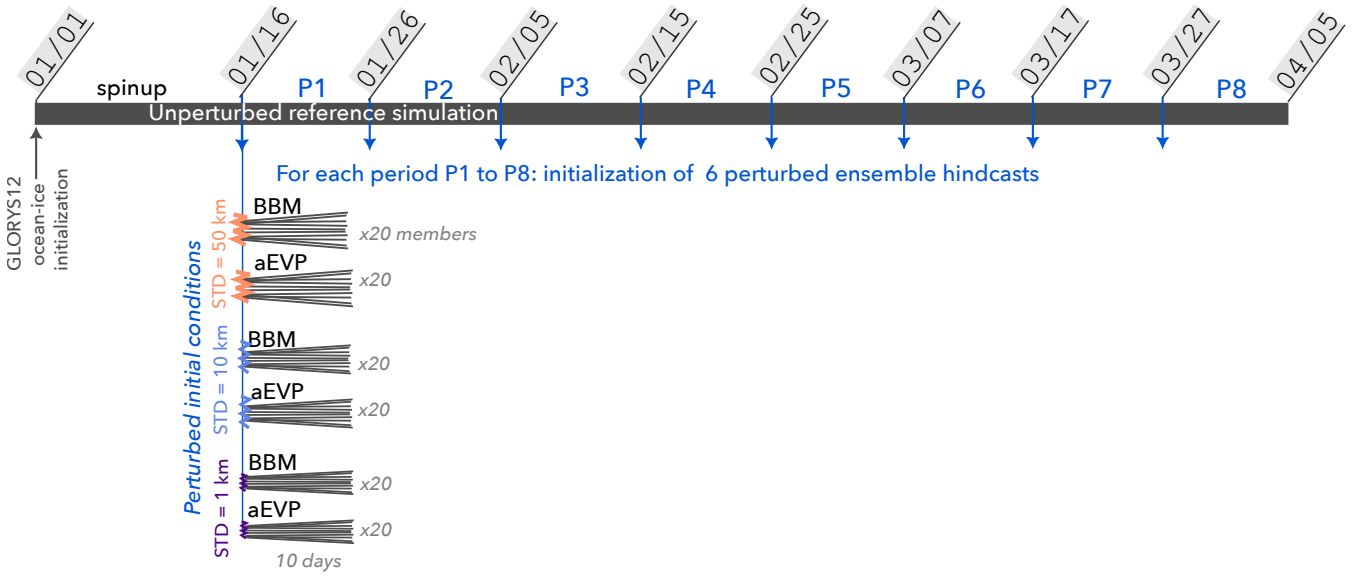

Figure 1. Summary of the experimental design, representing the unperturbed reference simulation from which are initialized, every 10 days, the 20-member ensemble hindcasts with perturbed ensemble conditions for both rheologies (BBM and aEVP) and three different magnitudes of the initial perturbation each.

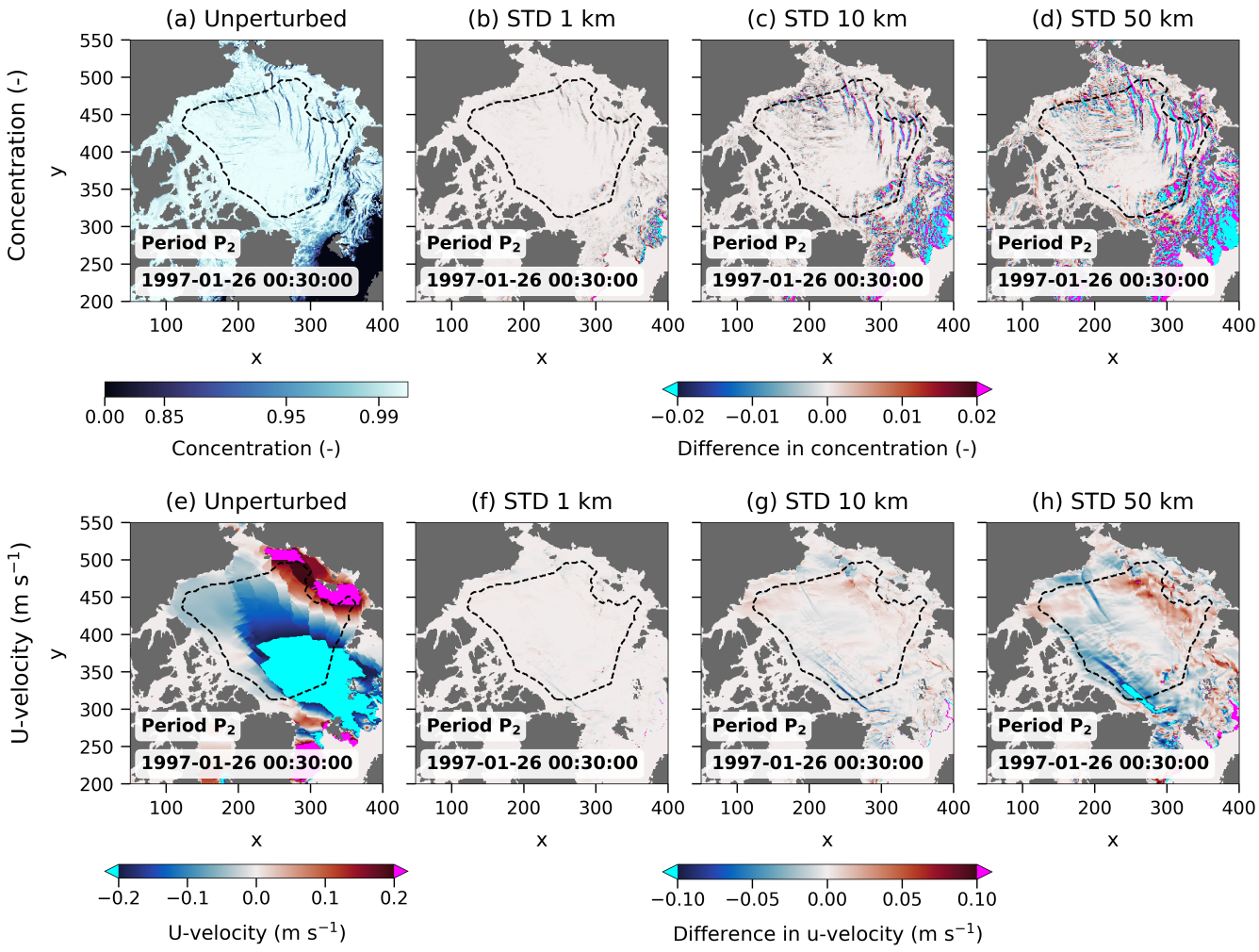

Figure 2. Unperturbed initial states in (a) sea ice concentration and (e) the U-component of sea ice velocity at the beginning of period P_2 (t_0: 1997-01-26) from the unperturbed reference experiment with the BBM model. Panels (b-d) and (f-h) show the difference in the initial states of two perturbed ensemble members for concentration and the U-component of sea ice velocity, respectively. Panels (b,f), (c,g), and (d,h) correspond to initial perturbations scaled for a standard deviation (STD) of 1 km, 10 km and 50 km, respectively. The black dashed line defines the boundaries of the region over which the ensemble scores for the pack ice (Sect. **??**) are computed. Axes indicate grid point indices on the model's native grid.

