Supplementary material to "Atmospheric geopotentials from ERA5 linked to the daily maximum temperature record-breaking in Spain (1960–2023)"

Elsa Barrio-Torres[1], Jesús Abaurrea[1], Jesús Asín[1], Jorge Castillo-Mateo[1], Ana Carmen Cebrián[1], and Zeus Gracia-Tabuenca[1]

[1]Department of Statistical Methods, University of Zaragoza, Pedro Cerbuna 12, 50009 Zaragoza, Spain

Correspondence: Elsa Barrio-Torres (e.barrio@unizar.es)

S.1 Extended Exploratory data analysis

S.1.1 Time trends of atmospheric variables

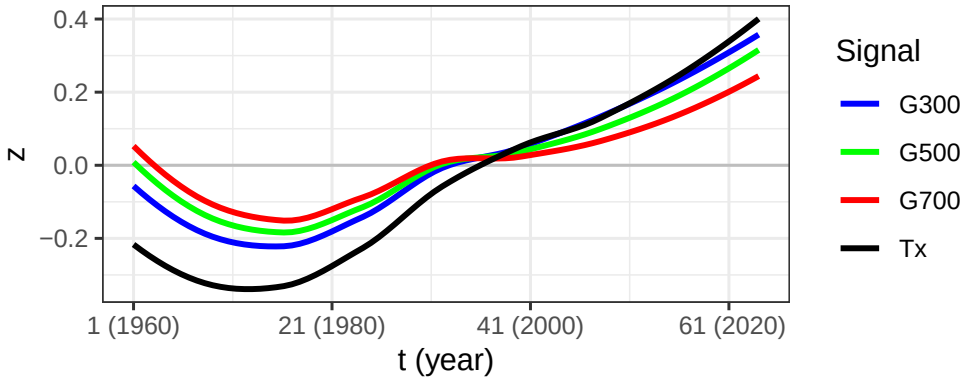

Figure S.1. LOESS curves of the spatial averages of the standardized T_x and geopotential variable series.

S.1.2 Correlations of atmospheric variables

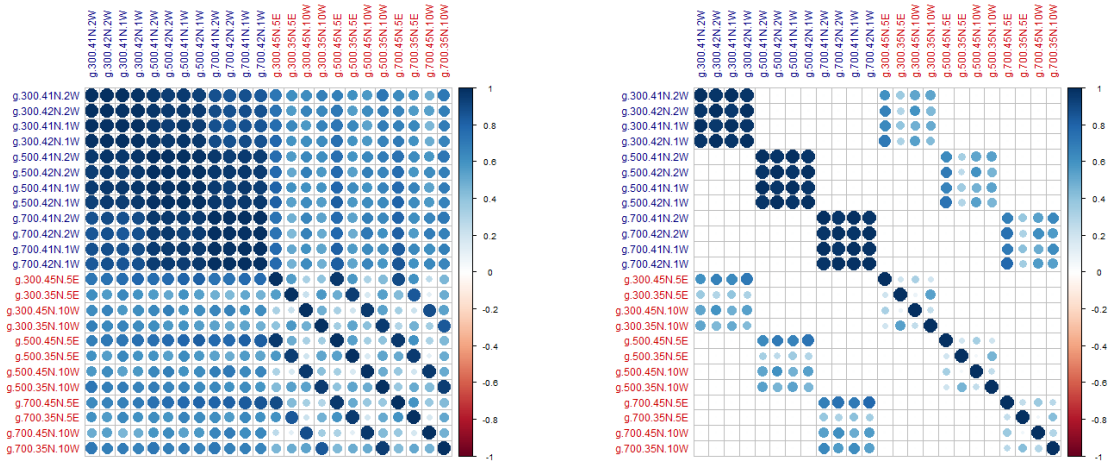

Figure S.2. Pearson's (left) and intra-class (right) cross-correlation plots of the geopotential altitude covariates at the three pressure levels in the four nearest and furthest grid points of Zaragoza station (41N,1W).

S.1.3 Boxplots of geopotential series conditioned on the occurrence or absence of record and geographical zone.

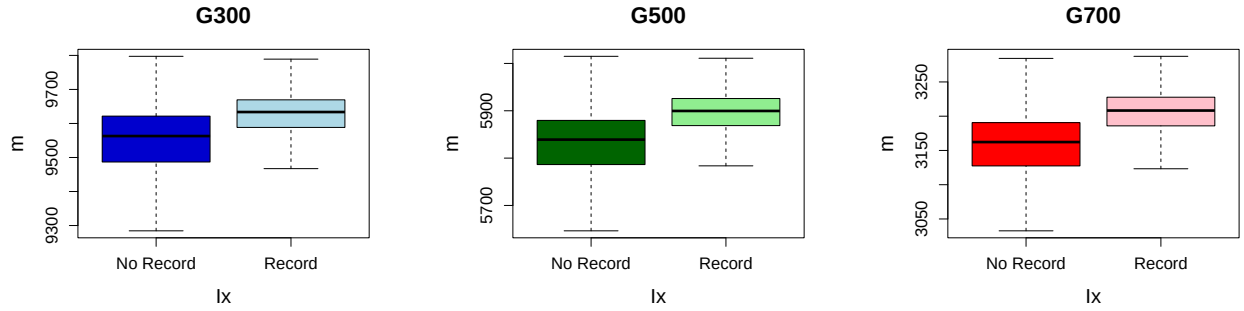

Figure S.3. Boxplots of geopotential variables at the 300, 500, and 700 hPa pressure levels (in geopotential meters) during the period 1984–2023, conditioned on the occurrence or non-occurrence of T_x records.

