This work examines the utility of probabilistic seasonal forecasts from the fifth generation ECMWF system combined with the Canadian FWI index for fire season forecasts over Greece, with a focus on the Attica region. The results are potentially of high value, given that this region is prone to regular fires. The general approach makes sense and the results are analysed using good quality standard assessment methods which give consistent results.

I have two main points, which relate to potentially improving forecast skill, rather than the quality of the study per se:

1) I'm not sure how the Greek fire service plans resource allocation, but, rather than attempting an aggregate forecast for the entire fire season, would it not be useful to, say, divide the fire season in two, and give forecasts for each half separately (e.g. for may-july initialised in march/april; and for july-sep initialised in may/june). This would allow forecasts with shorter lead times, which should in turn improve skill.

2) Related to this: the question of why the forecast skill seems to be so low for the longer-timescale components of the FWI system (those for the denser fuels). I guess this arises from two things: if I understand correctly, the authors do not use observations to spinup the FWI system. Since the BUI and DC have spinup timescales of the order of 15 and 50 days, so initialisation with obs would surely give some additional predictability for the latter in particular. This would be more relevant if my suggestion 1 is implemented.

Answer:

Comments 1 and 2 are somewhat related, thus, the experiments described below have been performed in order to examine whether the division of the May-June-July-August-September (MJJAS) period in two sub-periods, May-June-July (MJJ) and July-August-September (JAS) improves the predictability of FWI and/or its subcomponents:

- SEAS5 forecasts initialized in March and April for MJJ (two month and one month in advance of the target fire season, respectively) with no spin up and spin up performed from both SEAS5 and ERA5-Land data

- SEAS5 forecasts initialized in May and June for JAS (two month and one month in advance of the target fire season, respectively) with no spin up and spin up performed from both SEAS5 and ERA5-Land data

In all experiments bias correction is applied using daily data for the period of interest (MJJ, JAS) using a moving window width of 31-days, following the methodology of the first version of the manuscript. In addition, the results of the statistically downscaled temperature, relative humidity, wind speed and precipitation for each period are also presented. It should be noted that in the revised version the computations were performed and the respective maps/plots were constructed only for the Attica region in order to minimize the computational cost.

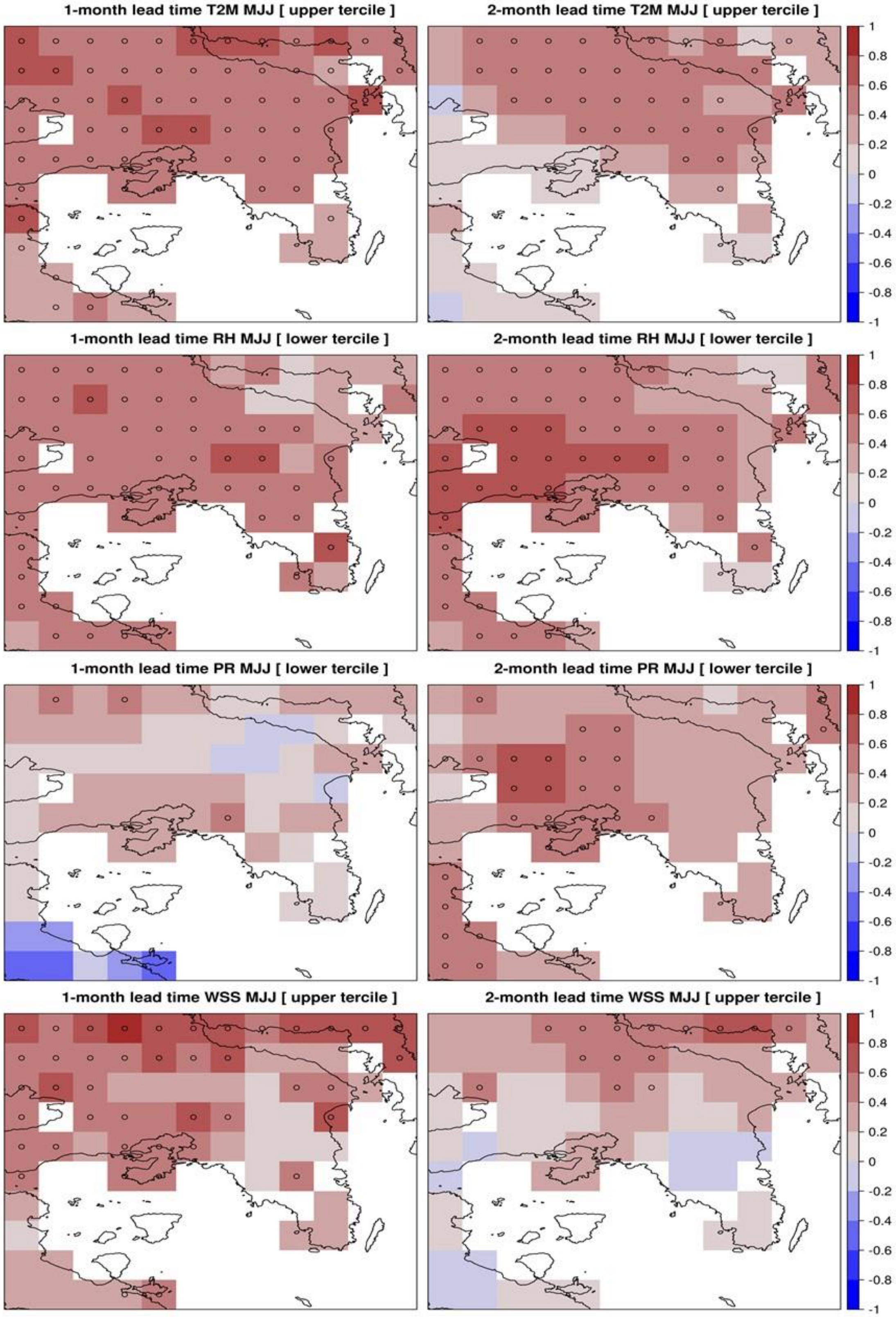
Regarding MJJ, the ROCSS for the upper tercile category of air temperature and wind speed, as well as the lower tercile category of relative humidity and total precipitation for the 1-