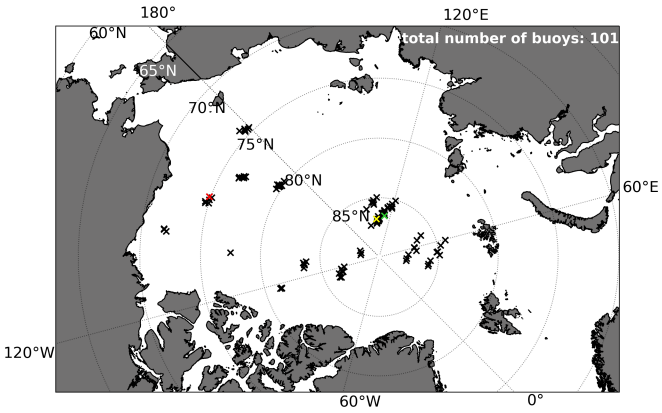

Figure 3. Location of the observed IABP buoys at initial time of the P1,...,P8 hindcast periods (start dates from the 16th of January to the 27th of March 1997, see Table **??**). 101 IABP buoys are considered in total. The colored "x" markers highlight the initial position of the few buoys taken as examples in the following (Figs. 5 and 14).

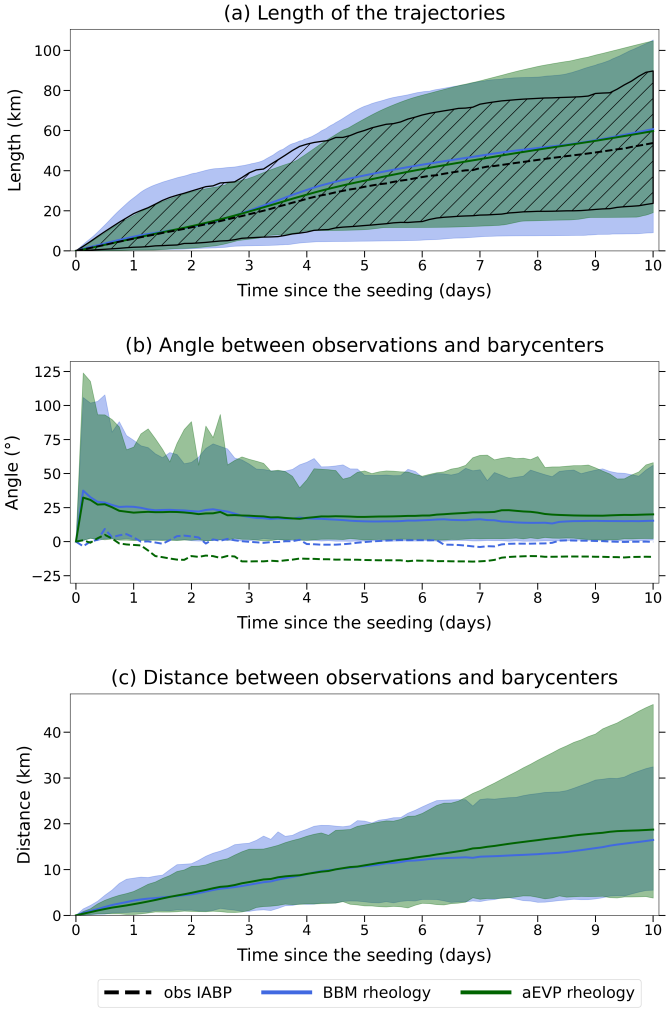

Figure 4. Temporal evolution of different properties of the simulated buoys compared to the observed IABP buoys: (a) length, i.e., the distance covered by the buoys, (b) angle between the two lines drawn from the seeding position to, respectively, the observed buoy and the barycentre of the N = 20 simulated buoys of each ensemble hindcast, and (c) the direct distance between the observed buoy and the barycentre. The results of BBM and aEVP are colored blue and green, respectively, with the thick lines for the ensemble means of all the buoys over all eight 10-day periods, while the shaded envelopes indicate the 5%-to-95% percentile ensemble distribution. In (a) the distance covered by the observed buoys is with the black line (mean) and the hatched envelope (5%-to-95% percentile). In (b), the envelopes and solid lines are computed considering the absolute value of the angle, while the dashed lines correspond to the mean value of the angle.

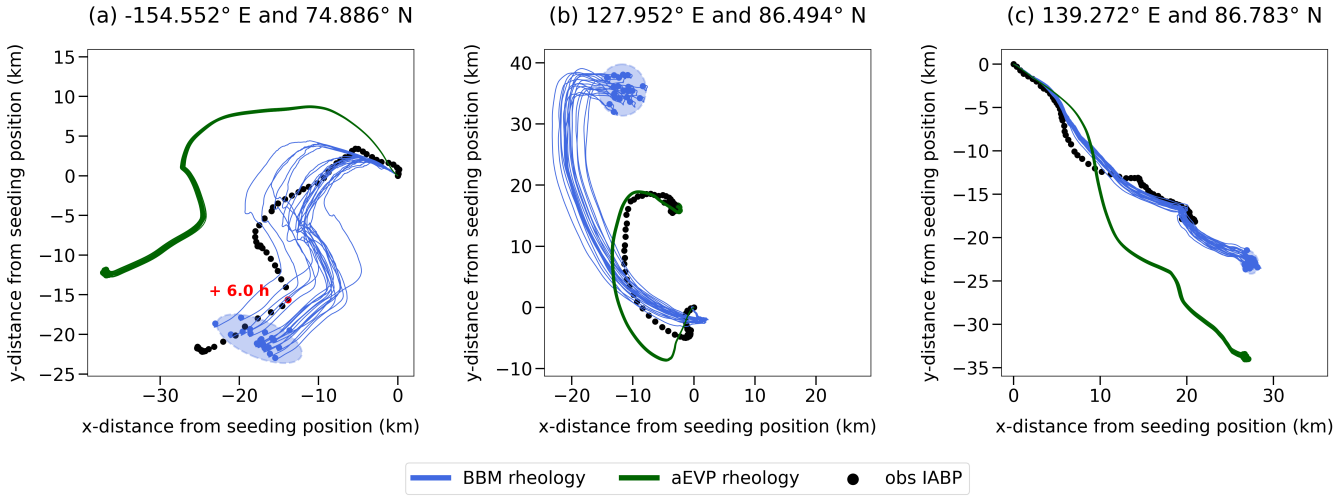

Figure 5. Three examples of observed IABP trajectories (black) and corresponding virtual trajectories simulated from the BBM ensemble hindcast (blue, 20 members) and aEVP ensemble hindcast (green, 20 members) seeded at midnight on: (a) March 7th, at -154.552° E;74.886° N (red cross in Fig. 3), (b) February 15th, at 127.952° E;86.494° N (green cross in Fig. 3), and (c) February 5th, at 139.272° E; 86.783° N (yellow cross in Fig. 3). The observed IABP trajectories are plotted with a plain black circle every 3 hours. In case of missing values, the circle following the gap is shown in red, with the corresponding time gap. The simulated ensemble trajectories are plotted as colored curves, and only the final positions after 10 days are marked as plain circles. The shaded ellipses represent the 95% confidence regions of the final positions, assuming a bivariate normal distribution.

