Appendix A: Forecasting performance for data frequencies and forecasting horizons: Point Forecasts (PFs)

This appendix provides the analytical results of the M4 Competition in terms of PFs accuracy. The results are presented per data frequency and forecasting horizon (short, medium and long-term), displaying the values of sMAPE, MASE and OWA (averages per case). For reasons of simplicity, only the methods that performed better than the Comb are considered, as well as the benchmarks and the standards for comparison. The order of the methods follows the final ranks of the competition. Yet, the last column indicates the rank of the methods for whole sample of the examined frequency (according to OWA) to facilitate comparisons.

Table A1: Forecasting performance achieved per forecasting horizon - Point Forecasts: Yearly data.

Method	sMAPE				MASE				OWA				Rank
	Short (1-2)	***Medium (3-4)***	***Long (5-6)***	***Total***	***Short (1-2)***	***Medium (3-4)***	***Long (5-6)***	***Total***	***Short (1-2)***	***Medium (3-4)***	***Long (5-6)***	***Total***	
Smyl	8.998	**13.428**	**17.101**	**13.176**	1.783	**2.984**	**4.173**	**2.980**	0.911	**0.768**	**0.735**	**0.778**	**1**
Montero-Manso, et al.	9.070	13.806	17.709	13.528	1.800	3.067	4.314	3.060	0.919	0.789	0.760	0.799	3
Pawlikowski, et al.	9.178	14.279	18.373	13.943	1.813	3.147	4.431	3.130	0.928	0.813	0.785	0.820	10
Jaganathan & Prakash	9.106	14.019	18.011	13.712	1.810	3.123	4.444	3.126	0.923	0.803	0.778	0.813	8
Fiorucci & Louzada	9.114	14.030	17.874	13.673	1.781	3.072	4.286	3.046	0.916	0.797	0.762	0.802	5
Petropoulos & Svetunkov	9.152	13.994	17.860	13.669	1.810	3.095	4.339	3.082	0.926	0.798	0.766	0.806	7
Shaub	9.201	14.026	17.811	13.679	1.791	3.061	4.261	3.038	0.924	0.795	0.758	0.801	4
Legaki & Koutsouri	**8.912**	13.750	17.436	13.366	**1.746**	3.055	4.227	3.009	**0.897**	0.786	0.747	0.788	2
Doornik, et al.	9.328	14.134	18.268	13.910	1.893	3.232	4.660	3.262	0.956	0.820	0.802	0.836	13
Pedregal, et al.	9.227	14.074	18.161	13.821	1.845	3.171	4.537	3.185	0.938	0.810	0.789	0.824	12
Spiliotis & Assimakopoulos	9.195	14.050	18.166	13.804	1.842	3.169	4.540	3.184	0.936	0.809	0.790	0.823	11
Roubinchtein	9.603	14.775	18.957	14.445	1.888	3.264	4.581	3.244	0.968	0.843	0.811	0.850	14
Ibrahim	9.107	14.076	17.848	13.677	1.790	3.123	4.312	3.075	0.918	0.804	0.763	0.805	6
Tartu M4 seminar	9.583	14.396	18.311	14.096	1.851	3.108	4.314	3.091	0.958	0.812	0.774	0.820	9
Waheeb	9.597	15.064	19.687	14.783	1.914	3.375	4.911	3.400	0.975	0.865	0.855	0.880	19
Darin & Stellwagen	9.575	14.931	19.482	14.663	1.925	3.395	4.900	3.406	0.976	0.863	0.849	0.877	18
Dantas & Cyrino Oliveira	9.846	15.035	19.358	14.746	1.940	3.289	4.652	3.294	0.994	0.853	0.826	0.866	15
Theta - Benchmark	9.316	15.034	19.430	14.593	1.838	3.437	4.871	3.382	0.941	0.872	0.846	0.872	17
Comb - Benchmark	9.528	15.201	19.815	14.848	1.841	3.300	4.699	3.280	0.953	0.860	0.840	0.867	16

ARIMA - Standard for comparison	10.037	15.553	19.914	15.168	1.967	3.417	4.821	3.402	1.011	0.885	0.852	0.892	21
Damped - Benchmark	9.902	15.573	20.119	15.198	1.933	3.392	4.812	3.379	0.995	0.882	0.856	0.890	20
ETS - Standard for comparison	9.809	15.744	20.516	15.356	1.934	3.456	4.943	3.444	0.991	0.895	0.876	0.903	22
Holt - Benchmark	10.299	16.646	22.116	16.354	1.974	3.537	5.138	3.550	1.026	0.932	0.928	0.947	23
SES - Benchmark	9.821	16.923	22.444	16.396	1.989	4.047	5.906	3.981	1.005	1.003	1.001	1.003	27
Naïve 2 - Benchmark	9.748	16.846	22.433	16.342	1.984	4.039	5.900	3.974	1.000	1.000	1.000	1.000	24
Naïve 1 - Benchmark	9.748	16.846	22.433	16.342	1.984	4.039	5.900	3.974	1.000	1.000	1.000	1.000	24
Naïve S - Benchmark	9.748	16.846	22.433	16.342	1.984	4.039	5.900	3.974	1.000	1.000	1.000	1.000	24
RNN - Benchmark	19.092	22.613	25.489	22.398	3.847	4.988	6.003	4.946	1.949	1.289	1.077	1.308	29
MLP - Benchmark	17.928	22.232	25.133	21.764	3.754	5.042	6.044	4.946	1.866	1.284	1.072	1.288	28

Table A2: Forecasting performance achieved per forecasting horizon - Point Forecasts: Quarterly data.

Method	**sMAPE**				**MASE**				**OWA**				**Rank**
	Short (1-2)	***Medium (3-5)***	***Long (6-8)***	***Total***	***Short (1-2)***	***Medium (3-5)***	***Long (6-8)***	***Total***	***Short (1-2)***	***Medium (3-5)***	***Long (6-8)***	***Total***	
Smyl	6.544	**9.355**	**12.092**	**9.679**	0.682	1.061	1.466	1.118	0.875	0.854	**0.833**	0.847	2
Montero-Manso, et al.	**6.454**	9.403	12.249	9.733	**0.666**	**1.053**	**1.466**	**1.111**	**0.859**	**0.854**	0.839	**0.847**	**1**
Pawlikowski, et al.	6.492	9.458	12.336	9.796	0.671	1.066	1.488	1.125	0.865	0.861	0.848	0.855	5
Jaganathan & Prakash	6.505	9.450	12.370	9.809	0.679	1.073	1.501	1.135	0.870	0.864	0.853	0.859	6
Fiorucci & Louzada	6.533	9.484	12.336	9.816	0.677	1.065	1.475	1.122	0.871	0.862	0.845	0.855	4
Petropoulos & Svetunkov	6.511	9.455	12.339	9.800	0.672	1.061	1.471	1.118	0.867	0.859	0.844	0.853	3
Shaub	7.130	10.063	12.858	10.378	0.746	1.140	1.557	1.198	0.955	0.919	0.886	0.908	20
Legaki & Koutsouri	6.813	9.821	12.717	10.155	0.731	1.138	1.570	1.198	0.924	0.907	0.884	0.898	18
Doornik, et al.	6.669	9.666	12.555	10.000	0.698	1.103	1.533	1.163	0.894	0.885	0.868	0.878	7
Pedregal, et al.	6.741	9.751	12.668	10.093	0.695	1.105	1.535	1.164	0.896	0.890	0.873	0.883	9
Spiliotis & Assimakopoulos	6.785	9.773	12.712	10.128	0.716	1.118	1.547	1.178	0.913	0.896	0.878	0.889	12
Roubinchtein	6.948	9.835	12.659	10.172	0.719	1.097	1.515	1.159	0.926	0.891	0.867	0.885	10
Ibrahim	6.808	9.729	12.636	10.089	0.727	1.125	1.549	1.185	0.922	0.897	0.875	0.890	14
Tartu M4 seminar	8.264	10.728	13.385	11.109	0.864	1.189	1.569	1.250	1.106	0.969	0.908	0.960	23
Waheeb	6.641	9.692	12.706	10.059	0.689	1.095	1.538	1.160	0.886	0.884	0.875	0.880	8
Darin & Stellwagen	6.648	9.754	12.892	10.155	0.688	1.097	1.558	1.168	0.886	0.887	0.887	0.887	11
Dantas & Cyrino Oliveira	6.833	9.834	12.955	10.254	0.712	1.102	1.544	1.170	0.914	0.893	0.886	0.892	16
Theta - Benchmark	6.855	9.933	12.993	10.311	0.737	1.165	1.629	1.232	0.931	0.922	0.910	0.917	21
Comb - Benchmark	6.764	9.758	12.866	10.175	0.712	1.106	1.548	1.173	0.909	0.891	0.883	0.890	13
ARIMA - Standard for comparison	6.908	10.079	13.131	10.431	0.695	1.105	1.538	1.165	0.908	0.906	0.890	0.898	19
Damped - Benchmark	6.814	9.830	12.926	10.237	0.712	1.107	1.547	1.173	0.912	0.895	0.885	0.893	17
ETS - Standard for comparison	6.795	9.870	13.042	10.291	0.694	1.093	1.540	1.161	0.899	0.891	0.887	0.891	15
Holt - Benchmark	6.986	10.407	14.021	10.907	0.718	1.124	1.592	1.198	0.928	0.929	0.937	0.932	22
SES - Benchmark	6.899	10.154	13.512	10.600	0.761	1.254	1.811	1.340	0.949	0.967	0.978	0.970	24

