Supplementary Material for "Possible mechanisms for four regimes associated with cold events over East Asia"

Zifan Yang[1], Wenyu Huang[1], Bin Wang[1,2], Ruyan Chen[1], Jonathon S. Wright[1], and Wenqian Ma[1]

[1] Ministry of Education Key Laboratory for Earth System Modeling, and Department of Earth System Science (DESS), Tsinghua University, Beijing 100084, China

[2] State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG), Institute of Atmospheric Physics (IAP), Chinese Academy of Sciences, Beijing 100029, China

Corresponding author:

Wenyu Huang

E-mail: huangwenyu@mail.tsinghua.edu.cn

1. Circulation regimes identified by the distance-based *k*-means clustering method

Here, we present the four circulation regimes identified by applying the distance-based *k*-means clustering method to the HGT anomalies on the 500 hPa surface over the region of 0-180°E, 30-89.5°N (Fig. S1). Compared to the AWPC-based *k*-means clustering method, the regimes identified by the distance-based *k*-means clustering method have the following characteristics:

(1) To compare the amplitude in the HGT anomaly associated with regimes identified by two clustering methods, we define a metric called the regime-averaged zonal mean amplitude of HGT anomaly at 500 hPa (RZAH) for each latitude. Its calculating formula is:

$$RZAH_j = \frac{1}{M \times N} \sum_{k=1}^{M} \sum_{i=1}^{N} \left| R_{i,j,k} \right| \ ,$$

where M is the number of the regimes, N is the number of grid points in the zonal direction, (i, j) are the indices for grid points along the zonal and meridional directions, k is the index for regimes, $R_{i,j,k}$ is the HGT anomaly of cell (i, j) for regime k, and $\left| R_{i,j,k} \right|$ is the absolute value of $R_{i,j,k}$. In the polar regions, the distance-based *k*-means method has composite HGT anomalies of larger amplitude (Fig. S2). But in the mid-latitudes, the distance-based *k*-means method has composite HGT anomalies of smaller amplitude. Therefore, using the distance-based *k*-means method without considering any area weight leads to over-weighting of the polar cells. As we mentioned in the manuscript (Sect. 3.2), the upstream and downstream circulation anomalies over the mid-latitudes play essential roles in the formation of cold events over East Asia. Therefore, using the distance-based *k*-means method will underestimate the effects from the upstream and downstream circulation anomalies.

(2) To evaluate the classification accuracy, we have defined a metric called the within-group AWPC, which is estimated by the average of the AWPCs between the samples and the corresponding centroids. A larger value of the within-group AWPC means a more accurate classification. The within-group AWPCs of the four regimes identified by the distance-based *k*-means method are 0.66, 0.45, 0.62, and 0.47 (Fig.

S1), two of which are less than 0.5. In contrast, the within-group AWPCs of the four regimes identified by the AWPC-based *k*-means method are 0.60, 0.59, 0.60, and 0.61, all of which are larger than 0.5 (Fig. 2a-d). Therefore, the classification accuracy is lower in the distance-based *k*-means method.

(3) The maximum AWPC between any two identified regimes by the distance-based *k*-means method is 0.75, while that by the AWPC-based *k*-means is only 0.43. Therefore, the regimes identified by the distance-based *k*-means method are not significantly different from each other. Further, based on the distance-based *k*-means clustering method, we also carry out a sensitivity analysis by varying the number of regimes. Even if the number of regimes is reduced to 3 (Fig. S3), the maximum AWPC between any two identified regimes is 0.66, which is still comparatively large. When the number of regimes is reduced to 2 (Fig. S4), although the AWPC between the two regimes is only 0.20, the corresponding within-group AWPCs are 0.53 and 0.40, suggesting that the similarities of circulation patterns belonging to the same regime are relatively low.

