A Statistical Method to Determine the Number of Lagrangian Elements for Optimal Gridded Field Representation of Random Displacement Model Outputs

Guillaume Marcotte*, Alain Malo, Éric Legault-Ouellet, Jean-Philippe Gauthier, and Gilles Mercier
Environment and Climate Change Canada, Meteorological Service of Canada, Environmental Emergency Response Section
Dorval, Québec, Canada
*guillaume.marcotte@canada.ca

Abstract

Gridded concentration fields are proposed as an improvement to the presentation of oil spill modelling results. Concentration calculations on fixed grids are common practice in other random displacement modelling domains and the added value of the results is no longer under debate. As the slick representation in gridded concentration fields is more sensitive to the number of Lagrangian elements used and to the grid mesh, a statistical sensitivity study of those parameters was performed. Considering the statistical analysis and other operational needs, a fast model runtime and a clean representation of the fields, it is proposed to use 10 000 Lagrangian elements with a 100 m grid mesh as a basic model configuration for elliptical spills covering 2.25 to 8.5 km^2.

1 Introduction

In oil spill modelling, several choices must be made by the modeller and by the development team to ensure a fast, accurate and credible modelled trajectory and oil fate. One of those choices is the selection of an appropriate number of Lagrangian elements (LEs) that will accurately represent the spill trajectory and the oil fate and behaviour. The modeller may specify as many LEs as he wishes, but this number is usually kept low since only a few hundred elements are needed to capture the oil fate and behaviour and a few thousand are judged sufficient to represent trajectories for a short forecast (6-24 h).

Keeping the number of LEs low ensures that the model runs quickly, thus suitable for operational responses. However, given the evolving computer power and the assumption that a very large number of elements might increases the accuracy of the representation, it is reasonable to assume that the more elements used, the more accurate the trajectories should become. Hence, there is no obvious reason to limit ourselves to rather low numbers of LEs. While this might not be true for particle representation of the trajectories in moderate currents, it becomes significant when the surface or three dimensional oil concentrations are outputted by the model.

Despite the obvious drawbacks of source uncertainties that limit the confidence in quantitative concentration fields, concentration contour plots are classical in the atmospheric Lagrangian modelling applications (D'Amours, 2015; Draxler and Hess, 1998; Stohl, 1998) because they show more information in a snapshot than scatter maps. It is hard to evaluate the LEs density when looking at a scatter; while a colour contour gives an easy interpretation of the element density. Furthermore, in a situation where the modelled source term is well defined, concentration fields add precious quantitative information that can be used for operational planning or toxicology purposes. Figure 1 shows an example of the added value of concentration fields.

Marcotte, G., A. Malo, E. Legault-Ouellet, J.-P. Gauthier, and G. Mercier, A Statistical Method to Determine the Number of Lagrangian Elements for Optimal Gridded Field Representation of Random Displacement Model Outputs, Proceedings of the Fortieth AMOP Technical Seminar, Environment and Climate Change Canada, Ottawa, ON, pp. 610-628, 2017.

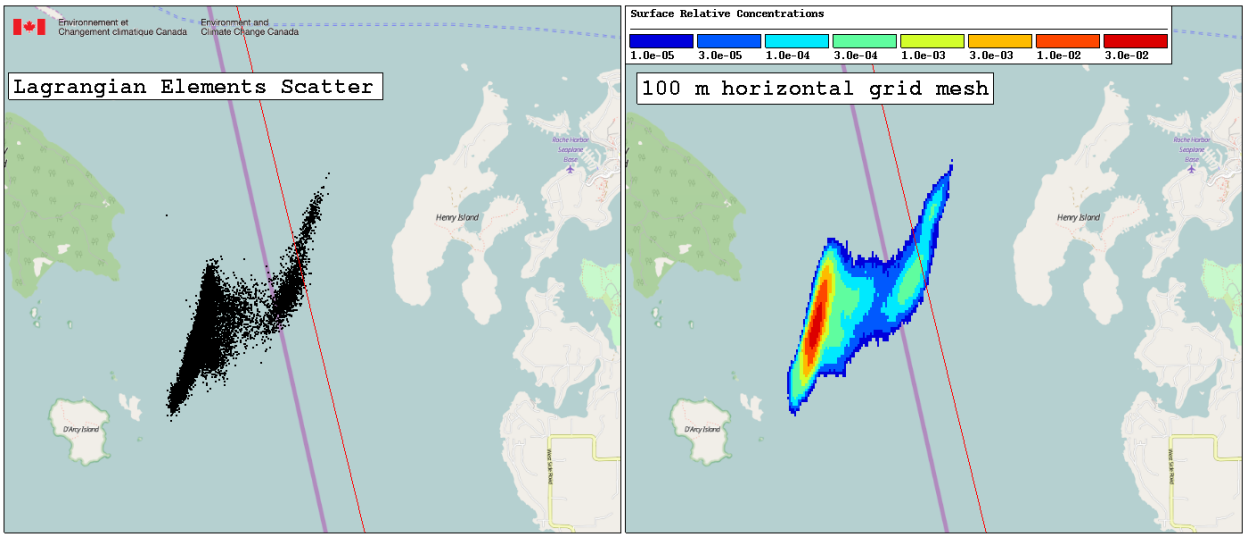

Figure 1 Comparison of the information contained in a snapshot of a hypothetical spill trajectory forecast. The LE density is easily captured in the gridded concentration field (right-hand side), but the location and extent of the maximum density of LEs is obscured by the overlapping of the LE at the displayed resolution (left-hand side).

Apart from producing 2D or 3D oil concentrations that might become a valuable indicator for field operations, gridded concentrations allow the use of statistical methods to validate the choice of number of LEs and grid mesh. In the context of operational modelling in support to field response, the evaluation of the model internal variability as a function of grid mesh and number of LEs allows the optimization of those parameters for a particular event scale. For that purpose, gridded concentrations allow a direct comparison between two numerical simulations. When projected in grid cells, LEs positions can be compared on the basis of the larger grid cell area and the LEs population of each grid cell. On the opposite, a direct element-based statistical comparison between two sets of LEs from different simulations is pointless due to the inherent stochastic diffusion of random displacement models.

It is important to emphasize that both parameters, number of LEs and grid mesh, are to be optimized together. Large-mesh grids require fewer LEs to keep a high statistical confidence in the plotted output reproducibility. However, it has to be remembered that the scale of the problem is equally important since oil spills in coastal or tortuous channel regions would better benefit from 50 m grid mesh than 1 km grid mesh. The latter would result in a degraded representation of the spill trajectory with respect to the elements scatter because of the averaging over large grid cells.

In this paper, we describe a statistical method that allows one to rationalize the choice of the number of LEs and grid mesh. In short, the Canadian Oil Spill Modelling Suite, COSMoS, (Marcotte et al., 2015) will be run with two different random number seeds in five simulation domains to generate surface concentrations using a combination of five different grid meshes with five different numbers of LEs at six different locations in each domain. This strategy produces a statistically valid sample for the evaluation of the model internal reproducibility regarding the variability introduced by the random number generation. The modelling and statistical methods are presented in the methodology section. A comprehensive selection of the statistical results is then shown and the discussion on the chosen number of LEs and grid mesh follows.

2 Methodology

The methodology presented in this paper was specifically applied to oil spill modelling. However, a similar validation could be done for any random displacement model calculations, e.g. dissolved chemicals or floating debris. No attention is given to oil fate and behaviour.

There are three criteria considered in this paper for the selection of the optimal number of LEs and grid mesh: the statistical validity of the results, the model runtime, and the spatial scale of the problem. The statistical validity of a simulation is tested using the analysis of chosen statistical indicators time series calculated with gridded concentration fields. These calculations require a pair of similar trajectories. A pair of trajectories contains a typical simulation and a replicate where the only difference lies in the random number generating function seed. In COSMoS, the random number seed is calculated using the machine time. A fixed, past machine time is used to generate one series of random number (the replicate) and the other random number series uses the present machine time (the typical simulation). Those two model replicates are run for every number of LEs and grid mesh pair.

The second criterion is the total runtime of the model. Considering the operational context where the oil spill models are used, the faster the model is, the better. The objective is to find a compromise where the simulation results will have the maximum quality in the minimum runtime acceptable.

The last criterion, the spatial scale of the problem, is equally important. It should be kept in mind that some grid mesh will not give results of higher quality than a LE cloud. A grid mesh appropriate for the scale of the spill and response should be chosen. This criterion forces the selection of optimized pairs, i.e. a number of LEs with a grid mesh. One metric to identify this scale would be the slick area at a chosen time. In the present experiments, the statistics are compared after 24 h. At this time, the oil slick area varies between 2.25 to 8.5 km^2 and is rather elliptic in shape. Thus, the recommended optimal parameters presented in this paper are representative of this range of slick size and shape only. Other situations where the slick shape is irregular (e.g. Figure 1) or where the slick area differs from the present range require another statistical study of the model reproducibility. The combination of those three concerns, statistical validity, model runtime and scale of the problem, allow a rational selection of the number of LEs to use with each relevant grid mesh.

