SUPPLEMENTARY MATERIAL
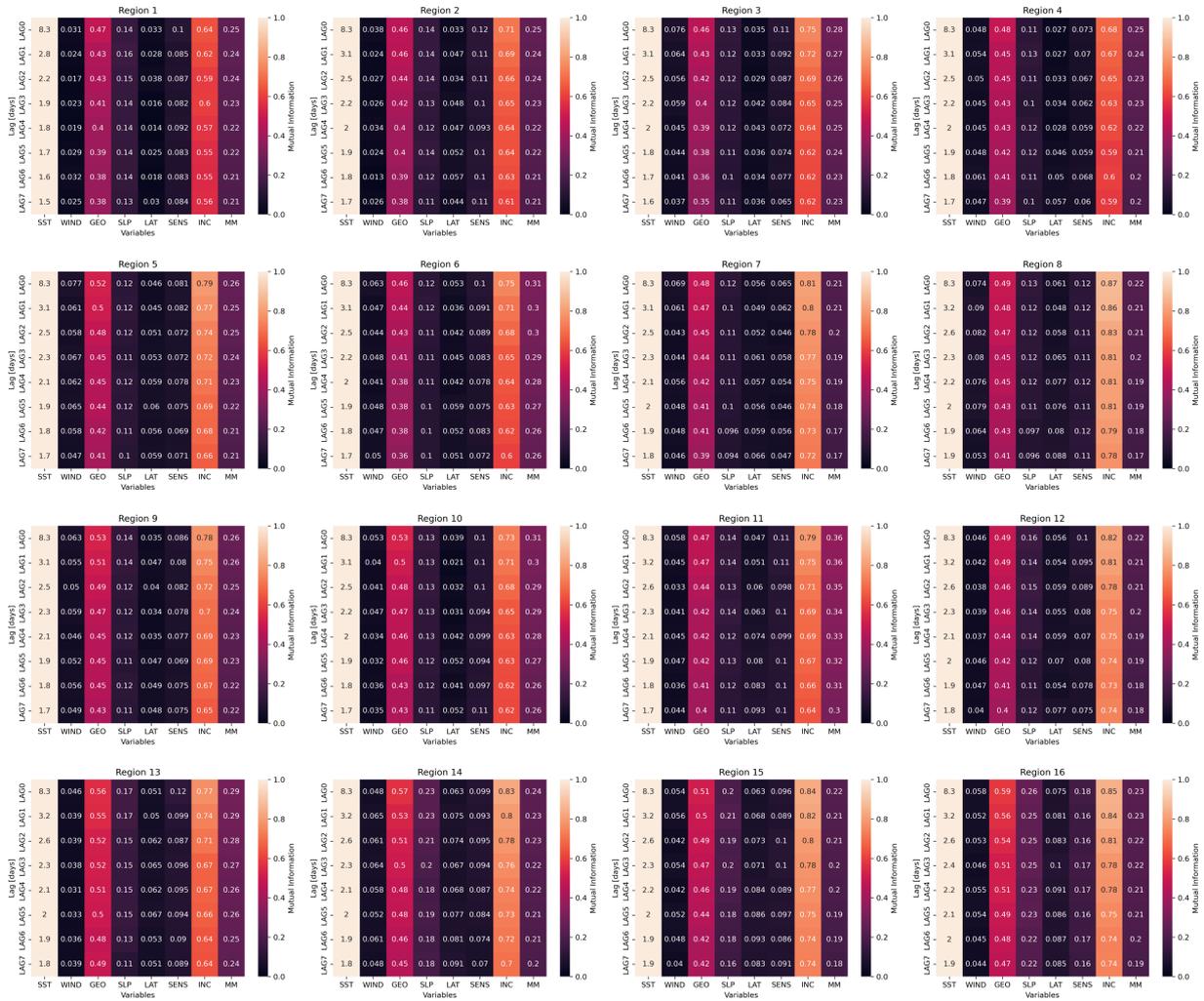
Figure S1: Heat maps of mutual information for each atmospheric variable and SST for all the regions at different days of lag.