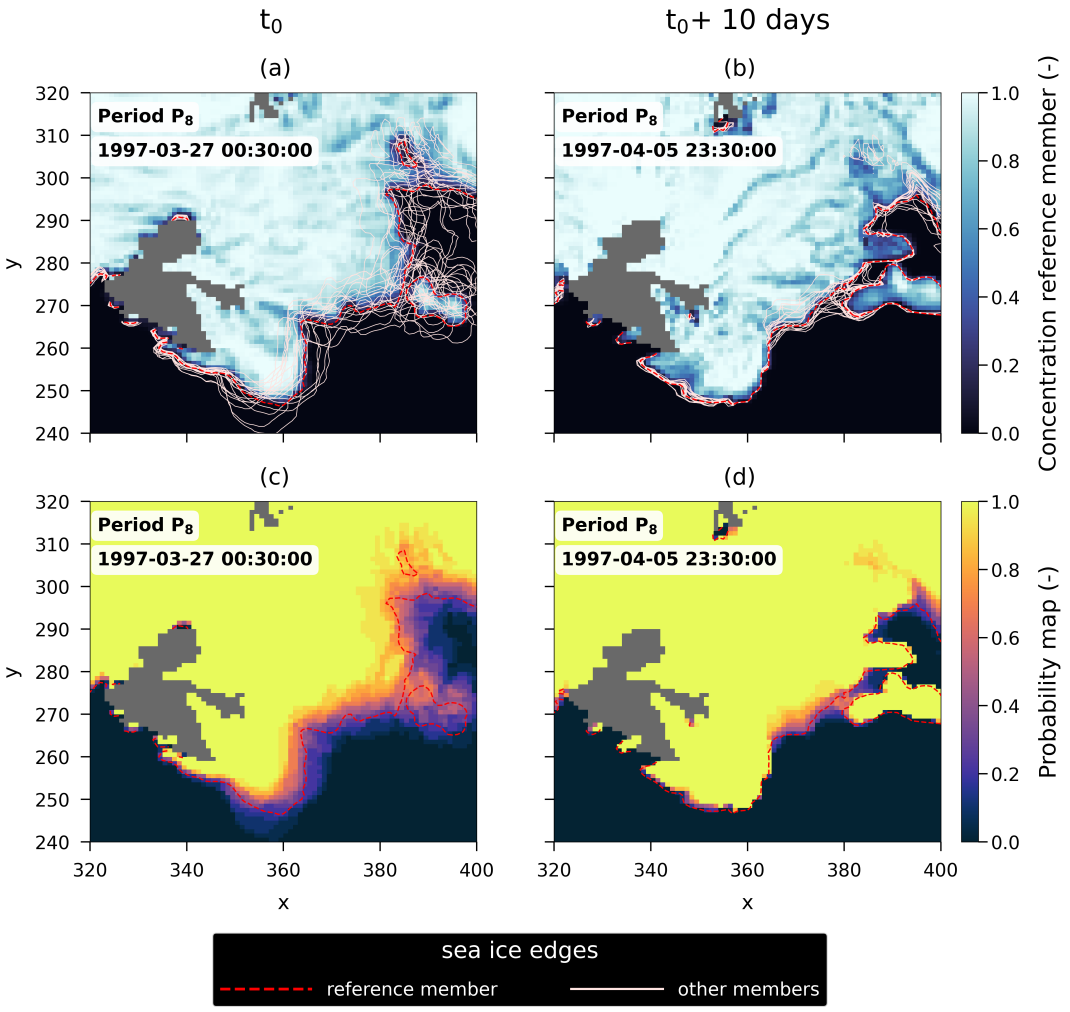

Figure 6. Example of the dispersion of the sea ice edge position (pink and red lines) in the Svalbard region in the BBM ensemble members for period P_8 and from the 50-km-scaled perturbation, at initial time t_0 and after 10 days (a,b respectively). The hourly sea ice concentration field of the reference member is shown in the background for their respective lead time (shading in a, b). The corresponding maps of probability $P[c > 0.15]$, computed from the 19 members excluding the chosen reference member (whose ice edge is shown in red) are also plotted at initial time and after 10 days (c,d resp.) to illustrate the methodology to compute the SPS metrics (see text for more details). Axes indicate grid point indices on the model's native grid.