(4) Ural blocking is one of the major causes for the occurrence of cold events over East Asia. Therefore, identifying the Ural blocking-like circulation pattern is very important. When four regimes are identified by the distance-based *k*-means method, the Ural blocking-like circulation pattern is split into two regimes (Fig. S1a and S1c). When three regimes are identified by the distance-based *k*-means method, there is only one regime with the Ural blocking-like circulation pattern that only contains 26 cold events (Fig. S3a). Although the two-regime classification by the distance-based *k*-means method has one regime with the Ural blocking-like circulation pattern that contains 60 cold events (Fig. S4a), the classification accuracy of the other regime is very low (Fig. S4b). In contrast, the four-regime classification by the AWPC-based *k*-means method recognizes the Ural blocking-like circulation pattern in one regime with a comparatively large number of cold events (43, accounting for 41.7%) and high classification accuracy for each regime.

According to these results, the AWPC-based *k*-means method is more appropriate for the classification of the circulation patterns associated with cold events over East Asia.

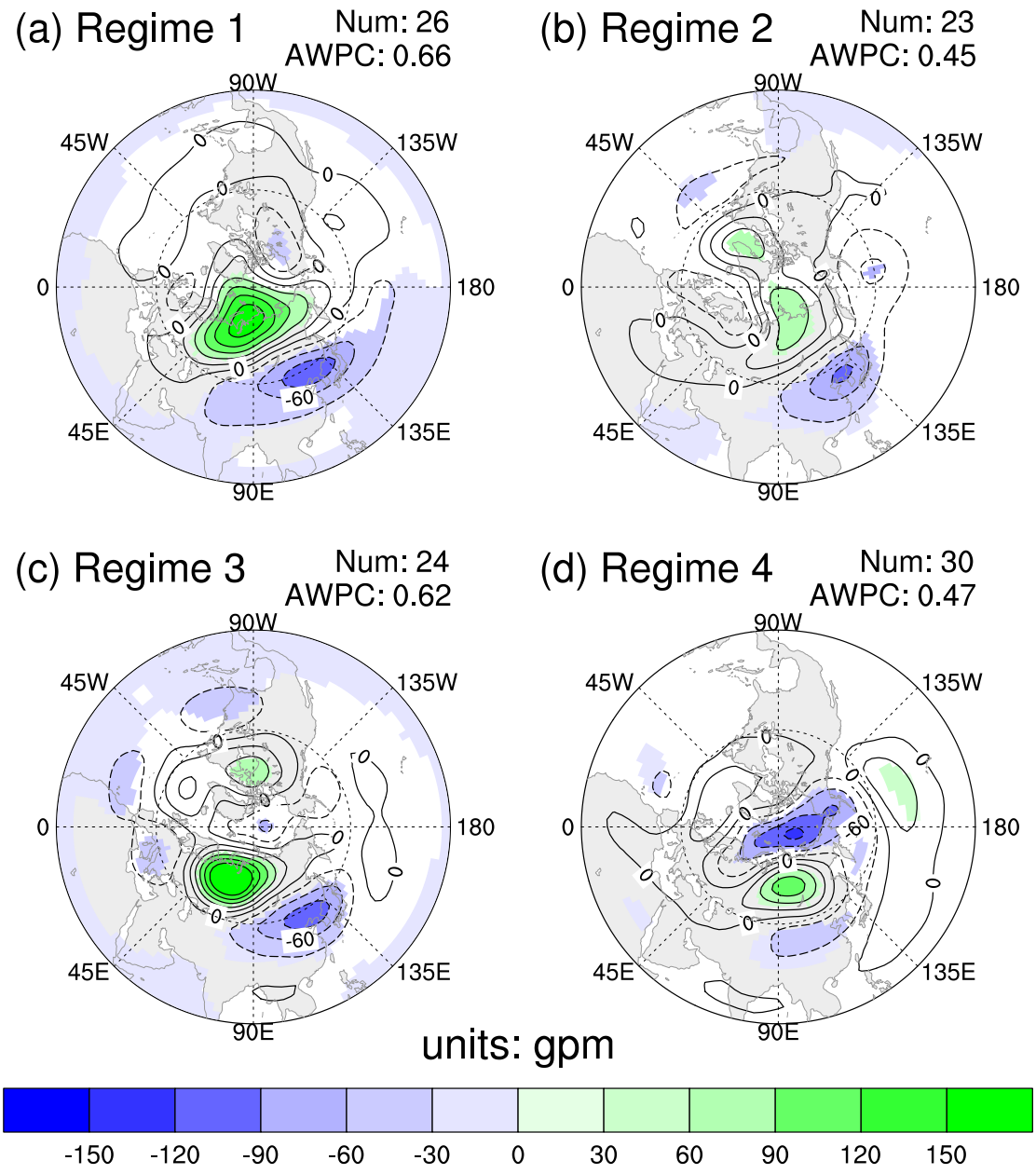

Fig. S1 Composite means of the HGT anomalies on the 500 hPa surface for the regimes identified by the distance-based *k*-means clustering method during DJFM 1948-2014 when the regime number is set to four. Shading indicates that the composite value meets the 95% confidence level based on two-tailed Student's *t* tests. The number of cold events belonging to each regime and the associated within-group AWPC are shown at the top right corner of each panel

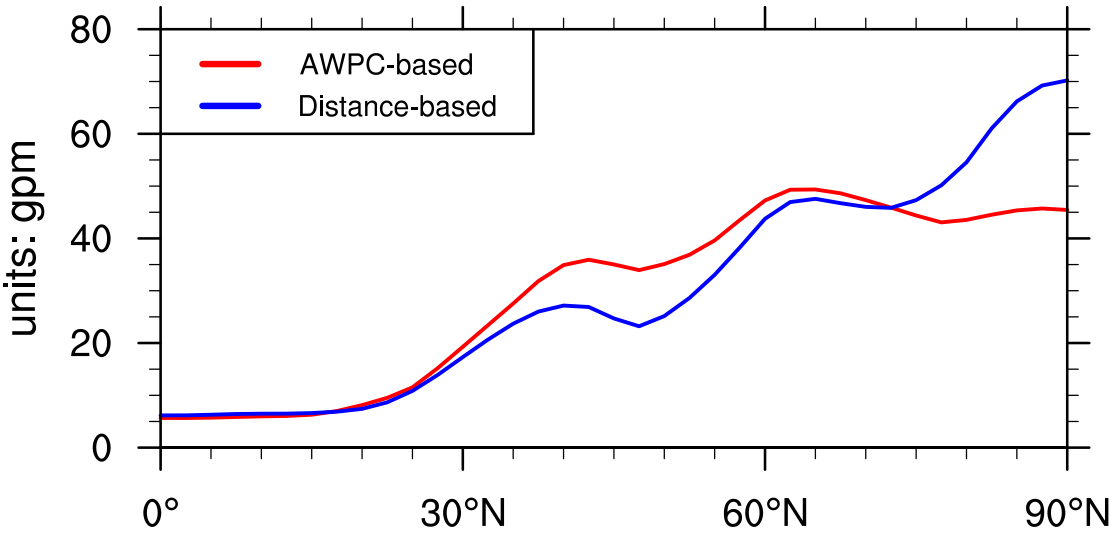

Fig. S2 The regime-averaged zonal mean amplitude of the HGT anomalies at 500 hPa (RZAH) for the regimes identified by the AWPC-based and distance-based *k*-means methods. Note that in both *k*-means clustering methods, the number of regimes is set to four

