Characterizing Drought Behavior using Unsupervised Machine Learning for Improved Understanding of Future Drought in the Colorado River Basin

Carl J. Talsma[1], Katrina E. Bennett[1], Velimir V. Vesselinov[1]

[1] Earth and Environmental Sciences, Los Alamos National Laboratory, Los Alamos, NM, 87545.

Corresponding authors: Carl J. Talsma(talsmac83@lanl.gov), Velimir V. Vesselinov (vvv@lanl.gov)

Figure 1: The domain of the Colorado River Basin with adjacent areas that receive Colorado River water. Adapted from USGS, 2012 (accessed Jan 11th, 2021; USBR, 2012).

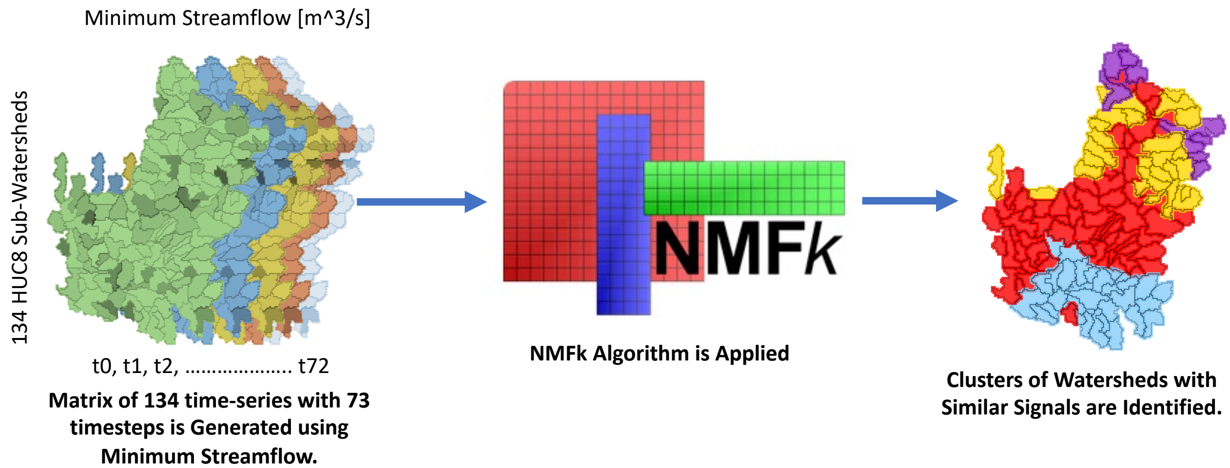


Figure 2: Process by which the NMFk algorithm is applied to the drough indicator data. A 2d matrix of minimum streamflow (*qn*) is created using the 134 HUC8 sub-watersheds with 73 5-day timesteps of streamflow throughout a year. This matrix is input into NMFk which clusters similar temporal signals of *qn* together.

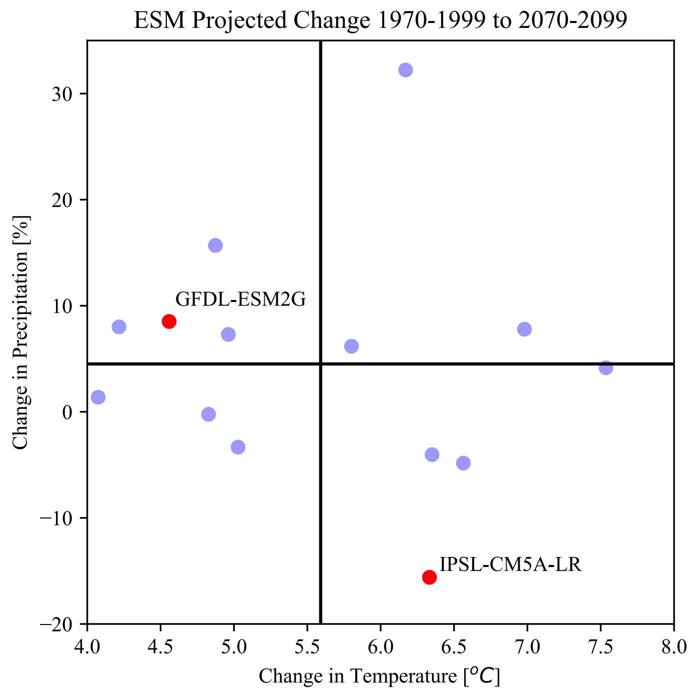


Figure 3: Average annual Precipitation changes (%) plotted against temperature changes (°C) for the CRB region for 14 different ESM's. GFDL-ESM2G and IPSL-CM5A-LR models are highlighted in red. Of the 14 ESMs, six were used in our analysis to cover the range of ESM results for precipitation and temperature change. Those six models are presented in the supplementary materials with the two highlighted models (GFDL-ESM2G and IPSL-CM5A-LR) discussed ind detail here. The vertical and horizontal black lines represent the multi-model mean of projected temperature and precipitation change, respectively.