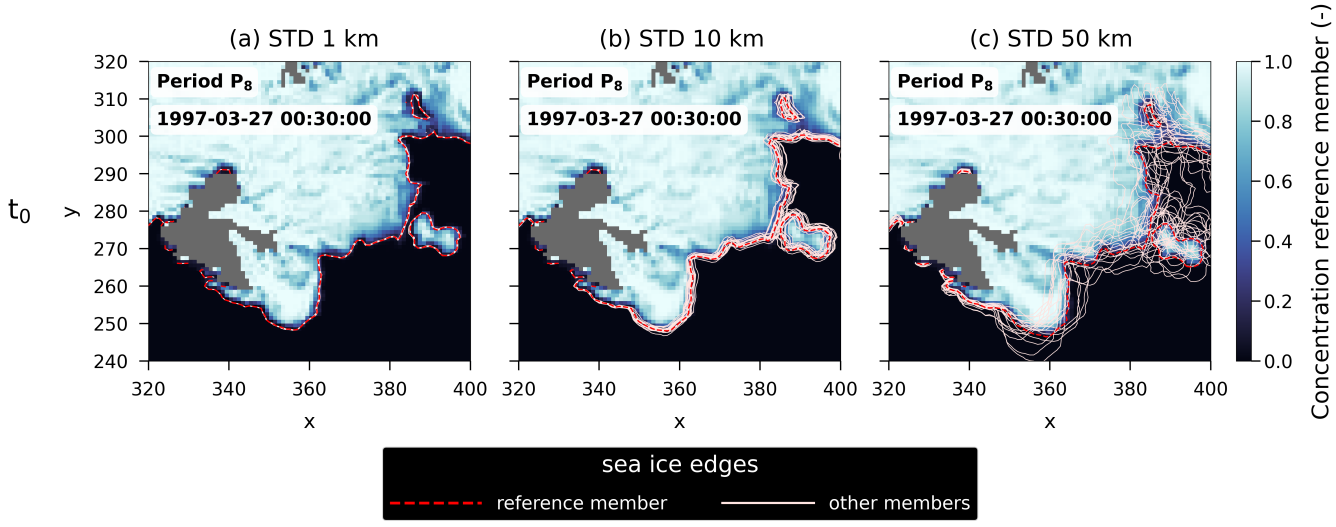

Figure 7. Initial spread of the local sea ice edge position in the Svalbard region during period P_8 (1997-03-27) for an initial perturbation scaled to a standard deviation (STD) of (a) 1 km, (b) 10 km, and (c) 50 km. The ice edge of the reference ensemble member (used here to illustrate the method) is indicated by a thick dashed red line, while the remaining 19 members are shown as thin pink lines. The corresponding hourly sea ice concentration field of the reference member is displayed in the background (shading). Axes indicate grid point indices on the model's native grid.

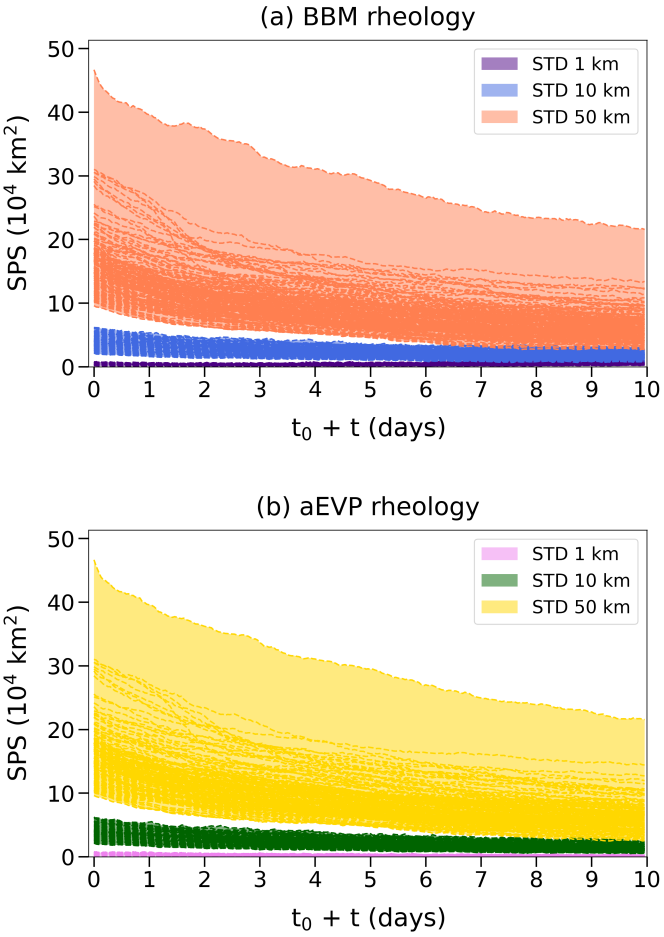

Figure 8. Temporal evolution of the SPS score computed over the whole Arctic, for BBM and aEVP hindcasts (a,b resp.). The plotted curves correspond to the SPS scores computed for each of the eight forecast periods, and for each ensemble member taken alternatively as the pseudo-truth, resulting in total in 8 Periods $\times$ 20 scores = 160 curves. The colors correspond to the three different amplitudes of initial perturbation.