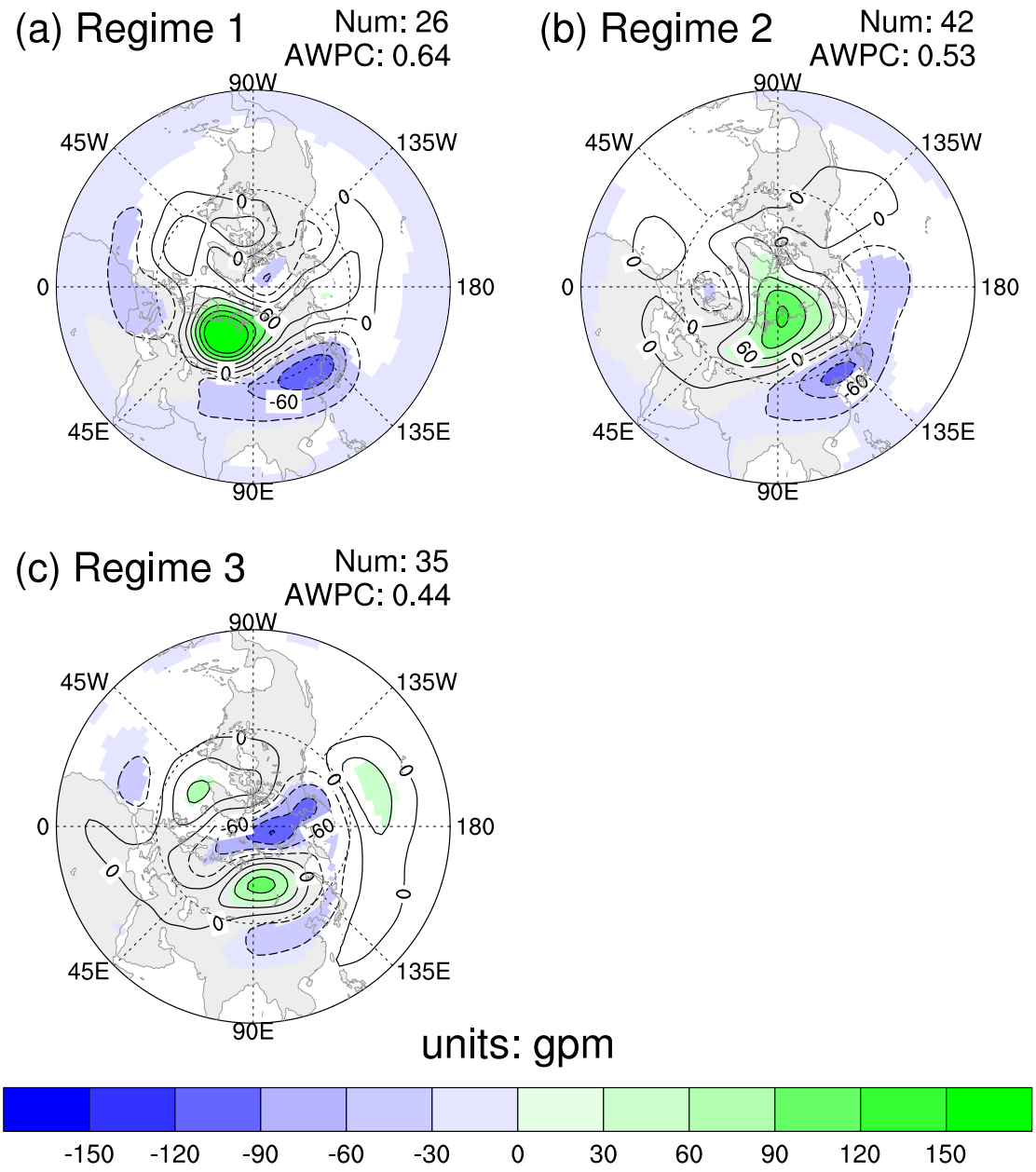

Fig. S3 Composite means of the HGT anomalies on the 500 hPa surface for the regimes identified by the distance-based *k*-means clustering method during DJFM 1948-2014 when the regime number is set to three. Shading indicates that the composite value meets the 95% confidence level based on two-tailed Student's *t* tests. The number of cold events belonging to each regime and the associated within-group AWPC are shown at the top right corner of each panel