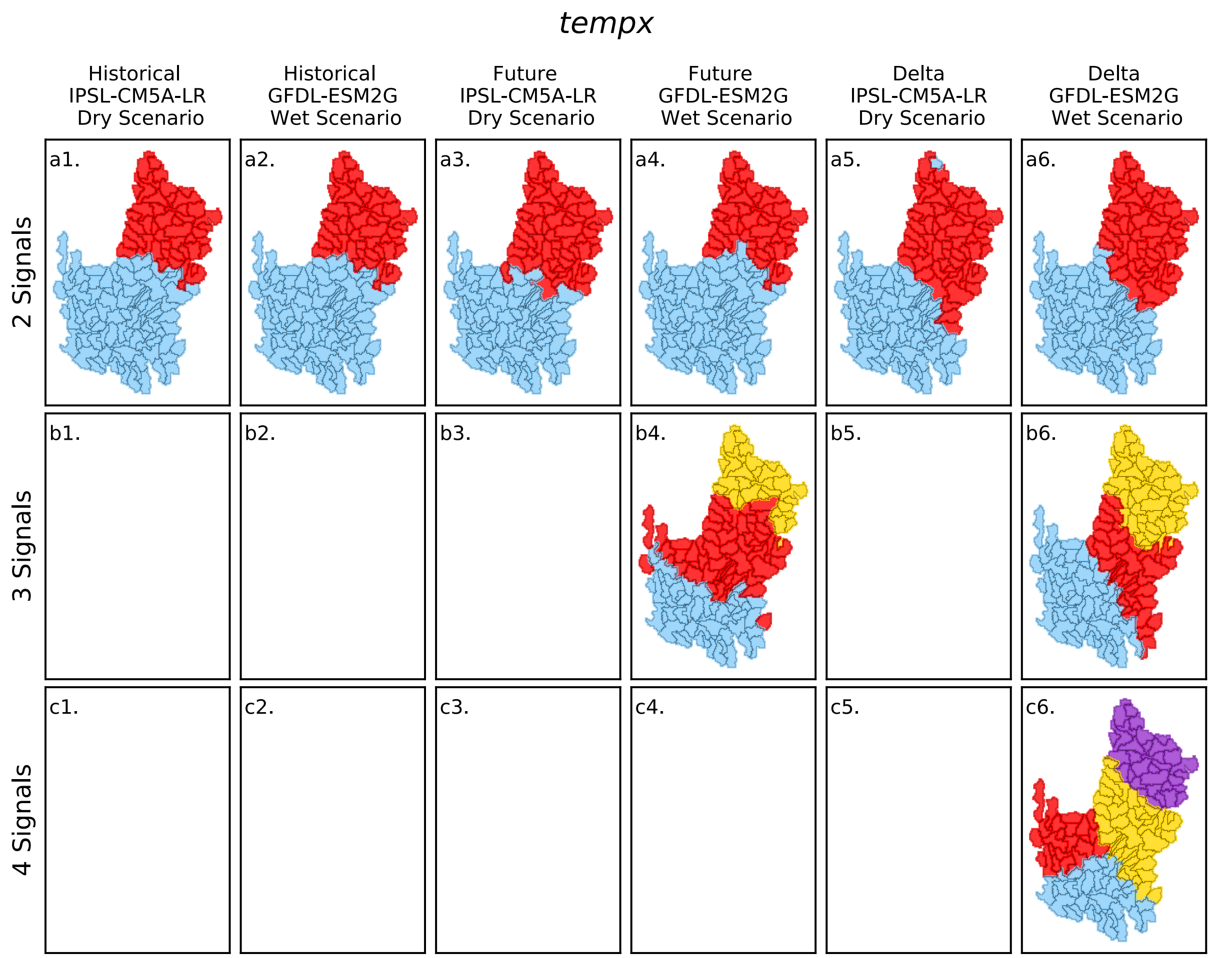


Figure 4: NFMk spatial grouping of HUC8 subsub-watersheds based on *tempx* dataset using solutions for 2, 3, and 4 extracted signals. The historical and future time periods, as well as the delta, are shown for both wet and dry scenarios. Each panel represents an independent NMFk clustering and the colors shown are not meaningful to one another across panels. Blank panels represent cases for each NMFk could not produce an acceptable solution.