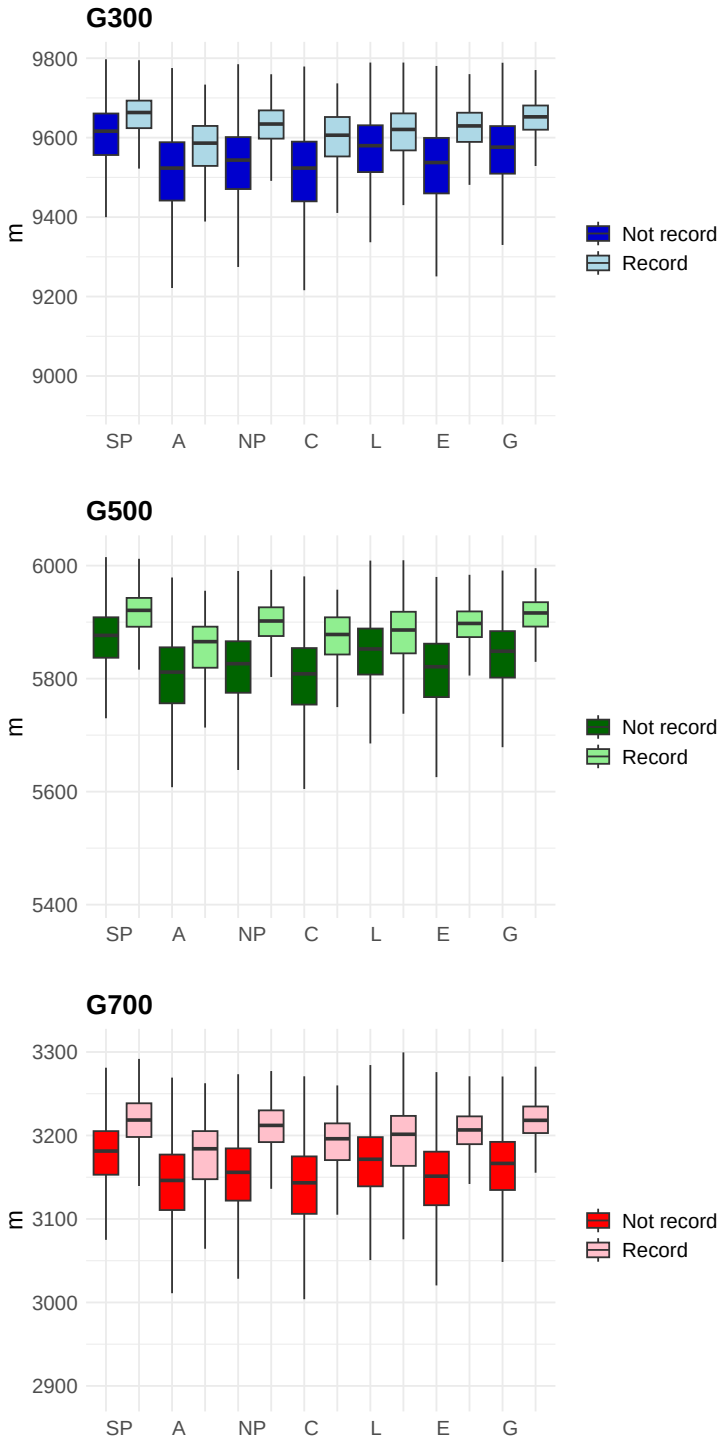

Figure S.4. Boxplots of geopotential series at pressure levels of 300, 500, and 700 hPa in the reference period 1984–2023 conditioned on occurrence or absence of T_x record and to the geographical region (A: Andalusia, C: Cantabrian coast, E: Ebro valley, G: Galicia, L: Levante, NP: North Plateau, and SP: South Plateau.

S.2 Extended Modeling Results

S.2.1 Local models

Table S.1. Number of parameters (k) and performance metrics (AUC, AIC) for each local model corresponding to each station. AIC was computed in the training period (1960–2010) and the AUC in the test period (2011–2023). Stations are sorted by distance to the coast in ascending order.