month and 2-month lead times are presented in Fig. R1. From Fig. R1 it is evident that forecasts initialized in April (1-month lead time) exhibit higher discrimination skill for temperature and wind speed, while forecasts initialized in March (2-month lead time) exhibit higher discrimination skill for relative humidity and precipitation. As far as the FWI is concerned, from the tercile plots (Fig. R2) is evident that 2-month lead time forecasts exhibit higher skill than the corresponding 1-month lead time. The highest upper tercile ROCSS, 0.7 is found for the 2-month lead time forecasts and when using ERA5-Land for the spin-up, classified to perfect reliability (according to the new tercile plots-not shown) while the rest of the experiments are classified as marginally useful (not shown).

Longer spin-up periods and implanting ERA5-Land in the spin-up, increases the DC upper tercile ROCSS compared to 1-month lead time spin-up (Fig. R3) with the predictions classified as marginally useful (not shown). In contrast, the highest BUI upper tercile ROCSS is found for the 2-month lead time forecast with spin-up from SEAS5 (Fig. R4) with the predictions classified as marginally useful (not shown). Concerning DMC, the highest upper tercile ROCSS is found for the 2-month lead time forecasts without spin-up (0.58) and with spin-up (0.49) when using SEAS5 data, while the predictions are classified as marginally useful (not shown). Finally, ISI results indicate that the specific sub-component is insensitive to the spin-up and its results are mostly controlled by the ROCSS of the meteorological variables used for the FWI calculations from the different lead time forecasts. In particular, from Fig. R6 it is evident that the highest ISI upper tercile ROCSS, 0.83 (perfect reliability), is found for all 1-month lead time experiments (with and without spin up) while lower ROCSS, 0.66 (marginally useful), is found for the corresponding 2-month lead time experiments. The highest ROCSSs are attributed to the higher discrimination power of temperature and wind speed over the area under study in the 1-month lead time forecasts.

For the JAS period the results are not found as encouraging as found in MJJ. In particular, from Figure R7 it is evident that both forecasts initialized in June (1-month lead time) and May (2-month lead time) exhibit very low discrimination skill for the majority of the meteorological variables used to drive the FWI calculations with the exception of high discrimination skill shown for relative humidity for 2-month lead forecasts and the relatively low skill shown for wind speed for both forecasts. As a result, poorer skill is found for FWI and ISI in JAS (Figs. R8 and R9, respectively) compared to MJJ, with the predictions for the 1-month lead time forecast characterized as dangerously useless and as not useful for FWI and ISI, respectively (not shown), while for the 2-month lead time forecast the predictions are characterized as dangerously useless and not useful (not shown) for the same components, respectively.

From the above analysis, we conclude that both the initialisation date of the forecasts, the length of the spin-up period as well as the way spin-up is implanted (with or without ERA5-Land) play an important role in the prediction of FWI and its sub-components. Therefore, in the revised manuscript we plan to maintain the MJJAS period but analyse and discuss the experiments as above (i.e., 1-month and 2- month lead time fire danger experiments with no spin up and spin up performed from both SEAS5 and ERA5-Land data).



1-month lead time T2M MJJ [upper tercile]

2-month lead time T2M MJJ [upper tercile]

1-month lead time RH MJJ [lower tercile]

2-month lead time RH MJJ [lower tercile]

1-month lead time PR MJJ [lower tercile]

2-month lead time PR MJJ [lower tercile]

1-month lead time WSS MJJ [upper tercile]

2-month lead time WSS MJJ [upper tercile]

Figure R1: ROC Skill Scores of the FWI input variables for 1-month and 2-moth lead time MJJ forecasts that correspond to high fire danger values, i.e., upper tercile of air temperature (T2M), upper tercile of wind speed (WSS), lower tercile of air relative humidity (RH) and lower tercile of total precipitation (PR). The grid points with significant ROCSS values are indicated by circles (α=0.05).