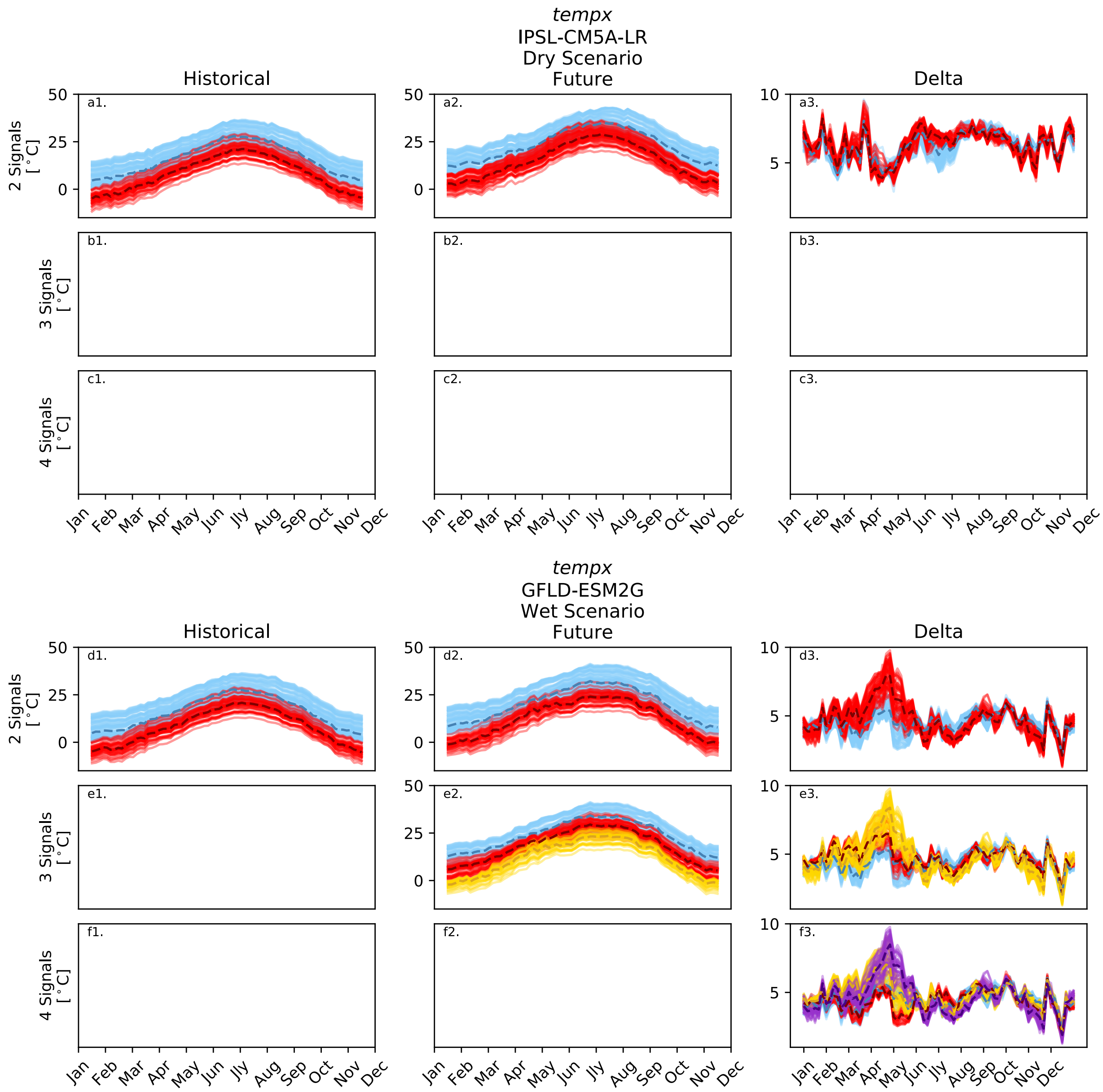


Figure 5: Temporal NFMk clustering of HUC8 subsub-watersheds based on the annual *tempx* signals for both IPSL-CM5A-LR (dry scenario) and GFDL-ESM2G (wet scenario) simulations. Solutions for 2, 3, and 4 extracted signals are presented for each time period. The clustering on this figure corresponds directly to the spatial clustering in the appropriate panels of Figure 3. Each line represents a single sub-watershed, while the dashed lines are representing the cluster medians at each time-step.

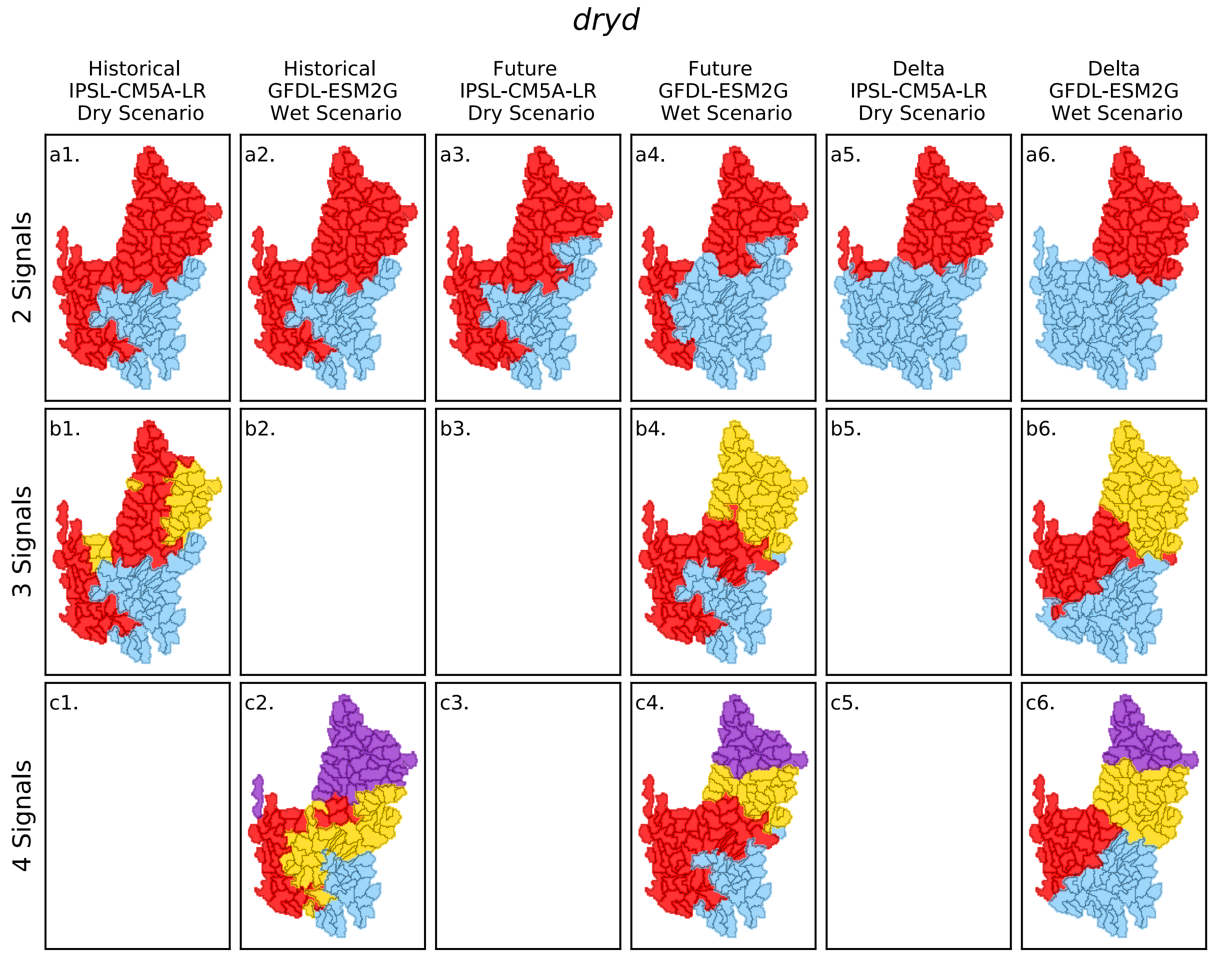


Figure 6: NFMk spatial grouping of HUC8 subsub-watersheds based on *dryd* dataset using solutions for 2, 3, and 4 extracted signals. The historical and future time periods, as well as the delta, are shown for both wet and dry scenarios. Each panel represents an independent NMFk clustering and the colors shown are not meaningful to one another across panels. Blank panels represent cases for each NMFk could not produce an acceptable solution.

