Supplementary information

SI1

The gaseous VOC precursors for organic aerosol (OA) formation in different chemical schemes used in this work are presented in table SI-1. Species with an L after their name are lumped species, meaning that they are surrogates for a group of species; in this case, a representative compound is given in the table. Note that the precursors for both the standard VBS schemes (with and without BSOA aging) and modified VBS scheme are the same.

	CHIMERE standard		VBS (standard and modified)	
Anthropogenic	TOL	Toluene (L)	TOL	Toluene (L)
	TMB	Trimethylbenzene (L)	TMB	Trimethylbenzene (L)
	NC4H10	n-butane (L)*	NC4H10	n-butane (L)
			OLE1	C4-C13 terminal alkenes (L)
			OLE2	C4-C13 internal alkenes (L)
			ARO1	Toluene (L)
			ARO2	Xylenes (L)
			ALK4	Branched C5-C6 alkanes (L)
			ALK5	C7-C13 n-alkanes
Biogenic	APINEN	α-pinene	APINEN	α-pinene
	BPINEN	β-pinene	BPINEN	β-pinene
	LIMON	Limonene	LIMON	Limonene
	OCIMEN	Ocemene (L)	OCIMEN	Ocemene (L)
	C5H8	Isoprene	C5H8	Isoprene
	HUMULE	Humulene (L)	HUMULE	Humulene (L)

Table SI-1. Precursors for used schemes: name used in the model and the name of the species. For lumped species which are signified with an L after their model name, an example of species lumped in that group is given.

*: only higher weight alkanes which are included in the lumped nCH_4H_{10} species form organic aerosol.

SI2

A simplified diagram of OA formation schemes from VOCs was presented in the main article; here, the same type of diagrams are shown for OA formation from semi-volatile primary OA for the VBS standard scheme and modified VBS scheme.