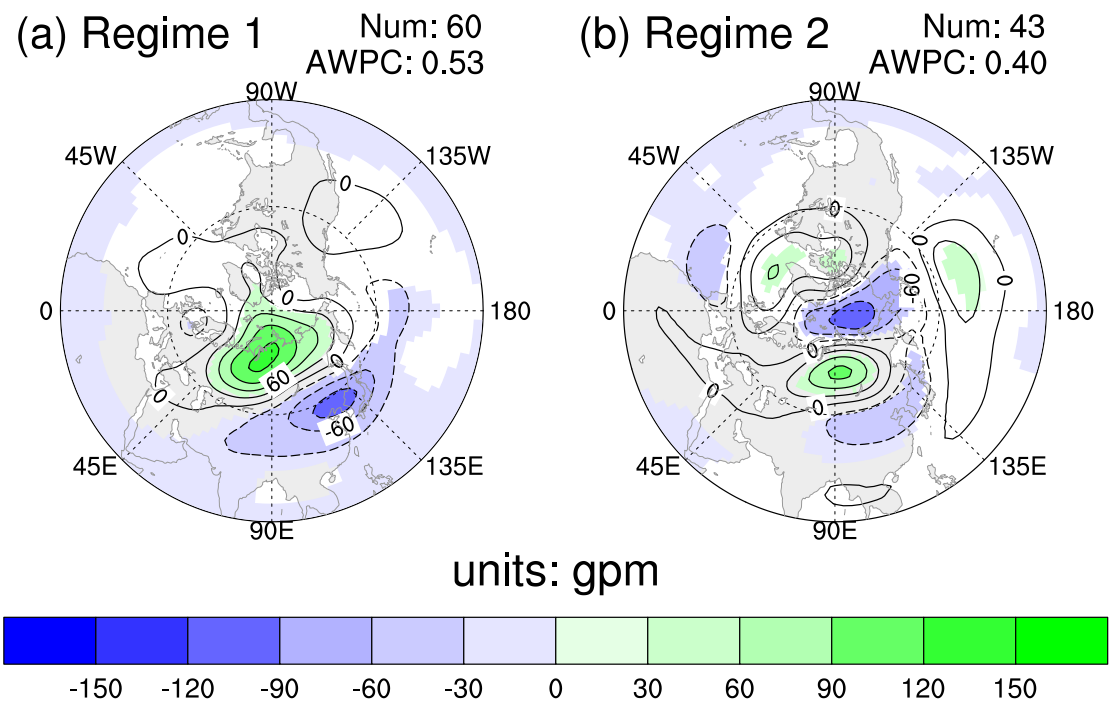

Fig. S4 Composite means of the HGT anomalies on the 500 hPa surface for the regimes identified by the distance-based *k*-means clustering method during DJFM 1948-2014 when the regime number is set to two. Shading indicates that the composite value meets the 95% confidence level based on two-tailed Student's *t* tests. The number of cold events belonging to each regime and the associated within-group AWPC are shown at the top right corner of each panel