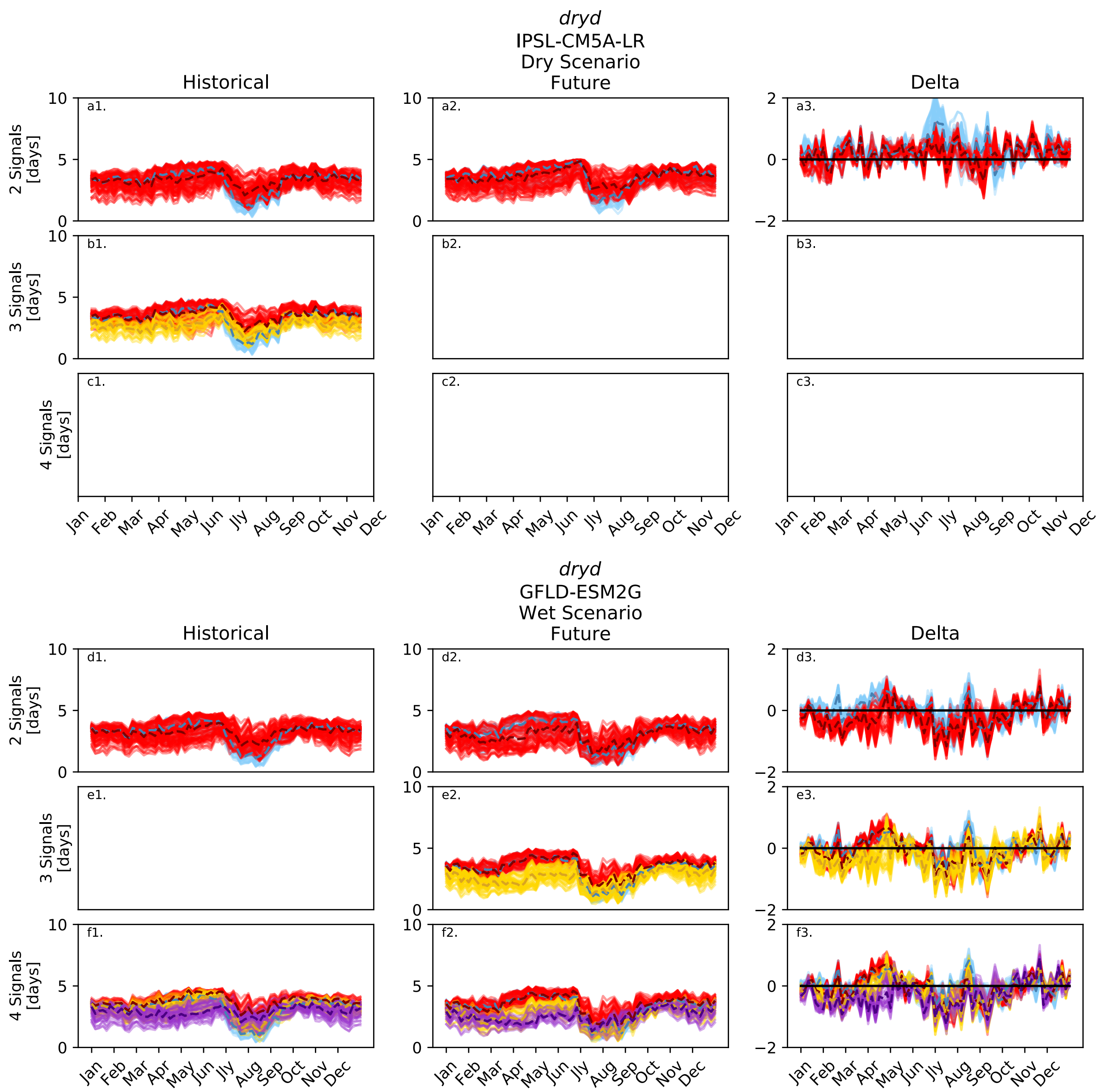


Figure 7: Temporal NFMk clustering of HUC8 subsub-watersheds based on the annual *dryd* signals for both IPSL-CM5A-LR (dry scenario) and GFDL-ESM2G (wet scenario) simulations. Solutions for 2, 3, and 4 extracted signals are presented for each time period. The clustering on this figure corresponds directly to the spatial clustering in the appropriate panels of Figure 3. Each line represents a single sub-watershed, while the dashed lines are representing the cluster medians at each time-step.