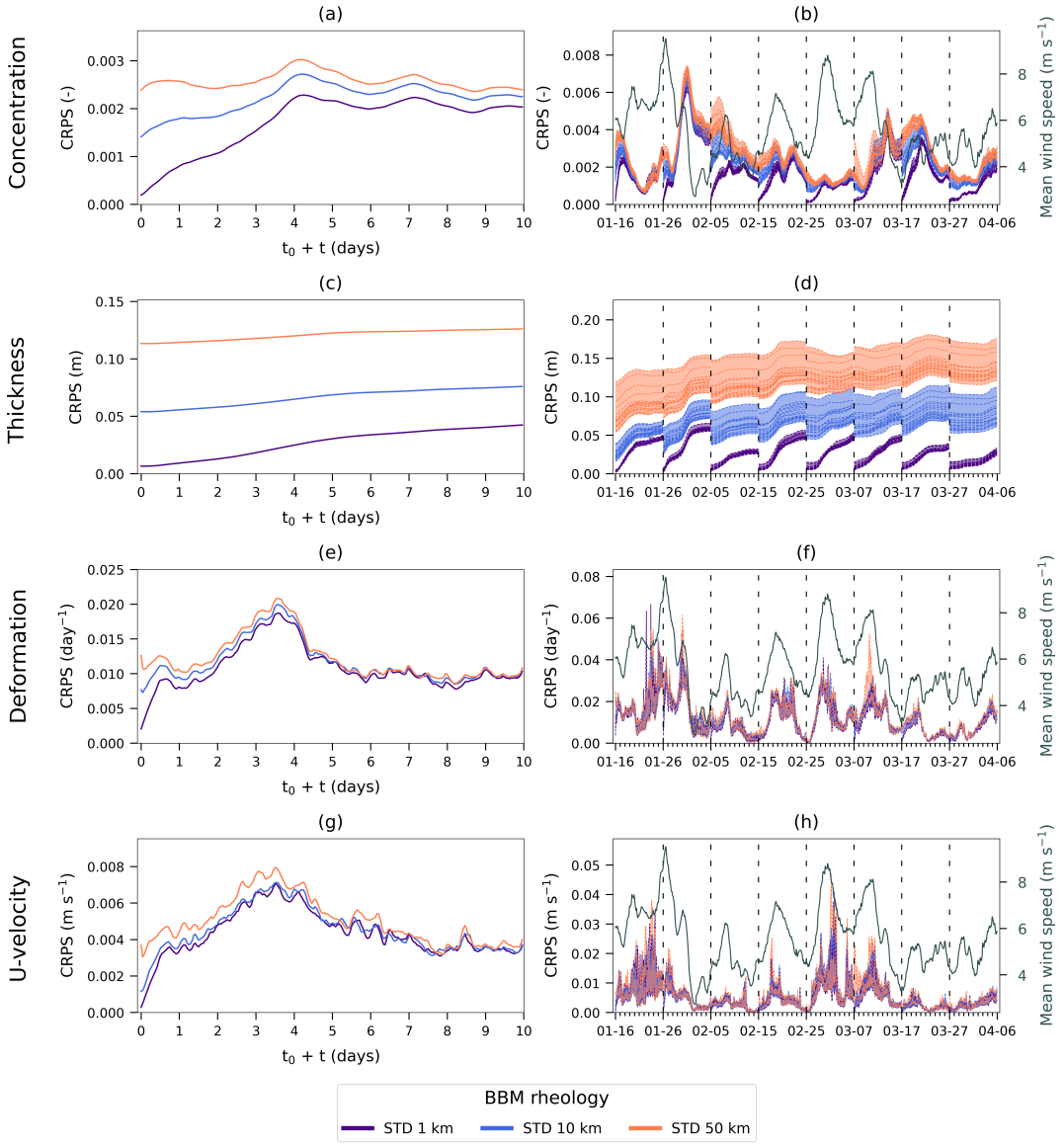

Figure 9. Temporal evolution of the CRPS metric in the BBM hindcasts for the three types of initial perturbations, shown in average for the eight periods P_1,...P_8 (left column) and separately for each period (right column). The CRPS metric is computed for (a,b) sea ice concentration, (c,d) thickness, (e,f) deformation and (g,h) velocity along the x-axis. The colors correspond to the three amplitudes of initial perturbation. The solid line on (b, f, h) stands for the mean wind speed averaged over the same domain as the CRPS metrics (cf Fig. 2). The colored envelopes (right column) correspond to the min-to-max of the CRPS scores computed for each member taken alternatively as the pseudo-truth. The thin curves on panels (b) and (d) show the scores corresponding to each member taken alternatively as the pseudo-truth.

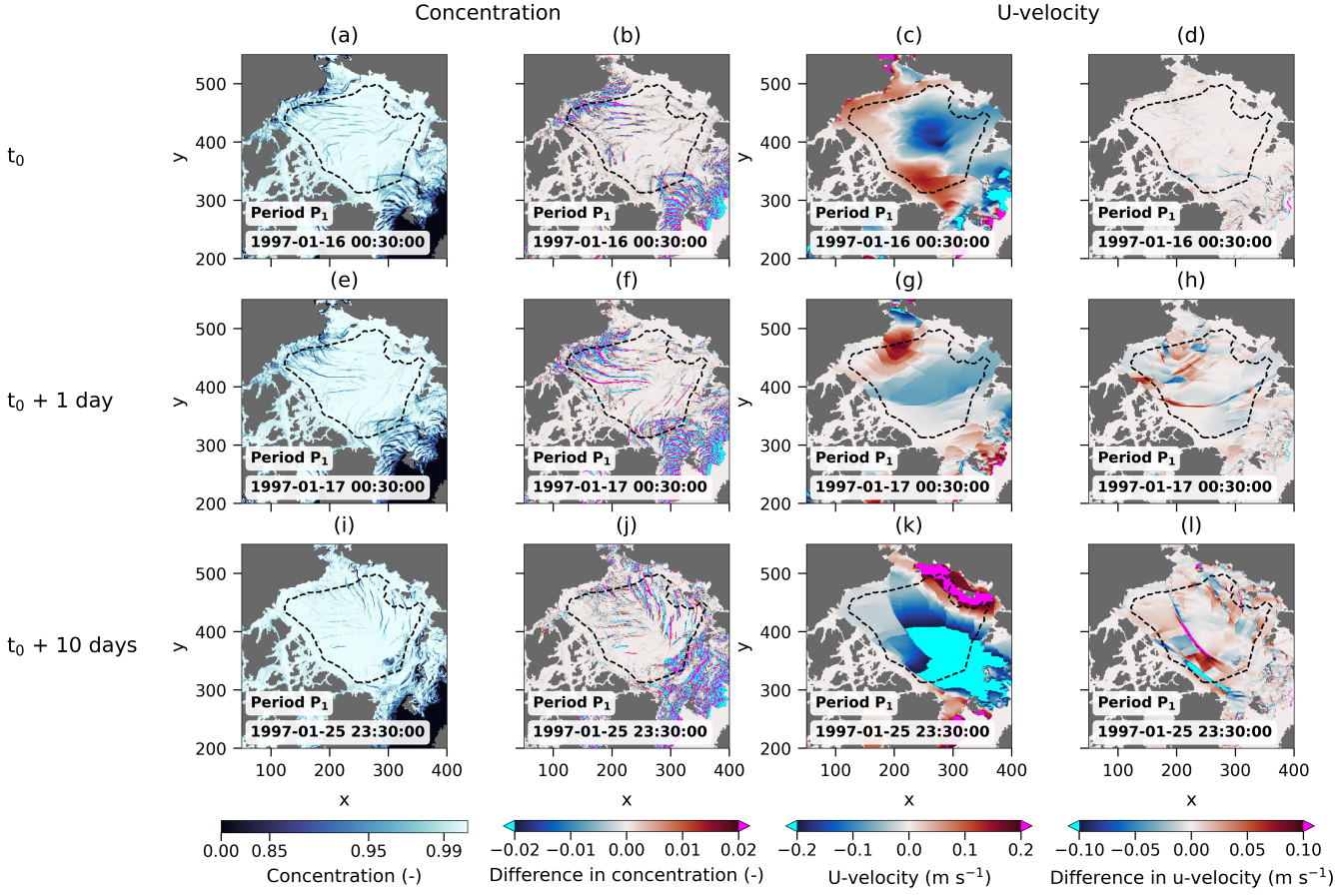

Figure 10. Sea-ice concentration and U-component of sea-ice velocity (a,e,i and c,g,k, resp.) from the BBM hindcasts initialized with the 10-km perturbation at three lead times during period P_1: (a–d) initial time, (e–h) after one day, and (i–l) after ten days. Panels (b,f,j) and (d,h,l) show, respectively, the differences in concentration and in the U-component of velocity between two perturbed ensemble members. The black dashed line defines the boundaries of the region over which the CRPS metric is computed. Axes indicate grid point indices on the model's native grid.