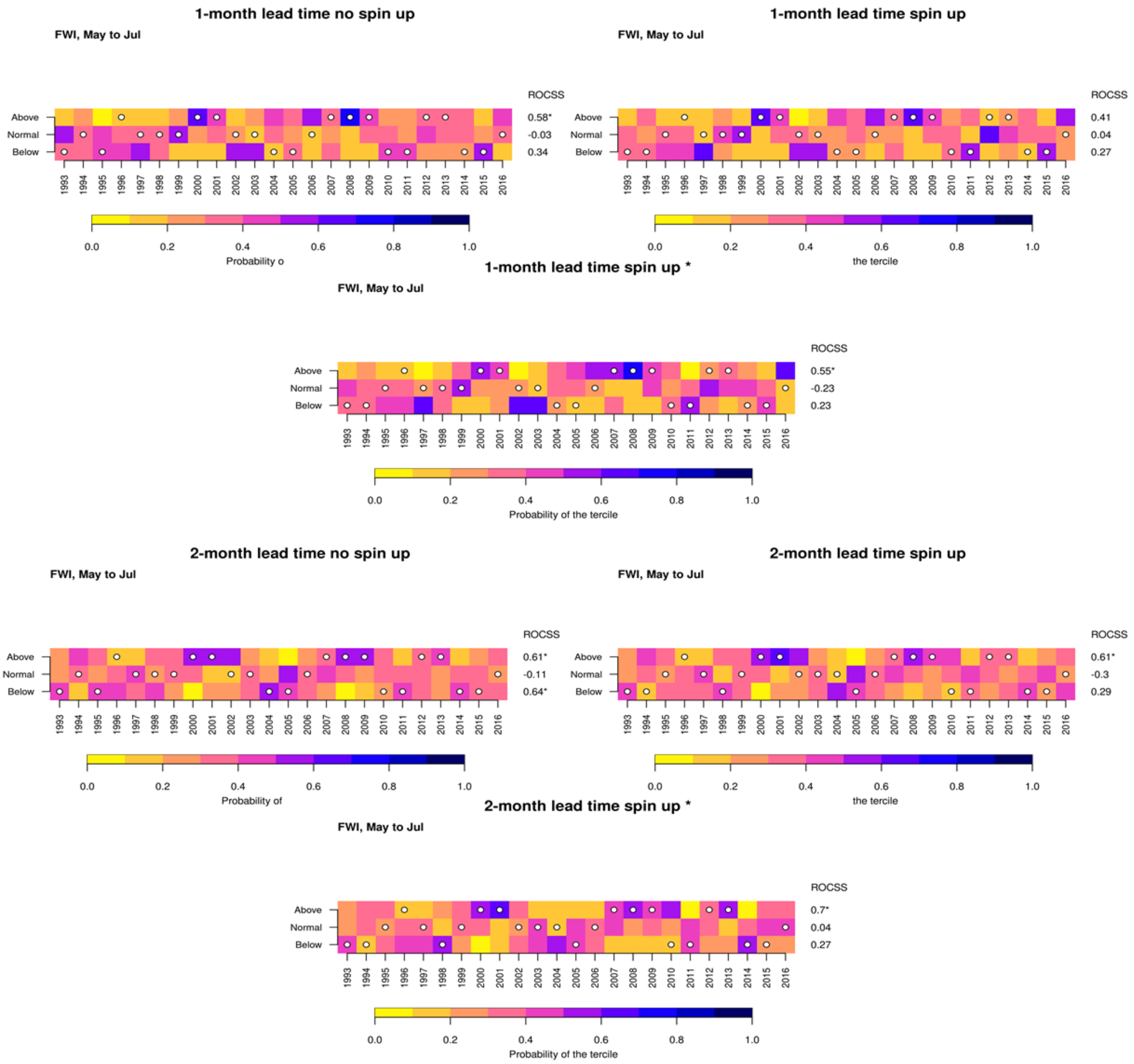
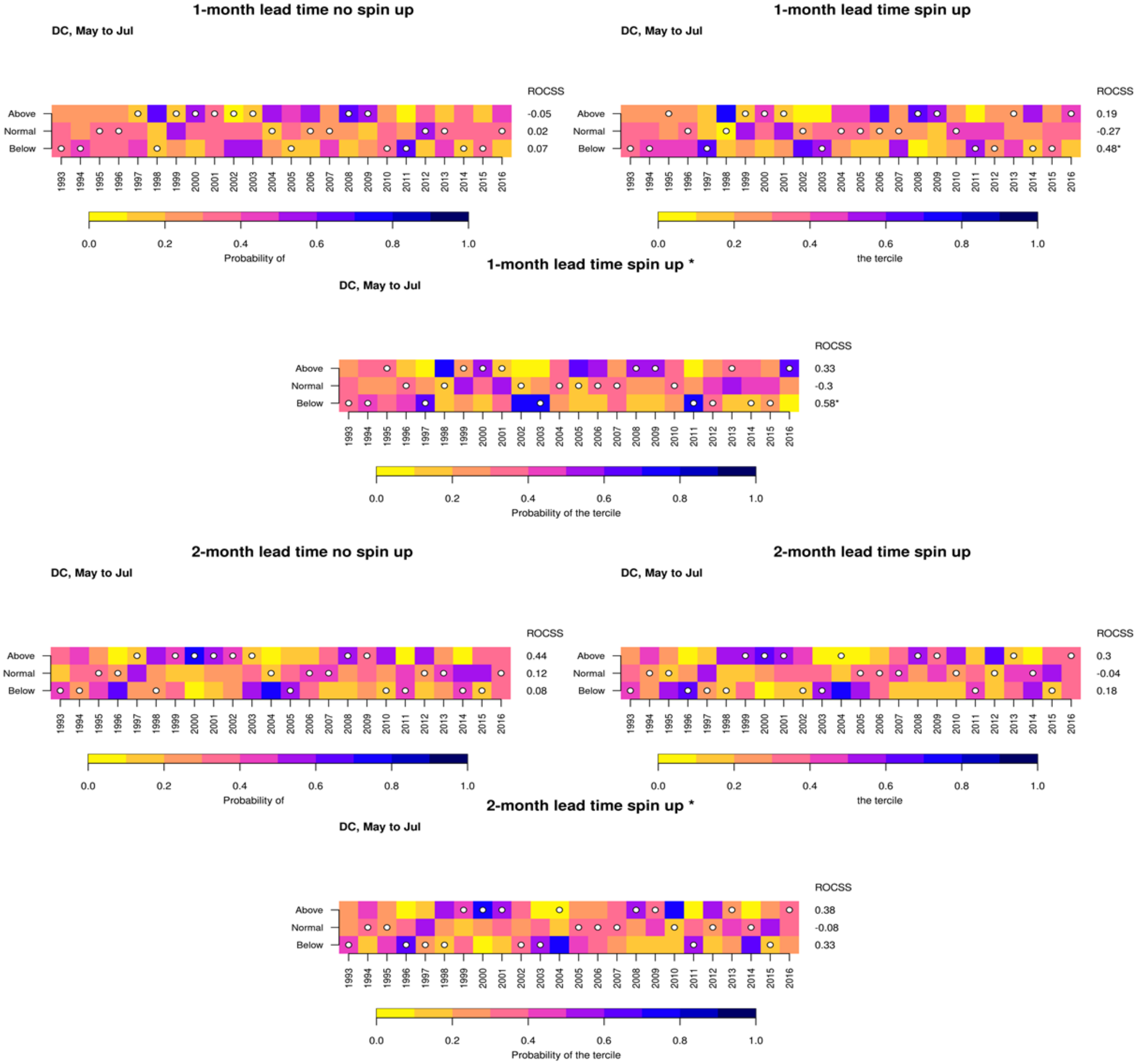