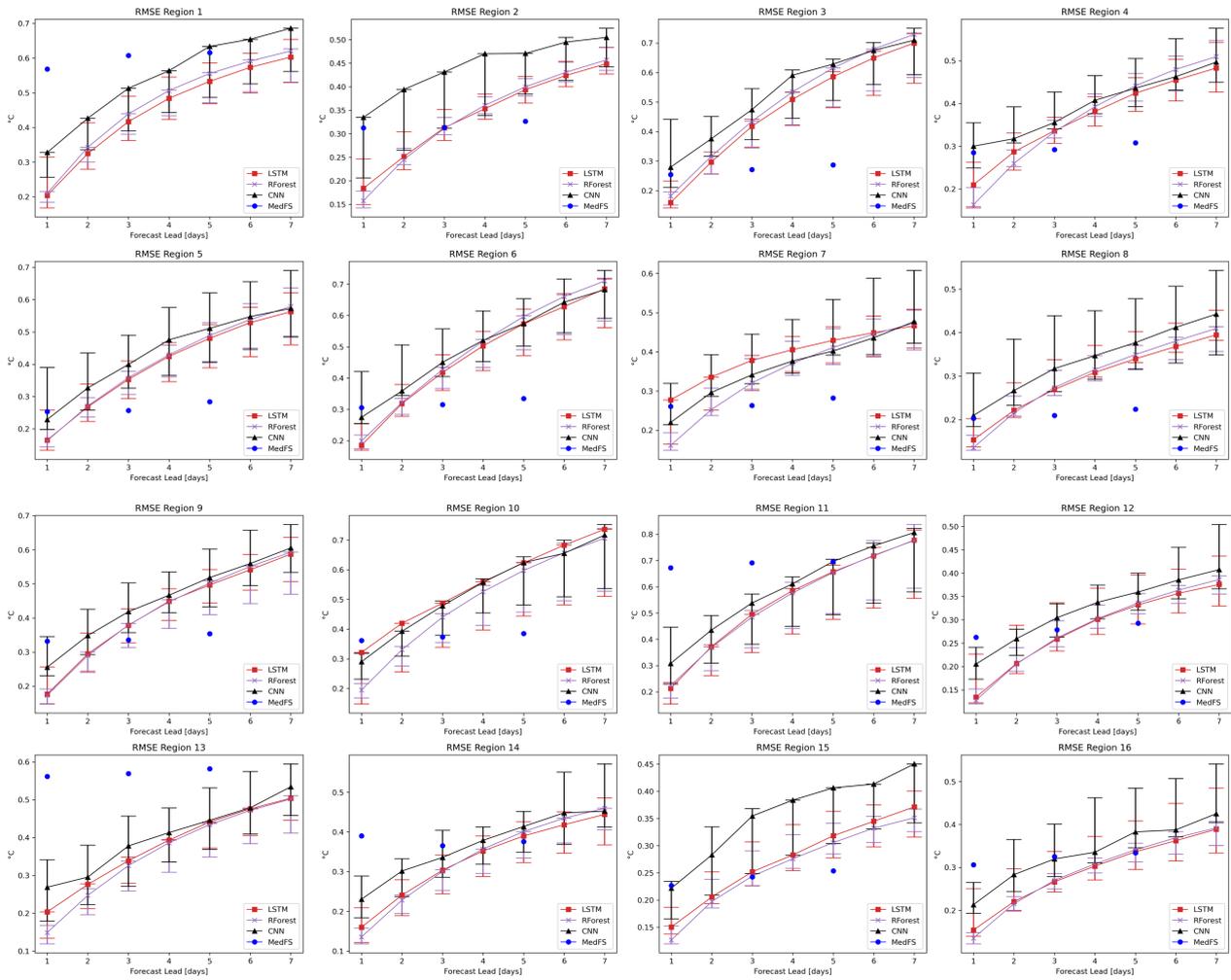
Figure S2: ML networks performance for the SST daily predictions in terms of Root Mean Square Error (RMSE) against the MedFS forecast system (MedFS) performance for each region.