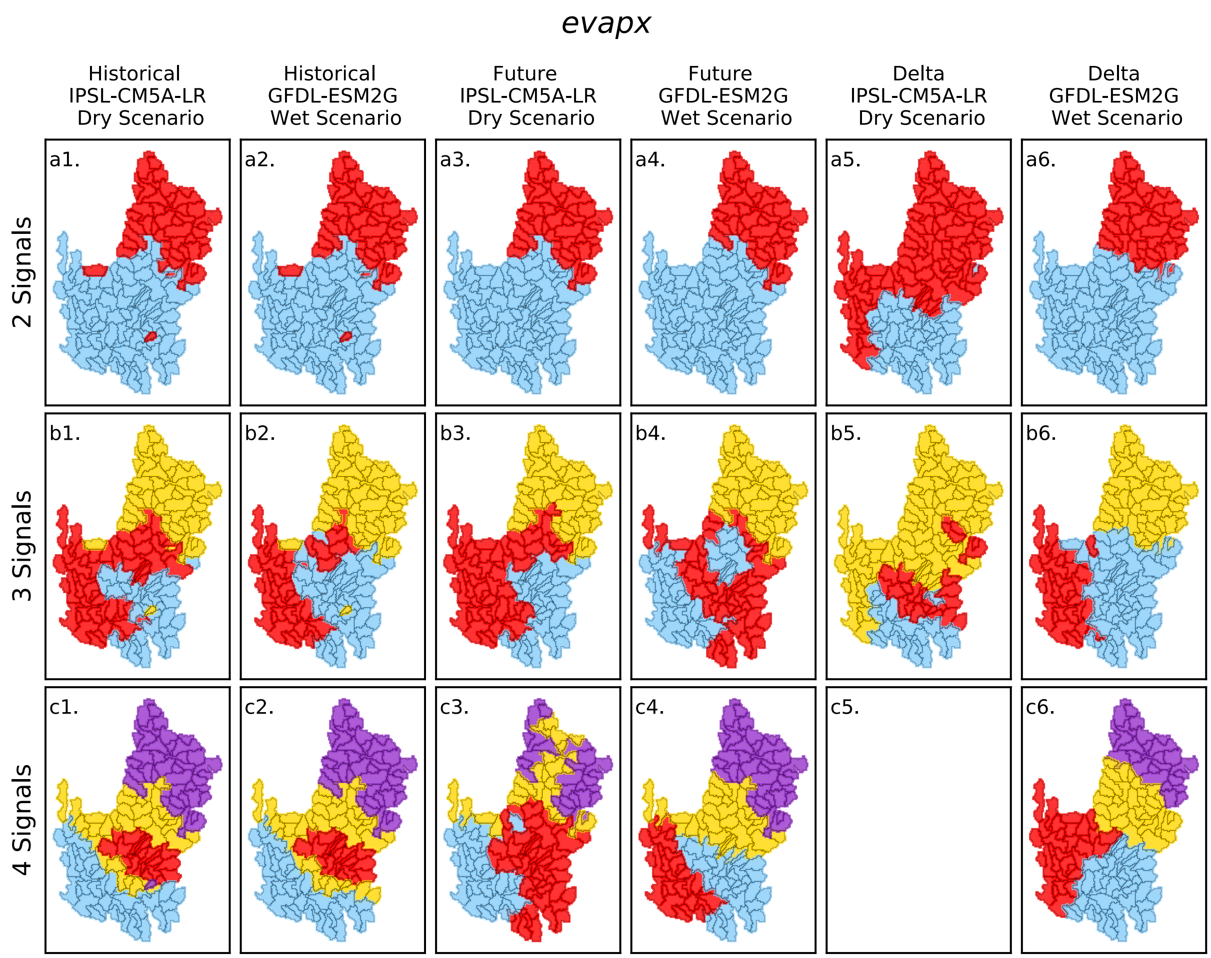


Figure 8: NFMk spatial grouping of HUC8 subsub-watersheds based on *evapx* dataset using solutions for 2, 3, and 4 extracted signals. The historical and future time periods, as well as the delta, are shown for both wet and dry scenarios. Each