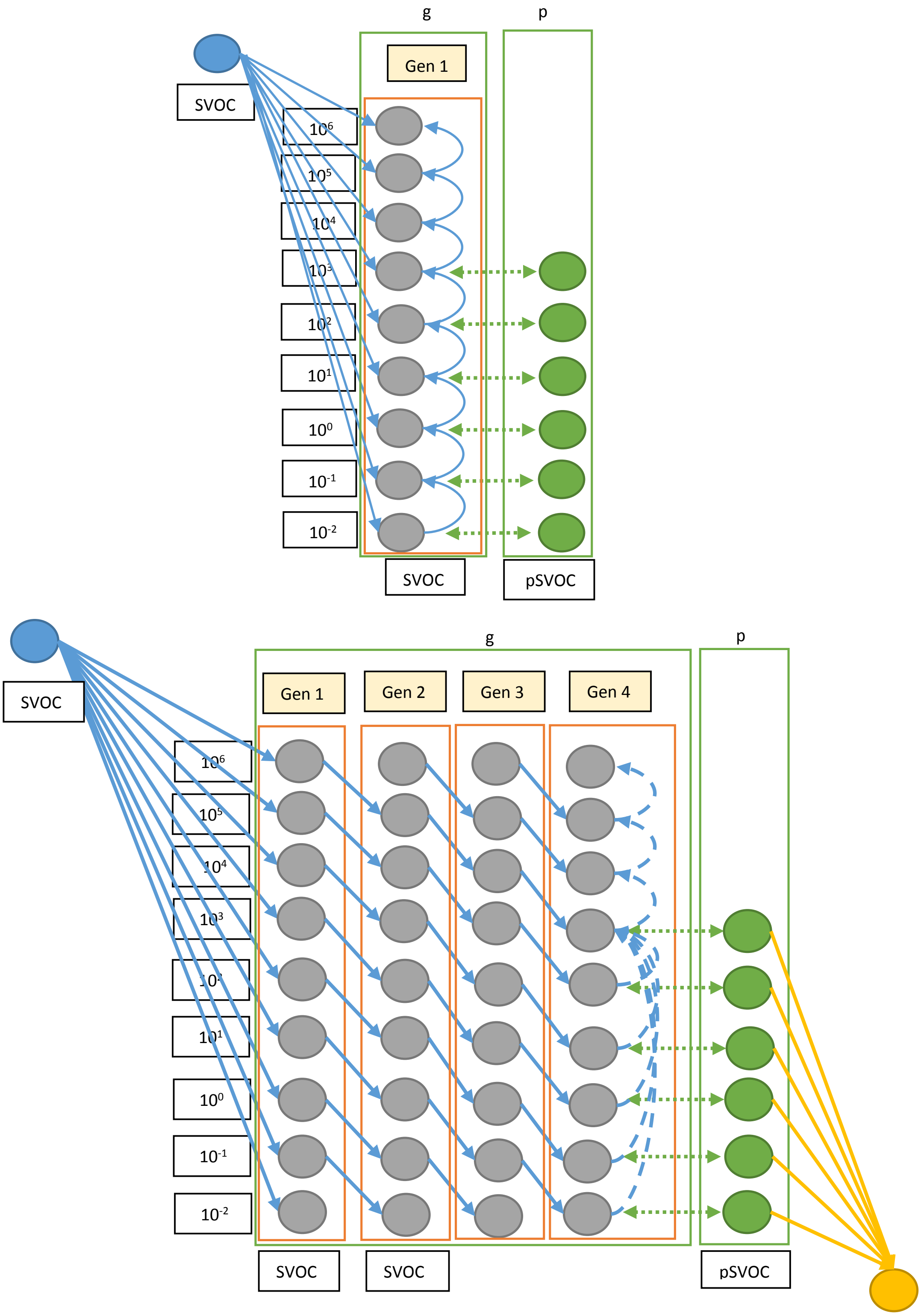

Figure SI-1. Schematic of oxidation of POAs in the Standard VBS scheme (upper panel) and modified VBS scheme (lower panel). POA and IVOC (intermediate volatility organic carbon) emissions are distributed into 9 SVOC/IVOC bins of saturation concentration ranging from 10^{-2} to 10^{6} µg.m^{-3}. In the standard VBS scheme (a), they undergo functionalization transferring species to classes with lower volatility. SVOC species are in equilibrium between the gas and condensed phase (SI-SOA); in the modified VBS scheme (b), the number of functionalization steps is accounted for (GEN). Highly functionalized SVOC's (starting from fourth generation) can undergo fragmentation, leading to transfer of species to classes with higher volatility. In addition, semi-volatile aerosol can be irreversibly transformed into non-volatile one (yellow-filled circle).

SI3

In the procedure to estimate a representativeness error described in section 4-1 of the article, a panel of 9 equations including 2 – 4 open parameters was tested for the non-linear regressions of simulated species concentration versus model altitude. Estimated species concentrations were required to be positive.

$$Eq\ 1.\ y = d + \frac{a-d}{1+\left(\frac{x}{c}\right)^b}\ , \qquad Eq\ 2.\ y = \left(d + \frac{a-d}{1+\left(\frac{x}{c}\right)^b}\right)^m\ , \qquad Eq\ 3.\ y = a \times \left(1 - e^{\frac{-x}{b}}\right) + c,$$

$$Eq\ 4.\ y = a \times \left(1 - e^{\frac{-x}{b}}\right) + \log c\ , \qquad Eq\ 5.\ y = a \times \left(1 - e^{\frac{-x}{b}}\right) + |c|\ ,$$

$$Eq\ 6.\ y = |a| \times \left(1 - e^{\frac{-x}{b}}\right) + |c|\ , \qquad Eq\ 7.\ y = a \times x^b + c\ , \qquad Eq\ 8.\ y = a \times x^b\ ,$$

$$Eq\ 9.\ y = a + b \times x + c \times x^2 + d \times x^3$$

The results for several equations were not used since the number of hourly time-steps for which convergence was not achieved were too frequent (more than 40% of all time-steps), and thus the results have been judged as not representative. Finally, the results for five equations were retained (eq. 1, 3, 5, 7 and 8).

For each of the five equations, regressions for all time steps were calculated. The results for each equation for different simulation times were filtered, by two factors: first convergence for a given simulation time step; second correlation coefficient of higher than 50% between the fitted data and the original points. If these two criteria were met, the results were retained. Then the results for all equations were put together and means, lower and upper confidence intervals calculated (intervals were calculated by using the maximum and minimum values seen for different equations, and the mean presents the average of all equations). Only simulation times with at least two converged equations were used in the calculation of orographic representativeness error.

SI4

Meteorological output of the mesoscale WRF model at different resolutions has been used as input to the CHIMERE CTM. The meteorological data used by CHIMERE was compared to various meteorological observations such as radiosoundings and surface observations at the measurement sites. Detailed results of these comparisons are given in this supplementary material (SI4), in addition to the short overview given in the main article.

First, we present here a comparison of simulations with observations from the European gridded dataset of daily meteorological data (E-Obs; Haylock et al., 2008; Hofstra et al., 2009) for all available stations in the European D10 domain. In addition, individual stations where data was available (EMEP stations in Europe, ChArMEx stations of Ersa and Es Pinar, QualitairCorse stations in Corsica) were also used for meteorological comparisons; among them, Ersa, France, Es Pinar, Spain, Corsica Isl., France (42°14'09.6", 9°10'46.1") stations are presented in this study since they are closest to sites where measurements for pollutants were done.