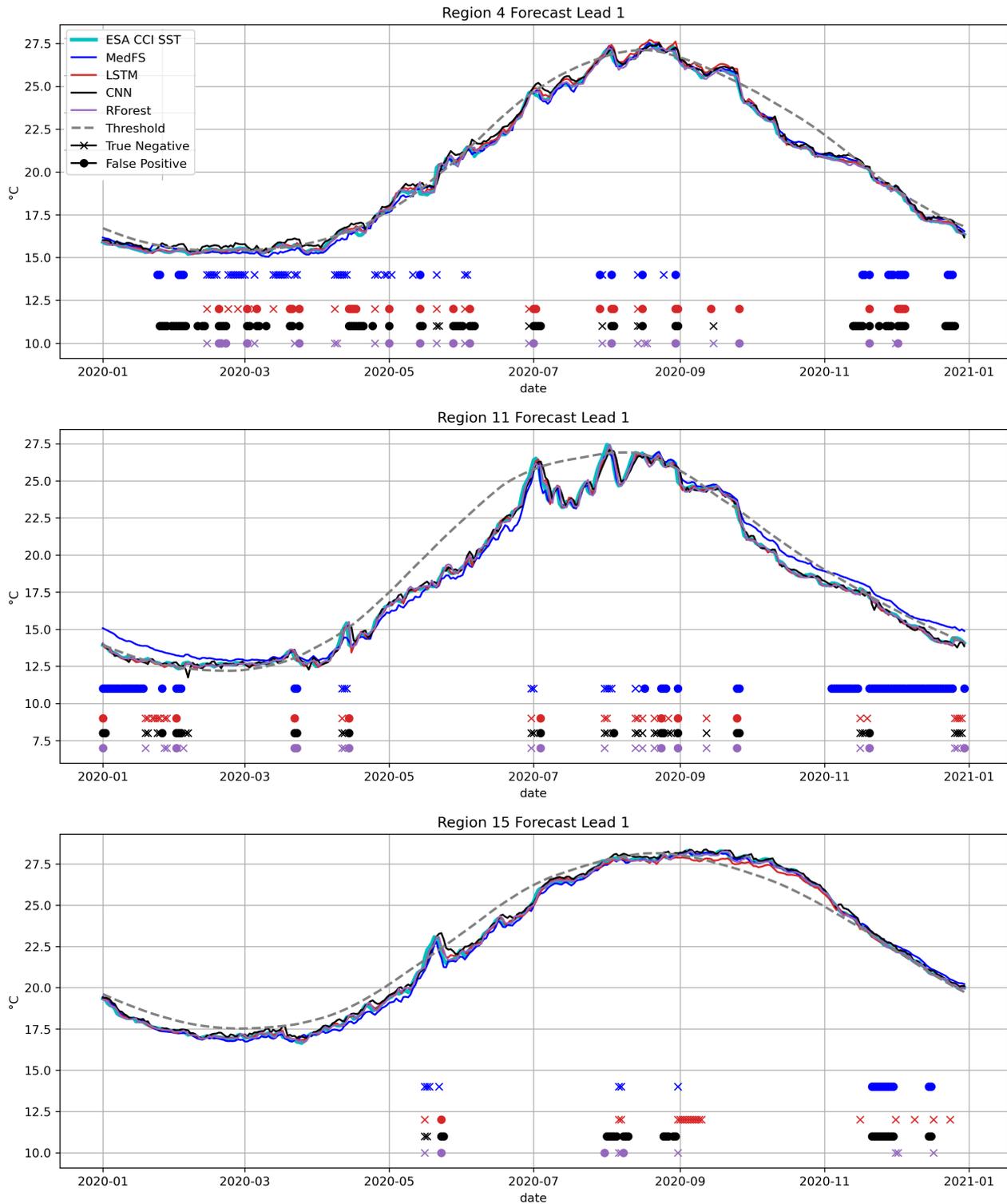
Figure S3: Time series of ESA CCI SST observed and predicted at forecast lead time 5 by the ML models (LSTM, CNN, RForest) and by MedFS during 2020 for (top panel) Western Mediterranean, (middle panel) Central Mediterranean and (bottom panel) Eastern Mediterranean. The 90th percentile threshold to define MHWs is represented in gray. Crosses correspond to misses alarms (False negative) and points to false alarms (False Positive) in the forecasts output in predicting MHWs. Colors refer to the different ML models.