Figure R2: Tercile plots for May to July FWI predictions covering the hindcast period (1993-2016) for Attica. Forecast probabilities for the three tercile categories are codified in a yellow (0, no member forecasts in one category) to blue (1, all the members in the same category) scale. The white bullets represent the observed category according to the ERA5-Land dataset. ROCSS values obtained from the hindcast period are shown on the right side of each category and the asterisk indicates significant values (α=0.05). The asterisk (*) in the title indicates that the ERA5-Land data are implanted in the spin-up experiment.



1-month lead time no spin up

1-month lead time spin up

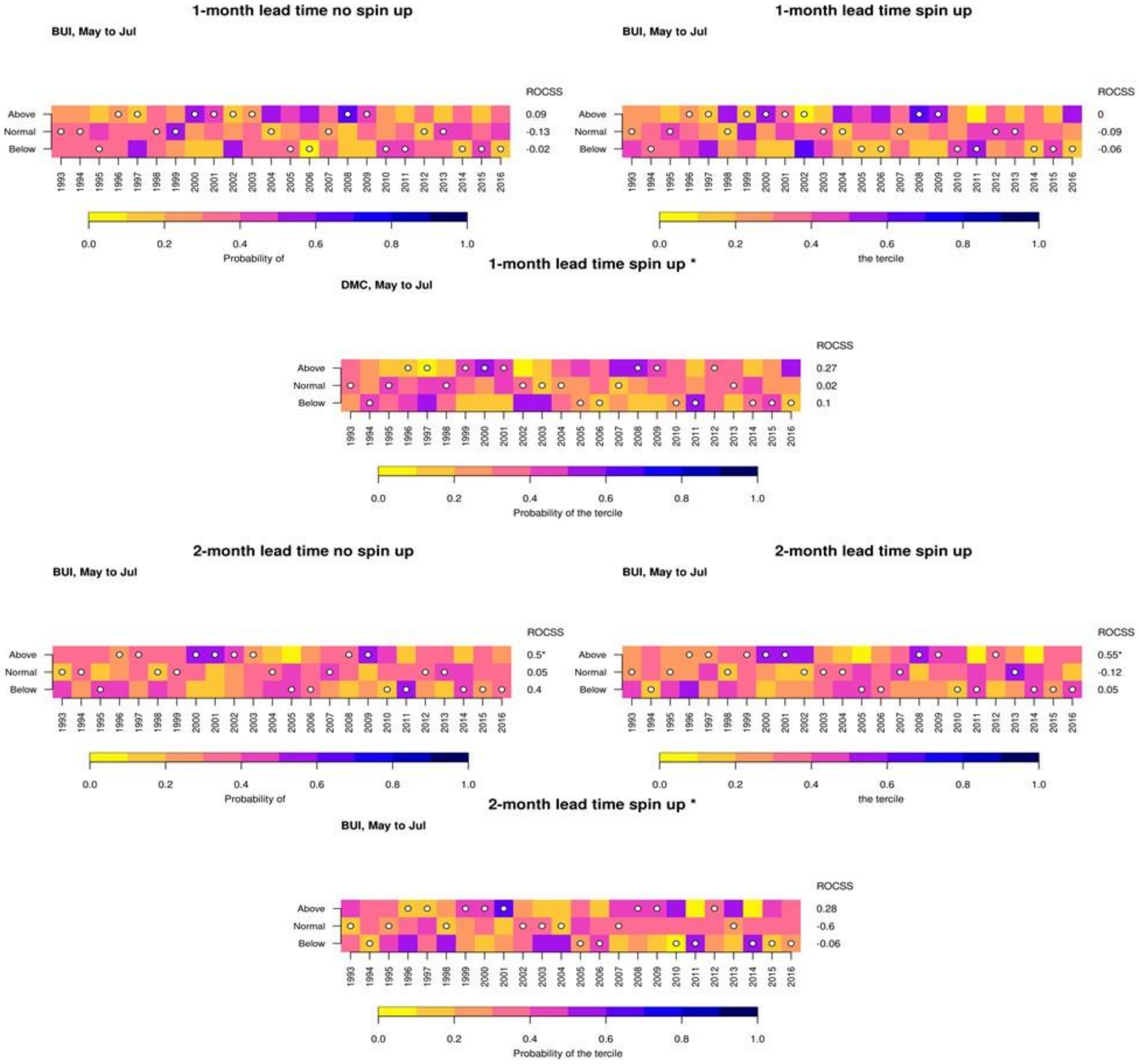
1-month lead time spin up *

2-month lead time no spin up

2-month lead time spin up

2-month lead time spin up *

Figure R3: Tercile plots for May to July DC predictions covering the hindcast period (1993-2016) for Attica. Forecast probabilities for the three tercile categories are codified in a yellow (0, no member forecasts in one category) to blue (1, all the members in the same category) scale. The white bullets represent the observed category according to the ERA5-Land dataset. ROCSS values obtained from the hindcast period are shown on the right side of each category and the asterisk indicates significant values (α=0.05). The asterisk (*) in the title indicates that the ERA5-Land data are implanted in the spin-up experiment.



1-month lead time no spin up

1-month lead time spin up

1-month lead time spin up *

2-month lead time no spin up

2-month lead time spin up

2-month lead time spin up *

Figure R4: Tercile plots for May to July DC predictions covering the hindcast period (1993-2016) for Attica. Forecast probabilities for the three tercile categories are codified in a yellow (0, no member forecasts in one category) to blue (1, all the members in the same category) scale. The white bullets represent the observed category according to the ERA5-Land dataset. ROCSS values obtained from the hindcast period are shown on the right side of each category and the asterisk indicates significant values (α=0.05). The asterisk (*) in the title indicates that the ERA5-Land data are implanted in the spin-up experiment.

1-month lead time no spin up

1-month lead time spin up

1-month lead time spin up *

2-month lead time no spin up

2-month lead time spin up

2-month lead time spin up *

Figure R5: Tercile plots for May to July DMC predictions covering the hindcast period (1993-2016) for Attica. Forecast probabilities for the three tercile categories are codified in a yellow (0, no member forecasts in one category) to blue (1, all the members in the same category) scale. The white bullets represent the observed category according to the ERA5-Land dataset. ROCSS values obtained from the hindcast period are shown on the right side of each category and the asterisk indicates significant values (α=0.05). The asterisk (*) in the title indicates that the ERA5-Land data are implanted in the spin-up experiment.

