and J. Sukys. Multi-level monte carlo finite volume method for shallow water equations with uncertain parameters applied to landslides-generated tsunamis. Technical Report 2014-24, Seminar for Applied Mathematics, ETH Zürich, 2014.

[13] J. E. Stone, D. Gohara, and G. Shi. Opencl: A parallel programming standard for heterogeneous computing systems. *IEEE Des. Test*, 12(3):66–73, May 2010.

[14] P. J. van Leeuwen. Particle filtering in geophysical systems. *Monthly Weather Review*, 137:4089–4114, 2009.

[15] P. J. van Leeuwen. Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Quart. J. Roy. Met. Soc.*, 2010.