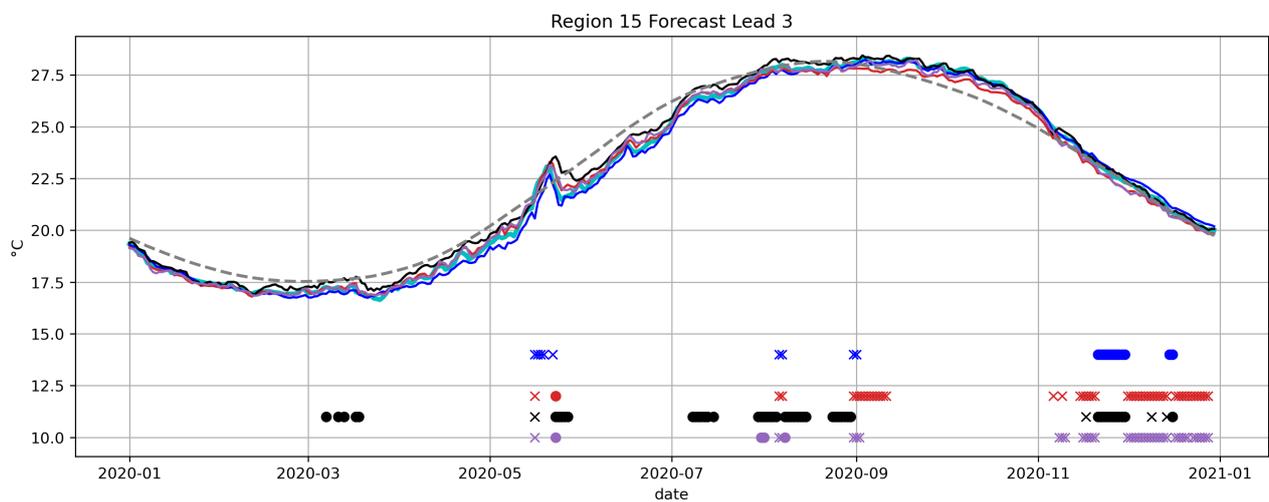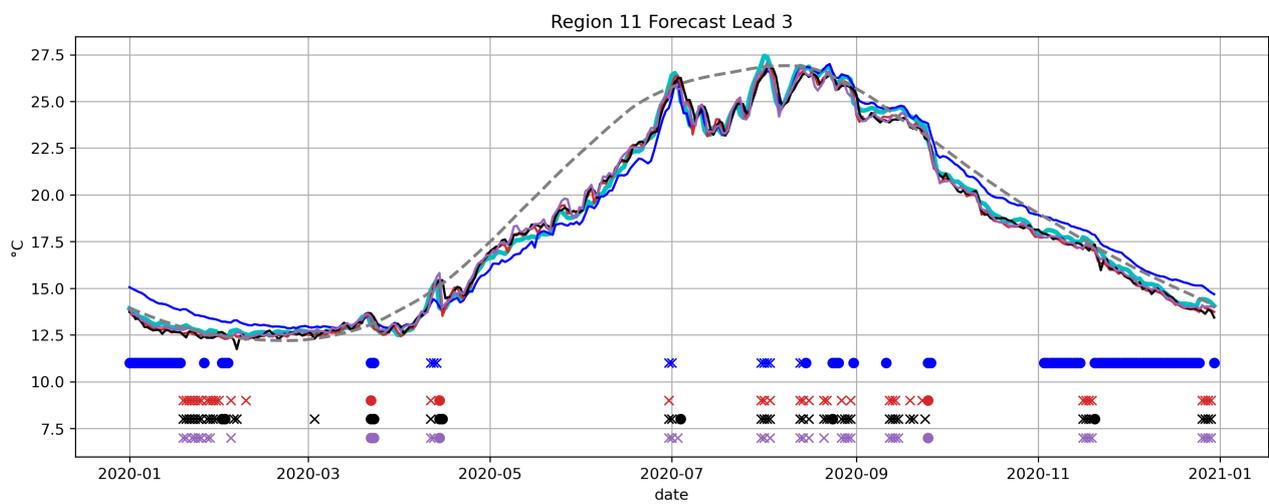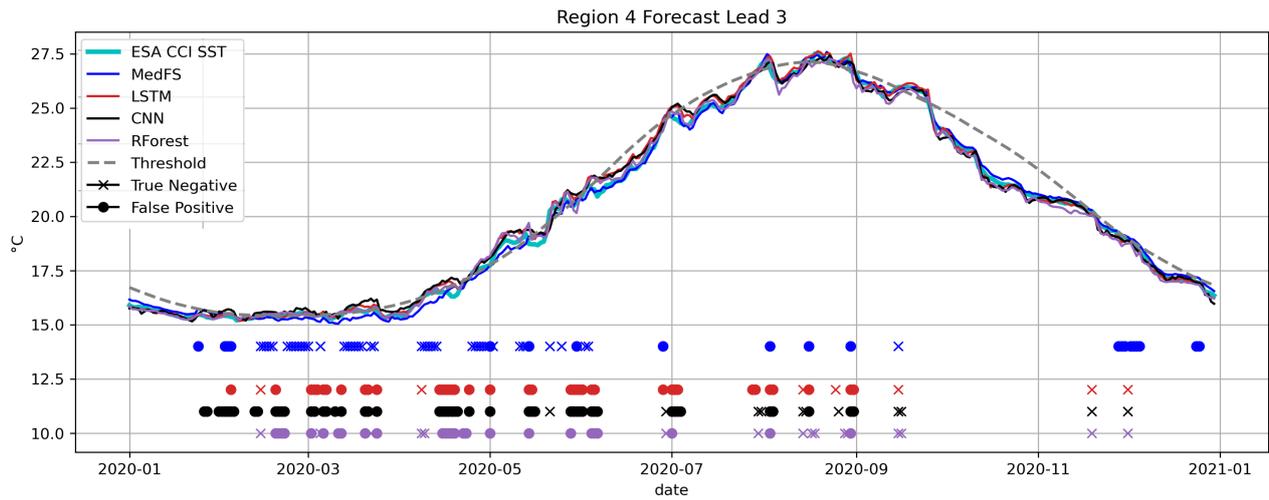
Figure S4: As Figure S4 but for forecast lead time 3.