panel represents an independent NMFk clustering and the colors shown are not meaningful to one another across panels. Blank panels represent cases for each NMFk could not produce an acceptable solution.

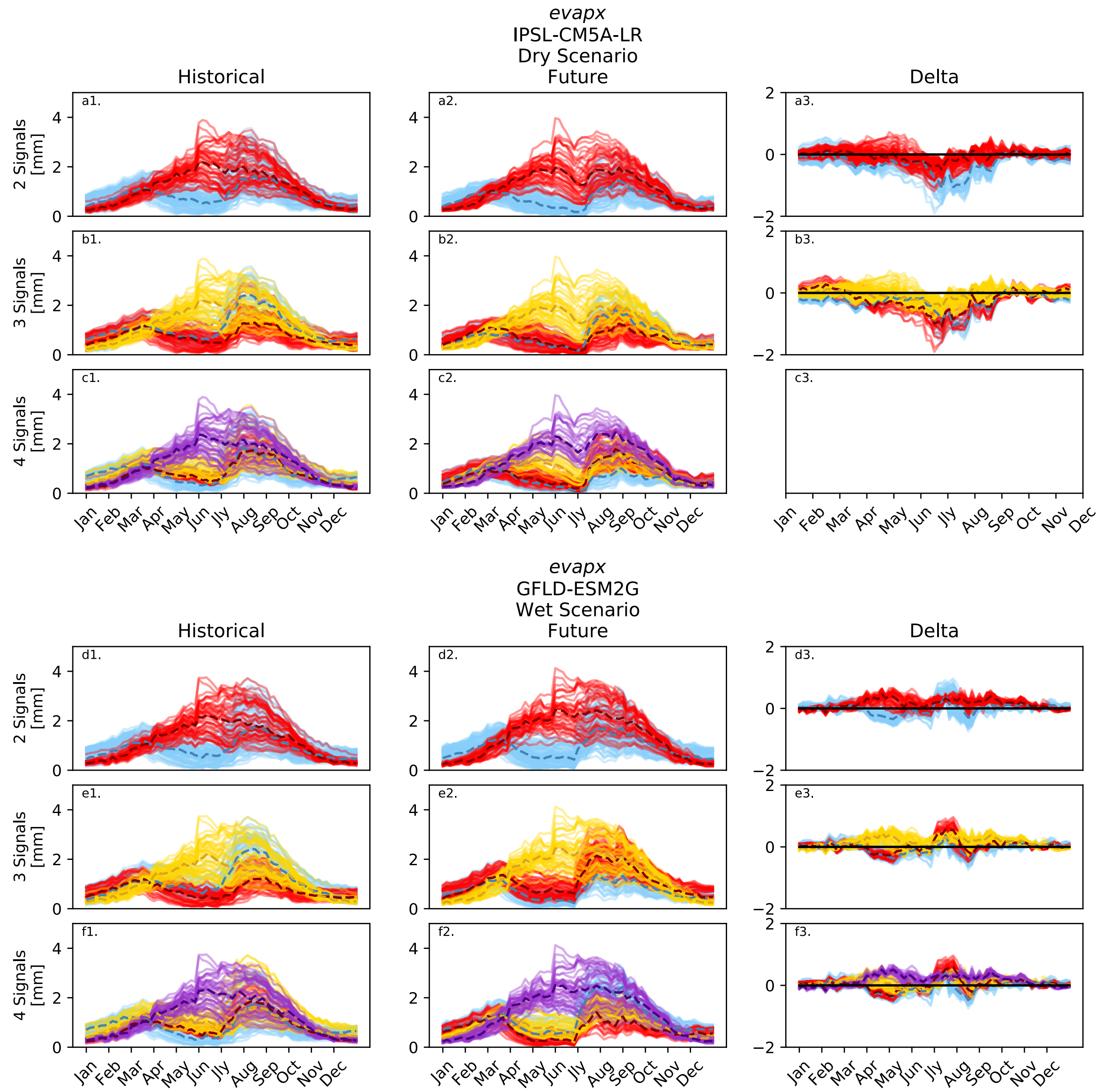
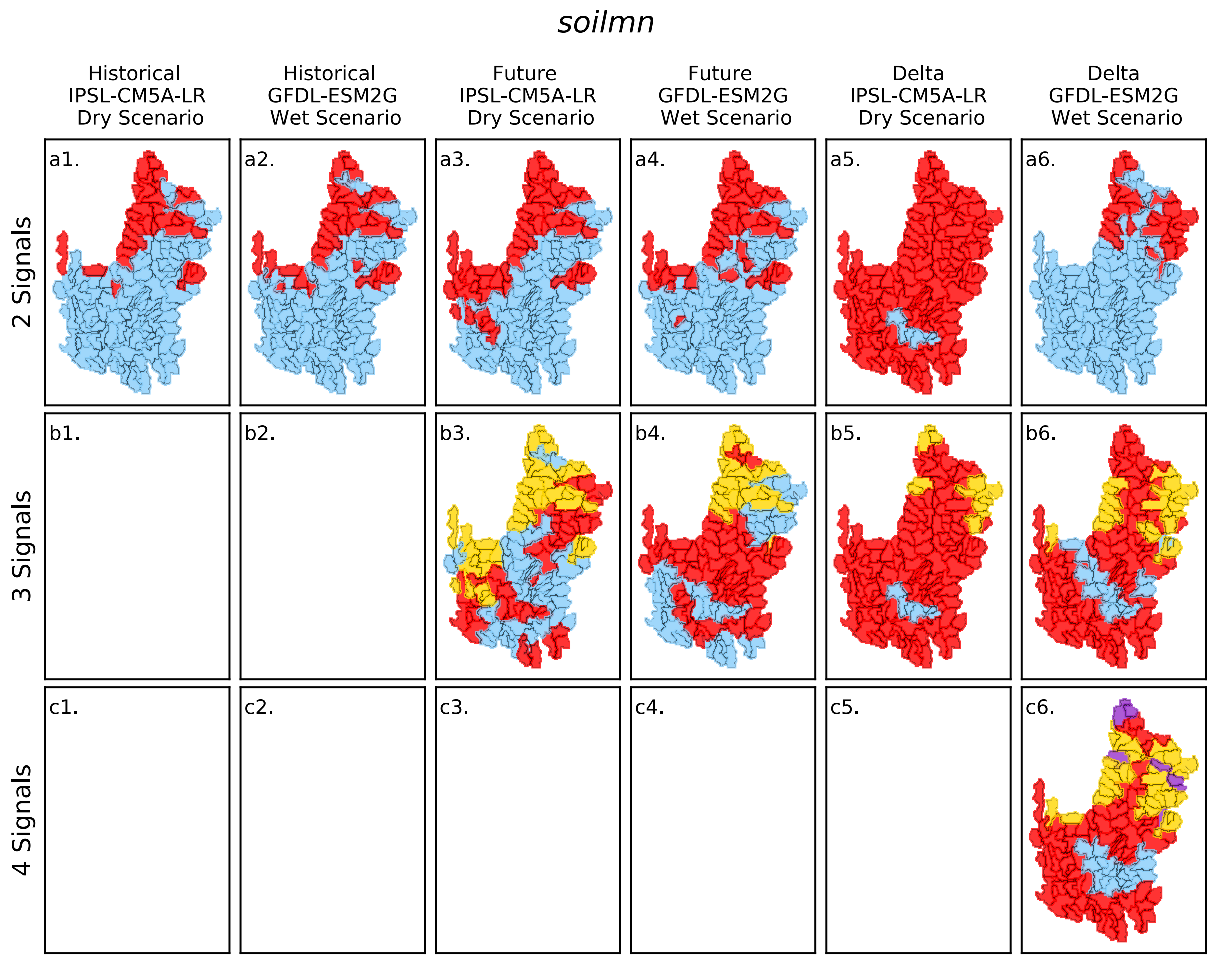