Station	k	AUC	AIC
Santander	20	0.89	2033
San Sebastián	17	0.92	1881
Almería	21	0.87	2325
A Coruña	25	0.82	2041
Barcelona-Airport	28	0.84	2352
Málaga	27	0.89	1961
Gijón	16	0.67	2160
Huelva	22	0.90	1873
Valencia	25	0.86	2227
Reus	21	0.81	2208
Castellón	29	0.86	2564
Barcelona-Fabra	20	0.86	2365
Bilbao	18	0.92	1967
Tortosa	22	0.83	2679
Santiago	15	0.93	1962
Murcia	26	0.89	1914
Vitoria	25	0.95	1871
Moron	19	0.95	1902
Sevilla	27	0.93	1711
Lleida	22	0.91	1976
Logroño	26	0.93	1950
León	23	0.94	1791
Ponferrada	15	0.95	1803
Burgos	22	0.96	2011
Albacete	17	0.92	1878
Badajoz	20	0.94	1971
Soria	34	0.94	1748
Daroca	23	0.94	1859
Zaragoza	28	0.90	2001
Valladolid	22	0.96	1812
Cáceres	29	0.95	1790
Zamora	23	0.95	1871
Ciudad Real	34	0.96	1763
Salamanca	25	0.94	1820
Navacerrada	26	0.95	1946
Madrid	27	0.94	1998

Table S.2. Number of surviving covariates with different restrictions.

	$s_t = 1/4(9)$	$s_t = 1/3(12)$	$s_t = 1/2(18)$
$z_t = 1.6$	37	30	19
$z_t = 2$	31	23	15
$z_t = 2.6$	19	16	8
$z_t = 3.2$	12	6	4

S.2.2 Summary of global model M1)

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-133.5485	8.6343	-15.47	0.0000
g700.	0.0011	0.0001	21.99	0.0000
g700.45N.10W	-0.0014	0.0001	-20.67	0.0000
poly(g300.35N.5E, 2)1	-181.7786	16.7657	-10.84	0.0000
poly(g300.35N.5E, 2)2	130.1354	5.3892	24.15	0.0000
g700..lag1	0.0018	0.0001	20.47	0.0000
g300.35N.10W.lag1	-0.0002	0.0000	-9.77	0.0000
poly(g700.35N.5E, 2)1	-31.4230	18.1093	-1.74	0.0827
poly(g700.35N.5E, 2)2	-135.7835	6.8580	-19.80	0.0000
g500.45N.10W	0.0007	0.0000	14.39	0.0000
g300.35N.10W	-0.0010	0.0000	-23.53	0.0000
poly(g700.45N.5E, 2)1	-367.4691	18.1819	-20.21	0.0000
poly(g700.45N.5E, 2)2	79.7605	4.0654	19.62	0.0000
g500.45N.5E	0.0012	0.0001	20.68	0.0000
g300.35N.5E.lag1	-0.0004	0.0000	-12.66	0.0000
g500.35N.10W	0.0016	0.0001	24.73	0.0000
g500..lag1	-0.0007	0.0001	-11.71	0.0000
g500.35N.5E	0.0014	0.0001	9.96	0.0000

As a complement, Figure S.5 in the Supplementary material summarizes the z-scores of the predictors included in the final local models.