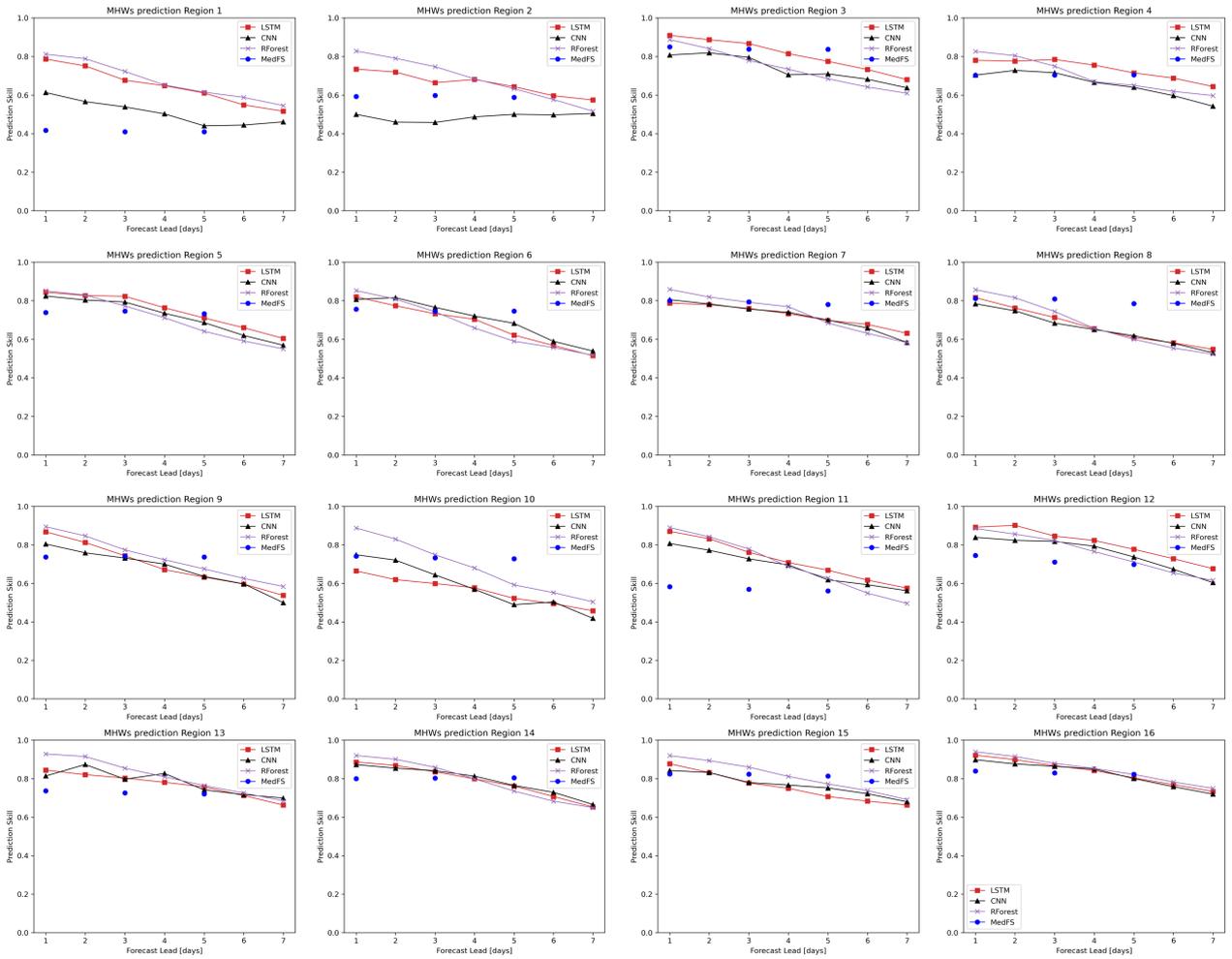
Figure S5: Variation of F1 score for MHWs occurrence prediction with the forecast lead time for each region.