All available stations from E-Obs were compared to model results starting from the D10 domain. In the two smaller domains (D3 and D1) no station is available. In D10, statistical results of the comparisons as well as the number of stations for each parameter are shown in table SI-2. Note that the E-OBS stations were not filtered with respect their orographic caracteristics, which might not well be taken into account by the model. The correlation between simulations and observations for average daily temperature and the daily maximum of temperature is high (0.82 with a +2.1°C bias), while it is lower for the daily minima is underestimated (correlation of 0.55 with a -3.3°C bias) in the model. Note that the model quite overestimates the daily dynamics of temperature by more than 5°C. The correlation for relative humidity is also acceptable, since this parameter is also very sensitive to altitude changes. The other parameters (wind speed, WS; wind direction, WD; and precipitation, P) are not very well correlated (r <0.5).

For the E-OBS network (Datasets provided by European Climate Assessment & Dataset project for monitoring and analyzing climate extremes, Haylock et al., 2008; Hofstra et al., 2009) comparisons also show a good correlation and a low bias for temperature (correlation of 0.79 with a bias of -0.54°C for mean temperature observed for 71 stations in D10), while the daily minima seems to be underestimated (bias of -3°C for daily minima observed for 71 stations).

	D10				
	stations	R	RMSE	Bias	Mean_obs
T_{max} (°C)	71	0.82	3.39	2.18	29.51
T_{min} (°C)	71	0.48	3.64	-3.17	17.14
T_{mean} (°C)	62	0.79	1.71	-0.54	22.85
RH (%)	59	0.63	11.15	4.7	64.13
WS ($m.s^{-1}$)	59	0.50	1.34	-0.91	2.39
WD(deg)	58	0.30	72.29	0.95	184.7
P ($mm.d^{-1}$)	71	0.35	4.86	1.83	1.88

Table SI-2. Statistical data for E-OBS dataset compared to D10 simulations over the 1-month simulation period from 10 July to 8 August 2013. Mean_obs shows the average of observations in all stations.

An example of radiosounding comparisons for the balloons launched on the afternoon of 25 july 2013 is shown in figure SI-2. The red line shows the observations while the blue one shows the simulations. A good correlation is seen for this day for all three stations for all parameters. Only the Ajaccio station is located in a smaller domain (D3): for the balloon of 25 July, a correlation of 0.99, 0.98 and 0.91 and a bias of -2, -18 and -9% is seen for temperature, wind speed and relative humidity respectively.

The radiosounding measurements stations have been listed in table 2. In figure SI-2, an example of model-observation comparisons is shown for D10, for all the three stations and for the three parameters, temperature, relative humidity and wind speed for the balloons of the afternoon radiosounding of 25 July 2013. There is an excellent correlation with respect to altitude and an acceptable bias (a minimum correlation of 0.8 and a maximum bias of less than -20%) for the compared parameters for all stations in this example. In figure SI-3, vertical profiles of the same variables are shown. In addition, for each vertical level of simulations, two boxplots show the whole variability of the measurements (in red) and simulations (in blue, for Ajaccio D10 in dark blue and D3 in light blue). The small white dot on the boxes shows the average for each level. The altitudes correspond to the average altitude represented in each vertical level. For temperature, bias at different sites and altitude levels is in general small, being on average -1.16, -0.61 and -0.39°C over the different altitudes for Palma, Nimes and Ajaccio respectively. Correlation (with respect to sounding time), is on average 0.65, 0.80 and 0.80, for Palma, Nimes and Ajaccio respectively. Looking at the box plots for each altitude, there is not much variation in the values seen in each level neither in the model nor in observations.

In figure SI-3, vertical profiles of the same variables are shown, but for a total of 32 noon balloon launches at each site in the period of 10 july to 30 july. In addition, for each vertical level of simulations, two boxplots show the variability of the 32 measurements (in red) and corresponding simulations (in blue, for Ajaccio D10 in dark blue and D3 in light blue). The small white dot on the boxes shows the average for each level. The altitudes correspond to the average altitude represented in each vertical level. For temperature, bias at different sites and altitude levels is in general small, being on average -1.16, -0.61 and -0.39°C over the different altitudes for Palma, Nimes and Ajaccio respectively. Correlation (with respect to sounding time), is on average 0.65, 0.80 and 0.80, for Palma, Nimes and Ajaccio respectively. Looking at the box plots for each altitude, there is not much variation in the values seen in each level neither in the model nor in observations.