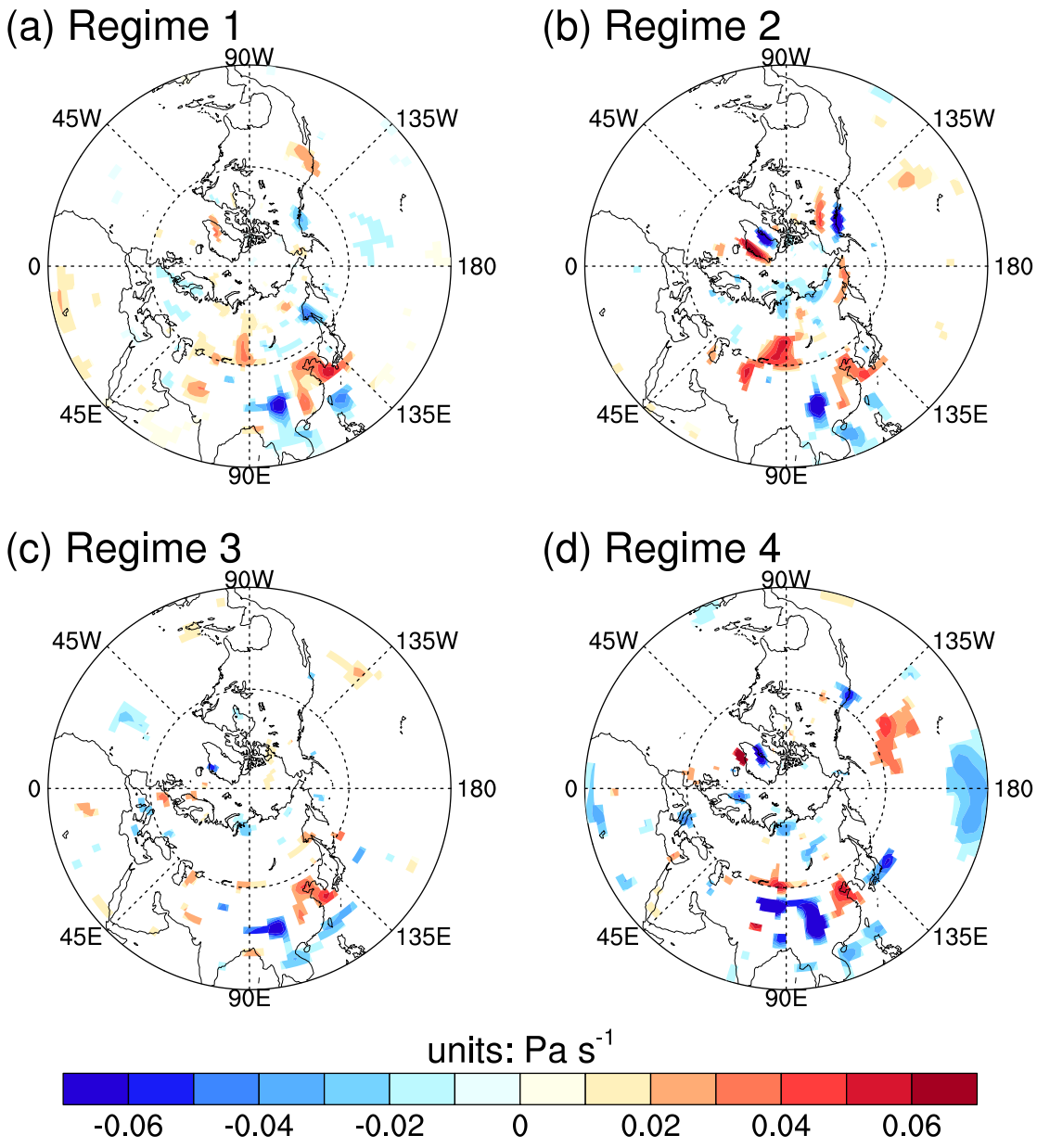

Fig. S5 Composite means of anomalies in vertical pressure velocity (units: Pa s^{-1}) on the 850 hPa surface for the four different regimes identified by the AWPC-based *k*-means clustering method during DJFM 1948-2014. Only composite values that meet the 95% confidence level based on two-tailed Student's *t* tests are plotted