Naïve 2 - Benchmark	7.312	10.578	13.912	11.012	0.798	1.287	1.839	1.371	1.000	1.000	1.000	1.000	25
Naïve 1 - Benchmark	8.310	11.005	14.416	11.610	0.973	1.363	1.928	1.477	1.178	1.050	1.042	1.066	26
Naïve S - Benchmark	10.481	11.735	14.668	12.521	1.241	1.471	1.974	1.602	1.494	1.126	1.064	1.153	27
RNN - Benchmark	14.861	16.708	18.789	17.027	1.692	1.958	2.291	2.016	2.077	1.551	1.298	1.508	28
MLP - Benchmark	16.546	18.264	20.039	18.500	2.013	2.263	2.567	2.314	2.393	1.743	1.418	1.684	29

Table A3: Forecasting performance achieved per forecasting horizon - Point Forecasts: Monthly data.

Method	**sMAPE**				**MASE**				**OWA**				**Rank**
	Short (1-6)	***Medium (7-12)***	***Long (13-18)***	***Total***	***Short (1-6)***	***Medium (7-12)***	***Long (13-18)***	***Total***	***Short (1-6)***	***Medium (7-12)***	***Long (13-18)***	***Total***	
Smyl	**8.749**	**12.614**	**15.015**	**12.126**	0.598	0.917	**1.136**	**0.884**	**0.798**	**0.860**	**0.839**	**0.836**	**1**
Montero-Manso, et al.	9.100	12.916	15.903	12.639	**0.597**	**0.915**	1.166	0.893	0.813	0.870	0.875	0.858	3
Pawlikowski, et al.	9.458	12.826	15.956	12.747	0.620	0.924	1.171	0.905	0.845	0.871	0.878	0.867	4
Jaganathan & Prakash	9.143	12.718	15.599	12.487	0.607	0.920	1.158	0.895	0.822	0.865	0.863	0.854	2
Fiorucci & Louzada	9.504	12.857	15.848	12.737	0.626	0.930	1.166	0.907	0.851	0.874	0.873	0.868	5
Petropoulos & Svetunkov	9.638	12.969	16.056	12.888	0.628	0.935	1.175	0.913	0.859	0.881	0.882	0.876	6
Shaub	9.735	13.024	15.758	12.839	0.651	0.953	1.182	0.929	0.878	0.891	0.876	0.882	9
Legaki & Koutsouri	9.899	13.142	15.966	13.002	0.678	0.990	1.231	0.966	0.904	0.912	0.900	0.905	13
Doornik, et al.	9.255	13.133	15.952	12.780	0.628	0.961	1.204	0.931	0.841	0.898	0.890	0.881	8
Pedregal, et al.	9.917	13.207	16.330	13.151	0.662	0.960	1.207	0.943	0.894	0.901	0.902	0.899	11
Spiliotis & Assimakopoulos	9.901	13.254	16.272	13.142	0.669	0.982	1.224	0.959	0.898	0.912	0.906	0.907	14
Roubinchtein	9.498	13.091	16.143	12.911	0.636	0.941	1.185	0.921	0.858	0.888	0.888	0.881	7
Ibrahim	10.062	13.364	16.536	13.321	0.684	0.999	1.249	0.977	0.915	0.924	0.923	0.921	19
Tartu M4 seminar	10.039	13.296	16.535	13.290	0.725	1.003	1.278	1.002	0.942	0.923	0.933	0.932	22
Waheeb	9.327	12.996	15.986	12.770	0.625	0.964	1.498	1.029	0.842	0.895	0.998	0.927	21
Darin & Stellwagen	9.438	13.259	16.477	13.058	0.623	0.947	1.203	0.924	0.846	0.896	0.904	0.887	10
Dantas & Cyrino Oliveira	10.056	13.529	16.800	13.462	0.671	0.963	1.222	0.952	0.906	0.913	0.920	0.914	16
Theta - Benchmark	9.883	13.179	15.945	13.002	0.679	0.994	1.236	0.970	0.904	0.915	0.901	0.907	15
Comb - Benchmark	10.062	13.464	16.776	13.434	0.678	0.984	1.238	0.966	0.911	0.920	0.925	0.920	18
ARIMA - Standard for comparison	9.596	13.540	17.193	13.443	0.625	0.954	1.212	0.930	0.855	0.909	0.928	0.903	12
Damped - Benchmark	10.136	13.506	16.778	13.473	0.682	0.989	1.246	0.972	0.917	0.924	0.928	0.924	20
ETS - Standard for comparison	9.933	13.637	17.005	13.525	0.650	0.967	1.227	0.948	0.886	0.919	0.928	0.915	17
Holt - Benchmark	10.318	14.807	19.312	14.812	0.686	1.022	1.319	1.009	0.928	0.985	1.027	0.988	24
SES - Benchmark	10.175	13.626	17.052	13.618	0.695	1.038	1.325	1.019	0.928	0.951	0.965	0.951	23
Naïve 2 - Benchmark	11.050	14.448	17.782	14.427	0.744	1.082	1.364	1.063	1.000	1.000	1.000	1.000	25
Naïve 1 - Benchmark	11.990	15.227	18.552	15.256	0.911	1.207	1.497	1.205	1.155	1.085	1.070	1.096	26
Naïve S - Benchmark	13.860	14.487	19.618	15.988	1.052	1.102	1.625	1.260	1.335	1.011	1.147	1.147	27
RNN - Benchmark	19.903	24.142	28.123	24.056	1.335	1.613	1.855	1.601	1.798	1.581	1.471	1.587	28
MLP - Benchmark	20.016	24.643	28.340	24.333	1.698	1.941	2.137	1.925	2.047	1.750	1.580	1.749	29

Table A4: Forecasting performance achieved per forecasting horizon - Point Forecasts: Weekly data.