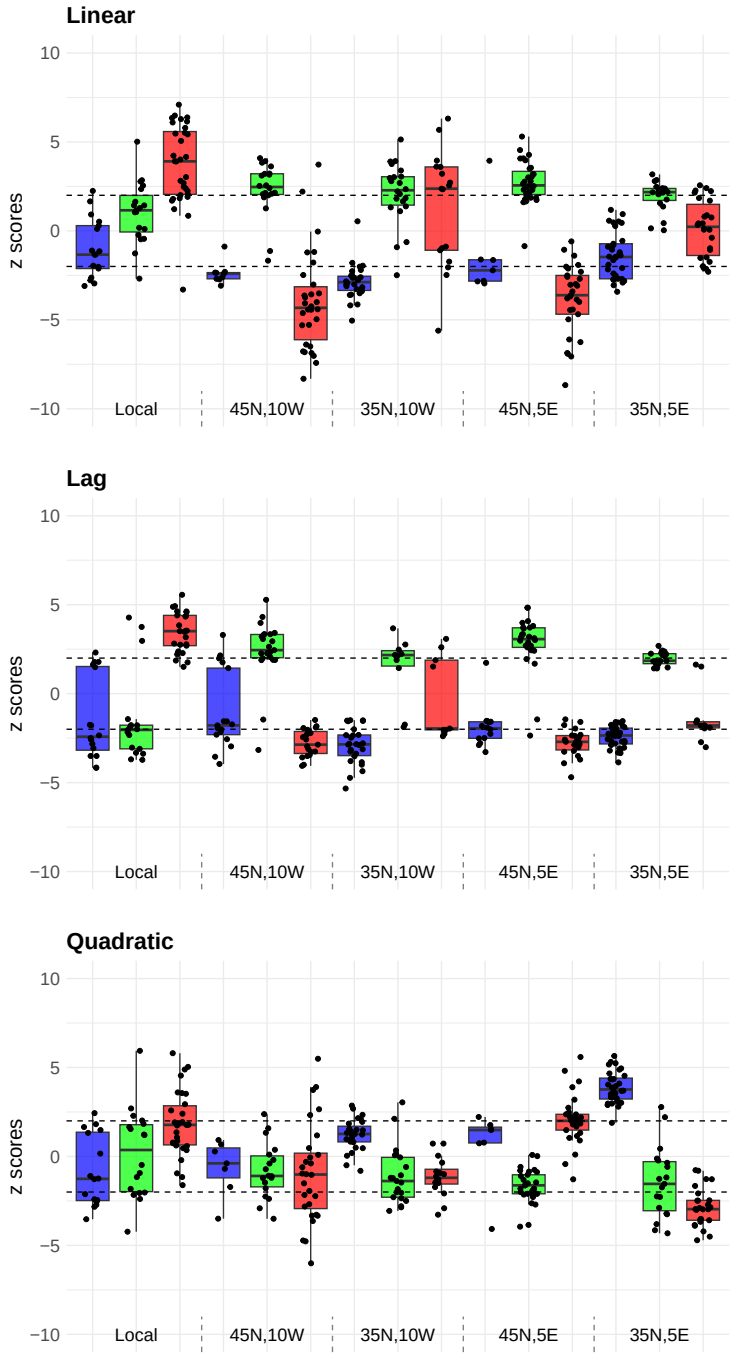

Figure S.5. Jittered boxplot of the z-scores of the estimated coefficients from the optimal local models. Color indicates preassure level in hPa: 300 (blue), 500 (green), 700 (red). x-axis indicates the five grid points: nearest (Local), north-west (45N.10W), south-west (35N.10W), north-east (45N.5E), south-east (35N.5E). Covariates: linear terms (top), 1-lagged terms (middle), quadratic terms (bottom). Two horizontal lines are shown at $y = 2$ and $y = -2$.

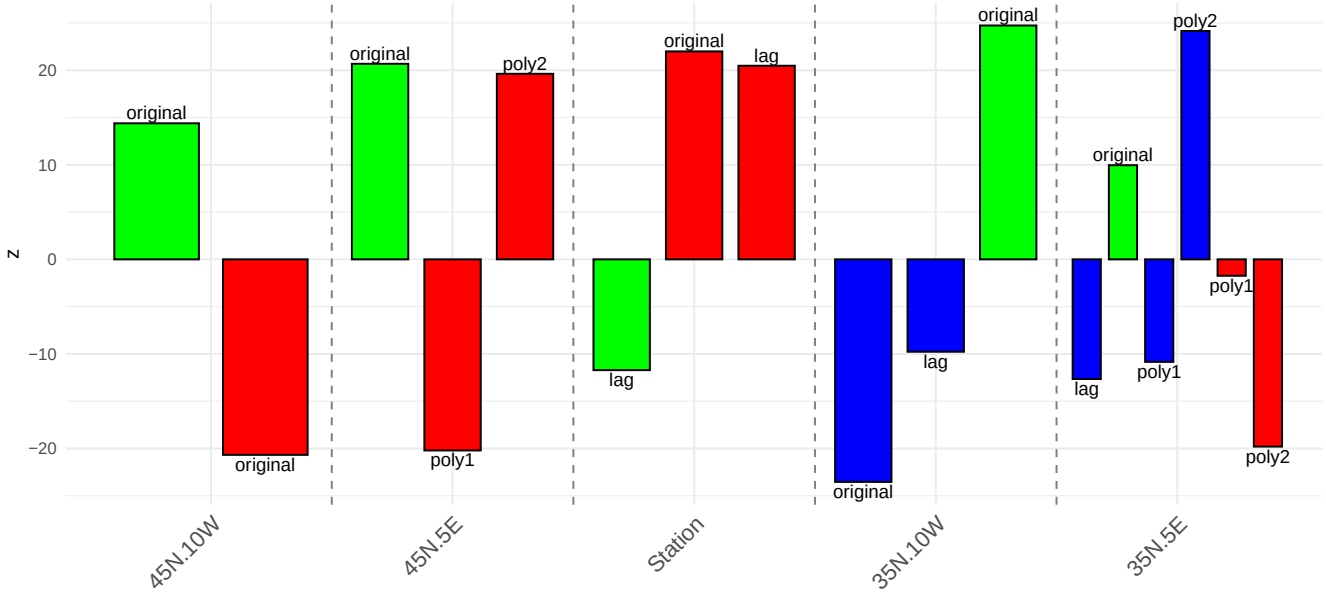

Figure S.6. Diagram with the 14 geopotential variables retained in M1 model. Variables followed by 'lag' represent one-day lagged terms, while those preceded by 'poly' correspond to the second-order polynomial of the corresponding variable. Top: pseudo-geographical representation of the four corners of the grid and the nearest point to a station. Bottom: standardized estimated coefficients (z-scores) of the model. Blue color is associated to 300 hPa, green to 500 hPa, and red to 700 hPa.