The present experiment is using 100, 1000, 10 000, 100 000, and 500 000 LEs in combination with 25, 50, 100, 500, and 1000 m grid meshes for 5 different domains in Canadian coastal waters (Figure 2; 2 x Gulf of Saint Lawrence; near Anticosti Island and between Newfoundland and Labrador, Bay of Fundy, Bay of Ungava, and Beaufort Sea). There is no domain located on the Canadian west coast because no modelled currents were available for this region at the Canadian Center for Meteorological and Environmental Prediction when this experiment was conducted. In each simulation domain, 6 spill locations are randomly chosen, giving a final set of 1500 simulations or 750 model run replicates.

The currents and winds used in the experiment are taken from the Regional Ice-Ocean Prediction System, RIOPS (Lemieux et al., 2016). These fields are space and time varying and the model typical grid mesh is around 7 km. RIOPS current and wind fields are valid at the same date for each spill location in a given geographical domain. The following parameters are fixed for every simulation: the simulation duration (36 h), the total mass spilled (30 m^3), the product spilled (Diesel fuel oil), and the spill duration (100 s). Special care was taken to use the same COSMoS executable. The overall simulation duration may happen to be shorter than the

prescribed 36 h because the LEs sometimes evade the simulation grid before the end of the simulation. This explains the rapid deterioration or the spiking of certain statistical indicators presented in the next section. To mitigate the influence of LE losses on the statistics, the indicators are compiled 24 h after the beginning of the simulation.

As COSMoS allows LE diving under the effect of entrainment, the concentrations are integrated from 100 m under the water surface up to the surface. This depth is far below the surface mixed layer where the LEs are usually dispersed. This high integration volume prevents the loss of LE in the vertical when the concentrations are calculated. In this experiment, the quantitative value of the surface concentrations is insignificant since the statistics are drawn from the variation in the concentrations between simulation duplicates with different random number seeds. The absolute values of the concentrations are therefore not reported in this paper. The calculation of the concentrations is done using the same methodology than in atmospheric dispersion modelling (D'Amours, 2015; Stohl, 1998).

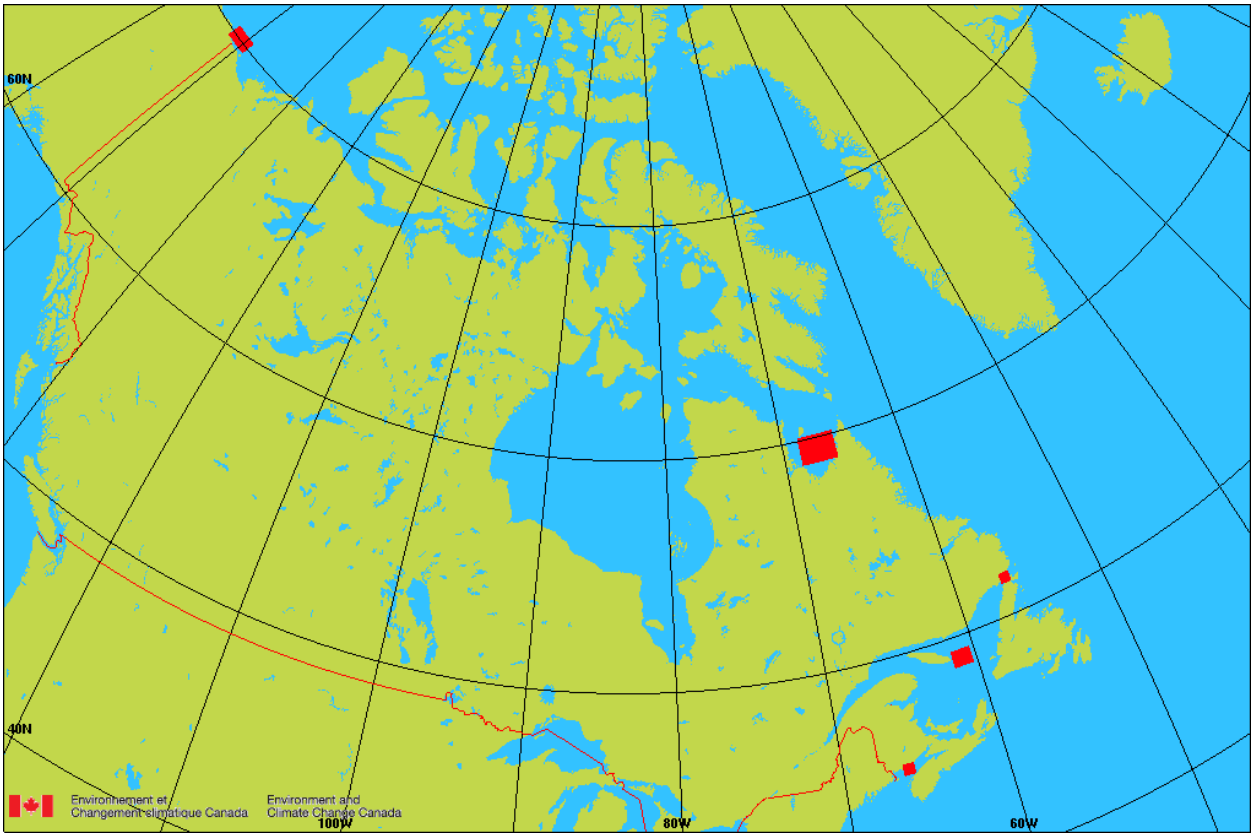

Figure 2 Simulation domains.

The statistical indicators used in this experiment are calculated using an in-house application called TheJudge. This validation tool allows the comparison of concentration fields and produces time series graphs for the selected statistical indicators. Figure 3 shows a simple flow chart of TheJudge. In brief, the users choose the statistical indicators that they want to be plotted, give TheJudge instructions on the plot format and the path to the files containing the concentration fields to compare. Once the results are generated, it is possible to combine similar simulations and average the indicators in order to enlarge the statistical sample and to remove features that are site or time dependent (e.g. tidal cycles). In this experiment, the statistics were

generated for every 750 duplicate model runs. Then, the indicators were averaged for each grid mesh or number of LEs. Each plotted curve is thus an average over 30 duplicate model runs.

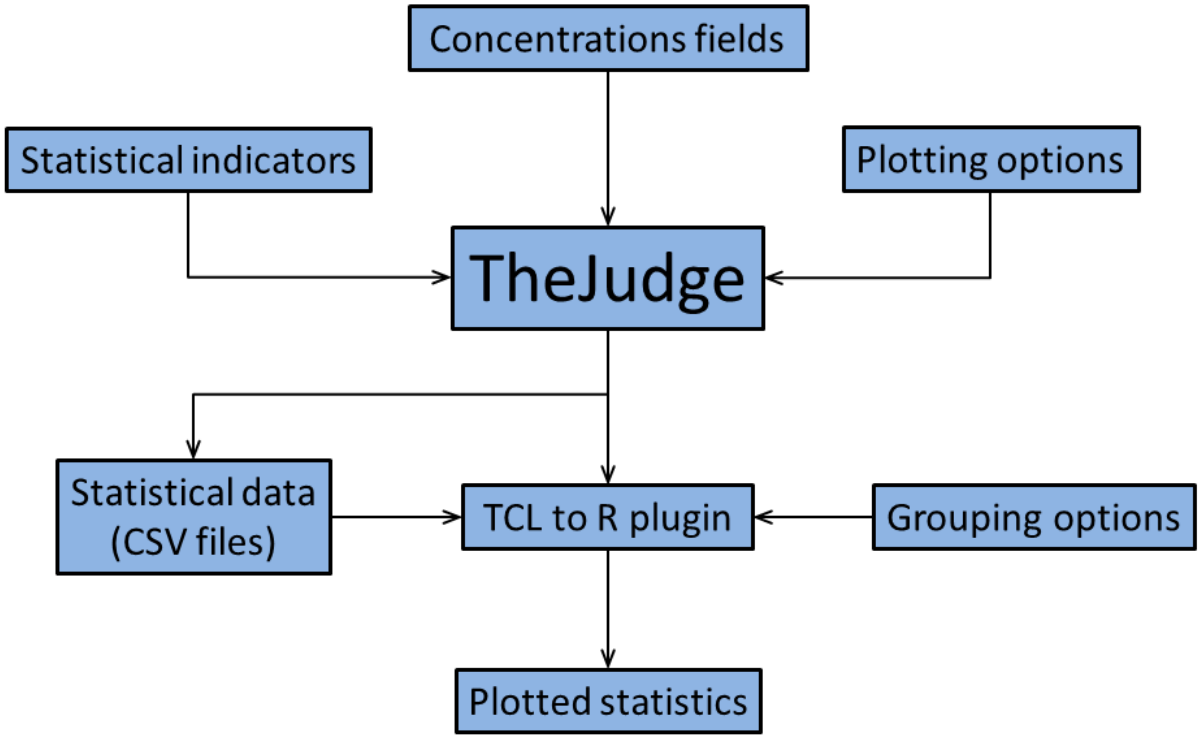

Figure 3 The Judge flowchart.

The indicators selected for analysis are Pearson's correlation coefficient, factor of two (Mosca, 1998), factor of exceedance (Mosca, 1998), fractional bias (Quélo, 2007), figure of merit in space (Quélo, 2007), and figure of merit in space (integrated). Pearson's correlation coefficient is a measure of the evolution of the gridded fields in the same direction. The optimal value is 1, where the two fields are moving at the same speed in all directions. However, this indicator alone does not inform on the quantitative likeness of the slicks. Combined with the factor of 2, the percentage of time the absolute gridded field values ratio is included between 0.5 and 2, gives a more complete picture of the situation. A high factor of two and correlation coefficient indicates that the clouds are very similar.