Method	sMAPE				MASE				OWA				Rank
	Short (1-4)	*Medium (5-9)*	*Long (10-13)*	*Total*	*Short (1-4)*	*Medium (5-9)*	*Long (10-13)*	*Total*	*Short (1-4)*	*Medium (5-9)*	*Long (10-13)*	*Total*	
Smyl	6.219	8.932	8.021	7.817	1.716	2.464	2.860	2.356	1.080	0.838	0.754	0.851	8
Montero-Manso, et al.	5.063	8.790	8.730	7.625	**1.272**	**2.230**	2.790	2.108	0.842	0.792	0.776	0.796	7
Pawlikowski, et al.	**4.886**	7.908	7.716	6.919	1.323	2.245	2.884	2.158	**0.841**	0.752	0.744	0.766	3
Jaganathan & Prakash	5.435	7.793	6.970	6.814	1.661	2.460	2.901	2.350	0.992	0.783	0.713	0.795	6
Fiorucci & Louzada	5.365	9.893	10.307	8.627	1.457	2.597	2.991	2.368	0.925	0.906	0.874	0.897	11
Petropoulos & Svetunkov	5.319	7.789	**6.806**	6.726	1.474	2.335	2.539	2.133	0.925	0.762	**0.656**	0.751	2
Shaub	6.625	8.897	7.662	7.818	2.342	3.003	3.480	2.947	1.304	0.926	0.825	0.957	19
Legaki & Koutsouri	5.729	10.373	11.037	9.148	1.616	2.785	3.357	2.601	1.005	0.961	0.957	0.968	22
Doornik, et al.	5.260	7.496	7.235	6.728	1.667	2.420	2.788	2.302	0.978	0.762	0.709	0.782	5
Pedregal, et al.	5.532	10.197	10.936	8.989	1.520	2.670	3.227	2.488	0.959	0.933	0.935	0.939	16
Spiliotis & Assimakopoulos	5.530	10.197	10.941	8.990	1.519	2.670	3.229	2.488	0.958	0.933	0.935	0.939	17
Roubinchtein	6.914	10.175	7.780	8.435	1.524	2.528	2.760	2.290	1.085	0.908	0.729	0.873	9
Ibrahim	5.683	10.303	10.979	9.089	1.554	2.776	3.371	2.583	0.983	0.956	0.956	0.961	20
Tartu M4 seminar	4.901	9.812	10.500	8.513	1.412	2.605	3.051	2.375	0.869	0.904	0.891	0.892	10
Waheeb	5.229	7.600	8.268	7.076	1.388	2.295	2.827	2.180	0.892	0.746	0.760	0.779	4
Darin & Stellwagen	5.153	**7.438**	6.940	**6.582**	1.450	2.305	**2.517**	**2.107**	0.903	**0.740**	0.658	**0.739**	**1**
Dantas & Cyrino Oliveira	5.579	10.101	10.630	8.873	1.591	2.770	3.182	2.534	0.984	0.945	0.915	0.941	18
Theta - Benchmark	5.578	10.324	11.070	9.093	1.563	2.831	3.468	2.637	0.976	0.966	0.974	0.971	23
Comb - Benchmark	5.557	10.185	10.779	8.944	1.501	2.633	3.111	2.432	0.955	0.926	0.911	0.926	13
ARIMA - Standard for comparison	5.063	9.735	10.891	8.653	1.442	2.738	3.441	2.556	0.893	0.922	0.962	0.932	15
Damped - Benchmark	5.587	10.093	10.610	8.866	1.491	2.607	3.062	2.404	0.955	0.918	0.897	0.917	12
ETS - Standard for comparison	5.558	9.954	10.362	8.727	1.594	2.756	3.174	2.527	0.983	0.936	0.902	0.931	14
Holt - Benchmark	5.726	10.790	12.337	9.708	1.499	2.597	3.119	2.420	0.970	0.949	0.982	0.966	21
SES - Benchmark	5.528	10.229	10.974	9.012	1.570	2.888	3.545	2.685	0.973	0.971	0.981	0.975	24
Naïve 2 - Benchmark	5.498	10.400	11.276	9.161	1.669	3.015	3.589	2.777	1.000	1.000	1.000	1.000	25
Naïve 1 - Benchmark	5.498	10.400	11.276	9.161	1.669	3.015	3.589	2.777	1.000	1.000	1.000	1.000	25
Naive S - Benchmark	5.498	10.400	11.276	9.161	1.669	3.015	3.589	2.777	1.000	1.000	1.000	1.000	25
RNN - Benchmark	12.505	15.988	16.974	15.220	4.280	5.044	6.094	5.132	2.420	1.605	1.602	1.755	28
MLP - Benchmark	22.039	21.638	20.298	21.349	13.159	13.491	14.072	13.568	5.948	3.278	2.861	3.608	29

Table A5: Forecasting performance achieved per forecasting horizon - Point Forecasts: Daily data.

Method	sMAPE				MASE				OWA				Rank
	Short (1-4)	*Medium (5-9)*	*Long (10-14)*	*Total*	*Short (1-4)*	*Mediu m (5-9)*	*Long (10-14)*	*Total*	*Short (1-4)*	*Medium (5-9)*	*Long (10-14)*	*Total*	

Smyl	1.825	3.036	4.379	3.170	2.070	3.325	4.668	3.446	1.101	1.031	1.039	1.046	22
Montero-Manso, et al.	1.665	3.003	4.338	3.097	1.897	3.269	4.578	3.344	1.006	1.016	1.024	1.019	20
Pawlikowski, et al.	**1.304**	**2.354**	**3.468**	**2.452**	**1.468**	**2.554**	**3.669**	**2.642**	**0.784**	**0.796**	**0.820**	**0.806**	**1**
Jaganathan & Prakash	1.658	2.926	4.251	3.037	1.860	3.182	4.454	3.258	0.995	0.990	1.000	0.996	9
Fiorucci & Louzada	1.623	2.891	4.169	2.985	1.842	3.129	4.339	3.194	0.979	0.976	0.977	0.977	3
Petropoulos & Svetunkov	1.640	2.889	4.184	2.995	1.860	3.157	4.397	3.229	0.989	0.980	0.986	0.984	5
Shaub	1.944	3.122	4.345	3.222	2.181	3.399	4.596	3.479	1.166	1.057	1.027	1.060	24
Legaki & Koutsouri	1.668	2.956	4.224	3.041	1.867	3.191	4.425	3.254	0.999	0.997	0.993	0.996	10
Doornik, et al.	1.679	2.963	4.241	3.053	1.883	3.214	4.475	3.284	1.007	1.001	1.001	1.002	19
Pedregal, et al.	1.665	2.937	4.203	3.026	1.858	3.168	4.396	3.232	0.996	0.990	0.988	0.990	6
Spiliotis & Assimakopoulos	1.666	2.939	4.206	3.027	1.858	3.168	4.396	3.232	0.996	0.990	0.988	0.990	7
Roubinchtein	2.031	3.167	4.366	3.270	2.306	3.547	4.777	3.632	1.226	1.087	1.050	1.091	25
Ibrahim	1.696	2.979	4.262	3.071	2.529	3.817	5.064	3.894	1.185	1.098	1.069	1.098	26
Tartu M4 seminar	1.523	2.743	4.024	2.852	1.709	2.947	4.156	3.025	0.914	0.922	0.940	0.930	2
Waheeb	1.687	2.877	4.164	2.997	1.939	3.236	4.512	3.321	1.024	0.990	0.996	0.999	13
Darin & Stellwagen	1.701	2.956	4.298	3.077	2.738	4.024	5.346	4.128	1.242	1.126	1.105	1.135	27
Dantas & Cyrino Oliveira	1.833	3.160	4.458	3.245	2.069	3.350	4.617	3.436	1.103	1.056	1.043	1.057	23
Theta - Benchmark	1.672	2.964	4.246	3.053	1.868	3.199	4.441	3.262	1.001	0.999	0.998	0.999	14
Comb - Benchmark	1.630	2.881	4.158	2.980	1.848	3.135	4.354	3.203	0.983	0.975	0.978	0.978	4
ARIMA - Standard for comparison	1.696	3.102	4.481	3.193	1.880	3.299	4.744	3.410	1.011	1.038	1.060	1.044	21
Damped - Benchmark	1.657	2.975	4.279	3.064	1.856	3.162	4.413	3.236	0.993	0.995	0.999	0.997	12
ETS - Standard for comparison	1.665	2.942	4.254	3.046	1.861	3.178	4.440	3.253	0.997	0.992	0.999	0.996	11
Holt - Benchmark	1.656	2.994	4.265	3.066	1.844	3.147	4.403	3.223	0.990	0.996	0.996	0.995	8
SES - Benchmark	1.667	2.953	4.238	3.045	1.874	3.214	4.473	3.281	1.001	1.000	1.001	1.000	18
Naïve 2 - Benchmark	1.667	2.957	4.236	3.045	1.871	3.212	4.471	3.278	1.000	1.000	1.000	1.000	15
Naïve 1 - Benchmark	1.667	2.957	4.236	3.045	1.871	3.212	4.471	3.278	1.000	1.000	1.000	1.000	15
Naïve S - Benchmark	1.667	2.957	4.236	3.045	1.871	3.212	4.471	3.278	1.000	1.000	1.000	1.000	15
RNN - Benchmark	4.637	5.741	7.248	5.964	5.097	6.042	7.331	6.232	2.754	1.911	1.675	1.930	28
MLP - Benchmark	8.665	9.255	9.913	9.321	12.072	12.868	13.799	12.973	5.826	3.568	2.713	3.509	29