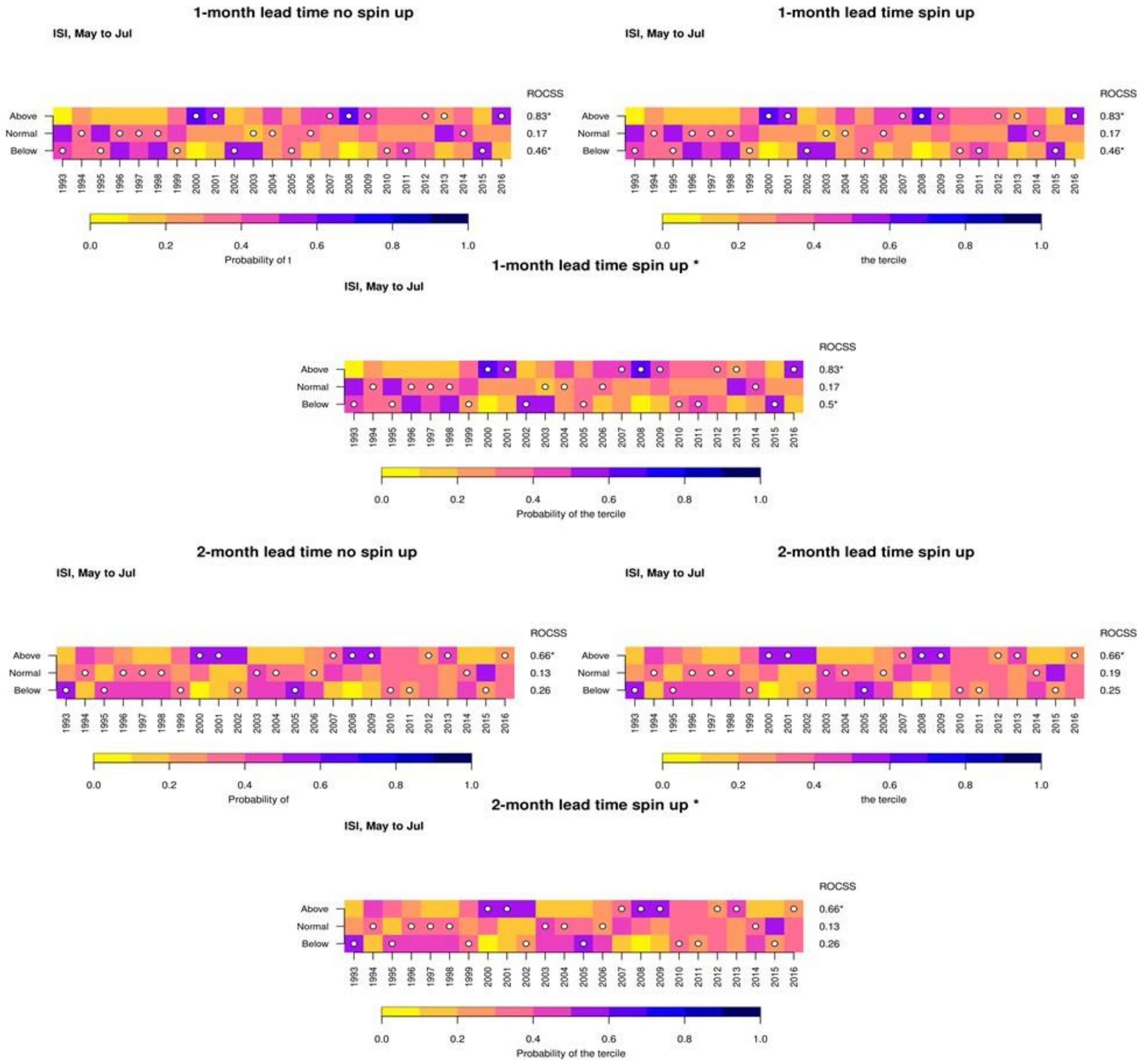


Figure R6: Tercile plots for May to July ISI predictions covering the hindcast period (1993-2016) for Attica. Forecast probabilities for the three tercile categories are codified in a yellow (0, no member forecasts in one category) to blue (1, all the members in the same category) scale. The white bullets represent the observed category according to the ERA5-Land dataset. ROCSS values obtained from the hindcast period are shown on the right side of each category and the asterisk indicates significant values (α=0.05). The asterisk (*) in the title indicates that the ERA5-Land data are implanted in the spin-up experiment.