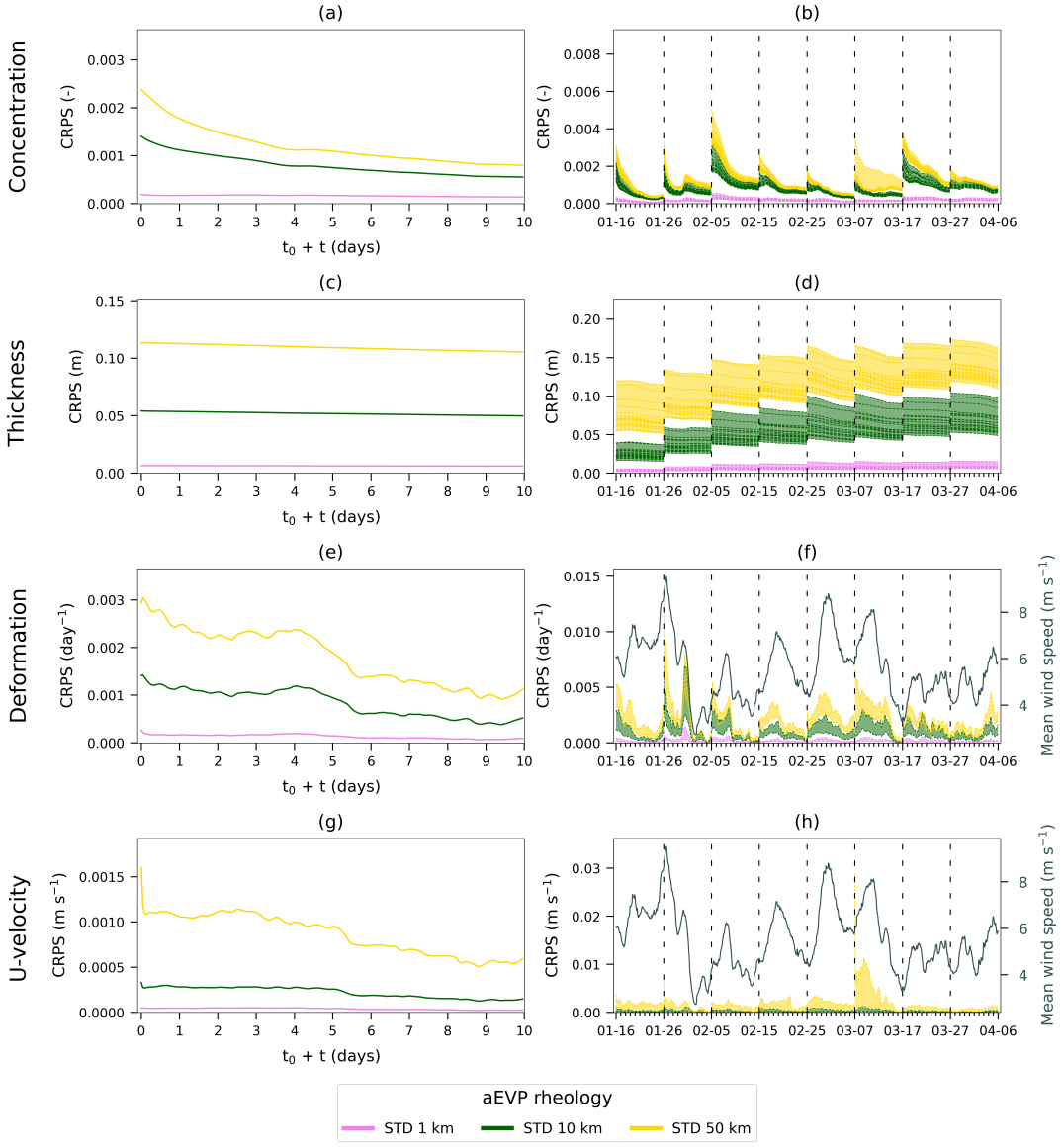

Figure 11. Same as Fig. 9 but for the CRPS metric from the aEVP hindcasts initialized with the 10-km perturbation.

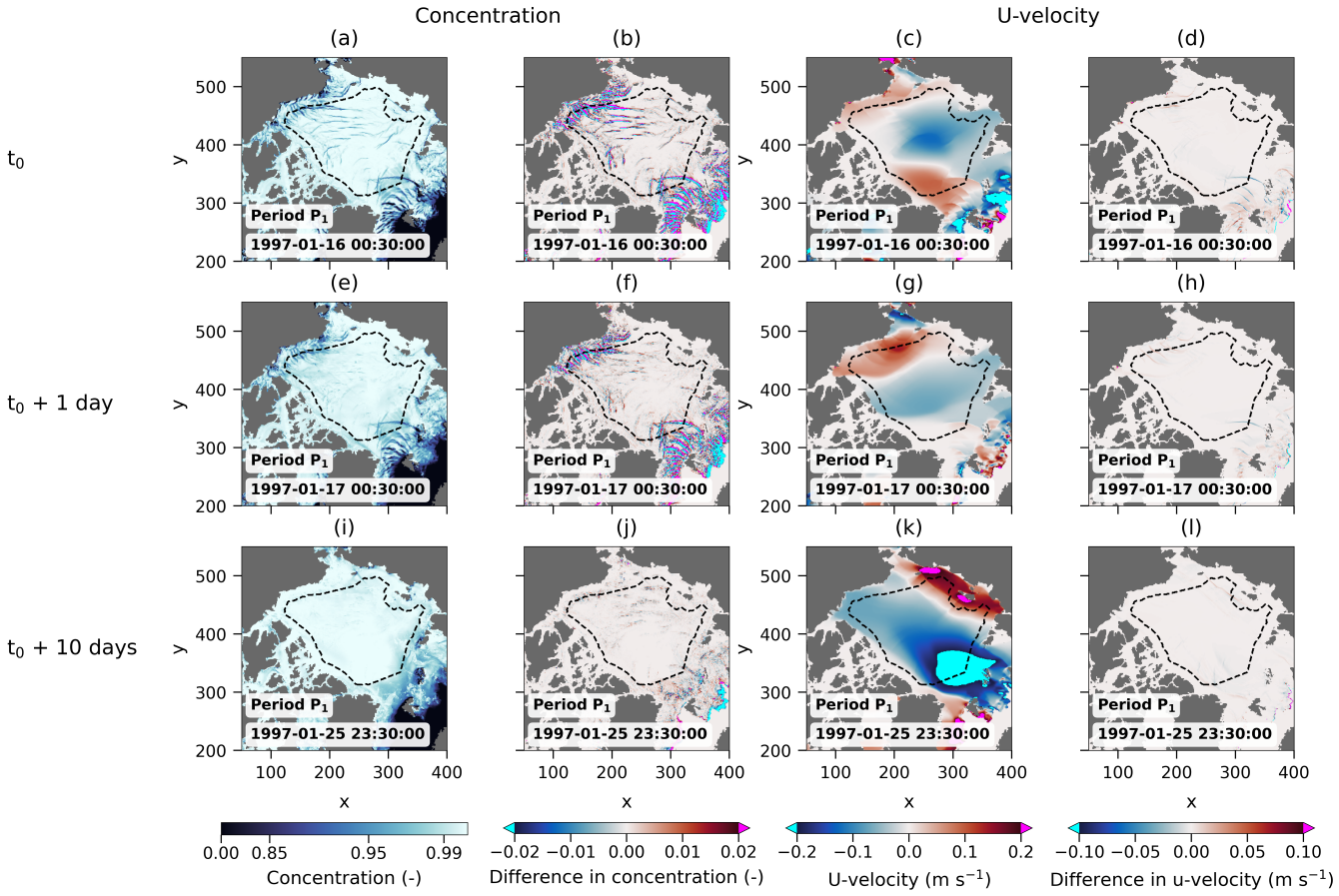

Figure 12. Same as Fig. 10 but from the aEVP hindcasts initialized with the 10-km perturbation.