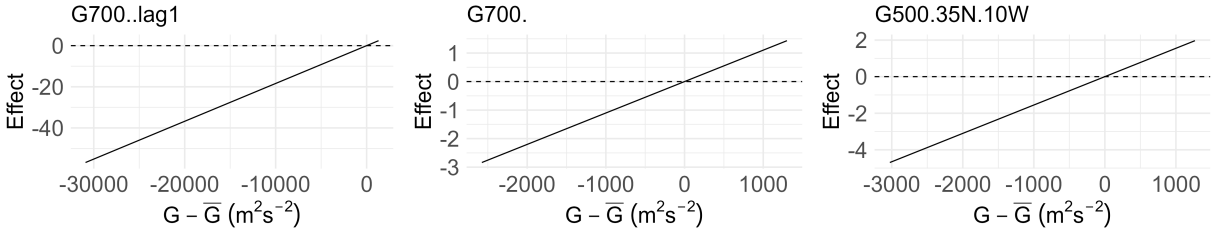

Figure S.7. Partial effects of the 700 hPa (column) and 500 hPa (35N.10W) predictors from the M1 model. The y-axis shows each variable multiplied by its corresponding regression coefficient and centered around the mean. The x-axis shows the observed values of each predictor centered around the mean.

S.2.3 Summary of atmospheric-geodetic model (M2)

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-485.3693	58.6888	-8.27	0.0000
g700.	0.0020	0.0001	22.20	0.0000
g700.45N.10W	0.0187	0.0014	13.61	0.0000
LAT	7.8238	1.4230	5.50	0.0000
poly(g300.35N.5E, 2)1	-204.4119	16.9461	-12.06	0.0000
poly(g300.35N.5E, 2)2	127.3015	5.4687	23.28	0.0000
g700..lag1	0.0019	0.0001	16.21	0.0000
g300.35N.10W.lag1	-0.0002	0.0000	-10.37	0.0000
poly(g700.35N.5E, 2)1	-61.6837	18.6022	-3.32	0.0009
poly(g700.35N.5E, 2)2	-128.0573	6.9334	-18.47	0.0000
poly(g700.45N.5E, 2)1	-3371.5517	334.9217	-10.07	0.0000
poly(g700.45N.5E, 2)2	-147.7684	78.3932	-1.88	0.0594
g500.45N.5E	0.0054	0.0011	4.79	0.0000
g300.35N.5E.lag1	-0.0003	0.0000	-10.57	0.0000
g500.35N.5E	-0.0022	0.0005	-4.25	0.0000
g300.35N.10W	-0.0007	0.0000	-15.21	0.0000
g500.45N.10W	-0.0049	0.0010	-5.07	0.0000
g500.35N.10W	0.0011	0.0001	16.32	0.0000
LON	-9.3112	1.0743	-8.67	0.0000
g700.45N.10W.lag1	-0.0006	0.0001	-7.17	0.0000
g500..lag1	-0.0008	0.0001	-10.83	0.0000
g500.45N.10W.lag1	-0.0034	0.0004	-8.18	0.0000
g500.45N.5E.lag1	0.0037	0.0004	9.08	0.0000
g700.45N.5E.lag1	-0.0003	0.0001	-2.62	0.0089
g700.45N.10W:LAT	-0.0005	0.0000	-14.21	0.0000
LAT:poly(g700.45N.5E, 2)1	69.1298	8.1977	8.43	0.0000
LAT:poly(g700.45N.5E, 2)2	5.6440	1.9189	2.94	0.0033
LAT:g500.45N.10W	0.0001	0.0000	5.25	0.0000
poly(g700.45N.5E, 2)1:LON	-73.6749	6.3212	-11.66	0.0000
poly(g700.45N.5E, 2)2:LON	5.9789	1.3493	4.43	0.0000
g300.35N.10W:LON	0.0000	0.0000	7.29	0.0000
g700.:LON	-0.0001	0.0000	-9.99	0.0000
g700.45N.10W:LON	0.0002	0.0000	9.05	0.0000
LON:g500.45N.10W.lag1	-0.0001	0.0000	-9.38	0.0000
LAT:g500.45N.10W.lag1	0.0001	0.0000	8.81	0.0000
g500.45N.5E:LON	0.0001	0.0000	6.75	0.0000
LAT:g500.35N.5E	0.0001	0.0000	7.25	0.0000
LAT:g500.45N.5E	-0.0001	0.0000	-3.79	0.0002
LON:g700.45N.5E.lag1	0.0001	0.0000	12.22	0.0000
LAT:g500.45N.5E.lag1	-0.0001	0.0000	-7.76	0.0000
g500.45N.10W:LON	-0.0001	0.0000	-4.25	0.0000

S.2.4 Alternative global model

Table S.3. Similar to Table 2, but for the alternative model to M1.

