



**Figure S1.** Example of cross-validation scores for two possible end-member reconstructions. The series shown have no significance other than this demonstration. In (a) a reconstruction exactly reproduces the variability of the target series but does not capture a change in the mean from the calibration period to the verification period. Here, RE and CE indicate a poor reconstruction but  $r^2$  indicates a perfect reconstruction. In (b) the situation is reversed. Here the reconstruction has captured the change in the mean but none of the variability. Hence,  $r^2$  and CE indicate a very poor reconstruction whereas RE indicates a skillful reconstruction because the change in the mean is captured. In (c) there is almost no change in the mean between calibration and reconstruction, and the reconstruction has perfectly captured the variability of the target series but at reduced variability (0.25 of the target series). RE and CE indicate a moderately skillful reconstruction whereas  $r^2$  indicates a perfect reconstruction.