In this Supplementary Material of appendices, all figure numbers are continued from the paper.

Appendix I: Sensitivity for Number of Histogram Bins

To see how sensitive our results are to the number of histogram bins, recall that our time-step selection is by dynamic programming whose cost function is based on **summing the VI values** (Eqs. (5), (6); see also Eqs. (3), (2) for VI). Thus we study the sensitivity of the VI values with respect to the number of histogram bins—if the VI values are not sensitive, then the selection results are not sensitive. We varied the number of histogram bins as 64, 128, 256, 512 and computed $VI(X_i;X_j)$, the VI between time steps i and j in the original dataset, for all pairs $i \neq j$. To compare such VI values, we *normalized* the VI values to the range [0,1] by *scaling with a constant factor*. Such scaling does *not* affect our time-step selection results, since the dynamic programming cost function is based only on *summing* the VI values (Eqs. (5) and (6)). The results on Isabel are shown in Fig. 17.

Consider a scatter plot in the lower triangle of the matrix in Fig. 17, say the one labeled 64 in column and 128 in row. For each time step pair (i, j), we computed the VI value using 64 bins, which gave the x-coordinate; similarly we computed the VI value using 128 bins, which gave the y-coordinate. We see that these scatter plots always show *linear* relationships, meaning that the time steps i, j in a pair are similar/dis-similar regardless of different numbers of bins used, and that our VI is a *robust* metric, not very sensitive to the number of bins. The scatter plots in the upper triangle would be symmetric, and thus we computed and show the corresponding *normalized root-mean-square difference (NRMSD)*, defined as $(\sqrt{(\sum(VI_{(i,j),b_1} - VI_{(i,j),b_2})^2)/n_p})/(1-0)$, where $VI_{(i,j),b_1}$ is the VI value of the time step pair (i, j) using b_1 bins (similarly for $VI_{(i,j),b_2}$) and n_p is the number of the (i, j) pairs. Since all VI values have range [0,1] due to normalization, we have $1-0=1$ in the denominator. Also, the scatter plots in the diagonal are not very interesting since the row and column have the same number (of bins). Thus we use the space to show something else more interesting: the corresponding distribution of the VI values. For example, for entry $(1, 1)$, we show the distribution of the VI values computed using 64 bins: the VI value range [0,1] (as the x-axis) is divided evenly into 20 bins (same for each diagonal entry); for each bin, we show the **count** (i.e, the number of time step pairs (i, j) whose VI values fall into this bin). The range of the count is [0,600], which is the range of the y-axis (labels 0, 600 not shown in Fig. 17) for each diagonal entry. We see that the distributions evolve gradually from 64 to 512 bins, with their differences shown by the NRMSDs. We fixed the bin number to 128 for all other experiments as it has the lowest NRMSDs.

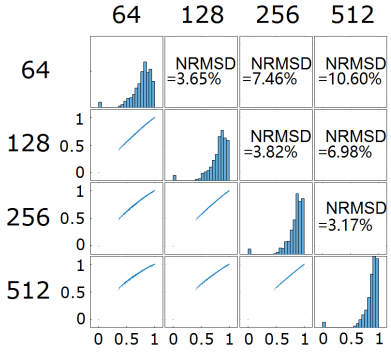

Figure 17: *Scatter-plot matrix for the VI values between different time steps i, j from the Isabel dataset, computed using 64, 128, 256, 512 histogram bins and normalized to the range [0,1].*

Appendix II: Comparing with Base-Line Methods on Synthetic Data

As mentioned in Sec. 4.4, we compared our accurate approach with the two base-line methods, uniform sampling and *Greedy*, all under VI, on our synthetic datasets.

Uniform Sampling: We compared with uniform sampling on dataset Synthetic-1.