The factor of exceedance is a measure of the spread of the data against the perfect correlation line between a measure and the model. The perfect score with this indicator is 0. A negative value is interpreted as an undershoot of the model and a positive score as an overshoot. The fractional bias is a normalized version of the bias. Its values fall in the range between -2 and 2 and the perfect score is 0. In the present context, the factor of exceedance and the fractional bias are used as controls of the statistical validity of the random number generating function since no measurement is used in this experiment.

The figure of merit in space is a grid cell area normalized measure of the spatial overlap of the slicks. The perfect value is 100% and represents two slicks perfectly overlaid. The figure of merit in space (integrated) is a different representation of the overlap of the modelled oil slicks as it puts more emphasis on the highly concentrated areas. It is calculated using the sum of the overlapped minimum concentration divided by the sum of the concentrations. It can be

interpreted as the figure of merit in space weighted by the fraction of the gridded value contained in both concentration fields over the total of both gridded values. It gives an indication of the relative positions of the maxima in both slicks. For example, a high score in the figure of merit in space indicates that the shape of the slicks is similar, whereas a high score in the figure of merit in space (integrated) indicates that the shape and position of the slick maximum concentration is similar.

The oil spill kernel of COSMoS is parallelised using the Message Passing Interface (MPI) and OpenMP (OMP) standards. Considering the high number of simulations to execute, they were run using only 1 MPI and 4 OMP threads to allow a reasonable usage of the available computational power. One simulation representing each number of LEs and grid mesh pair was rerun in the Ungava domain using 4 MPI and 8 OMP threads. A second timing experiment was done using smaller grid sizes (i.e., smaller domain) for the input fields. In the first one, the environmental fields were read directly from the native grid of the Regional Ice-Ocean Prediction System (2 870 860 grid points over the Arctic and North Atlantic Oceans). In the second, those results were first interpolated onto the output gridded domain and the simulations were run using those inputs. The number of grid cells in the latter timing experiment is therefore grid mesh dependent. The results are compared and discussed in the following sections.

3 Results

This section is organised as follows: the statistical results are first described and shown, and then the timing experiment results are presented. For the statistical results, only a representative sample of each indicator is given and only the plots showing the same number of LEs with varying grid meshes are shown. The inverse graphs were also produced (i.e. same grid mesh, varying number of LEs) but they are more difficult to interpret due to the large change in the vertical scale. This change of scale masks details like the difference between 500 000 and 100 000 LEs at the top of the plots. For conciseness reasons, only the cases with 100, 1000, and 10 000 LEs are shown. However, results at 24 h for all numbers of LEs are compiled in Tables 1 to 4. The legends are uniformly formatted into “Nb LEs”_ “grid mesh (m)”. Throughout this section, the reader is advised to pay attention to the variation in the vertical axis of the graphs of the same indicators for different number of LEs. Two of the chosen indicators, the fractional bias and factor of exceedance, show a similar spread for all numbers of LEs. Thus, only the 100 LEs case is shown (Figures 4 and 5).

Figures 6 to 8 present the effect of grid mesh on Pearson’s correlation coefficient time evolution for 100 (Figure 6), 1000 (Figure 8), and 10 000 (Figure 8) LEs. In general, Pearson’s correlation coefficient increases with larger grid cells and with an increasing number of LEs. However, it is seen in Figures 6 to 8 that this statistical indicator does not allow the distinction between the two largest grids mesh (500 and 1000 m). The score stays in the upper limit, i.e. above 0.99. The usefulness of this score in the establishment of a clear statistical distinction of the different grid meshing is debatable as it does not show any significant difference before 24 h. The results of Pearson’s correlation coefficients time evolution are compiled after 24 h in Table 1.

Figures 9 to 11 show the impact of grid mesh on the figure of merit in space time evolution for 100 (Figure 9), 1000 (Figure 10), and 10 000 (Figure 11) LEs. In comparison with Pearson’s correlation coefficient (Figures 6 to 8), it is seen that this indicator is more sensitive to the choice of grid mesh and number of LEs. As expected, the difference in the slick shape traduced by the figure of merit in space scores deteriorates rapidly when the grid mesh gets finer.

The scores are also following the expected trend as they get considerably higher with a larger number of LEs. A comparison of the scores after 24 h is given in Table 2. Another interesting observation in Figures 9 to 11 is that the time series of the larger grid cells are noisier. This is explained by the fact that the figure of merit in space is a function of the total area of the slick. A lonely LE on the edge of a grid cell, constantly shifting from a cell to another, will have a great influence on the total gridded area of the slick computed using large grid cells. This difference is lowered when the slick spreads out or when the grid cell area lowers.

Figures 12 to 14 present the effect of grid mesh on the time evolution of the figure of merit in space (integrated) for 100 (Figure 12), 1000 (Figure 13), and 10 000 (Figure 14) LEs. The overall behaviour of this indicator is similar to the figure of merit in space (Figures 9 to 11) but the scores for 10 000 or more LEs are better; while the scores for 100 and 1000 LEs are degraded. This difference is best seen by comparing the results at a chosen grid mesh in Table 2, figure of merit in space, against Table 3, figure of merit in space (integrated).

Figures 15 to 17 shows the influence of grid mesh over the factor of two time evolution for 100 (Figure 15), 1000 (Figure 16), and 10 000 (Figure 17) LEs. Again, the score increases with larger grid cells and larger number of LEs. Results after 24 h are compiled in Table 4.

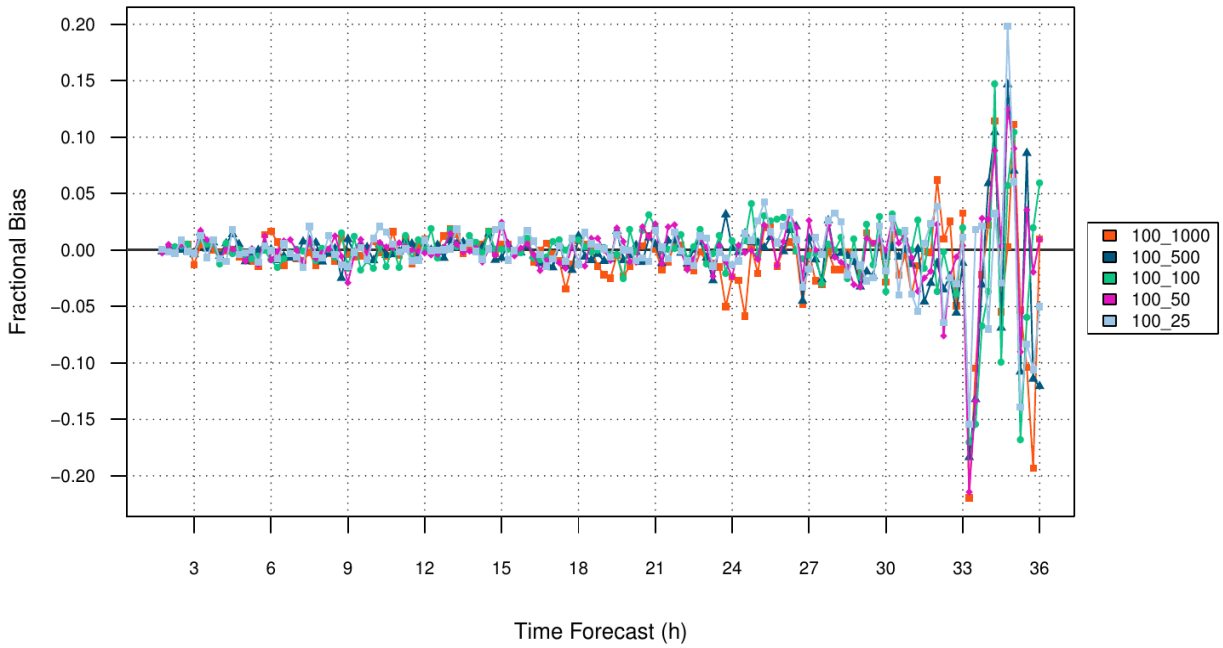

Figure 4 Fractional bias for 100 LEs with varying grid mesh.