Table A6: Forecasting performance achieved per forecasting horizon - Point Forecasts: Hourly data.

Method	sMAPE				MASE				OWA				Rank
	Short (1-16)	***Medium (17-32)***	***Long (33-48)***	***Total***	***Short (1-16)***	***Medium (17-32)***	***Long (33-48)***	***Total***	***Short (1-16)***	***Medium (17-32)***	***Long (33-48)***	***Total***	
Smyl	8.211	10.084	9.690	9.328	0.697	0.864	1.118	0.893	0.365	0.489	0.469	0.440	2
Montero-Manso, et al.	10.000	12.123	12.394	11.506	**0.637**	**0.757**	1.062	0.819	0.402	0.520	0.530	0.484	5
Pawlikowski, et al.	9.025	9.978	9.830	9.611	0.765	0.827	1.026	0.873	0.401	0.477	0.455	0.444	3
Jaganathan & Prakash	8.939	10.693	10.170	9.934	0.798	0.952	1.179	0.976	0.405	0.527	0.493	0.474	4
Fiorucci & Louzada	14.409	15.669	16.610	15.563	1.120	1.033	1.457	1.203	0.619	0.685	0.717	0.674	15
Petropoulos & Svetunkov	12.754	13.352	13.395	13.167	1.428	1.280	1.666	1.458	0.637	0.680	0.671	0.663	14

Shaub	12.185	14.651	13.561	13.466	1.288	1.286	1.542	1.372	0.593	0.717	0.652	0.653	13
Legaki & Koutsouri	17.287	17.201	18.213	17.567	2.604	2.116	2.952	2.557	0.998	0.988	1.043	1.012	24
Doornik, et al.	**8.036**	**9.729**	**8.973**	**8.913**	0.637	0.762	**1.003**	**0.801**	**0.348**	**0.454**	**0.428**	**0.410**	**1**
Pedregal, et al.	8.483	10.693	10.118	9.765	0.805	1.007	1.336	1.049	0.394	0.540	0.522	0.485	6
Spiliotis & Assimakopoulos	17.683	17.233	18.351	17.756	1.842	1.477	2.105	1.808	0.855	0.835	0.886	0.860	18
Roubinchtein	11.802	13.632	13.179	12.871	0.977	1.058	1.354	1.129	0.519	0.634	0.606	0.586	10
Ibrahim	18.084	17.614	18.581	18.093	2.463	2.073	2.627	2.388	0.991	0.989	0.991	0.991	21
Tartu M4 seminar	12.777	14.072	14.703	13.851	0.927	0.984	1.262	1.058	0.536	0.629	0.629	0.598	11
Waheeb	10.136	12.618	13.386	12.047	0.673	0.783	1.127	0.861	0.413	0.540	0.569	0.507	8
Darin & Stellwagen	10.514	12.457	12.079	11.683	0.692	0.797	1.079	0.856	0.427	0.539	0.525	0.496	7
Dantas & Cyrino Oliveira	16.162	16.908	17.752	16.941	1.470	1.431	1.893	1.598	0.738	0.815	0.830	0.794	16
Theta - Benchmark	18.095	17.605	18.714	18.138	2.490	2.065	2.809	2.455	0.997	0.987	1.029	1.006	23
Comb - Benchmark	18.842	21.165	26.153	22.053	3.017	3.577	7.152	4.582	1.124	1.449	2.051	1.556	27
ARIMA - Standard for comparison	12.070	15.110	14.760	13.980	0.704	0.860	1.264	0.943	0.471	0.628	0.631	0.577	9
Damped - Benchmark	18.480	18.800	20.516	19.265	2.730	2.488	3.649	2.956	1.056	1.122	1.236	1.141	26
ETS - Standard for comparison	15.301	17.138	19.481	17.307	1.410	1.611	2.450	1.824	0.703	0.865	0.982	0.852	17
Holt - Benchmark	20.955	29.331	37.462	29.249	4.211	7.830	16.025	9.356	1.423	2.699	4.036	2.749	28
SES - Benchmark	18.086	17.618	18.579	18.094	2.461	2.074	2.619	2.385	0.991	0.989	0.990	0.990	20
Naïve 2 - Benchmark	18.335	17.940	18.873	18.383	2.472	2.081	2.633	2.395	1.000	1.000	1.000	1.000	22
Naïve 1 - Benchmark	46.269	37.334	45.406	43.003	11.931	8.838	14.054	11.608	3.675	3.164	3.872	3.593	29
Naïve S - Benchmark	11.930	14.476	15.331	13.912	0.915	1.122	1.542	1.193	0.511	0.673	0.699	0.627	12
RNN - Benchmark	13.393	15.167	15.532	14.698	2.253	2.586	4.305	3.048	0.821	1.044	1.229	1.036	25
MLP - Benchmark	12.428	14.894	14.203	13.842	2.383	2.407	3.030	2.607	0.821	0.994	0.952	0.921	19

Appendix B: Forecasting performance for data frequencies and forecasting horizons: Prediction Intervals (PIs)

This appendix provides the analytical results of the M4 Competition in terms of the precision of PIs. The results are presented per data frequency and forecasting horizon (short, medium and long-term), displaying the values of MSIS and ACD (averages per case). For reasons of simplicity, only the methods that performed better than the Naïve 1 are considered, as well as the standards for comparison. The order of the methods follows the final ranks of the competition. Yet, the last column indicates the rank of the methods for whole sample of the examined frequency (according to MSIS) to facilitate comparisons.

Table B1: Forecasting performance achieved per forecasting horizon – Prediction Intervals: Yearly data.