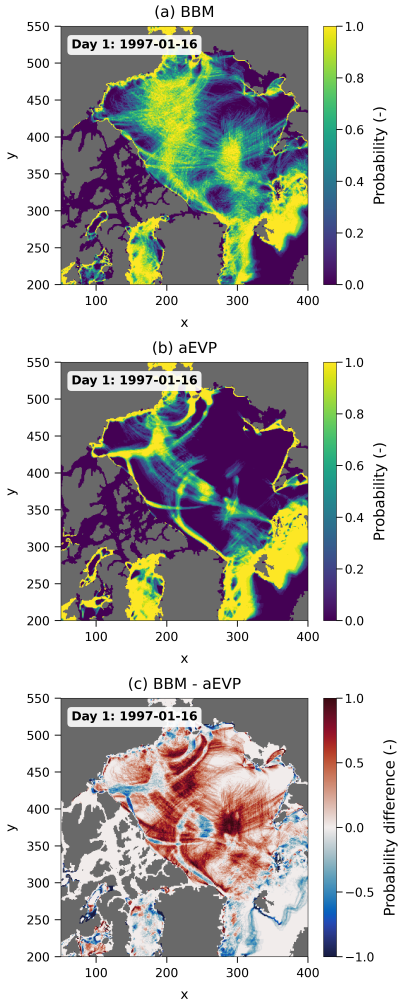

Figure 13. Maps illustrating the likelihood of experiencing at least one high-deformation event within 24 hours (here plotted for the first day of period P_1 as an illustration) in each model cell in the BBM and aEVP ensemble experiments (a,b resp.) initialized with STD = 50 km positional perturbations. The high-deformation events are defined as those when the hourly deformation in a model grid cell exceeds a threshold set at the 95th percentile of the hourly deformation distribution over the winter season in the pack-ice region (see text for details). Panel (c) shows the difference in probability between panels (a) and (b). Axes indicate grid point indices on the model's native grid.

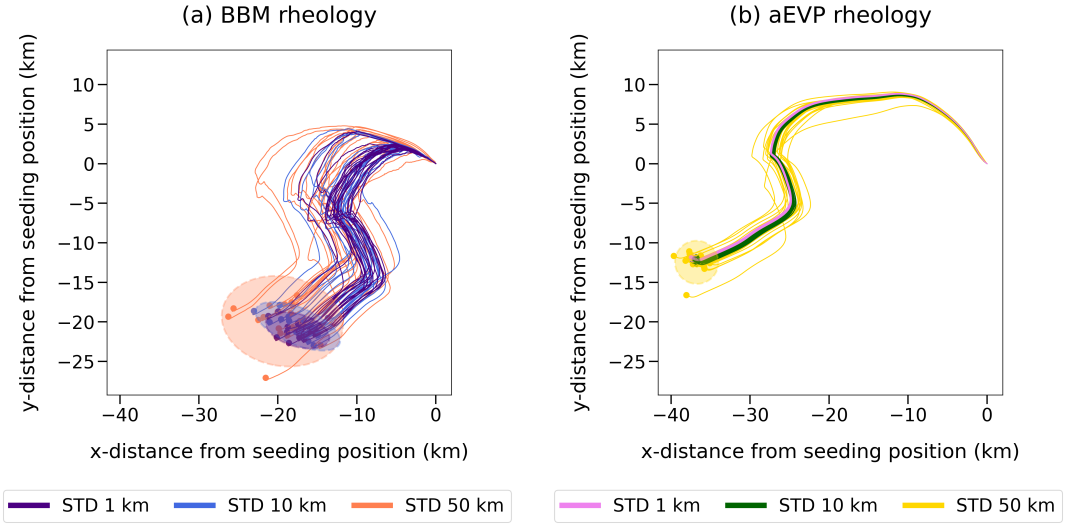

Figure 14. Example of virtual buoy trajectories simulated from the BBM and aEVP ensemble hindcasts (a,b, respectively) from an seeding on 7th March 1997 at midnight at -154.552° E and 74.886° N (red cross in Fig. 3). The coloured circles highlight the final positions after 10 days and the shaded ellipses represent the 95% confidence regions of the final positions, assuming a bivariate normal distribution. The colors correspond to the rheology and amplitude of the initial perturbation of the experiments.

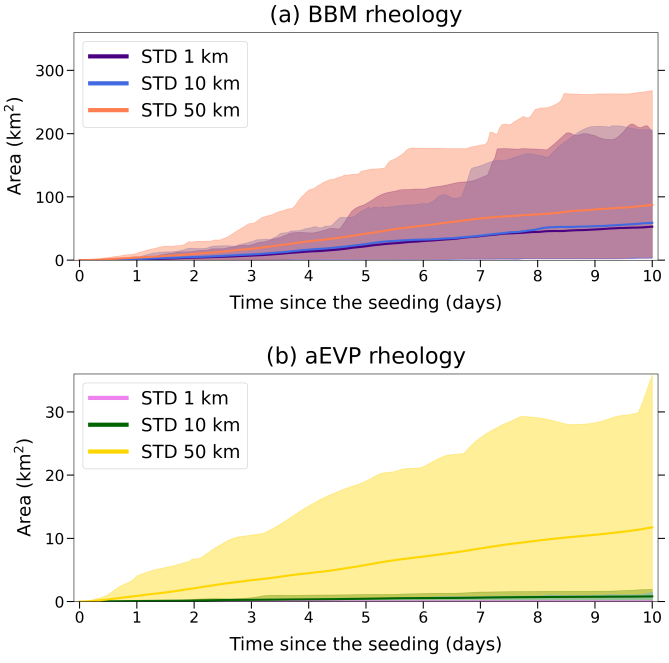

Figure 15. Temporal evolution of the spread of the virtual buoys from the BBM and aEVP ensemble hindcasts (a,b, resp.), as measured by the area of the ellipse defined by the 95% confidence contour of the distribution (assumed to be bivariate normal). The colors correspond to the rheology and amplitude of the initial perturbation of the experiment. The thick lines show the ensemble mean of all the buoys over all the eight 10-day periods, while the shaded envelopes indicate the 5%-to-95% percentile ensemble distribution. Panel (a) uses a y-axis range 10 times larger (area in km^2) than panel (b).