Figure 9: Temporal NFMk clustering of HUC8 subsub-watersheds based on the annual *evapx* signals for both IPSL-CM5A-LR (dry scenario) and GFDL-ESM2G (wet scenario) simulations. Solutions for 2, 3, and 4 extracted signals are presented for each time period. The clustering on this figure corresponds directly to the spatial clustering in the appropriate panels of Figure 3. Each line represents a single sub-watershed, while the dashed lines are representing the cluster medians at each time-step.



83
84
85 Figure 10: NFMk spatial grouping of HUC8 subsub-watersheds based on *soilmn* dataset using solutions for 2, 3, and
86 4 extracted signals. The historical and future time periods, as well as the delta, are shown for both wet and dry
87 scenarios. Each panel represents an independent NMFk clustering and the colors shown are not meaningful to one
88 another across panels. Blank panels represent cases for each NMFk could not produce an acceptable solution..
89

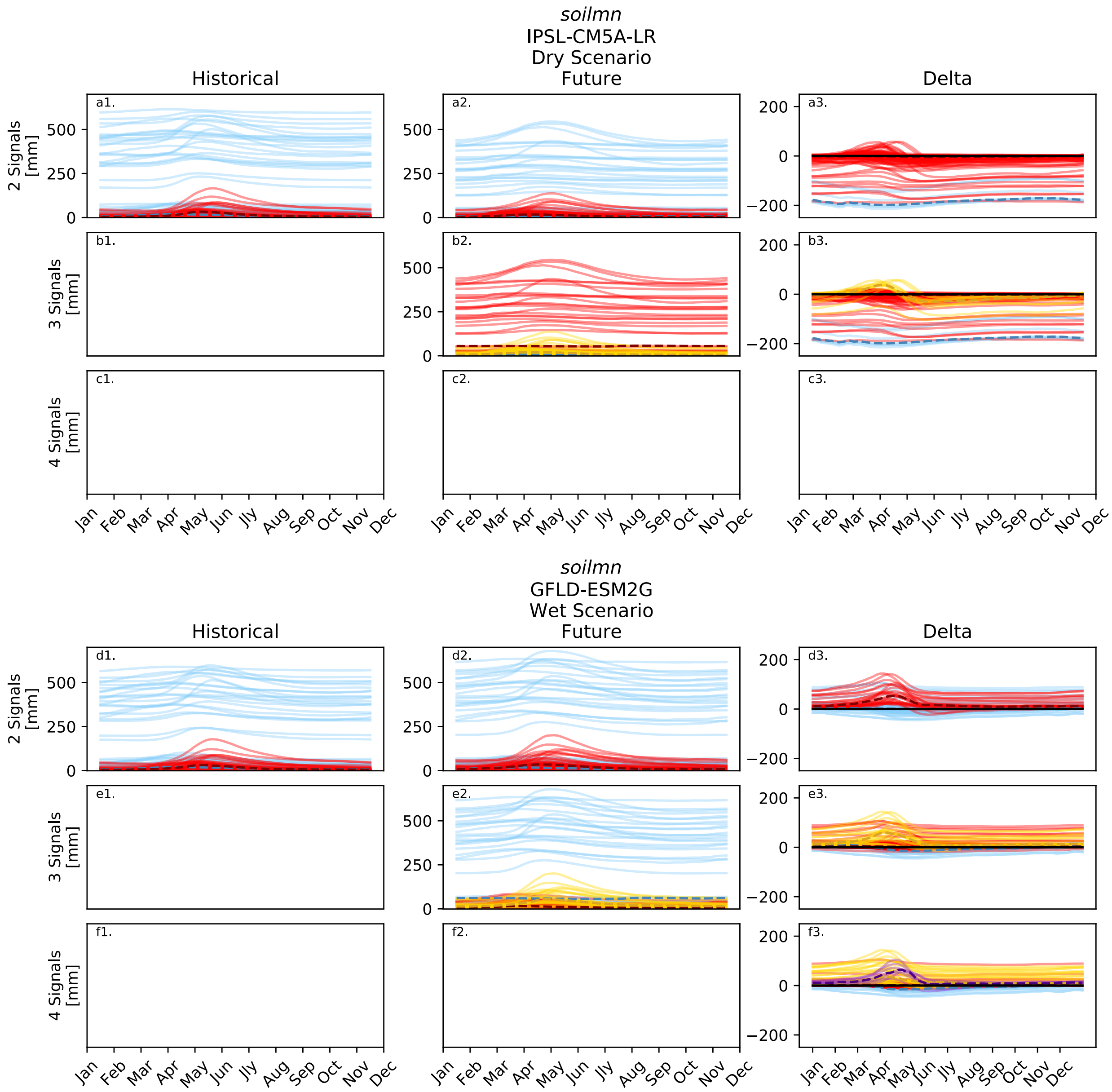


Figure 11: Temporal NFMk clustering of HUC8 subsub-watersheds based on the annual *soilmn* signals for both IPSL-CM5A-LR (dry scenario) and GFDL-ESM2G (wet scenario) simulations. Solutions for 2, 3, and 4 extracted signals are presented for each time period. The clustering on this figure corresponds directly to the spatial clustering in the appropriate panels of Figure 3. Each line represents a single sub-watershed, while the dashed lines are representing the cluster medians at each time-step.