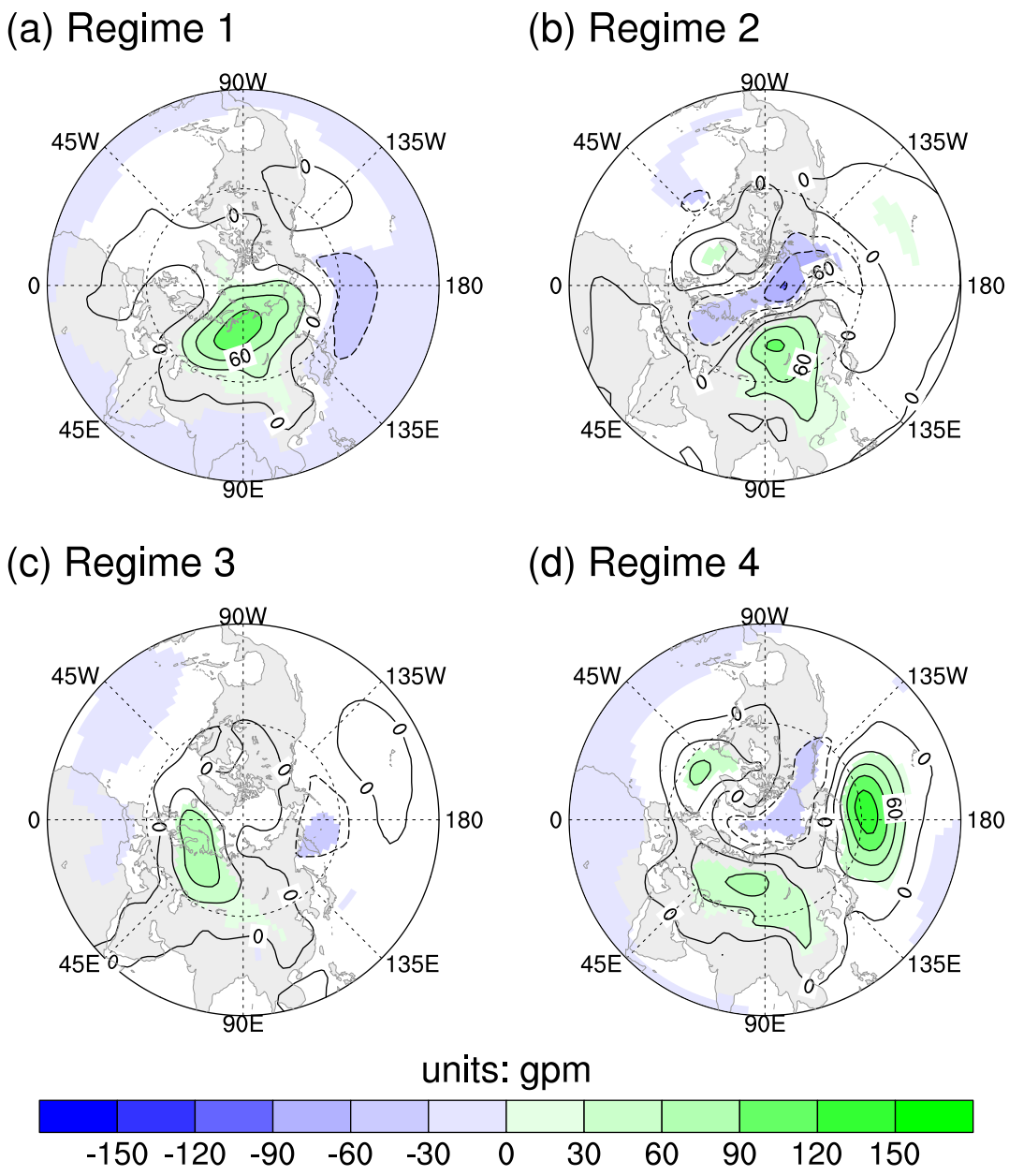

Fig. S6 Composite means of the HGT anomalies (units: gpm) on the 850 hPa surface for the four different regimes identified by the AWPC-based *k*-means clustering method during DJFM 1948-2014. Shading indicates that the composite value meets the 95% confidence level based on two-tailed Student's *t* tests