Method	MSIS				ACD				Rank
	Short (1-2)	*Medium (3-4)*	*Long (5-6)*	*Total*	*Short (1-2)*	*Medium (3-4)*	*Long (5-6)*	*Total*	
Smyl	**14.255**	**23.959**	**33.482**	**23.898**	**0.015**	**0.002**	**0.004**	**0.003**	**1**
Montero-Manso, et al.	16.530	27.434	38.467	27.477	0.032	0.005	0.006	0.014	2
Doornik, et al.	17.368	28.767	44.466	30.200	0.070	0.016	0.024	0.037	3
ETS - Standard for comparison	18.985	34.775	50.938	34.900	0.106	0.111	0.115	0.111	4
Fiorucci & Louzada	20.587	35.871	51.074	35.844	0.162	0.162	0.168	0.164	5
Petropoulos & Svetunkov	21.675	35.747	50.412	35.945	0.171	0.165	0.176	0.171	6
Roubinchtein	21.633	37.953	53.531	37.706	0.153	0.168	0.180	0.167	7
Talagala, et al.	22.945	39.939	56.494	39.793	0.163	0.162	0.170	0.165	8
ARIMA - Standard for comparison	27.976	44.712	62.333	45.007	0.235	0.219	0.225	0.226	9
Ibrahim	22.329	44.953	67.801	45.028	0.175	0.215	0.244	0.211	10
Iqbal, et al.	25.295	51.622	83.491	53.469	0.181	0.213	0.249	0.214	12
Reilly	32.610	54.423	74.960	53.998	0.303	0.285	0.291	0.293	13
Wainwright, et al.	39.433	58.319	78.036	58.596	0.325	0.287	0.274	0.295	15
Segura-Heras, et al.	30.028	50.828	76.091	52.316	0.062	0.141	0.194	0.132	11
Naïve 1 - Benchmark	22.208	55.020	92.435	56.554	0.137	0.249	0.317	0.234	14

Table B2: Forecasting performance achieved per forecasting horizon – Prediction Intervals: Quarterly data.

Method	MSIS				ACD				Rank
	Short (1-2)	*Medium (3-5)*	*Long (6-8)*	*Total*	*Short (1-2)*	*Medium (3-5)*	*Long (6-8)*	*Total*	
Smyl	**5.320**	**8.193**	**11.062**	**8.551**	0.013	**0.007**	**0.004**	**0.004**	**1**

Montero-Manso, et al.	5.620	9.038	12.239	9.384	0.007	0.026	0.013	0.016	2
Doornik, et al.	5.769	9.197	13.219	9.848	**0.003**	0.022	0.057	0.029	5
ETS - Standard for comparison	5.473	8.792	12.765	9.452	0.005	0.018	0.026	0.018	4
Fiorucci & Louzada	5.601	8.814	12.572	9.420	0.037	0.054	0.069	0.056	3
Petropoulos & Svetunkov	5.811	9.238	13.269	9.893	0.042	0.062	0.080	0.064	6
Roubinchtein	5.906	9.251	13.333	9.945	0.026	0.046	0.068	0.049	7
Talagala, et al.	6.649	10.387	15.153	11.240	0.056	0.069	0.092	0.074	9
ARIMA - Standard for comparison	6.463	10.270	14.911	11.059	0.066	0.079	0.101	0.084	8
Ibrahim	6.388	10.394	15.439	11.284	0.055	0.084	0.108	0.086	10
Iqbal, et al.	7.473	10.362	14.776	11.295	0.033	0.044	0.060	0.047	11
Reilly	7.693	12.298	18.671	13.537	0.082	0.097	0.119	0.102	13
Wainwright, et al.	7.907	11.520	16.346	12.426	0.094	0.096	0.108	0.100	12
Segura-Heras, et al.	10.165	14.046	19.333	15.058	0.008	0.040	0.059	0.039	15
Naïve 1 - Benchmark	7.764	12.675	19.678	14.073	0.051	0.077	0.114	0.084	14

Table B3: Forecasting performance achieved per forecasting horizon – Prediction Intervals: Monthly data.

Method	**MSIS**				**ACD**				**Rank**
	Short (1-6)	***Medium (7-12)***	***Long (13-18)***	***Total***	***Short (1-6)***	***Medium (7-12)***	***Long (13-18)***	***Total***	
Smyl	**4.644**	**7.584**	**9.386**	**7.205**	0.003	**0.011**	**0.008**	**0.005**	**1**
Montero-Manso, et al.	5.586	8.673	11.710	8.656	0.023	0.016	0.008	0.016	6
Doornik, et al.	5.725	9.880	12.877	9.494	0.025	0.053	0.083	0.054	9
ETS - Standard for comparison	5.152	8.552	11.188	8.297	0.019	0.016	0.013	0.016	5
Fiorucci & Louzada	5.151	8.435	10.502	8.029	0.028	0.031	0.025	0.028	2
Petropoulos & Svetunkov	5.253	8.663	10.773	8.230	0.031	0.039	0.035	0.035	3
Roubinchtein	5.187	8.566	10.948	8.233	0.014	0.025	0.025	0.021	4
Talagala, et al.	6.107	10.182	13.186	9.825	0.042	0.059	0.069	0.056	11
ARIMA - Standard for comparison	5.478	9.159	11.626	8.754	0.032	0.043	0.043	0.039	7
Ibrahim	5.914	9.809	12.446	9.390	0.039	0.055	0.056	0.050	8
Iqbal, et al.	7.778	11.201	14.498	11.159	0.040	0.050	0.053	0.048	13
Reilly	6.175	10.171	14.686	10.344	0.048	0.057	0.065	0.057	12
Wainwright, et al.	6.350	10.027	12.765	9.714	0.053	0.056	0.058	0.056	10
Segura-Heras, et al.	7.322	11.463	14.875	11.220	**0.002**	0.020	0.019	0.014	14
Naive - Benchmark	7.654	12.683	16.564	12.300	0.021	0.022	0.025	0.023	15

Table B4: Forecasting performance achieved per forecasting horizon – Prediction Intervals: Weekly data.

Method	**MSIS**				**ACD**				**Rank**
	Short (1-4)	***Medium (5-9)***	***Long (10-13)***	***Total***	***Short (1-4)***	***Medium (5-9)***	***Long (10-13)***	***Total***	
Smyl	17.551	22.971	25.345	22.034	0.041	0.035	0.017	0.021	11
Montero-Manso, et al.	14.291	23.020	26.910	21.531	0.033	0.047	0.039	0.040	10
Doornik, et al.	**11.286**	17.373	20.522	16.469	0.042	0.069	0.043	0.053	2

ETS - Standard for comparison	12.783	20.982	27.243	20.386	0.022	0.008	0.014	0.014	8
Fiorucci & Louzada	12.004	18.546	21.351	17.396	0.008	**0.000**	0.019	0.008	5
Petropoulos & Svetunkov	12.560	**17.144**	**18.660**	**16.200**	0.003	0.004	0.031	0.012	**1**
Roubinchtein	11.549	19.086	21.590	17.538	0.011	0.021	0.026	0.003	6
Talagala, et al.	13.394	21.655	27.280	20.844	**0.000**	0.024	0.007	0.011	9
ARIMA - Standard for comparison	12.726	20.431	25.354	19.575	0.005	0.015	**0.006**	0.005	7
Ibrahim	14.949	24.721	30.091	23.366	0.006	0.016	0.017	0.013	13
Iqbal, et al.	11.414	18.356	22.158	17.390	0.013	0.011	0.038	0.012	4
Reilly	12.880	17.159	20.602	16.902	0.020	0.003	0.010	0.004	3
Wainwright, et al.	22.372	25.376	35.090	27.441	0.038	0.034	0.018	0.030	15
Segura-Heras, et al.	14.948	23.338	28.561	22.364	0.018	0.037	0.047	0.034	12
Naïve 1 - Benchmark	17.205	27.958	33.510	26.358	0.014	0.002	0.014	**0.001**	14

Table B5: Forecasting performance achieved per forecasting horizon – Prediction Intervals: Daily data.