Region ID	Models	FPR			FNR			F1 Score		
		L1	L3	L5	L1	L3	L5	L1	L3	L5
1	LSTM	0.02	0.02	0.04	0.29	0.44	0.47	0.79	0.68	0.61
	RForest	0.03	0.03	0.04	0.22	0.36	0.46	**0.81**	**0.72**	**0.62**
	CNN	0.01	0.01	0.02	0.54	0.61	0.68	0.61	0.54	0.44
	MedFS	0.01	0.02	0.02	0.68**	0.68**	0.68**	0.42	0.41	0.41
2	LSTM	0.01	0.03	0.06	0.39	0.43	0.40	0.73	0.66	0.64
	RForest	0.02	0.04	0.07	0.21	0.3	0.39	**0.83**	**0.75**	**0.63**
	CNN	0.01	0.01	0.04	0.65**	0.68**	0.60**	0.50	0.46	0.50
	MedFS	0.10	0.09	0.09	0.33	0.36	0.36	0.59	0.60	0.59
3	LSTM	0.02	0.03	0.04	0.09	0.15	0.27	**0.91**	**0.87**	0.77
	RForest	0.02	0.03	0.05	0.14	0.28	0.37	0.89	0.78	0.68
	CNN	0.02	0.04	0.04	0.26**	0.24**	0.35**	0.81	0.8	0.71
	MedFS	0.06	0.06	0.07	0.09	0.11	0.09	0.85	0.84	**0.84**
4	LSTM	0.10	0.10	0.14	0.10	0.06	0.12	0.78	**0.78**	0.71
	RForest	0.03	0.06	0.09	0.19	0.24	0.34	**0.83**	0.75	0.65
	CNN	0.15**	0.12**	0.14**	0.1	0.17	0.25	0.7	0.72	0.64
	MedFS	0.07	0.07	0.06	0.29	0.31**	0.31**	0.70	0.70	0.70
5	LSTM	0.02	0.06	0.07	0.2	0.15	0.3	**0.84**	**0.82**	0.71
	RForest	0.02	0.05	0.08	0.20**	0.24**	0.38**	0.84	0.77	0.64
	CNN	0.05	0.09	0.12	0.16	0.11	0.22	0.82	0.79	0.69
	MedFS	0.09	0.08	0.09	0.18	0.18	0.19	0.74	0.75	**0.73**
6	LSTM	0.01	0.01	0.02	0.27	0.39	0.51	0.82	0.73	0.62
	RForest	0.02	0.02	0.05	0.18	0.35	0.5	**0.85**	0.74	0.59
	CNN	0.08	0.04	0.06	0.09	0.27	0.34	0.81	**0.77**	0.68
	MedFS	0.12**	0.12**	0.12**	0.09	0.11	0.12	0.76	0.75	**0.75**
7	LSTM	0.12**	0.14**	0.17**	0.04	0.05	0.11	0.79	0.76	0.7
	RForest	0.03	0.04	0.07	0.17**	0.24**	0.33**	**0.86**	**0.79**	0.68
	CNN	0.04	0.07	0.13	0.21	0.23	0.18	0.81	0.76	0.7
	MedFS	0.08	0.08	0.08	0.16	0.18	0.19	0.80	0.79	**0.78**
8	LSTM	0.05	0.09	0.13	0.17	0.25	0.33	0.82	0.71	0.61
	RForest	0.03	0.05	0.11	0.15	0.29**	0.39**	**0.86**	0.74	0.60
	CNN	0.08	0.11	0.17	0.14	0.25	0.25	0.78	0.68	0.62
	MedFS	0.06	0.06	0.06	0.17	0.18	0.21	0.81	**0.81**	**0.78**
9	LSTM	0.02	0.02	0.04	0.19	0.37**	0.49**	0.87	0.74	0.63
	RForest	0.03	0.03	0.05	0.12	0.31	0.42	**0.89**	**0.77**	0.68
	CNN	0.04	0.04	0.06	0.26**	0.36	0.45	0.8	0.73	0.64
	MedFS	0.23**	0.22**	0.23**	0.04	0.04	0.04	0.74	0.74	**0.74**
10	LSTM	0.00	0.00	0.01	0.50**	0.57**	0.64**	0.66	0.6	0.52