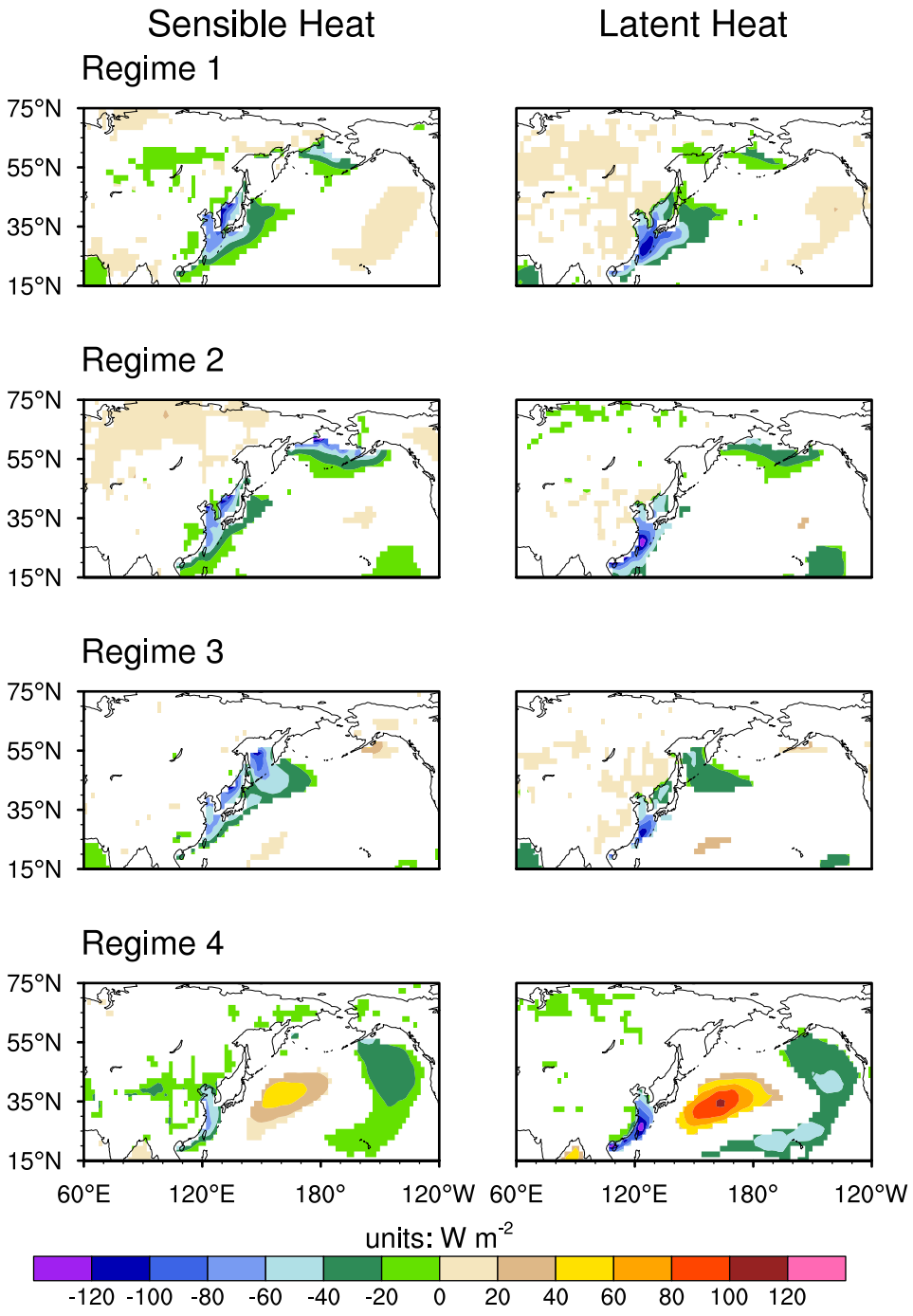

Fig. S7 Composite means of the anomalies in surface sensible and latent heat fluxes for the four different regimes identified by the AWPC-based *k*-means clustering method during DJFM 1948-2014. Only composite values that meet the 95% confidence level based on two-tailed Student's *t* tests are plotted