Method	**MSIS**				**ACD**				**Rank**
	Short (1-4)	***Medium (5-9)***	***Long (10-14)***	***Total***	***Short (1-4)***	***Medium (5-9)***	***Long (10-14)***	***Total***	
Smyl	14.705	**24.156**	**37.672**	**26.283**	0.005	0.011	0.005	0.004	**1**
Montero-Manso, et al.	18.411	33.376	48.156	34.379	0.023	0.032	0.023	0.026	11
Doornik, et al.	15.544	26.585	42.551	29.133	0.064	0.025	0.038	0.041	3
ETS - Standard for comparison	15.439	27.363	43.445	29.700	0.006	0.002	**0.004**	**0.001**	5
Fiorucci & Louzada	15.543	26.895	42.409	29.192	0.012	0.008	0.012	0.010	4
Petropoulos & Svetunkov	16.481	27.929	43.324	30.156	0.021	0.015	0.019	0.018	6
Roubinchtein	18.190	29.631	47.610	32.783	**0.004**	**0.000**	0.006	0.004	9
Talagala, et al.	21.094	33.371	51.569	36.363	0.029	0.034	0.051	0.039	12
ARIMA - Standard for comparison	16.851	29.742	47.640	32.451	0.021	0.016	0.023	0.020	7
Ibrahim	44.005	55.910	72.086	58.286	0.013	0.010	0.014	0.012	15
Iqbal, et al.	**14.462**	26.384	41.745	28.464	0.024	0.029	0.044	0.033	2
Reilly	22.220	35.221	54.623	38.436	0.064	0.050	0.045	0.052	14
Wainwright, et al.	19.491	31.265	47.062	33.543	0.038	0.026	0.030	0.031	10
Segura-Heras, et al.	22.324	34.237	50.034	36.475	0.039	0.062	0.073	0.059	13
Naïve 1 - Benchmark	17.374	30.432	46.813	32.552	0.023	0.018	0.021	0.021	8

Table B6: Forecasting performance achieved per forecasting horizon – Prediction Intervals: Hourly data.

Method	**MSIS**				**ACD**				**Rank**
	Short (1-16)	***Medium (17-32)***	***Long (33-48)***	***Total***	***Short (1-16)***	***Medium (17-32)***	***Long (33-48)***	***Total***	
Smyl	6.143	7.579	10.062	7.928	0.024	0.042	0.055	0.040	3
Montero-Manso, et al.	10.595	14.108	30.823	18.509	0.043	0.041	0.004	0.029	11
Doornik, et al.	4.992	5.832	7.592	**6.139**	**0.000**	**0.002**	0.008	0.002	**1**
ETS - Standard for comparison	11.685	15.107	25.669	17.487	0.022	0.021	0.005	0.001	10
Fiorucci & Louzada	7.509	8.651	12.154	9.438	0.015	0.040	0.044	0.033	5

Petropoulos & Svetunkov	11.519	13.067	16.090	13.559	0.004	0.034	0.041	0.024	8
Roubinchtein	8.720	9.988	13.617	10.775	0.043	0.005	**0.002**	0.015	6
Talagala, et al.	61.105	49.938	69.558	60.200	0.229	0.130	0.168	0.176	13
ARIMA - Standard for comparison	5.754	7.514	9.344	7.538	0.008	0.012	0.016	0.001	2
Ibrahim	20.370	16.233	23.218	19.940	0.062	0.031	0.040	0.003	12
Iqbal, et al.	88.347	71.559	108.528	89.478	0.265	0.248	0.242	0.251	15
Reilly	9.953	11.774	13.501	11.743	0.037	0.005	0.025	0.022	7
Wainwright, et al.	6.845	8.332	10.491	8.556	0.002	0.009	0.011	**0.000**	4
Segura-Heras, et al.	16.104	12.781	16.175	15.020	0.024	0.017	0.005	0.001	9
Naïve 1 - Benchmark	55.402	69.398	88.935	71.245	0.124	0.043	0.047	0.011	14

Appendix C: Best Performing methods: Data frequencies & application domains

This appendix provides information regarding the methods displaying the best results for data frequencies and domains. The Tables list the top 3 performing methods for both the accuracy of PFs and the precision of PIs.

Table C1: Methods providing the best results per data frequency (OWA and MSIS are used to evaluate the PFs and PIs, respectively)

Accuracy measure	Yearly (23,000)	Quarterly (24,000)	Monthly (48,000)	Weekly (359)	Daily (4,227)	Hourly (414)
OWA	Smyl	Montero-Manso, et al.	Smyl	Darin & Stellwagen	Pawlikowski, et al.	Doornik, et al.
	Legaki & Koutsouri	Smyl	Jaganathan & Prakash	Petropoulos & Svetunkov	Tartu M4 seminar	Smyl
	Montero-Manso, et al.	Petropoulos & Svetunkov	Montero-Manso, et al.	Pawlikowski, et al.	Fiorucci & Louzada	Pawlikowski, et al.
MSIS	Smyl	Smyl	Smyl	Petropoulos & Svetunkov	Smyl	Doornik, et al.
	Montero-Manso, et al.	Montero-Manso, et al.	Fiorucci & Louzada	Doornik, et al.	Iqbal, et al.	ARIMA
	Doornik, et al.	Fiorucci & Louzada	Petropoulos & Svetunkov	Reilly	Doornik, et al.	Smyl

Table C2: Methods providing the best results per data domain. (OWA and MSIS are used to evaluate the PFs and PIs, respectively)

Accuracy measure	Macro (19,402)	Micro (25,121)	Demographic (8,708)	Industry (18,798)	Finance (24,534)	Other (3,437)
OWA	Smyl	Smyl	Montero-Manso, et al.	Montero-Manso, et al.	Smyl	Smyl
	Jaganathan & Prakash	Legaki & Koutsouri	Smyl	Smyl	Montero-Manso, et al.	Pawlikowski, et al.
	Montero-Manso, et al.	Pawlikowski, et al.	Pawlikowski, et al.	Jaganathan & Prakash	Fiorucci & Louzada	Montero-Manso, et al.
MSIS	Smyl	Smyl	Smyl	Smyl	Smyl	Smyl
	Montero-Manso, et al.	Montero-Manso, et al.	Fiorucci & Louzada	Montero-Manso, et al.	Doornik, et al.	Montero-Manso, et al.
	ETS	Fiorucci & Louzada	Svetunkov, et al.	Doornik, et al.	Montero-Manso, et al.	Doornik, et al.