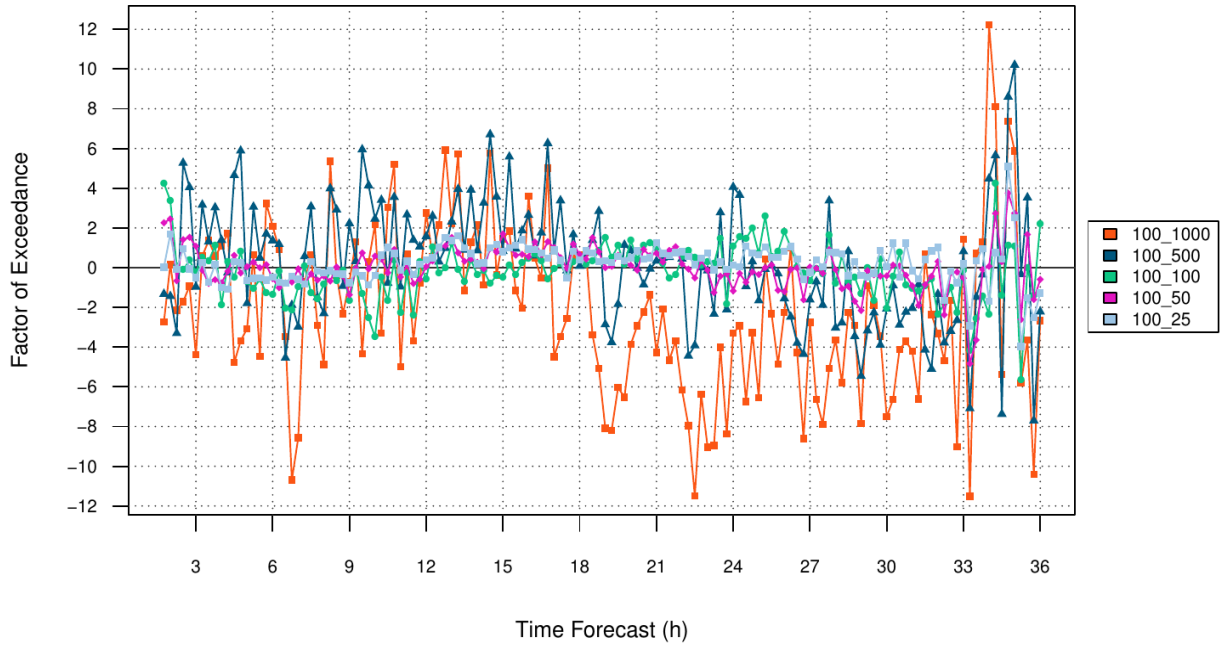

Figure 5 **Factor of exceedance for 100 LEs with varying grid mesh.**

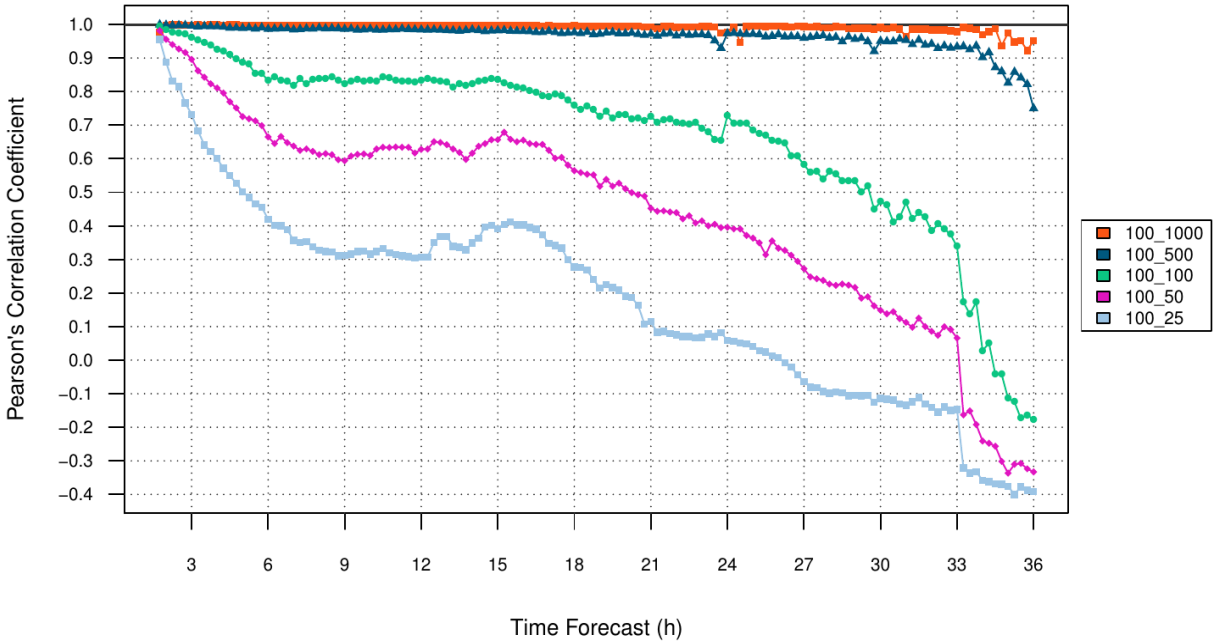

Figure 6 **Pearson's correlation coefficient for 100 LEs with varying grid mesh.**

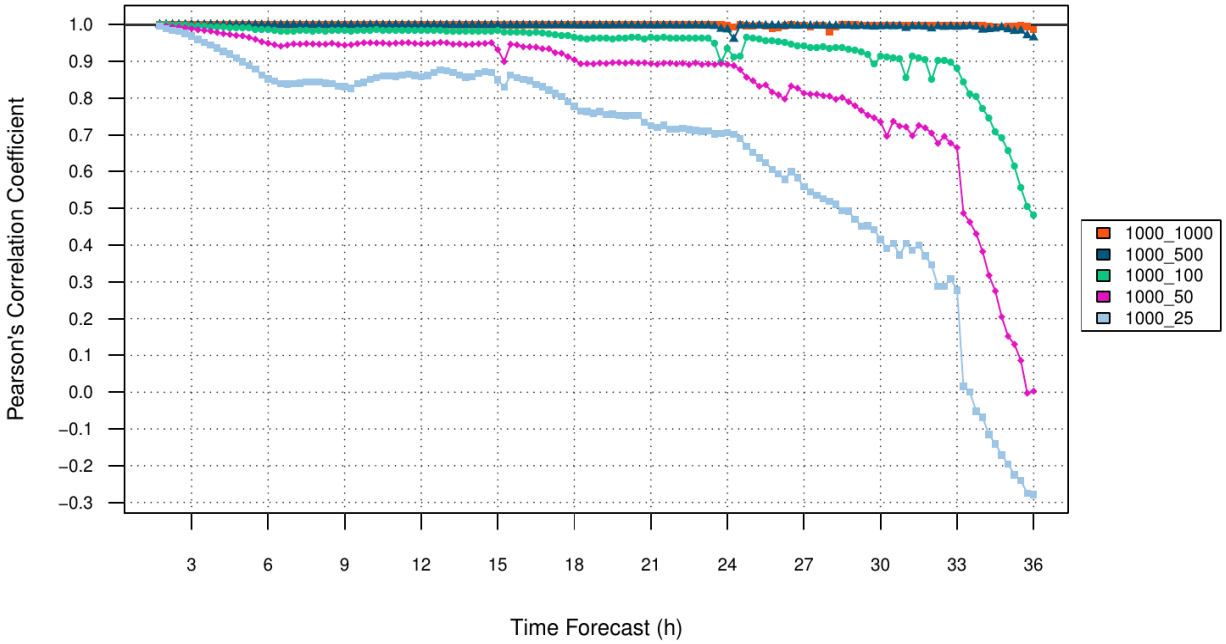

Figure 7 Pearson's correlation coefficient for 1000 LEs with varying grid mesh.

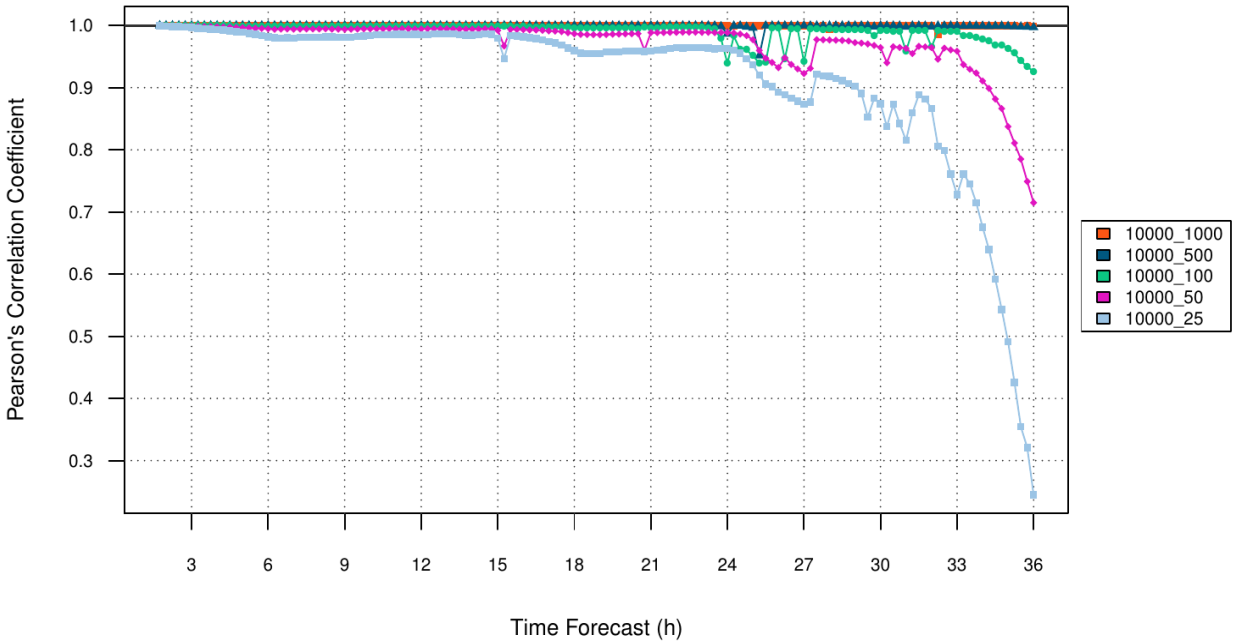

Figure 8 Pearson's correlation coefficient for 10 000 LEs with varying grid mesh.

