

S1 Wilcoxon rank-sum test to identify statistically significant cluster-mean differences

In order to derive statistically robust conclusions from cluster composites, a two-sided Wilcoxon rank-sum test (Wilks, 2011; p.159ff.) for each cluster pair and field was applied. Using a p-value of 0.05 regions can be identified where the clusters are statistically significantly different. The p-values can be used to identify where and when the differences in the clusters start to emerge in the ensemble forecast.

S1.1 P-values for Figure 6

Figure S1 shows p-values of QG ω and geopotential height on 850 hPa for clusters 2 and 3 during the amplification and propagation of the forecast uncertainty over the North Atlantic. The region and timesteps shown correspond to Fig. 7 (e-h). The results show that significant differences of QG ω are located in the region of strong QG ω on 1800 UTC 24 Sep 2018. Differences in geopotential height on 850 hPa are significant further to the east, but not in the center of the cyclone.

S1.2 P-values for Figure 8

Figure S2 shows p-values of PV on 325 K computed for the comparison of clusters 2 and 1, as well as clusters 3 and 1, during the diverging synoptic development over the Mediterranean. The region and timesteps shown correspond to Fig. 8. Differences are clearly significant in most parts of the Central Mediterranean. Figure S3 shows p-values of geopotential height at 850 hPa for the same clusters, regions and timesteps. Also for this variable, differences over the Mediterranean are clearly significant.

S1.3 P-values for Figure 9

Figure S4 shows p-values of accumulated precipitation from 1800 UTC 26 Sep to 0000 UTC 30 Sep 2018 for all cluster combinations. Differences are significant in large parts of the precipitation areas in the domain.

References

20 Wilks, D.S., 2011: Statistical methods in the atmospheric sciences, Third Edition. Elsevier Ltd. ISBN: 978-0-12-385022-5, 704p.