Model	k_{coeff}	n_v	AIC	AUC	AUC_I	AUC_C	Min_C	$Q50_C$	$Q10\Delta$
m1	18	15	97808.4	0.86	0.91	0.78	0.59	0.79	0.22
Step	18	12	98337.7	0.84	0.90	0.75	0.56	0.77	0.26

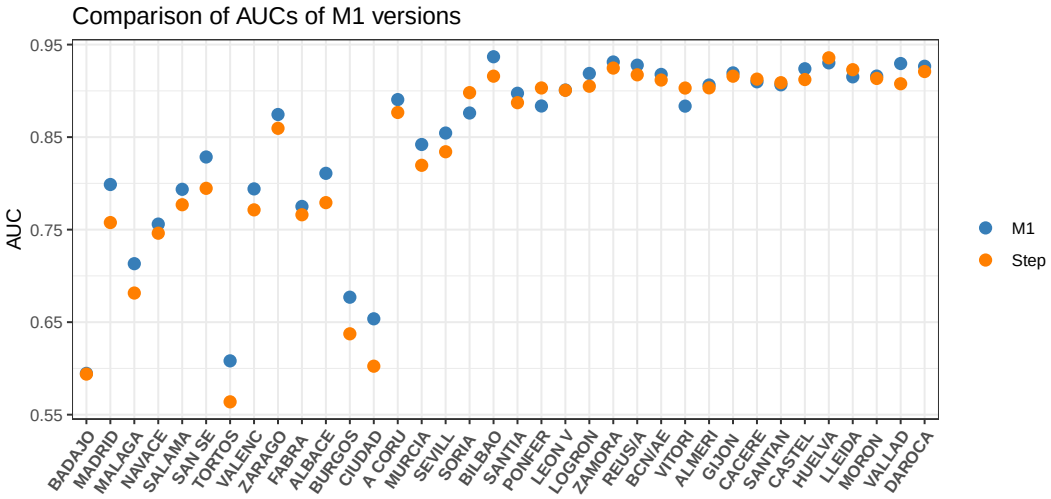

Figure S.8. Comparison of AUC values across stations for M1 and Step. M1 is a global model that includes only geopotential variables identified during Step 2 of the modeling algorithm. Step is an unsupervised model constructed by skipping Step 2.