Table C3: Methods providing the best results for data frequencies and domains (OWA is used to evaluate the PFs)

Accuracy measure	Macro (19,402)	Micro (25,121)	Demographic (8,708)	Industry (18,798)	Finance (24,534)	Other (3,437)
Yearly (23,000)	Smyl	Legaki & Koutsouri	Montero-Manso, et al.	Smyl	Smyl	Smyl
	Legaki & Koutsouri	Ibrahim	Smyl	Montero-Manso, et al.	Montero-Manso, et al.	Montero-Manso, et al.
	Petropoulos & Svetunkov	Shaub	Petropoulos & Svetunkov	Doornik, et al.	Petropoulos & Svetunkov	Legaki & Koutsouri
Quarterly (24,000)	Montero-Manso, et al.	Smyl	Montero-Manso, et al.	Montero-Manso, et al.	Smyl	Smyl
	Smyl	Montero-Manso, et al.	Pawlikowski, et al.	Petropoulos & Svetunkov	Montero-Manso, et al.	Jaganathan & Prakash

	Pawlikowski, et al.	Petropoulos & Svetunkov	Petropoulos & Svetunkov	Fiorucci & Louzada	Petropoulos & Svetunkov	Montero-Manso, et al.
Monthly (48,000)	Smyl	Smyl	Montero-Manso, et al.	Montero-Manso, et al.	Pawlikowski, et al.	Montero-Manso, et al.
	Jaganathan & Prakash	Doornik, et al.	Shaub	Jaganathan & Prakash	Jaganathan & Prakash	Smyl
	Montero-Manso, et al.	Jaganathan & Prakash	Smyl	Pawlikowski, et al.	Montero-Manso, et al.	Waheeb
Weekly (359)	Petropoulos & Svetunkov	Tartu M4 seminar	Smyl	Bandara, et al.	Fritschi	Doornik, et al.
	Darin & Stellwagen	Montero-Manso, et al.	Trotta	Shaub	Darin & Stellwagen	Smyl
	Jaganathan & Prakash	Svetunkov, et al.	Doornik, et al.	Svetunkov, et al.	Montero-Manso, et al.	Jaganathan & Prakash
Daily (4,227)	Pawlikowski, et al.	Pawlikowski, et al.	Pawlikowski, et al.	Pawlikowski, et al.	Pawlikowski, et al.	Pawlikowski, et al.
	Tartu M4 seminar	Tartu M4 seminar	Tartu M4 seminar	Fiorucci & Louzada	Fritschi	Doornik, et al.
	Jaganathan & Prakash	Alves Santos Junior	Nikzad	Petropoulos & Svetunkov	Bandara, et al.	Pedregal, et al.
Hourly (414)	-	-	-	-	-	Doornik, et al.
	-	-	-	-	-	Smyl
	-	-	-	-	-	Pawlikowski, et al.

Table C4: Methods providing the best results for data frequencies and domains (MSIS is used to evaluate the PIs)

Accuracy measure	Macro (19,402)	Micro (25,121)	Demographic (8,708)	Industry (18,798)	Finance (24,534)	Other (3,437)
Yearly (23,000)	Smyl	Smyl	Smyl	Smyl	Smyl	Smyl
	Montero-Manso, et al.	Montero-Manso, et al.	Montero-Manso, et al.	Montero-Manso, et al.	Doornik, et al.	Montero-Manso, et al.
	Doornik, et al.	Doornik, et al.	Doornik, et al.	Doornik, et al.	Montero-Manso, et al.	Talagala, et al.
Quarterly (24,000)	Smyl	Smyl	Smyl	Smyl	Smyl	Smyl
	Montero-Manso, et al.	Montero-Manso, et al.	ETS	Fiorucci & Louzada	Montero-Manso, et al.	Svetunkov, et al.
	Fiorucci & Louzada	ETS	Fiorucci & Louzada	Svetunkov, et al.	Svetunkov, et al.	Fiorucci & Louzada
Monthly (48,000)	Smyl	Smyl	Smyl	Smyl	Smyl	Smyl
	Roubinchtein	Fiorucci & Louzada	Fiorucci & Louzada	Fiorucci & Louzada	Fiorucci & Louzada	Fiorucci & Louzada
	ETS	Petropoulos & Svetunkov	Svetunkov, et al.	Montero-Manso, et al.	Petropoulos & Svetunkov	Petropoulos & Svetunkov
Weekly (359)	Petropoulos & Svetunkov	Reilly	Smyl	Talagala, et al.	Reilly	Doornik, et al.
	Roubinchtein	Petropoulos & Svetunkov	Trotta	Petropoulos & Svetunkov	Iqbal, et al.	Petropoulos & Svetunkov
	Fiorucci & Louzada	Doornik, et al.	Doornik, et al.	Svetunkov, et al.	Petropoulos & Svetunkov	Selamlar
Daily (4,227)	Smyl	Doornik, et al.	Svetunkov, et al.	Smyl	Smyl	Smyl
	Iqbal, et al.	Smyl	Montero-Manso, et al.	Montero-Manso, et al.	Fiorucci & Louzada	Doornik, et al.
	Petropoulos & Svetunkov	Iqbal, et al.	Wainwright, et al.	ETS	Iqbal, et al.	Iqbal, et al.
Hourly (414)	-	-	-	-	-	Doornik, et al.
	-	-	-	-	-	ARIMA
	-	-	-	-	-	Smyl

Appendix D: The performance of methods in comparison to the Comb and Naïve 1 benchmarks.

The top performing methods are compared to the Comb benchmarks used for PFs and the Naïve 1 used for the PIs. The comparisons indicate the percentage of series where the single benchmarks provide better forecasts than the examined methods, either in terms of PFs (using sMAPE) or PIs (using MSIS).

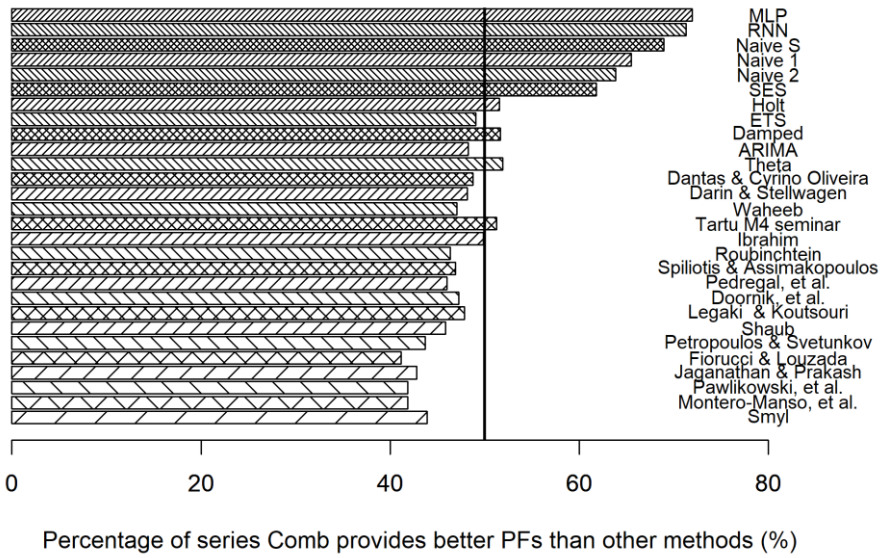

Figure D1: Percentage of series where Comb is better than other methods (%): Evaluation of Point Forecasts using sMAPE.

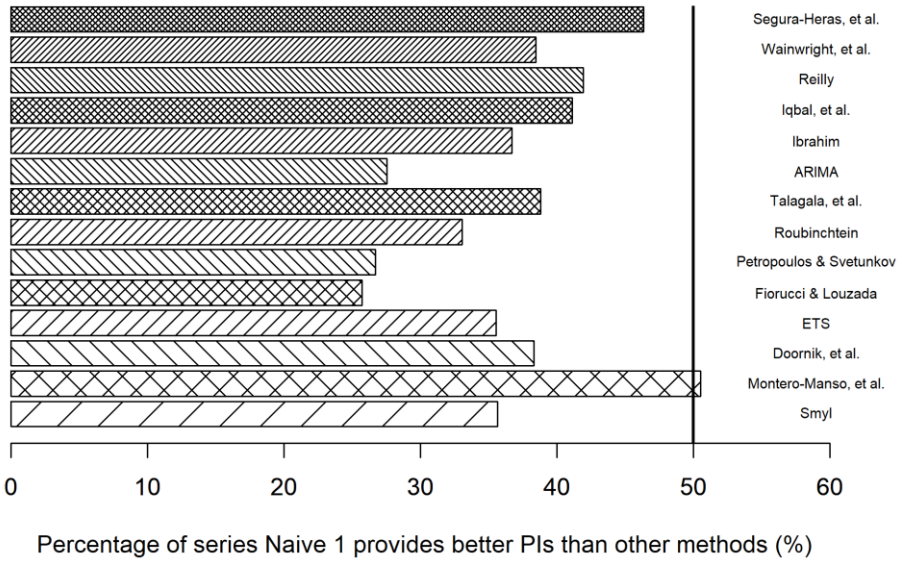

Figure D2: Percentage of series where Naïve is better than other methods (%): Evaluation of Prediction Intervals using MSIS.