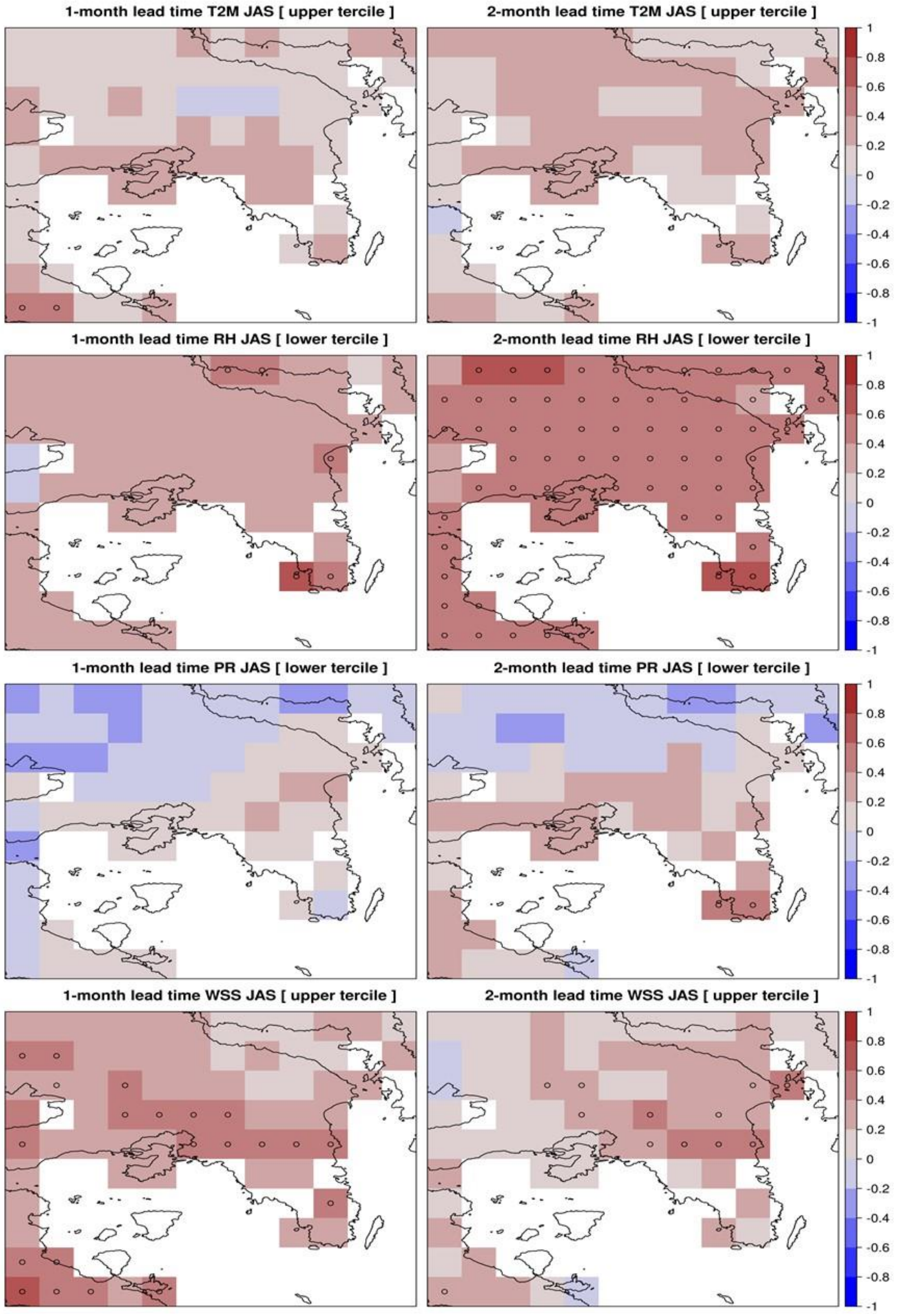


Figure R7: ROC Skill Scores of the FWI input variables for 1-month and 2-moth lead time JAS forecasts that correspond to high fire danger values, i.e., upper tercile of air temperature (T2M), upper tercile of wind speed (WSS), lower tercile of air relative humidity (RH) and lower tercile of total precipitation (PR). The grid points with significant ROCSS values are indicated by circles (α=0.05).

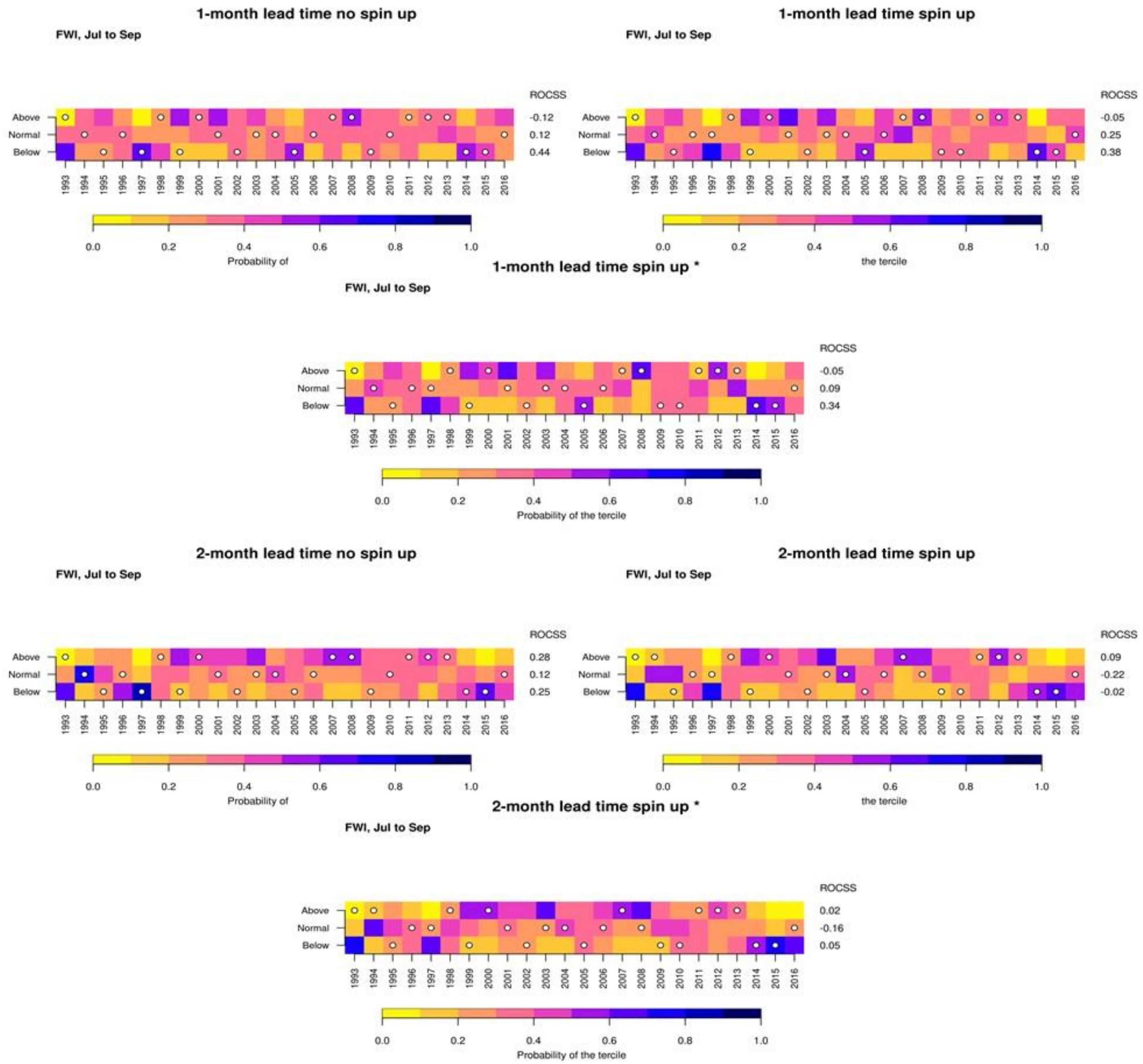


Figure R8: Tercile plots for July to September FWI predictions covering the hindcast period (1993-2016) for Attica. Forecast probabilities for the three tercile categories are codified in a yellow (0, no member forecasts in one category) to blue (1, all the members in the same category) scale. The white bullets represent the observed category according to the ERA5-Land dataset. ROCSS values obtained from the hindcast period are shown on the right side of each category and the asterisk indicates significant values (α=0.05). The asterisk (*) in the title indicates that the ERA5-Land data are implanted in the spin-up experiment.