	RForest	0.03	0.02	0.03	0.14	0.36	0.54	**0.89**	**0.75**	0.59
	CNN	0.02	0.02	0.03	0.37	0.5	0.65	0.75	0.64	0.49
	MedFS	0.16**	0.16**	0.17**	0.12	0.14	0.14	0.74	0.73	**0.73**
11	LSTM	0.02	0.02	0.01	0.17	0.35	0.48	0.87	0.76	**0.67**
	RForest	0.03	0.02	0.03	0.12	0.31	0.49	**0.89**	**0.78**	0.63
	CNN	0.05	0.03	0.02	0.21	0.37**	0.52**	0.81	0.73	0.62
	MedFS	0.27**	0.27**	0.27**	0.17	0.19	0.2	0.58	0.57	0.56
12	LSTM	0.04	0.06	0.1	0.09	0.15	0.18	**0.89**	**0.85**	**0.78**
	RForest	0.03	0.03	0.06	0.14	0.25	0.35	0.89	0.82	0.71
	CNN	0.06	0.05	0.09	0.16	0.21	0.26	0.84	0.82	0.74
	MedFS	0.01	0.01	0.01	0.39**	0.43**	0.44**	0.74	0.71	0.70
13	LSTM	0.01	0.01	0.04	0.26	0.31**	0.35**	0.84	0.80	0.76
	RForest	0.03	0.04	0.06	0.08	0.19	0.31	**0.93**	**0.85**	**0.76**
	CNN	0.01	0.03	0.08	0.30**	0.29	0.32	0.81	0.80	0.74
	MedFS	0.35**	0.34**	0.35**	0.07	0.10	0.11	0.74	0.73	0.72
14	LSTM	0.02	0.03	0.05	0.18**	0.24**	0.32	0.89	0.84	0.76
	RForest	0.02	0.03	0.05	0.11	0.20	0.35**	**0.92**	**0.86**	0.74
	CNN	0.11	0.09	0.16	0.06	0.14	0.18	0.87	0.84	0.76
	MedFS	0.28**	0.26**	0.26**	0.02	0.04	0.04	0.80	0.80	**0.80**
15	LSTM	0.03	0.04	0.07	0.18**	0.31**	0.38**	0.88	0.78	0.71
	RForest	0.03	0.04	0.07	0.10	0.19	0.28	**0.92**	**0.86**	0.77
	CNN	0.17**	0.26**	0.27**	0.02	0.02	0.07	0.84	0.78	0.75
	MedFS	0.13	0.11	0.12	0.12	0.15	0.16	0.82	0.82	**0.81**
16	LSTM	0.02	0.04	0.06	0.11	0.18	0.25**	0.92	0.87	0.80
	RForest	0.03	0.05	0.07	0.07	0.15	0.21	**0.94**	**0.88**	**0.82**
	CNN	0.08	0.15	0.2	0.06	0.04	0.10	0.90	0.86	0.80
	MedFS	0.18**	0.16**	0.17**	0.07	0.10	0.11	0.84	0.83	0.82

Table S1: ML networks' performance for the MHWs daily predictions in terms of False Positive rates (FPR), False Negative rates (FNR) and F1 score for Long-short term memory (LSTM), Random Forest (RForest), Convolutional Neural Network (CNN), and the Mediterranean Forecasting System (MedFs) at the first day of forecast (Forecast Lead 1, L1), at the third

days of forecast (Forecast Lead 3, L3) and at the 5th days of forecast (Forecast Lead 5, L5). Bold values identify the best F1 scores at L1, L3 and L5. Stars (**) identify the highest rates of FPR and FNR at L1, L3 and L5.