Dataset: Synthetic-1 Synthetic-1 is a small regular-grid dataset, with dimensions 10x10x10 and 24 time steps. The value at point (x,y,z) and time $t \in [1,24]$ is generated by a function $f(x,y,z,t) = x+y+z$ if $t \in \{8,18,20\}$, otherwise $f(x,y,z,t) = 0$. So it is a static field only appearing at time steps 8, 18 and 20. The results for uniform sampling is given in Fig. 18. An illustration of the time series is shown on top of Fig. 18 (a) and (b). The dataset has three peaks at 8, 18 and 20. Uniform sampling would start from time step 1, select every dth time step where d is a uniform distance. With different d, the selections of time steps are shown in Fig. 18(a). We see that the critical time steps 18 and 20 were never selected. The selection for 5 time steps was {1,8,15,22,24} which included the critical time step 8. However this selection resulted in a big error according to our error function. As shown by the plot on top of Fig. 18(a), the missing time steps were reconstructed by linear interpolation. The reconstruction (red dotted line) was actually very different from the original dataset (blue lines). On the other hand, our accurate method made globally optimal selections for each k. Note that to select a peak at q it is necessary to also select $q-1$ and $q+1$ to avoid big error(s) from interpolation, which requires $k \geq 5$ (2 more steps for the first and last ones). For $k=5$, this is exactly what our method did, selecting 1,7,8,9,24 ($q=8$). For $k=7$, the best choice is to use 5 steps for the last two peaks ($q=18,20$) which share a common neighbor 19, as did our method (17–21, plus 1 and 24). For $k \geq 10$, it is possible to include one more peak ($q=8$) to bring down error to 0, as our method did. The curves of the total information difference for the two methods are plotted in Fig. 18(c).

Greedy Method: We compared our accurate method with *Greedy* on dataset Synthetic-2.

Dataset: Synthetic-2 Synthetic-2 is a regular grid dataset with dimensions 10x10x10 and 24 time steps. The value at point (x,y,z) and time $t \in [1,24]$ is generated by the function: $f(x,y,z,t) = (x+y+z) \cdot \sin(2\pi\frac{t-1}{23})$. This function maps a sin function period from $[0,2\pi]$ to $[1,24]$, then multiplies it with a static linear field.

Greedy did not make good selections especially for small k. As shown in Fig. 19(c), the curves of total information difference of the two methods have a big difference. The maximum margin is at $k=4$. Our method selected steps 1,7,18,24 where 7 and 18 are

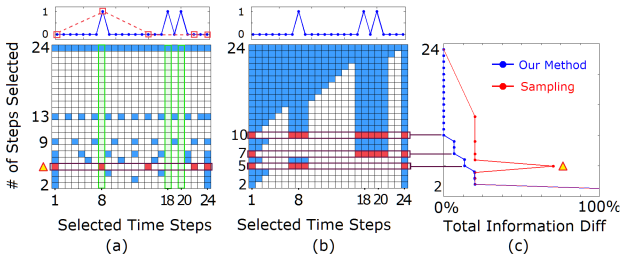

Figure 18: *Results for dataset Synthetic-1: (a) Selection Table of uniform sampling; (b) Selection Table of our accurate method; (c) the curves of total information difference.*

at the maximum and minimum of the sin function, but *Greedy* selected 1,11,23,24 which cannot represent the dataset well. This can be explained with the Selection Table in Fig. 19(a). At very early iterations, *Greedy* chose to discard the maximum and minimum at 7 and 18, marked with the red circle. Because at that time, the maximum and minimum were at flat regions and could be easily reconstructed from their neighboring time steps. However when selecting only a few time steps, the actual errors of the reconstruction at the maximum and minimum time steps are very high. So an optimal selection should include these time steps.

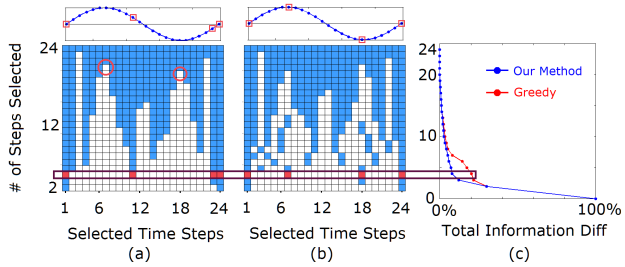

Figure 19: *Results for dataset Synthetic-2: (a) Selection Table of* Greedy*; (b) Selection Table of our accurate method; (c) the curves of total information difference.*

To conclude, while uniform sampling and *Greedy* are simple, the qualities of their results are unstable and can be very bad, with no guarantee. On the other hand, our accurate method always provides a guarantee on the optimality.