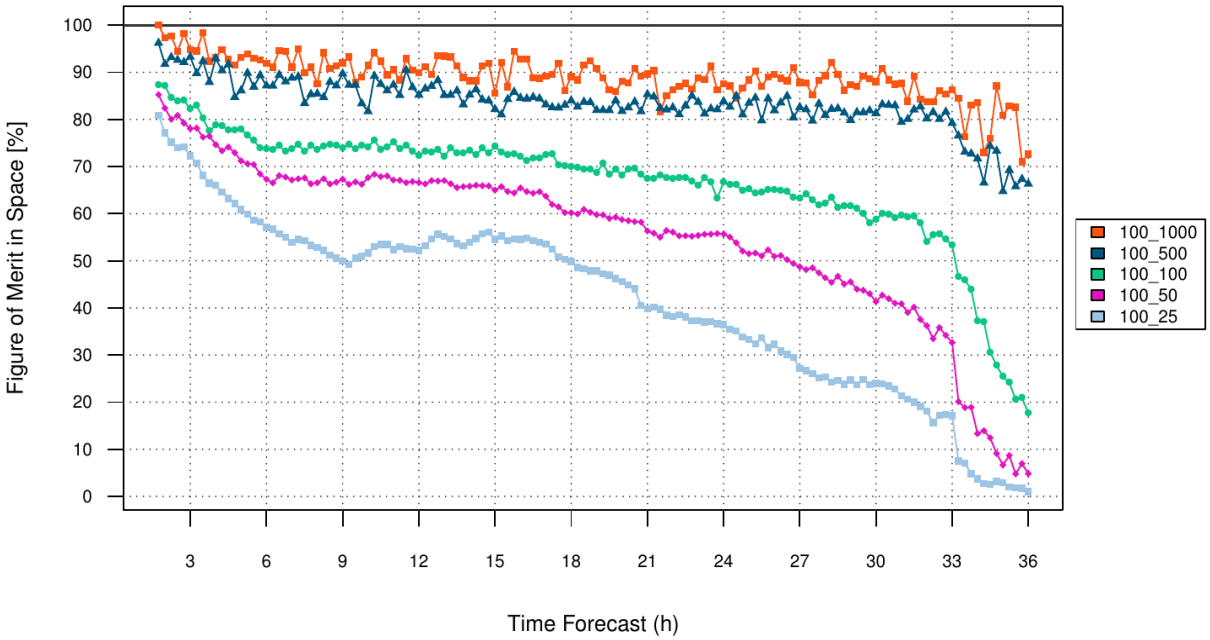

Figure 9 **Figure of merit in space for 100 LEs with varying grid mesh.**

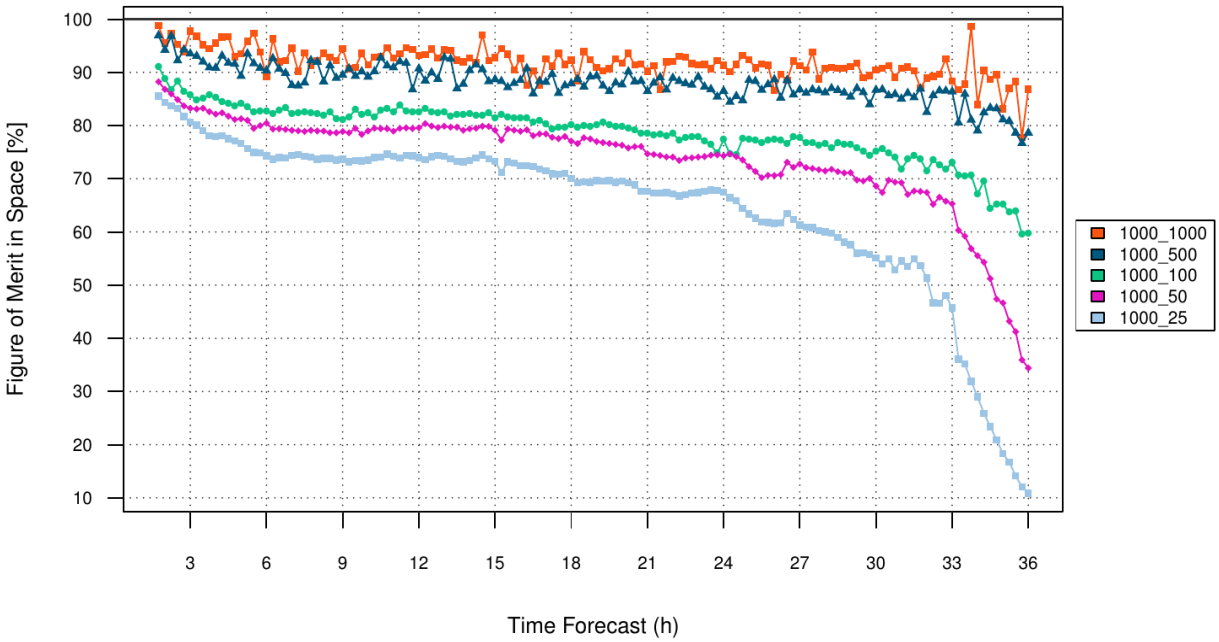

Figure 10 **Figure of merit in space for 1000 LEs with varying grid mesh.**

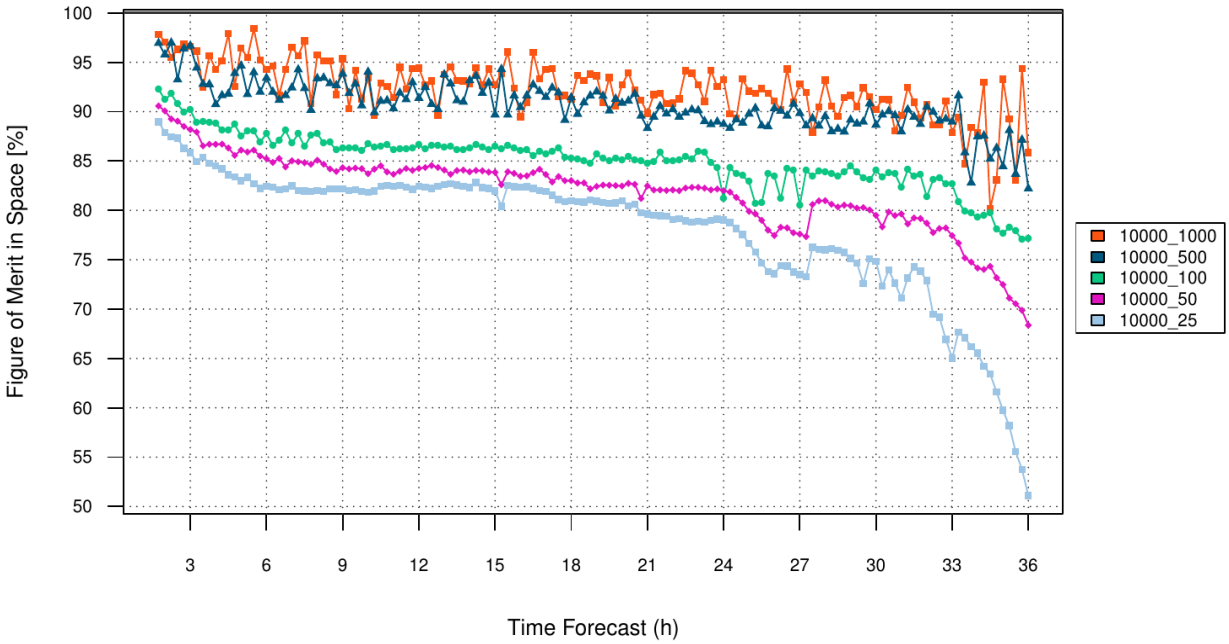

Figure 11 Figure of merit in space for 10 000 LEs with varying grid mesh.

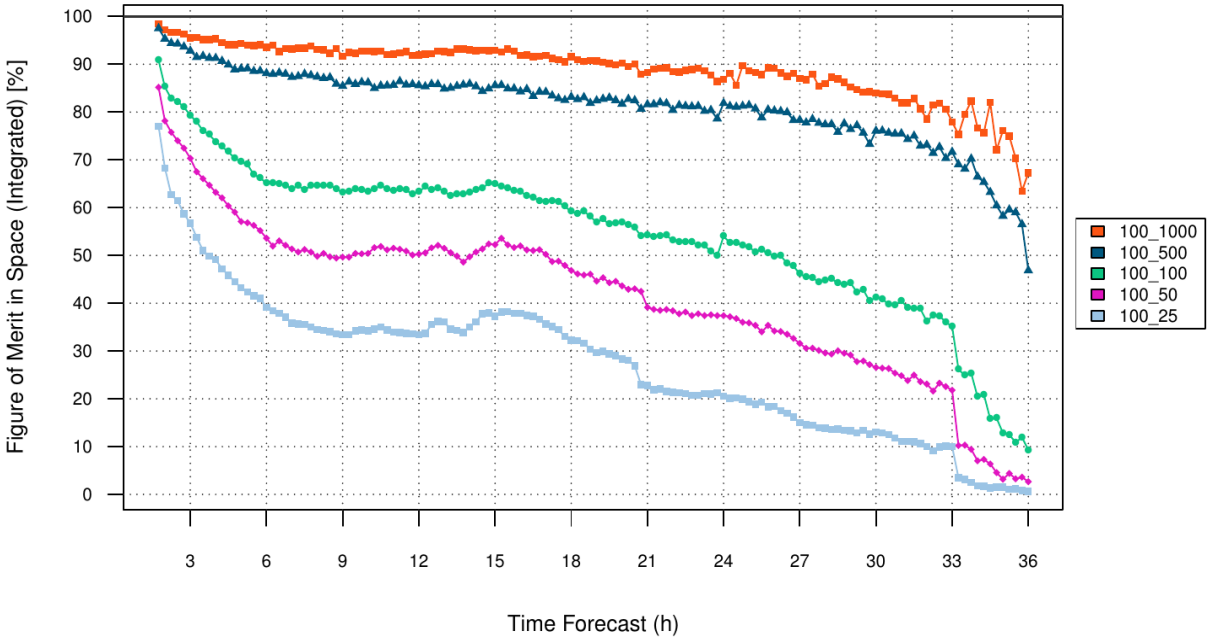

Figure 12 Figure of merit in space (integrated) for 100 LEs with varying grid mesh.