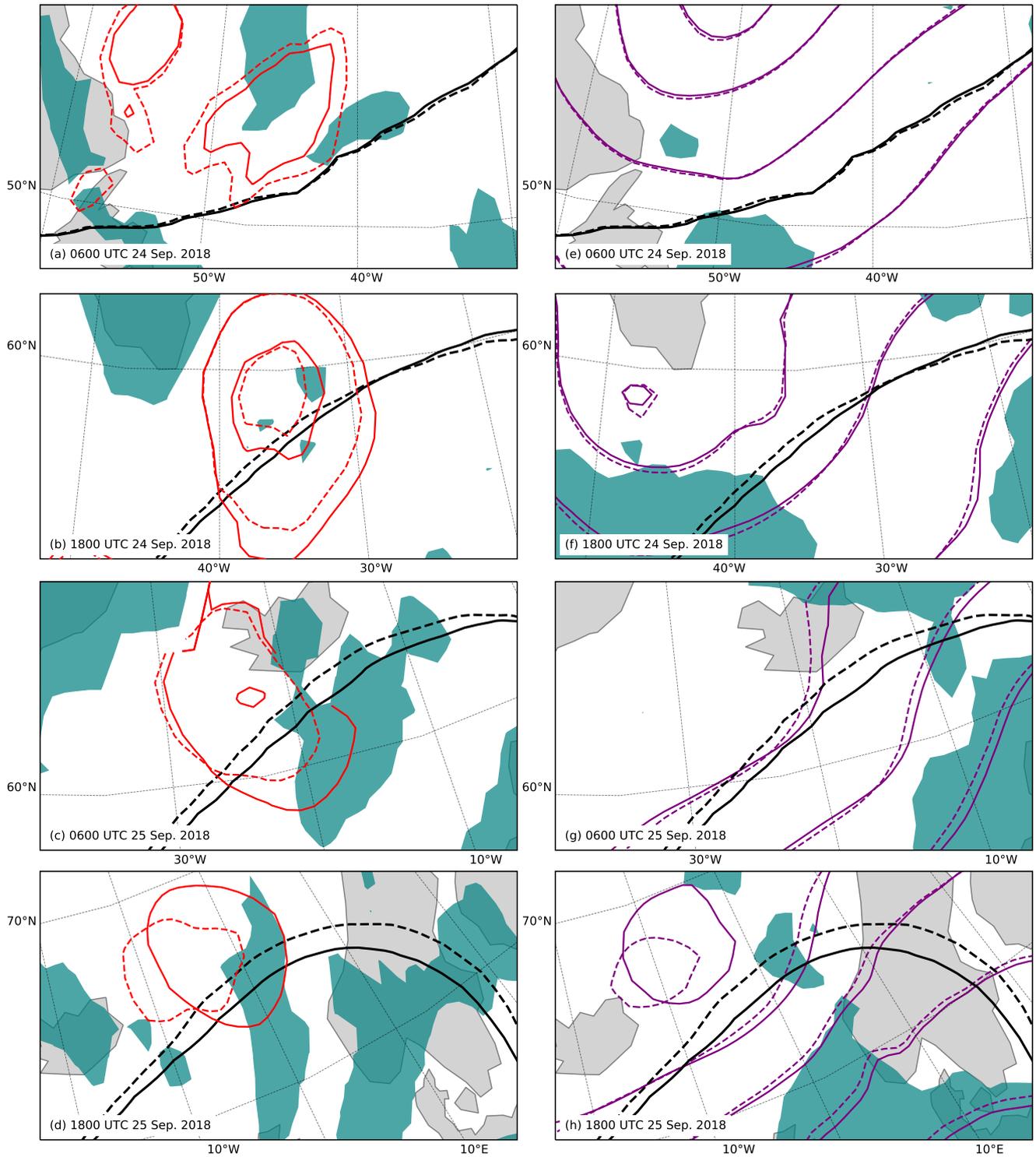


Figure S 1. (a-d) P-values of QG ω on 850 hPa (teal patches, ≤ 0.05) and QG ω (red contours; -0.01,-0.03 and -0.05 m/s) for cluster 3 (solid) and cluster 2 (dashed) and (e-h) p-values of geopotential height on 850 hPa (teal patches, ≤ 0.05) and geopotential height (purple contours; every 10 gpdm) for cluster 3 (solid) and cluster 2 (dashed) for the same timesteps and regions as Fig.7 e-h. 2-PVU contours on 325 K are shown in black (cluster 3: solid, cluster 2: dashed).