S.2.5 State of the atmosphere

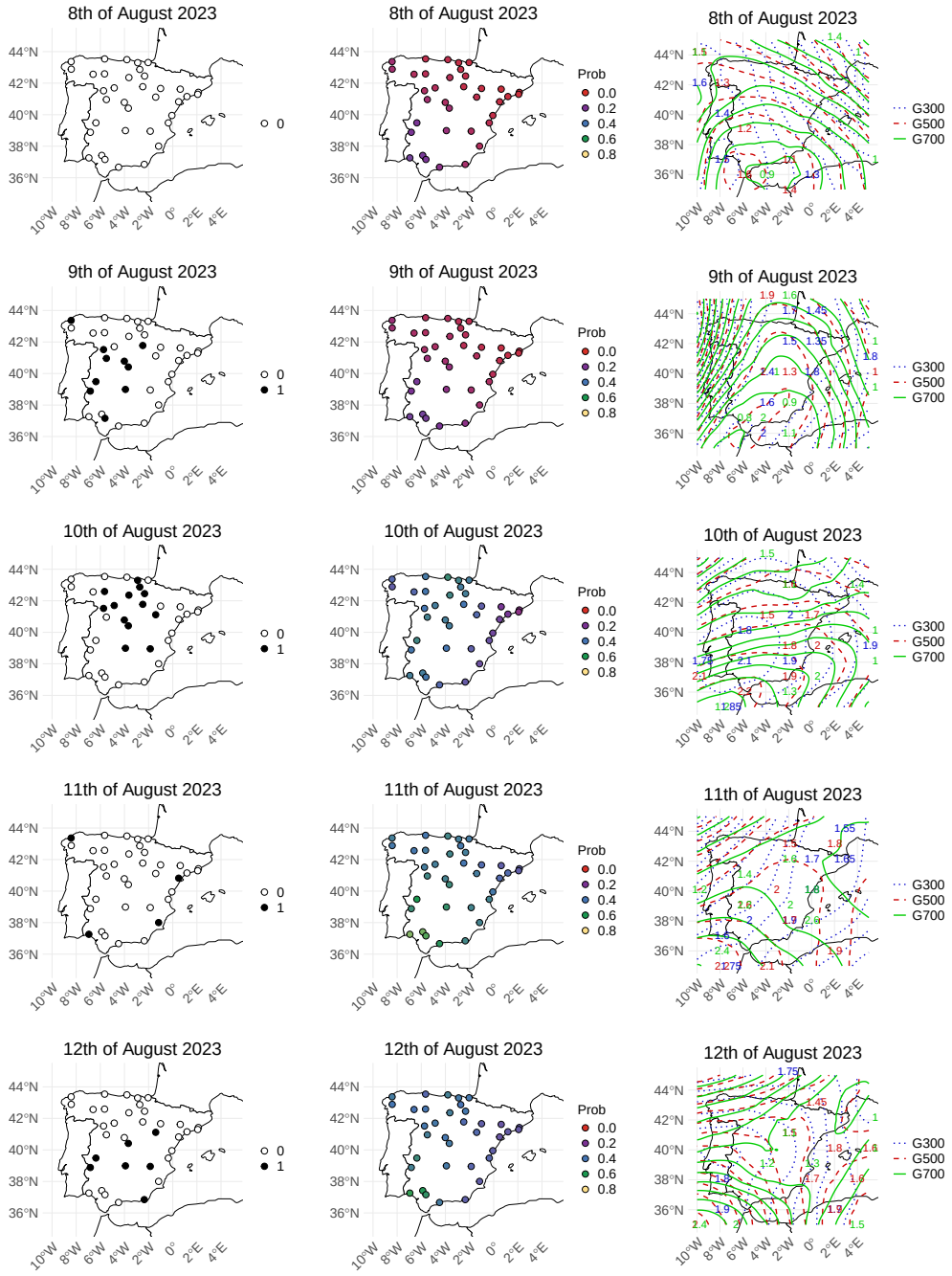

Figure S.9. Each row corresponds to a day of the heatwave from 08–12 August 2023 (validation period). Left column: indicates whether each station experienced a record (black points = records, white points = no record). Middle column: shows the record probability predicted by model M2 at each station. Right column: displays geopotential composites for the specific day at pressure levels of 300 hPa, 500 hPa, and 700 hPa.

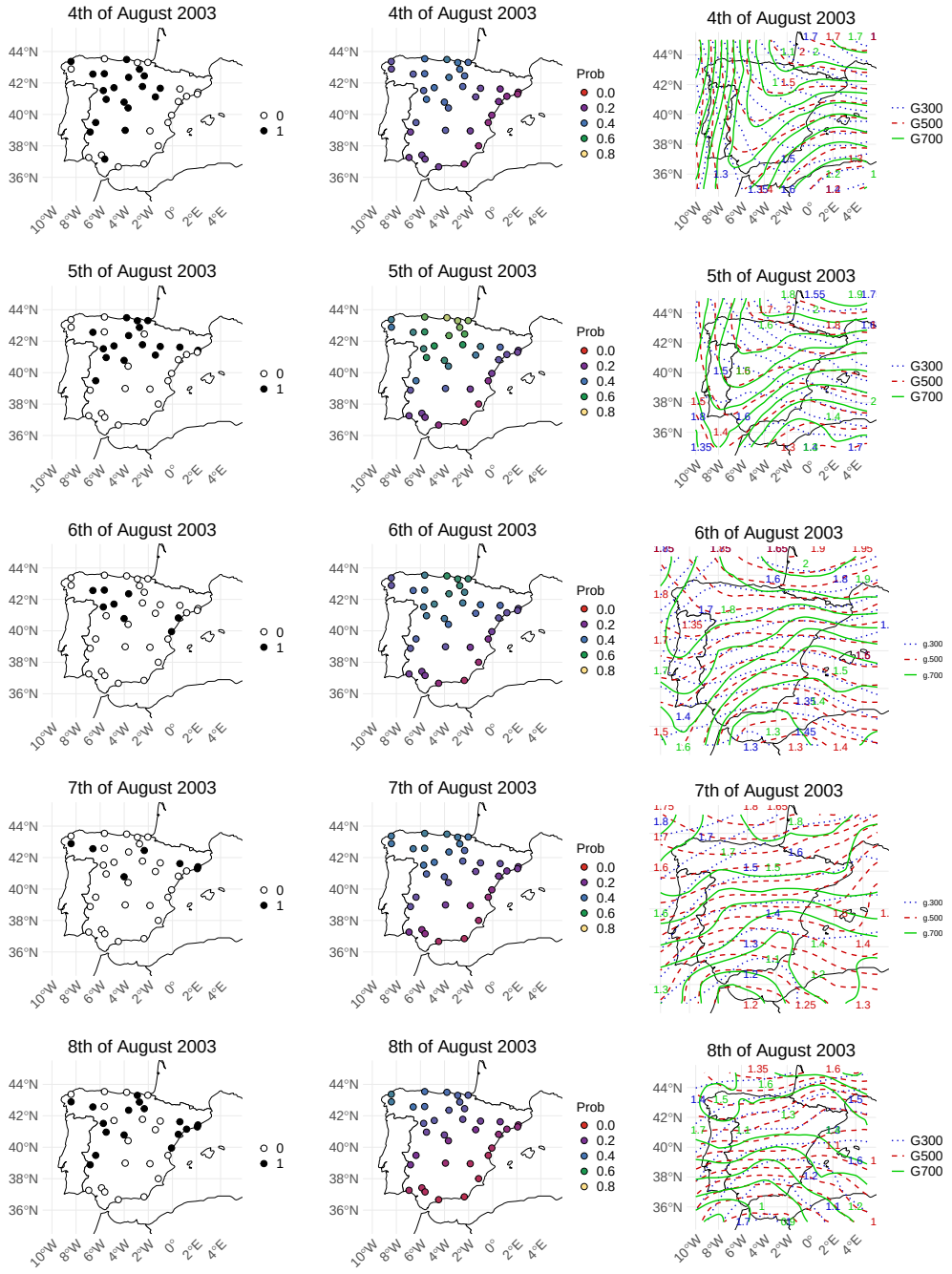

Figure S.10. Each row corresponds to a day of the heatwave from 04–08 August 2003 (training period). Left column: indicates whether each station experienced a record (black points = records, white points = no record). Middle column: shows the record probability predicted by model M2 at each station. Right column: displays geopotential composites for the specific day at pressure levels of 300 hPa, 500 hPa, and 700 hPa.