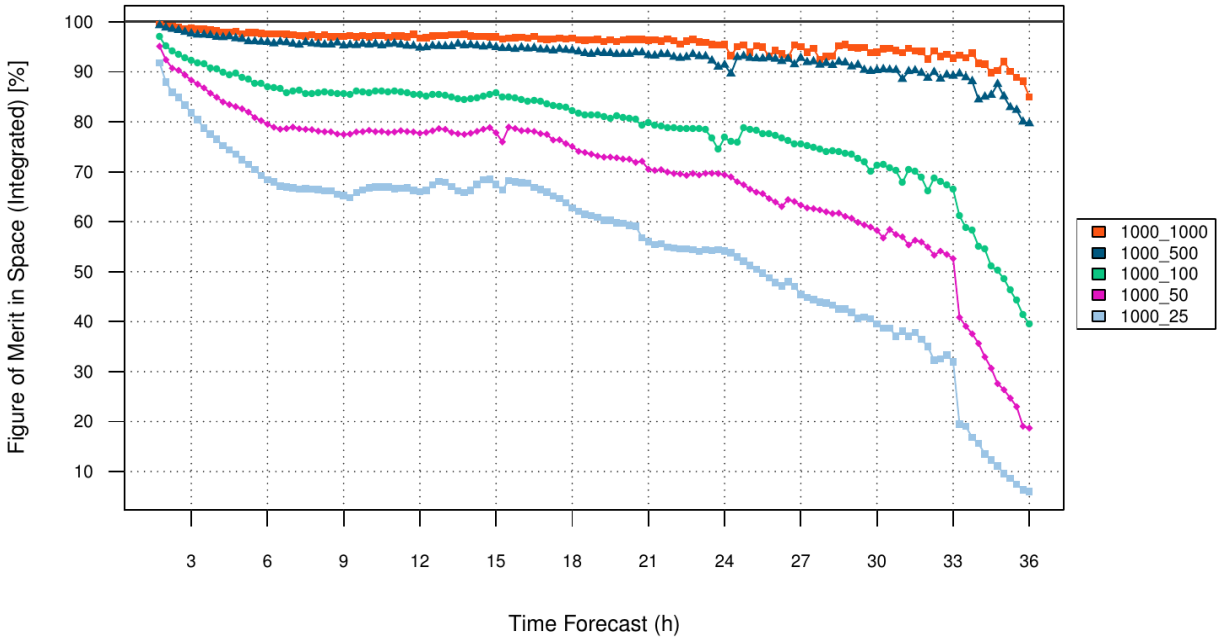

Figure 13 Figure of merit in space (integrated) for 1000 LEs with varying grid mesh.

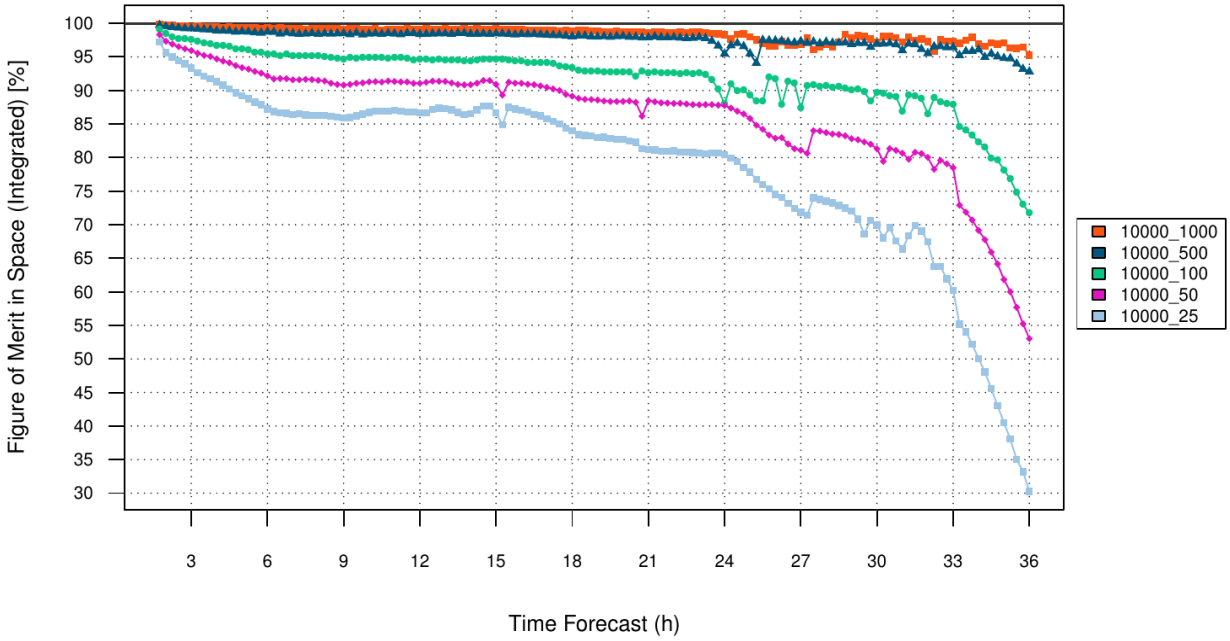

Figure 14 Figure of merit in space (integrated) for 10 000 LEs with varying grid mesh.

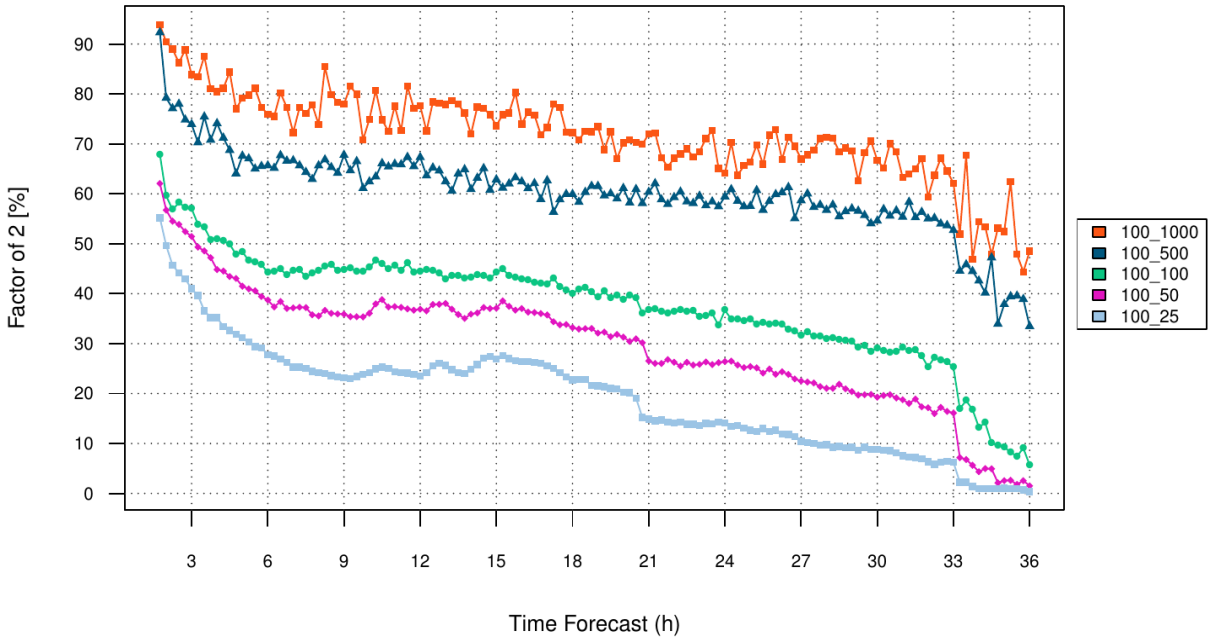

Figure 15 Factor of 2 for 100 LEs with varying grid mesh.