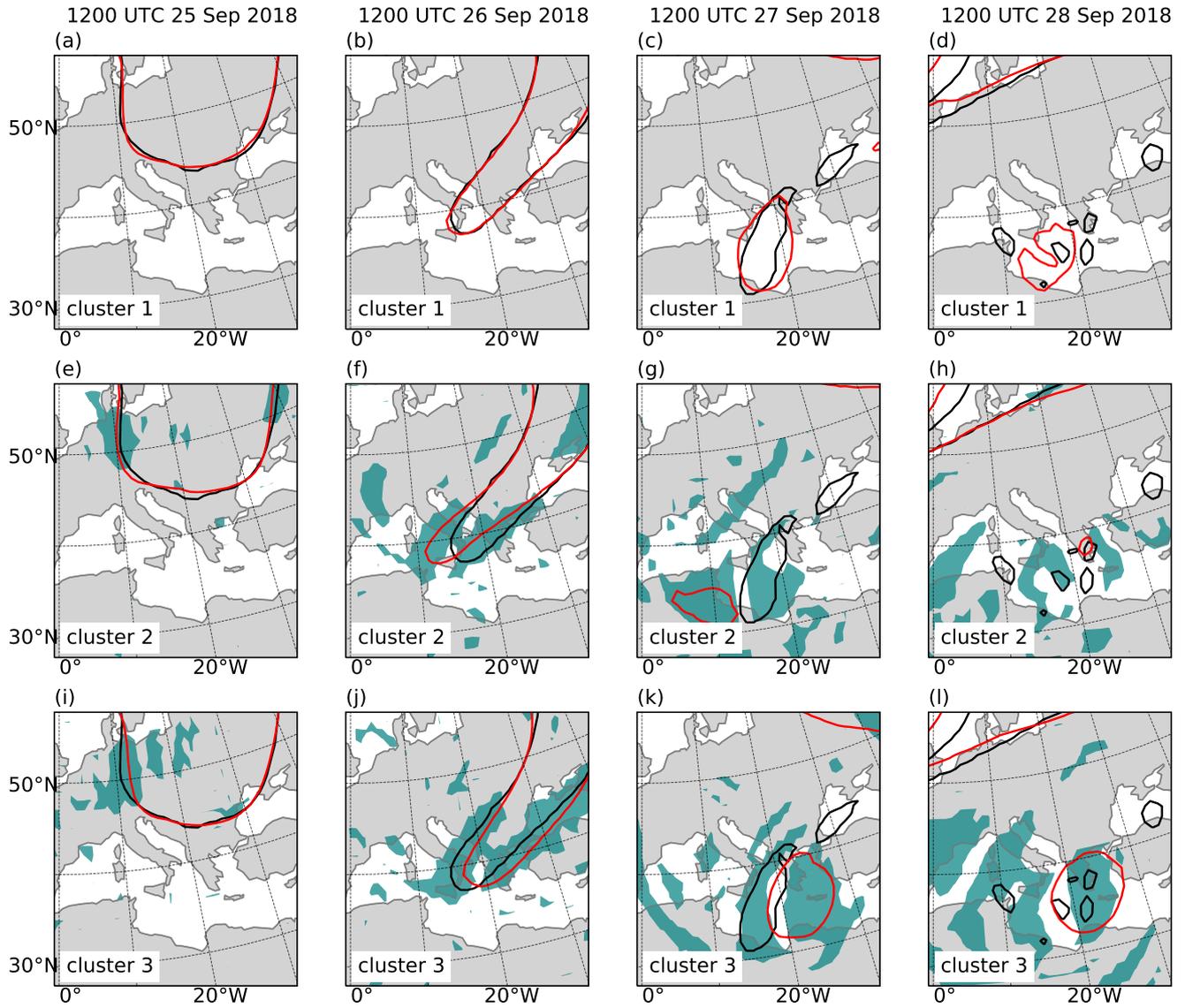


Figure S 2. P-values of PV on 325 K (teal patches, ≤ 0.05) resulting from comparison with cluster 1, cluster-mean 2-PVU contour on 325 K (red) and analysis 2-PVU contour on 325 K (black contour) for the same timesteps and regions as Fig. 8.