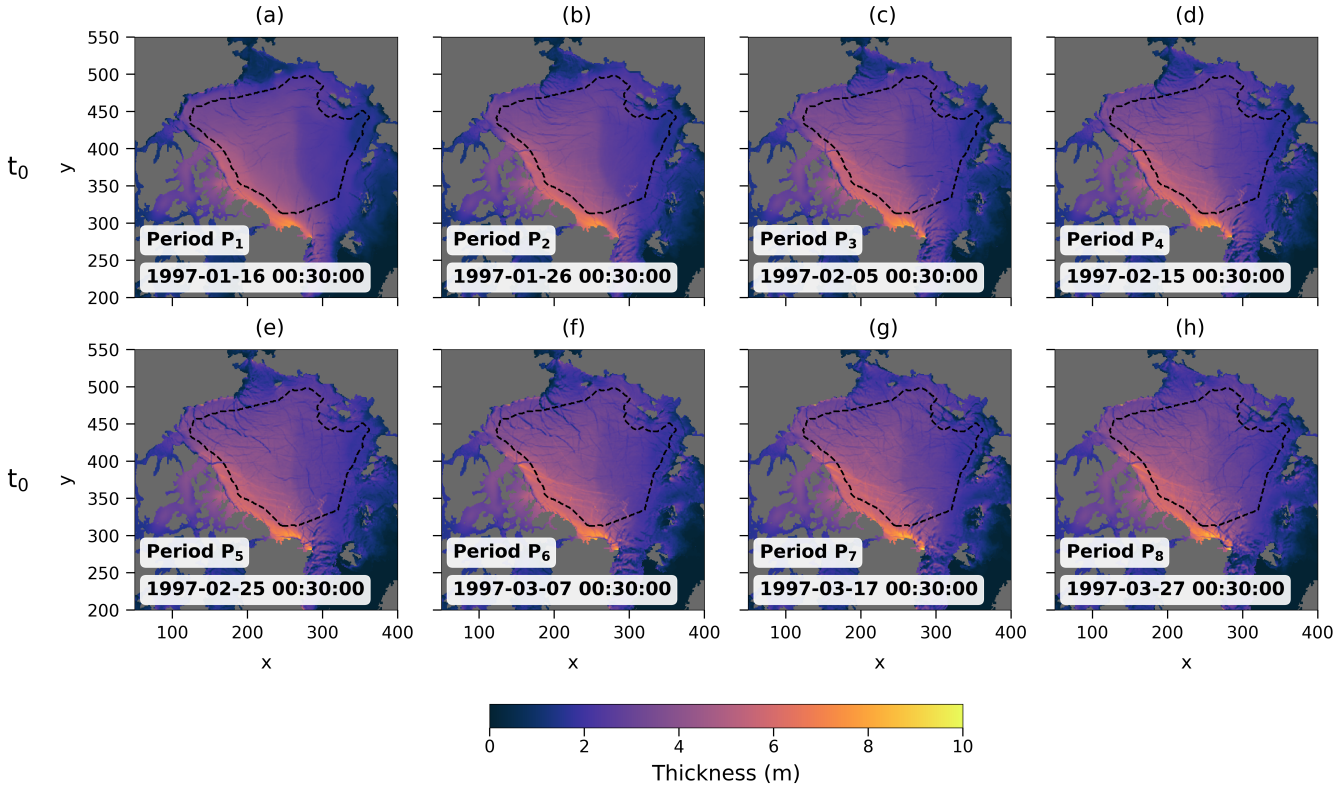

Figure 1. Sea ice thickness maps at initial time (t_0) of each period P_1,...,P_8 from the reference (unperturbed) experiment (see Sect. **??** for more details). The black dashed line defines the boundaries of the region over which the CRPS metric is computed in Sect. **??**. Axes indicate grid point indices on the model's native grid.

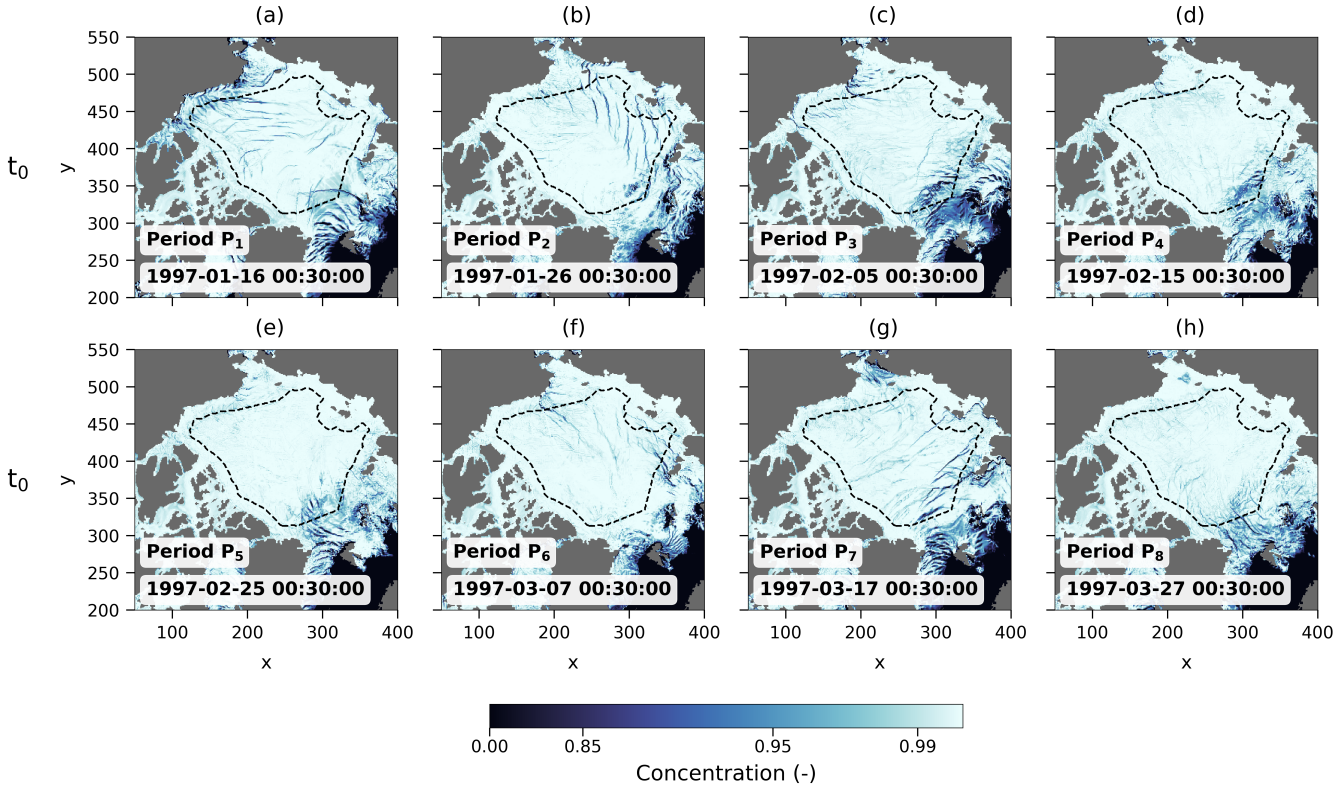

Figure 2. Sea ice concentration maps at initial time (t_0) of each period P_1,...,P_8 from the reference (unperturbed) experiment (see Sect. **??** for more details). The black dashed line defines the boundaries of the region over which the CRPS metric is computed in Sect. **??**. Axes indicate grid point indices on the model's native grid.