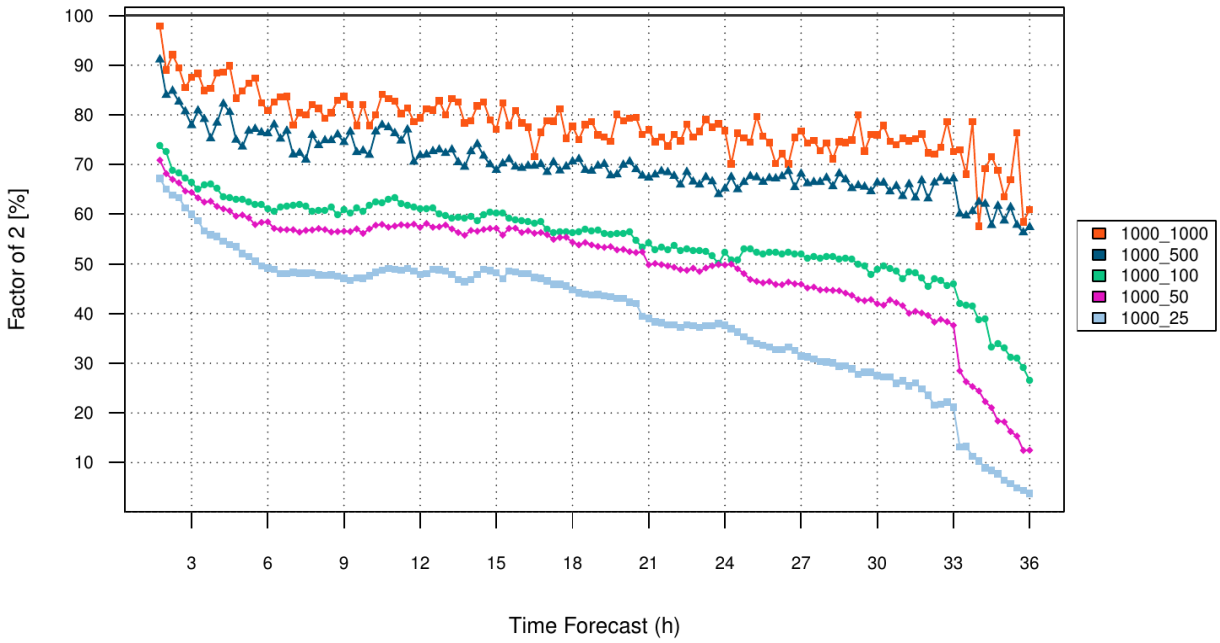

Figure 16 Factor of 2 for 1000 LEs with varying grid mesh.

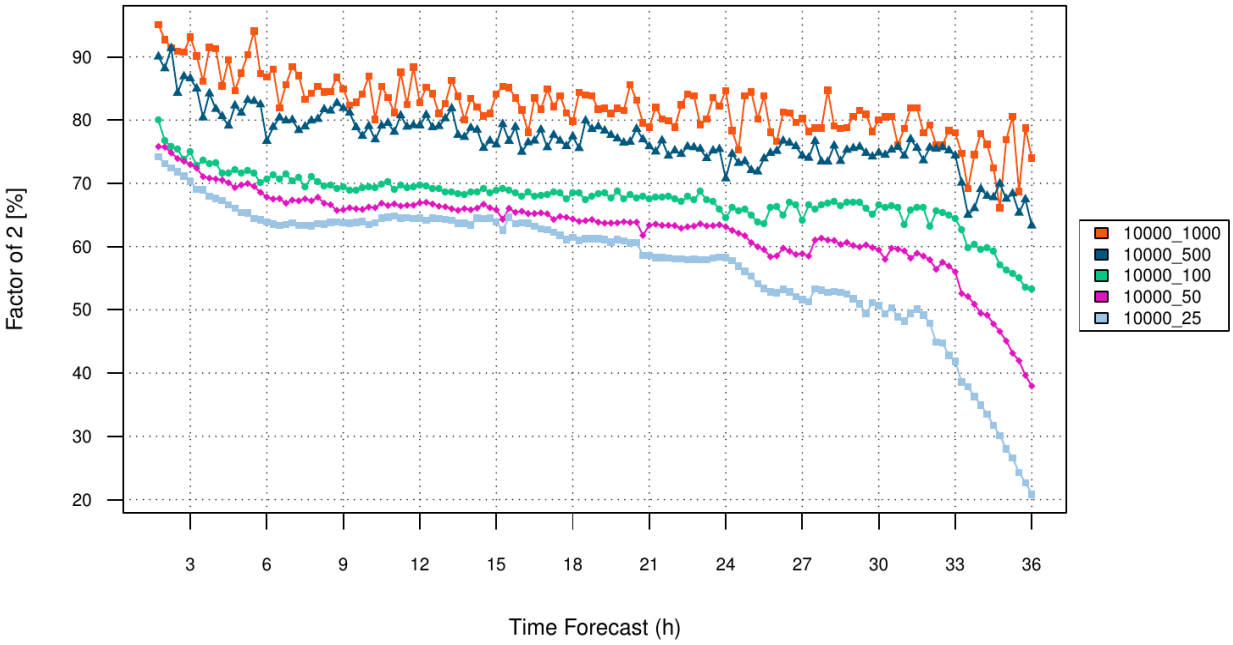

Figure 17 **Factor of 2 for 10 000 LEs with varying grid mesh.**

Table 1 **Pearson's correlation coefficient after 24 h.**

	Grid spacing (m)				
Nb. LEs	**25**	**50**	**100**	**500**	**1000**
100	0.06	0.4	0.7	0.97	0.99
1000	0.7	0.9	0.94	0.99	>0.99
10 000	0.97	0.99	>0.99	>0.99	>0.99
100 000	>0.99	>0.99	>0.99	>0.99	>0.99
500 000	>0.99	>0.99	>0.99	>0.99	>0.99

Table 2 **Figure of merit in space (%) after 24 h.**

	Grid spacing (m)				
Nb. LEs	**25**	**50**	**100**	**500**	**1000**
100	36	55	66	83	87
1000	68	75	77	86	92
10 000	79	83	84	89	93
100 000	84	87	87	91	93
500 000	87	89	90	93	95

Marcotte, G., A. Malo, E. Legault-Ouellet, J.-P. Gauthier, and G. Mercier, A Statistical Method to Determine the Number of Lagrangian Elements for Optimal Gridded Field Representation of Random Displacement Model Outputs, Proceedings of the Fortieth AMOP Technical Seminar, Environment and Climate Change Canada, Ottawa, ON, pp. 610-628, 2017.

Table 3 Figure of merit in space (integrated) (%) after 24 h.

	Grid spacing (m)				
Nb. LEs	**25**	**50**	**100**	**500**	**1000**
100	20	38	52	81	87
1000	53	70	76	92	95
10 000	80	87	92	96	98
100 000	93	96	97	98	99
500 000	96	98	99	>99	>99

Table 4 Factor of two (%) after 24 h.

	Grid spacing (m)				
Nb. LEs	**25**	**50**	**100**	**500**	**1000**
100	14	26	31	60	67
1000	38	50	51	65	77
10 000	59	64	66	72	81
100 000	69	72	73	79	83
500 000	74	77	77	81	86

Table 5 Model runtime (h) using the native grid resolution (1MPI x 4OMP, 2.3 GHz).

	Grid spacing (m)				
Nb. LEs	**25**	**50**	**100**	**500**	**1000**
100	1.13	1.18	1.04	1.20	1.04
1000	1.23	1.19	1.20	1.16	1.22
10 000	1.42	1.13	1.38	1.36	1.20
100 000	3.84	3.64	3.71	3.56	3.95
500 000	11.06	15.01	11.27	16.61	11.15

Table 6 Model runtime (h) using input fields projected on the modelling grid (4MPI x 8OMP, 2.3 GHz).

	Grid spacing (m)			
Nb. LEs	**25**	**100**	**500**	**1000**
100	0.67	1.07	0.06	0.02
1000	0.91	1.18	0.09	0.02
10 000	1.03	1.43	0.12	0.06
100 000	1.45	1.47	0.44	0.41
500 000	2.26	2.82	1.94	2.06

Marcotte, G., A. Malo, E. Legault-Ouellet, J.-P. Gauthier, and G. Mercier, A Statistical Method to Determine the Number of Lagrangian Elements for Optimal Gridded Field Representation of Random Displacement Model Outputs, Proceedings of the Fortieth AMOP Technical Seminar, Environment and Climate Change Canada, Ottawa, ON, pp. 610-628, 2017.

4 Discussion

Fractional bias (Figure 4) and factor of exceedance (Figure 5) were not presented for each grid mesh and number of LEs scenario. It was found the spread is rather similar for all combinations of grid mesh and number of LEs. Therefore, only the lower limit cases with 100 LEs are shown in Figures 4 and 5. It is seen in those figures that the spread over the optimal value for both indicator is small. In the context of this experiment, those indicators confirm that the random number generation function used in the model has the expected behaviour.

As shown in Figures 6 to 17 and summarized in Tables 1 to 4, the indicators' scores are following the expected trend: the statistical scores improve as the grid mesh or the number of LEs increases. This is explained by a better sampling of LEs within a given grid cell as the grid resolution gets coarser or because the number of LEs is greater in larger grid cells. The first criterion for the exclusion of a grid mesh and number of LEs combination is Pearson's correlation coefficient (Figures 6 to 8, Table 1). The variation induced by the change of the random generator series should, in principle, have no impact on the slick trajectory or shape. Thus, a degradation of Pearson's correlation coefficient is interpreted as an inappropriate reproducibility of the changes in the slick over time. Consequently, 100 and 1000 LEs should not be used with grid mesh under 500 m because the modelled results are not self-consistent. However, this issue vanishes for 10 000 or more LEs where a correlation coefficient is kept higher than 0.9 for all configurations.