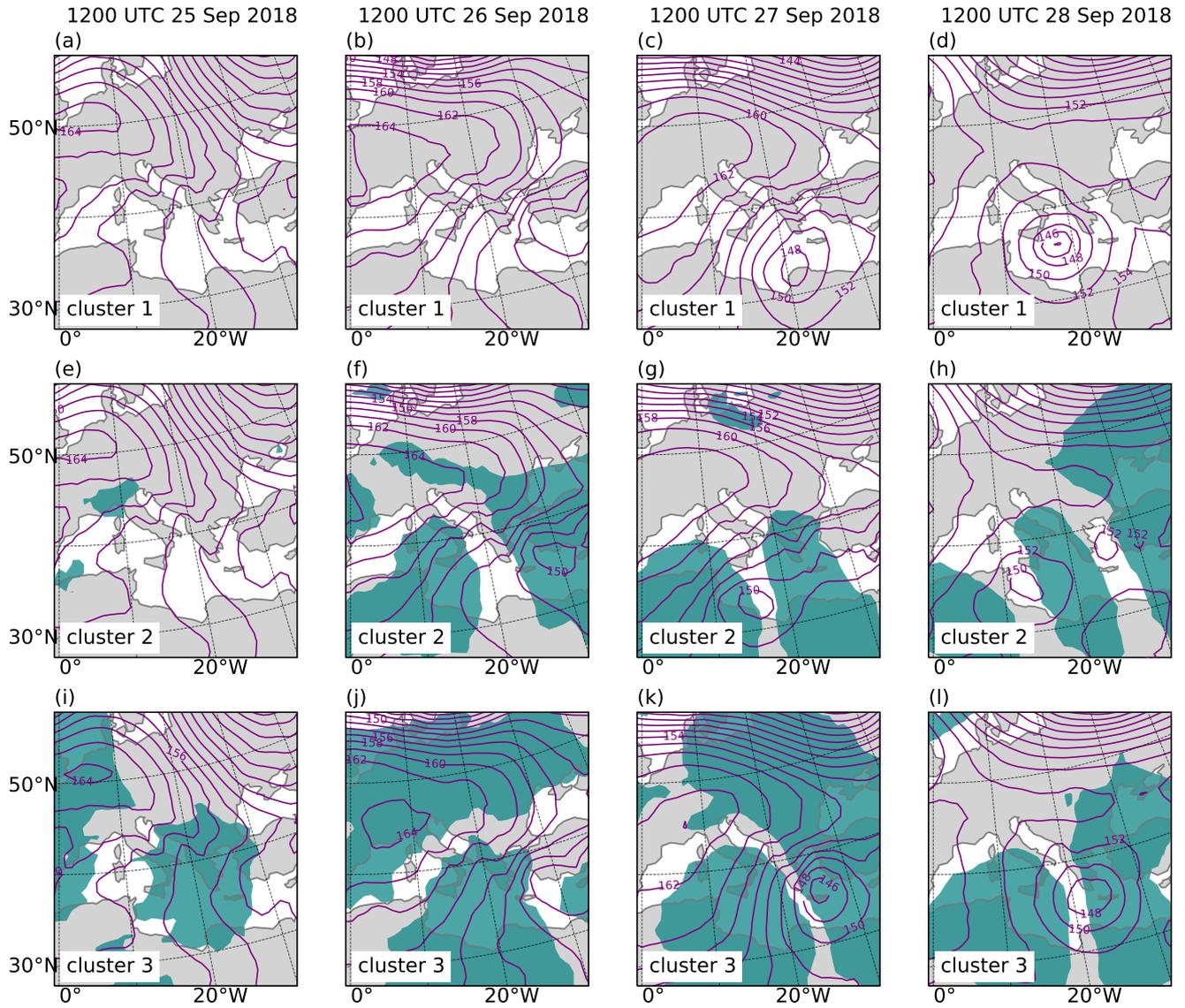


Figure S 3. P-values of geopotential height on 850 hPa (teal patches, ≤ 0.05) resulting from comparison with cluster 1 and cluster-mean geopotential height on 850 hPa (purple) for the same timesteps and regions as Fig. 8.

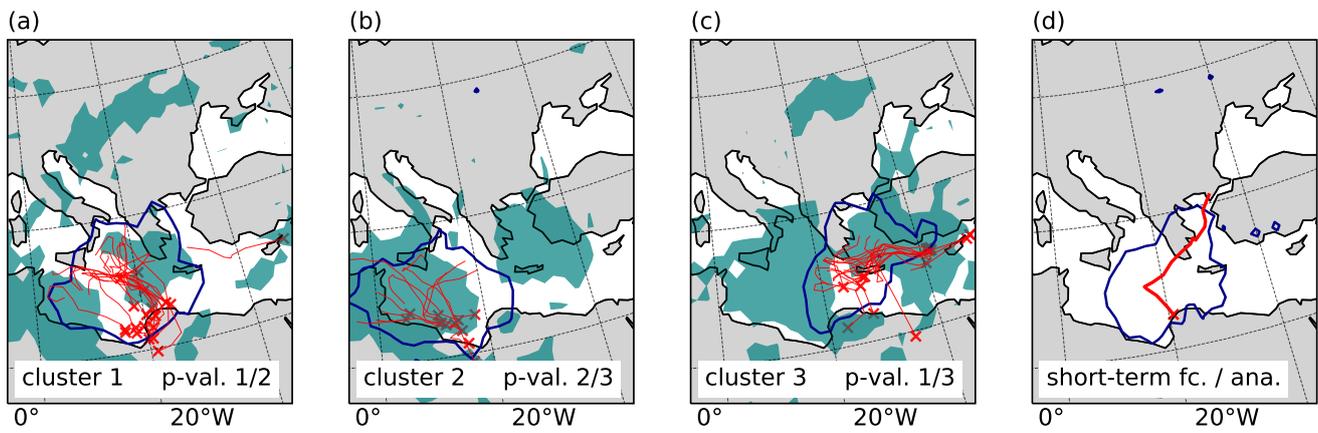


Figure S 4. P-values of accumulated precipitation from 1800 UTC 26 Sep to 0000 UTC 30 Sep 2018 for cluster-means of (a) clusters 1 and 2, (b) clusters 2 and 3, and (c) clusters 1 and 3. For the remaining parts of the Figure see Fig. 9.