Appendix III: Additional Figures

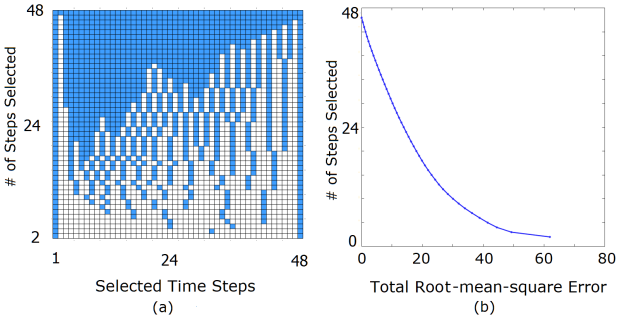

Figure 21: *Results of our in-core/accurate method (under RMSE) on Isabel-TC: (a) the Selection Table; (b) the curve of Total RMSE.*

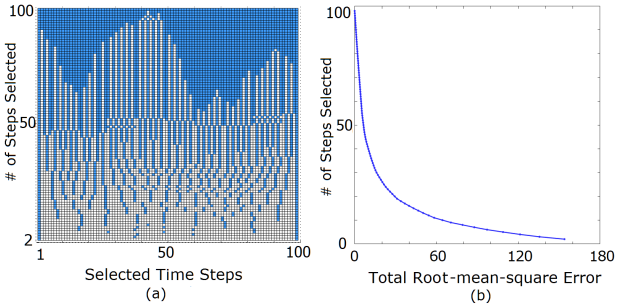

Figure 23: *Results of our in-core/accurate method (under RMSE) on Vortex.*

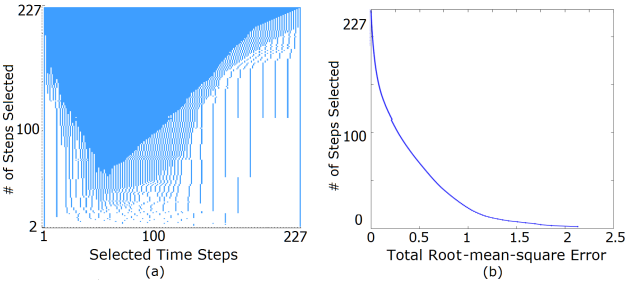

Figure 24: *Results of our approximate method under RMSE ($t = 12$) on TeraShake.*

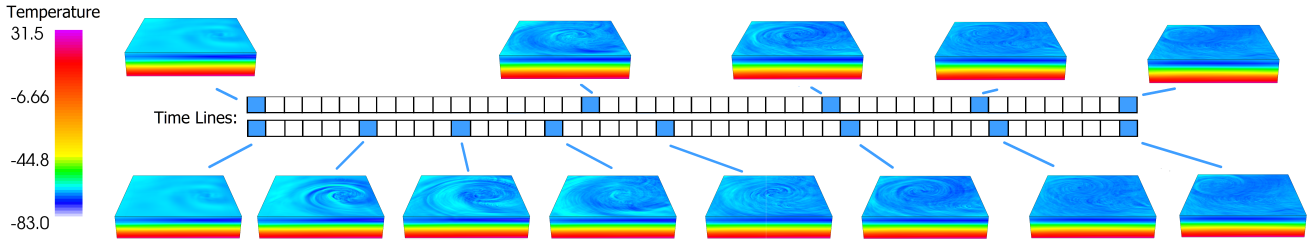

Figure 20: *Selecting 5 and 8 time steps from dataset Isabel-TC using our in-core/accurate method under VI: {1, 19, 30, 40, 48} for the top row, and {1, 7, 12, 17, 23, 33, 41, 48} for the bottom row.*