As stated before, Pearson's correlation coefficient should not be the only indicator used as it does not give information on the preservation of the quantities and is less sensitive to the spatial extent of the slick than the figure of merit in space (Figures 9 to 11, Table 2). The drop in figure or merit scores with the decrease in grid mesh is spectacular. This is explained by the increased sensitivity to the random component of the movement when the grid cells are smaller. Therefore, the analysis of the figure of merit in space confirms that smaller grid mesh requires higher numbers of LEs. Based on those results, it is suggested to avoid using less than 10 000 LEs in combination with grid mesh less than 500 m.

The information contained in the figure of merit in space (integrated, Figures 12 to 14, Table 3) is a useful complement to the figure of merit in space (Figures 9 to 11). A figure of merit in space (integrated) greater than the corresponding figure of merit in space indicates that the maximum of the gridded concentration fields is better reproduced than the overall extent of the slick. This appears when the edges of the slick are ill defined (noisiness induced by a reduced number of LEs at the edges) but the maxima of the two duplicate simulations are spatially overlaying at a given model time step. A lower figure of merit in space (integrated) with respect to the figure of merit in space indicates that the location and extent of the slick maximum concentration is harder to capture than the overall shape of the slick (ex. 100 LEs, grid mesh 50 m, Tables 2 and 3). When comparing the values for all grid meshes with a specific number of LEs in Table 2 with the matching values in Table 3, it is seen that the maximum in the slick concentration is better reproduced in the duplicate simulation for 10 000 or more LEs.

Finally, the factor of two (Figures 15 to 17, Table 4) is used as an indication of the reproducibility of quantitative fields. This indicator is very sensitive to grid cell area because of the spatial averaging. Thus, when the area diminishes, this score rapidly decreases. Considering the usual uncertainty about the source terms during operational response, this indicator is not as critical as it may seem. However, in a situation where the source term is well defined, it becomes interesting to have this score as high as possible. Considering a case with a defined source, the number of LEs used should be 100 000 or higher in order to keep the modelled concentration

fields in a factor of two after 24 h of difference more than 60 % of time, unless a very large grid mesh is used (e.g. 1 km, Table 4). The apparent noisiness of the score for larger grid cell areas is explained by the lower number of grid cell for numerical integration, as for the figure of merit in space.

Those statistical conclusions point towards the use of the largest number of LEs possible. Using the values reported in Tables 1 to 4 for 100 m grid mesh, a relative improvement of about 10% to 30% in the scores is found when the number of LEs increases from 1000 to 10 000. However, the limited time allowed to produce model output during a real-time response to a spill might not give the modeller the leisure to use several hundred thousand of LEs.

Tables 5 and 6 show the model runtime for oil spill modelling. It is important to point out that the calculations in Table 5 were not done using an operational parallel computing configuration due to the limited availability of the computational resources and the large volume of runs planned. Thus, those times should be seen as an indication of the trends in model runtime. Interestingly, the runtime is independent of grid mesh and of the number of LEs below 10 000 LEs. This demonstrates that the version of COSMoS used in this experiment is limited by the computational cost of data reading and writing. Nevertheless, it is interesting to note that the total time increases by a factor slightly larger than three between 100 and 100 000 LEs.

The runtimes reported in Table 6 were established using more computational power (4 MPI jobs with 8 OMP threads) per simulation. However, except for the small numbers of LEs combined with a large grid mesh, the gain in model runtime is somewhat marginal. This confirms that COSMoS is limited by its input/output (I/O) components. When the native concentration fields are interpolated onto the output grid, the total number of grid points is sometimes larger than the number of native grid points. In this case, the model runtime is therefore biased by the large grid used in the experiment. This observation points out that the modelling grid definition is linked to the scale and location of the spill. In areas of high wind speed or current speed, the grid needs to be larger to avoid losing LEs during the simulation. In those cases, it is impractical to retain a fine grid mesh since the number of grid cells will not allow the model to run in a reasonable time (5-15 min).

The finest grid cell spacing is thus advised to be used only in areas of complex currents structures where the oceanic or hydrodynamic model can resolve them. Otherwise, the grid mesh and size should be chosen to 1) avoid LE loss during simulation; 2) allow a smooth representation of the gridded fields in the model products; and 3) retain a statistically valid, quantitative representation of the computed concentration when needed. Usually, a grid mesh of 100 m is sufficient for a clean graphical representation of the fields (Figure 1).

5 Conclusions

In introduction, it was proposed to take advantage of the atmospheric Lagrangian dispersion modelling methods to apply a similar Lagrangian (particles) to Eulerian (gridded fields) representation of the model results. The concentration fields allow a more detailed interpretation of the model results than the usual scattered elements by adding information about the LEs density (contours) and by allowing the computation of oil surface concentrations or other quantitative information. Using well-established statistical methods (Draxler, 2006; Mosca, 1998; Quélo, 2007; Stohl, 1998), an experiment was designed to determine a combination of grid mesh and number of LEs offering a minimum model runtime with a maximum statistical quality of the results. The results presented in this paper apply in the limits of the scale, 2.25 to 8.5 km^2, and shape, elliptical, of the oil slick used in this experiment. A total of 1500 COSMoS runs were done to build a statistically valid sample of model outputs. In this experiment, no attention was given to fate and behaviour results.

A comprehensive statistical analysis of the gridded concentration fields was done to attempt to determine some optimal matches of number of LEs and grid mesh. It was found that the higher the number of LEs, the better the simulation statistical validity. When reducing the grid mesh, an increasing number of LEs is required to maintain the statistical confidence in the gridded fields.

Regarding model runtimes, it was seen that an extreme number of LEs is not appropriate in the context of operations. Finally, considering the scale of the problem (i.e., elliptical slicks covering 2.25 to 8.5 km^2), it was found that the number of LEs used should lie between 1000 and 10 000 while the grid mesh should be at most 100 m. When providing quantitative information, it is recommended to use at least 100 000 LEs in combination with a 100 m grid mesh to reproduce the values within a factor of two more than 60 % of the time.

Even though the proposed method is not perfect, it successfully allowed the determination of a set of modelling parameters, 10 000 LEs with a 100 m grid mesh, that give an accurate and reproducible representation of elliptical oil slicks with an area of 2.2 to 8.5 km^2. It is important to recall that this particular set of modelling parameters (100 grid mesh, 10 000 LEs) may not be optimal for slick total areas outside the explored range or for slick showing a complicated shape (e.g., Figure 1). A statistical validation should be done for problems with different scales, from several hundred square meters to a few hundred square kilometers, and in different contexts (e.g. river flow) to validate a wider range of grid mesh and number of LEs recommended values.

6 Acknowledgement

The authors wish to acknowledge the contribution of Mr. Pierre Bourgouin, a strong believer and supporter of the COSMoS project. The authors also thank Mr. Gregory Smith, Mr. Kristjan Onu, and Mr. Pierre Pellerin for giving access to the oceanic fields used in this study.

7 References

D'Amours, R., A . Malo, T. Flesch, J. Wilson, J.-P. Gauthier and R. Servanckx, "The Canadian Meteorological Center's Atmospheric Transport and Dispersion Modelling Suite", *Atmosphere-Ocean*, 53:1029-1070, 2015.

Draxler, R. R. “The Use of Global and Mesoscale Meteorological Model Data to Predict the Transport and Dispersion of Tracer Plumes over Washington, D.C.”, *Weather and Forecasting*, 21:383-394, 2006.

Draxler, R. R. and G. D. Hess, “An overview of the HYSPLIT_4 modelling system for trajectories, dispersion and deposition”, *Australian Meteorological Magazine*, 47:295-308, 1998.
Lemieux, J.-F., C. Beaudoin, F. Dupont, F. Roy, G. C. Smith, A. Shlyaeva, M. Buegner, A. Caya, J. Chen, T. Carrieres, L. Pogson, P. De Repentigny, A. Plante, P. Pestieau, P. Pellerin, H. Ritchie, G. Garric, and N. Ferry, “The Regional Ice Prediction System: verification of forecast sea ice concentration”, *Quarterly Journal of the Royal Meteorological Society*, 142:656-671, 2016.

Marcotte G., P. Bourgouin, G. Mercier, J.-P. Gauthier, P. Pellerin, G. Smith, K. Onu, and C. E. Brown, “Canadian Oil Spill Modelling Suite: An Overview”, in *Proceedings of the Thirty-ninth AMOP Technical Seminar on Environmental Contamination and Response*, Environment and Climate Change Canada, Ottawa, ON, 1:1026-1034, 2015.

Mosca, S., G. Graziani, W. Klug, R. Bellasio, and R. Bianconni, “A Statistical Methodology for the Evaluation of Long-range Dispersion Models: An Application to the ETEX Exercise”, *Atmospheric Environment*, 32:4307-4324, 1998.

Quélo, D., M. Krysta, M. Bocquet, O. Isnard, Y. Minier, and B. Sportisse, “Validation of the Polyphemus platform on the ETEX, Chernobyl and Algeciras cases”, *Atmospheric Environment*, 41:5300-5314, 2007.

Stohl, A., M. Hittenberg, and G. Wotawa, “Validation of the Lagrangian Particle Dispersion Model FLEXPART against Large-Scale Tracer Experiment Data”, *Atmospheric Environment*, 32:4245-4264, 1998.