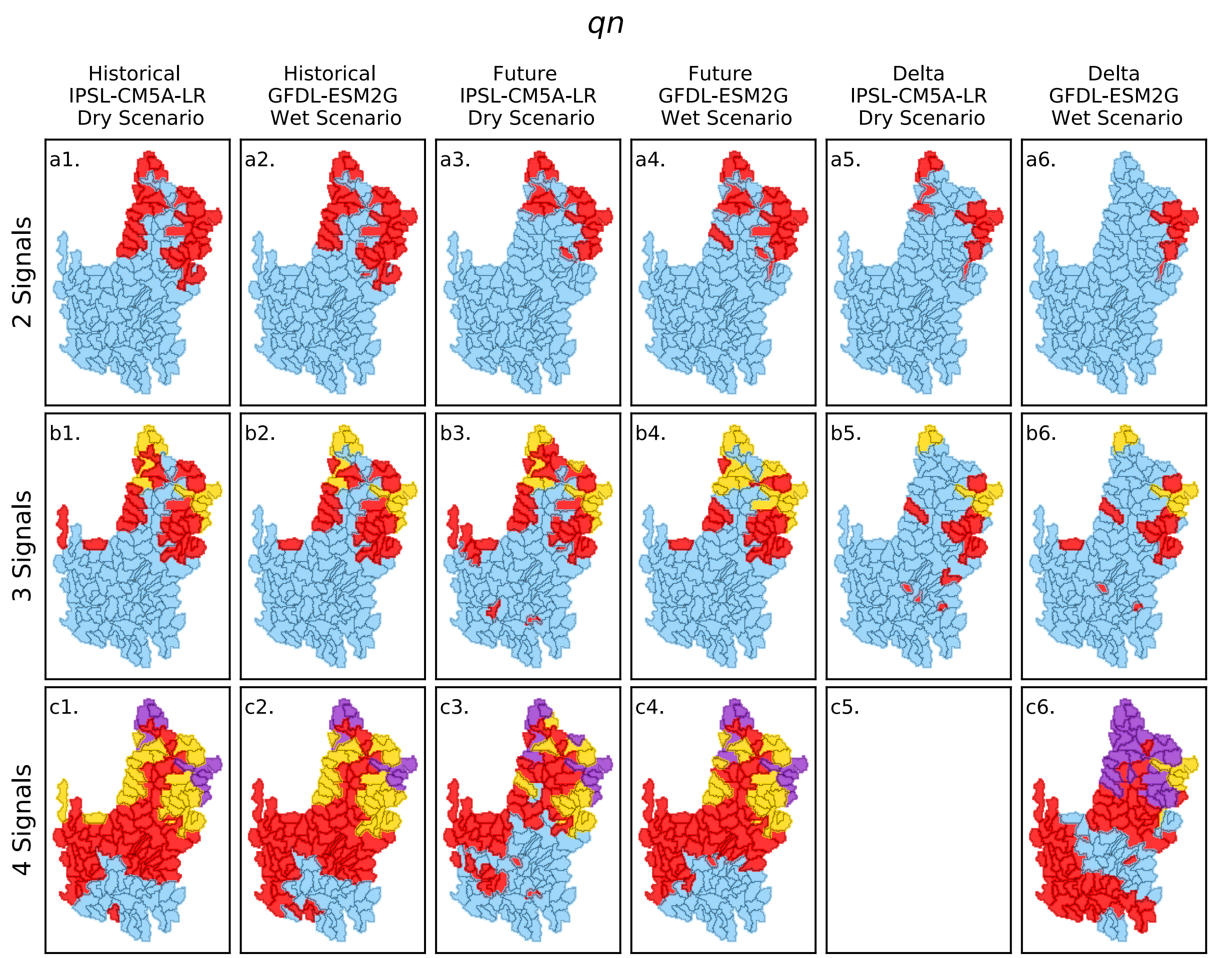


Figure 12: NFMk spatial grouping of HUC8 subsub-watersheds based on *qn* dataset using solutions for 2, 3, and 4 extracted signals. The historical and future time periods, as well as the delta, are shown for both wet and dry scenarios. Each panel represents an independent NMFk clustering and the colors shown are not meaningful to one another across panels. Blank panels represent cases for each NMFk could not produce an acceptable solution.

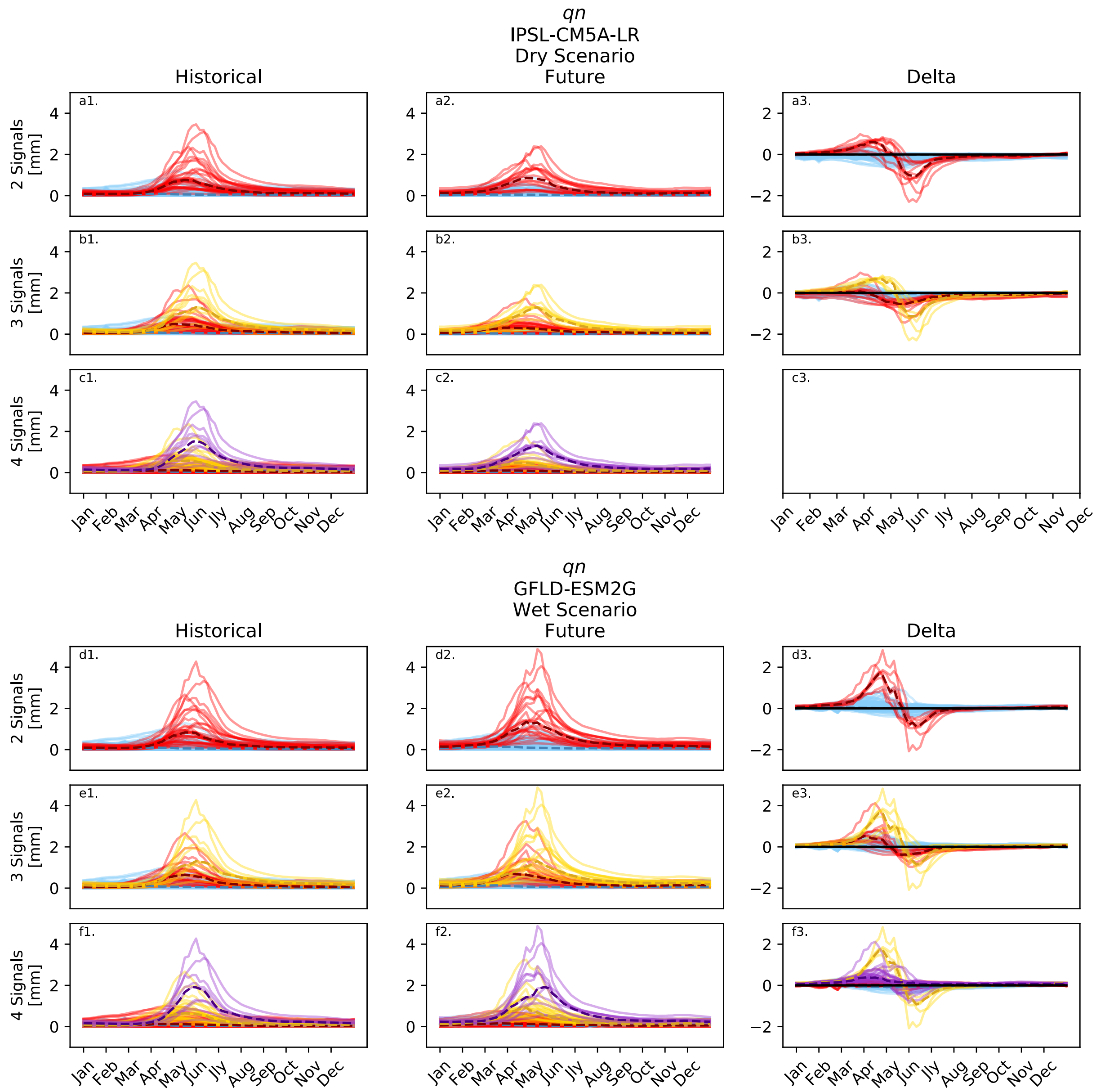


Figure 13: Temporal NFMk clustering of HUC8 subsub-watersheds based on the annual *qn* signals for both IPSL-CM5A-LR (dry scenario) and GFDL-ESM2G (wet scenario) simulations. Solutions for 2, 3, and 4 extracted signals are presented for each time period. The clustering on this figure corresponds directly to the spatial clustering in the appropriate panels of Figure 3. Each line represents a single sub-watershed, while the dashed lines are representing the cluster medians at each time-step.

	ΔT (℃)	ΔP (%)
IPSL-CM5A-LR	**6.33**	**-15.60**
HadGEM2-ES365	6.35	-4.04
MPI-ESM-LR	5.03	-3.33
GFDL-ESM2M	4.07	1.38
MIROC-ESM	6.98	7.79
GFDL-ESM2G	**4.56**	**8.51**

Table 1: Projected change in mean annual temperature and precipitation in the CRB simulated using the six ESM models used in this study. IPSL-CM5A-lR and GFDL-ESM2G are in bold and are discussed in detail in this study while the other models are presented in the supplementary materials.