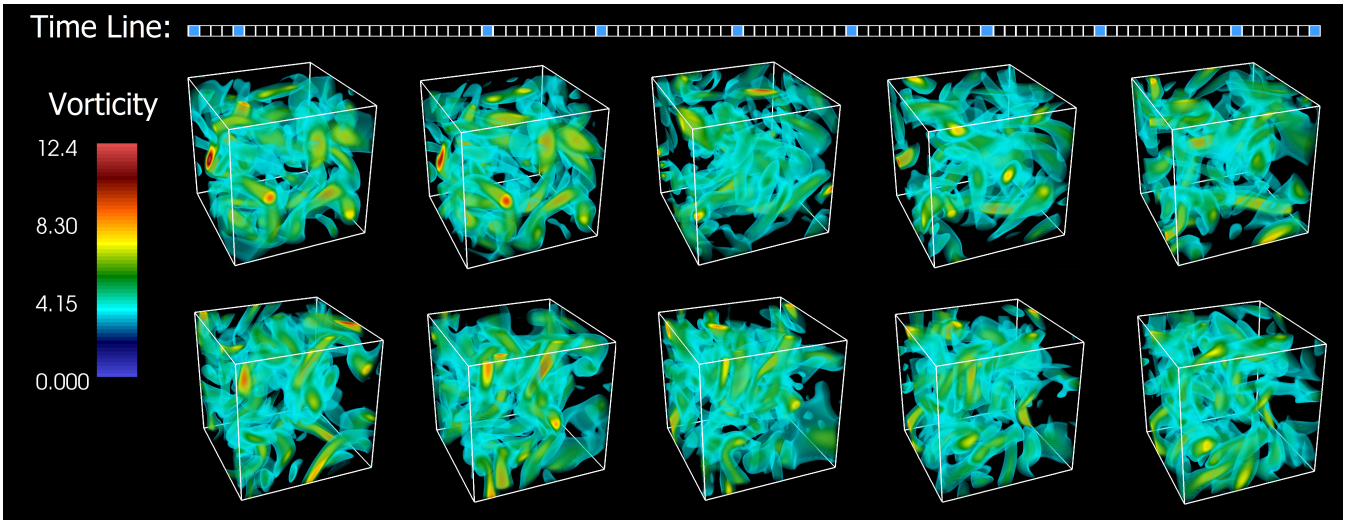

Figure 22: *Selecting 10 time steps from dataset Vortex using our in-core/accurate method under VI: {1, 5, 27, 37, 49, 59, 71, 81, 93, 100}.*

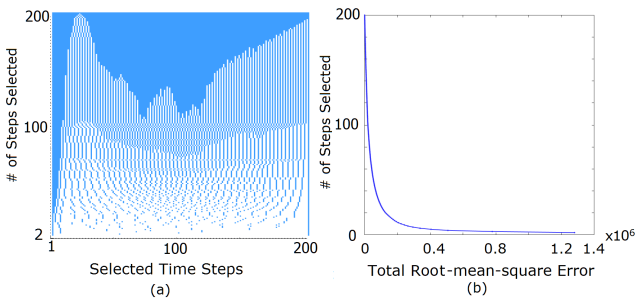

Figure 25: *Results of our approximate method under RMSE ($t = 12$) on Radiation.*

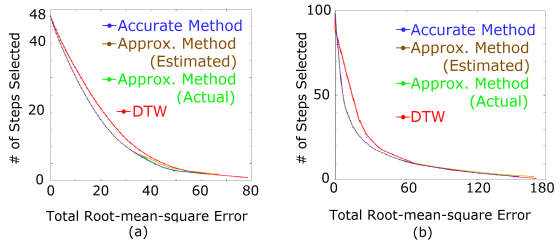

Figure 26: *The curves of total **RMSE** of our accurate method, approximate method (estimated total RMSE and actual total RMSE; $t = 8, 10$ for (a), (b)), and DTW: (a) Isabel-TC; (b) Vortex.*

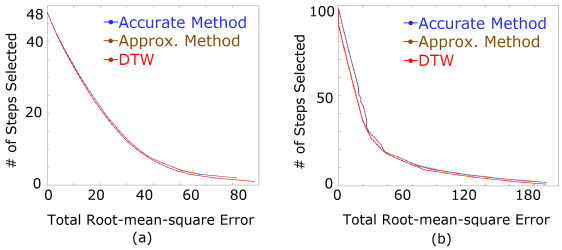

Figure 27: *The curves of total RMSE of our accurate method, approximate method ($t = 8, 10$ for (a), (b)) and DTW for (a) Isabel-TC and (b) Vortex, all under* ***DTW's metric****, which* favors DTW. *For our approximate method they are actual total RMSE. For DTW they are the same as those in Fig. 26.*

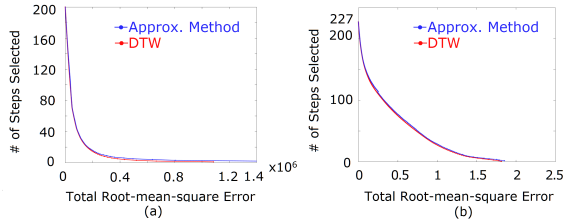

Figure 28: *The curves of* actual *total RMSE of our approximate method ($t = 12$) and DTW for (a) Radiation and (b) TeraShake, all under* ***DTW's metric****, which* favors DTW. *For DTW they are the same as those in Fig. 15(c)(d).*