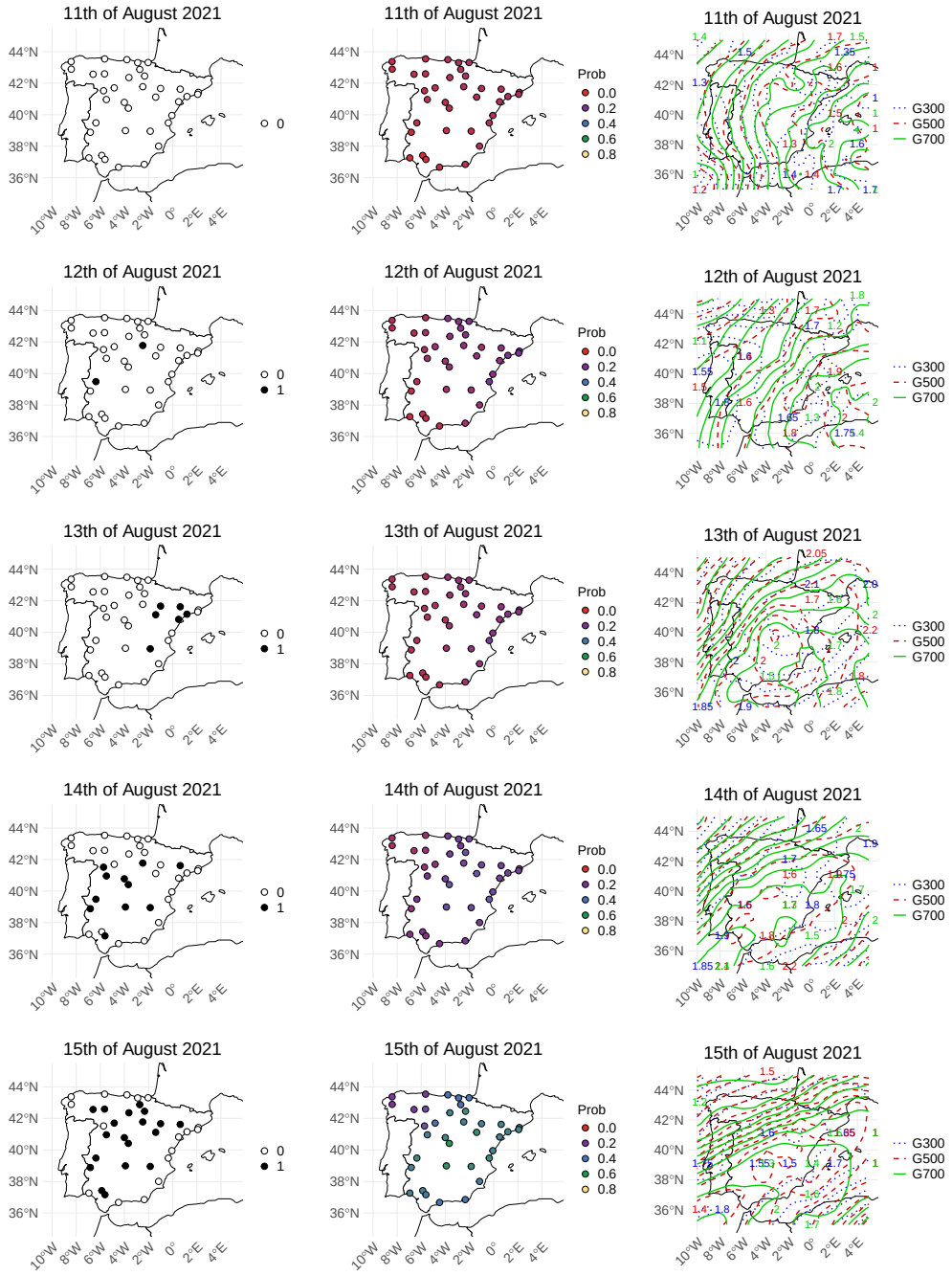

Figure S.11. Each row corresponds to a day of the heatwave from 11–15 August 2021 (training period). Left column: indicates whether each station experienced a record (black points = records, white points = no record). Middle column: shows the record probability predicted by model M2 at each station. Right column: displays geopotential composites for the specific day at pressure levels of 300 hPa, 500 hPa, and 700 hPa.