Appendix E: The benchmarks of the M4 Competition.

This appendix provides additional information about the statistical and ML methods used in the M4 Competition. This includes the software, packages and functions utilized for their estimation, as well as the pre-processing options considered.

The statistical benchmarks, along with the two standards of comparison, were estimated using the v8.2 of the *forecast* package for R (Hyndman, 2017). The ML benchmarks were developed in Python using the Scikit v0.19.1, the Keras v2.0.9 and the TensorFlow v1.4.0 libraries.

- **Naïve 1:** This is equivalent to a random walk model, assuming that future values are all the same as that of the last known observation. The method was implemented using the naïve() function of the *forecast* package for R (default settings). Naïve 1 is expressed by the following equation

$$F_t = Y_{t-1}$$

where Y_t and F_t are the actual and forecasted values at point *t*, respectively.

- **Naïve S:** The forecasts of the Seasonal Naïve method at point *t* are equal to the last known observation of the same period. The method was implemented using the *naive_seasonal()* function of the M4 GitHub repository, which is equivalent to the *snaive()* one of the forecast package or an ARIMA(0,0,0)(0,1,0)m model, where *m* is the seasonal period. Naïve S is expressed by the following equation

$$F_t = Y_{t-m}$$

Note that *m* is defined in Section 3.1 as being 12 for monthly, 4 for quarterly, 24 for hourly and 1 for yearly, weekly and daily data.

- **Naïve 2:** This method is equivalent to the Naïve 1 method, applied on seasonally adjusted data. In order to decide whether a series should be seasonally adjusted or not, a seasonality test is first performed to check for autocorrelation significance on the m_{th} term of the ACF. Thus, given a series of $n \geq 3m$ observations, frequency $m > 1$ and a confidence level of 90%, seasonal adjustments are considered only if the following rule is fulfilled:

$$|ACF_m| > 1.645 \sqrt{\frac{1 + 2(ACF_1 + \sum_{i=2}^{m-1} ACF_i^2)}{n}}$$

Non-seasonal series (m=1) and series where observations are fewer in number than three seasonal periods are not tested and assumed as not being seasonal.

Having identified a seasonal series, the classical multiplicative decomposition is then applied to estimate its seasonal component. To do so, the *decompose()* function of the *stats* package for R is used. More specifically, the trend component is first determined by applying a moving average of *m*. Then, the seasonal component is computed (seasonal indexes) by dividing the data from the trend and averaging the results, for each time unit, over all periods.

The seasonally adjusted series is computed by simply dividing the original series with the seasonal indexes of the respective period. After extrapolating the seasonally adjusted series, the non-seasonal forecasts can be re-seasonalized by simply multiplying their values with the respective indexes.

- **SES:** Single Exponential Smoothing was implemented through the *ses()* function of the *forecast* package for R (default settings). The initial level of the method was optimized along with the smoothing coefficient (Hyndman, 2017). This is equivalent to the ANN *ets* model (Hyndman et al., 2002), seasonally adjusted as Naïve 2. SES is expressed by the following equations

$$F_t = l_{t-1}$$
$$l_t = l_{t-1} + \alpha e_t$$

where l_t is the level of the series at point t, e_t the one-step-ahead error and α the smoothing coefficient.

- **Holt:** Holt exponential smoothing was implemented through the *holt()* function of the *forecast* package. As done with SES, the default settings were used and all the parameters (initial levels and smoothing coefficients) were optimized simultaneously. This is equivalent to the AAN *ets* model, seasonally adjusted as Naïve 2. Holt is expressed by the following equations

$$F_t = l_{t-1} + b_{t-1}$$
$$l_t = l_{t-1} + b_{t-1} + \alpha e_t$$
$$b_t = b_{t-1} + \beta e_t$$

where l_t and b_t are the level and slope of the series at point t, respectively, and α *and* β are the smoothing coefficients.

- **Damped:** Damped exponential smoothing uses the same function with Holt, but with a *damped*=TRUE argument. This is equivalent to the AAdN *ets* model, seasonally adjusted as Naïve 2. Damped is expressed by the following equations

$$F_t = l_{t-1} + \varphi b_{t-1}$$
$$l_t = l_{t-1} + \varphi b_{t-1} + \alpha e_t$$
$$b_t = \varphi b_{t-1} + \beta e_t$$

where φ is the dampening parameter.

- **Theta:** Theta is implemented as originally proposed by Assimakopoulos & Nikolopoulos (2000) using the *Theta.classic()* function of the M4 GitHub repository. Two Theta lines with ϑ coefficients of 0 and 2 are first calculated. The first line is forecasted by extrapolating the linear regression line in time, while the second one using SES. The

forecasts are then combined using equal weights. Seasonal adjustments are performed like Naïve 2. Note that this method is not equivalent to the *thetaf()* function of the forecast package, as the latter implements a "SES with drift" approach (Hyndman & Billah, 2003). The Theta line θ is estimated as follows

$$Y_t^{\theta} = \theta Y_t + (1-\theta)(b - at)$$

where *b* and *a* are the slope and intercept of the linear regression in time, respectively. The Theta method can then be expressed by the following equation

$$F_t = 0.5 * Y_t^0 + 0.5 * Y_t^2$$

- **Comb:** This is the simple arithmetic average of SES, Holt and Damped exponential smoothing, as implemented above.
- **MLP:** This is a single hidden layer NN of three input nodes, six hidden nodes and one output node implemented using scikit-learn. The Adam optimizer is used for estimating the optimal weights. The initial learning rate is 0.001 and is set to adapt during the training. Random initial weights are used for starting the training process with a maximum of 100 iterations. The activation function of all the layers is the linear one. Iterative forecasts are produced to generate predictions across the requested forecasting horizon. Seasonal adjustments are performed as in Naive 2, while de-trending is performed by subtracting the linear regression line in time from the deseasonalized data. The MLP is trained per series, being an ensemble of 30 individual networks, combined using the median operator. The architecture and the hyper-parameters of the model were selected based on the results of the Semenoglou et al. (2018) study.
- **RNN:** A sequential model from Keras is used to implement the RNN, comprised of two layers, a hidden one containing three recurrent nodes of six recurrent units and an output one containing one linear node. Regarding the hyper-parameters used, a number of 100 epochs was chosen and the learning ratio was set to 0.001, with the linear activation function being used in all nodes. Once again, iterative forecasts are produced to generate predictions across the requested forecasting horizon. Seasonal adjustments are performed as in Naive 2, while de-trending is performed by subtracting the linear regression line in time from the deseasonalized data. A single RNN model is trained per

series in order to accelerate the whole process. The architecture and the hyper-parameters of the model were selected based on the results of the Semenoglou et al. (2018) study.

References for Appendices

Hyndman, R. J., & Billah, B. (2003). Unmasking the theta method. *International Journal of Forecasting*, *19*(2), 287-290.

Semenoglou, A. A., Spiliotis, E., & Assimakopoulos, V. (2018). Batch forecasting utilizing ANNs: best practices and ways forward. *38th International Symposium on Forecasting*, June 17-20, 2018, Boulder Colorado, USA.