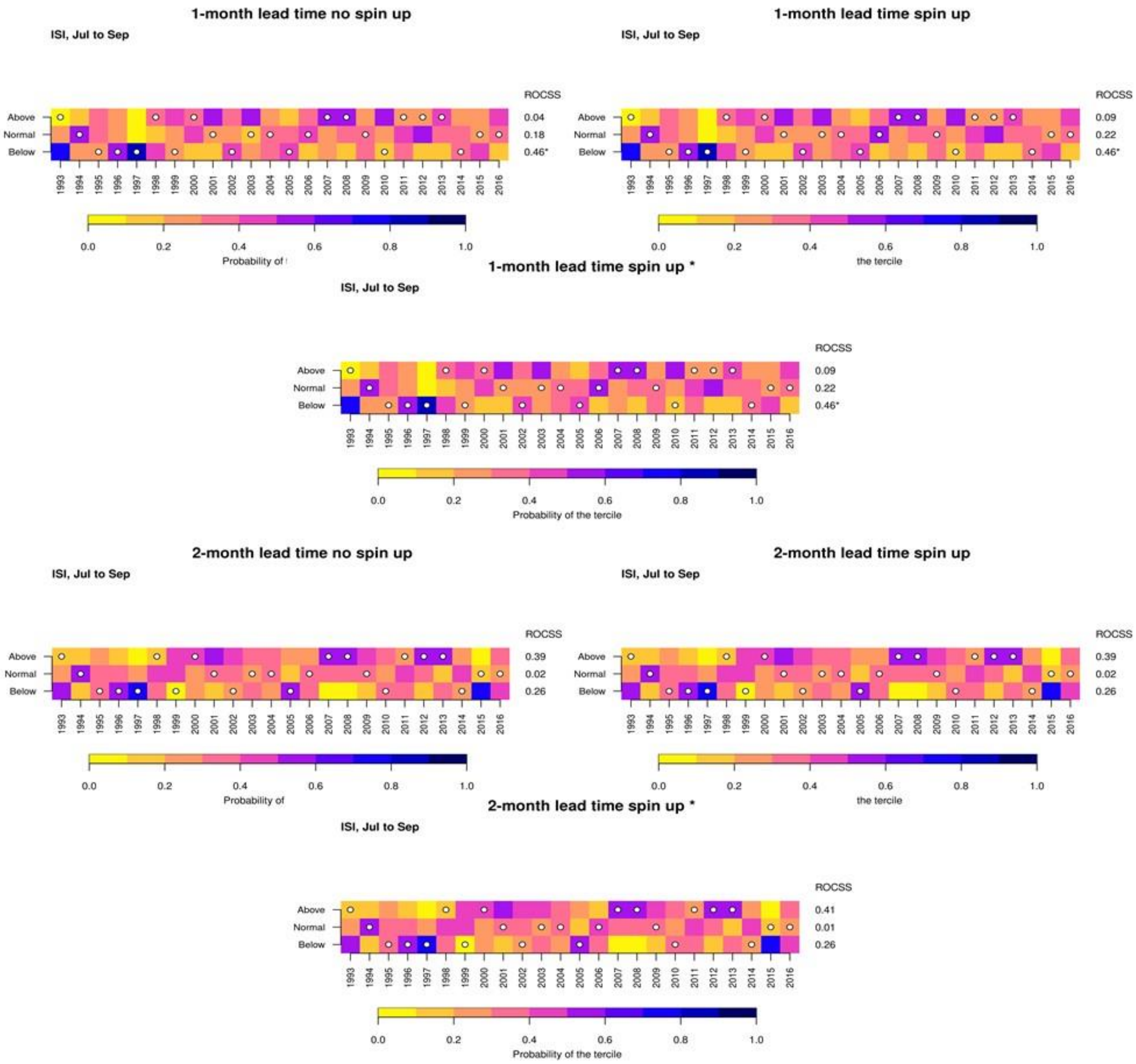


Figure R9: Tercile plots for July to September ISI predictions covering the hindcast period (1993-2016) for Attica. Forecast probabilities for the three tercile categories are codified in a yellow (0, no member forecasts in one category) to blue (1, all the members in the same category) scale. The white bullets represent the observed category according to the ERA5-Land dataset. ROCSS values obtained from the hindcast period are shown on the right side of each category and the asterisk indicates significant values (α=0.05). The asterisk (*) in the title indicates that the ERA5-Land data are implanted in the spin-up experiment.

----------minor points

The reliability diagrams are useful in that they're an alternative way of valdiating the forecasts, but perhaps could be in supplementary material, as they seem to largely just backup the ROCSS results.

Answer: Following the reviewer’s suggestions all reliability diagrams will be moved to the Supplementary Material.

I find the LM0/LM1 acronyms rather unnecessary and confusing. Suggest using e.g. '1 month lead' as it's not much longer, and much clearer.

Answer: Acronyms will be changed following the reviewer’s suggestion.