For wind speed, the correlation coefficients become better in many cases at higher levels, where the variability is also larger . General bias of 0.44%, -0.83% and -1.19% is seen for a general correlation of 0.55, 0.65 and 0.56 in Palma, Nimes and Ajaccio respectively. This bias drops to 5%, 0.2% and -7% if only altitudes above 3000 m are considered, thus lower layers are more biased. A reasonable explanation for the strong general bias observed at the Ajaccio station may be because of incorrect sea/land breezes representation in the model, however this cannot explain the bias for Nimes, since it is not close enough to the sea. Model errors in wind speed might directly affect advection to the measurement sites.

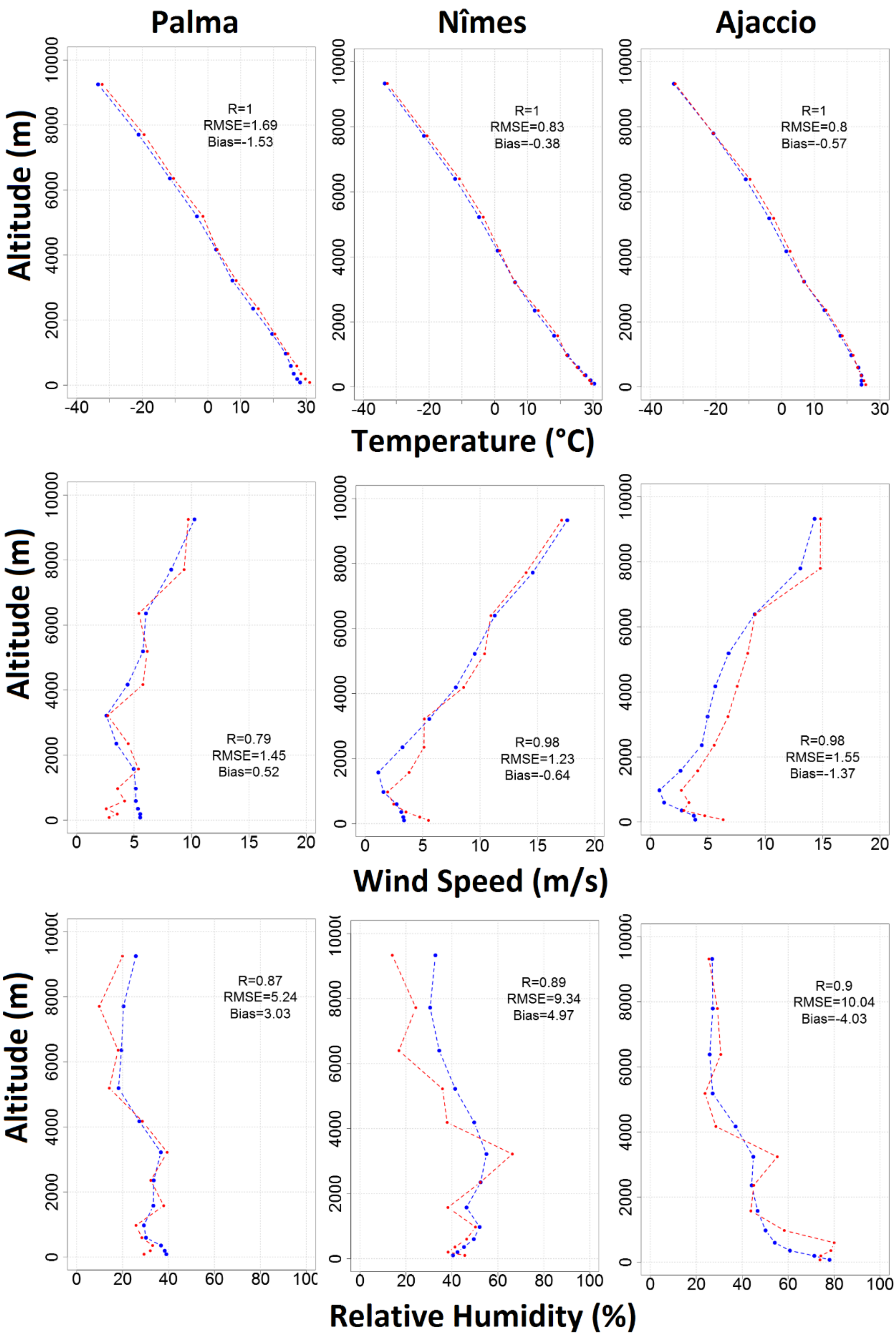

Figure SI-2. Comparison of radio sounding measurements for balloon soundings on the afternoon of 25 July. In each panel, the blue line shows the simulations and the red line shows the observations, points on each line show the vertical levels in CHIMERE.

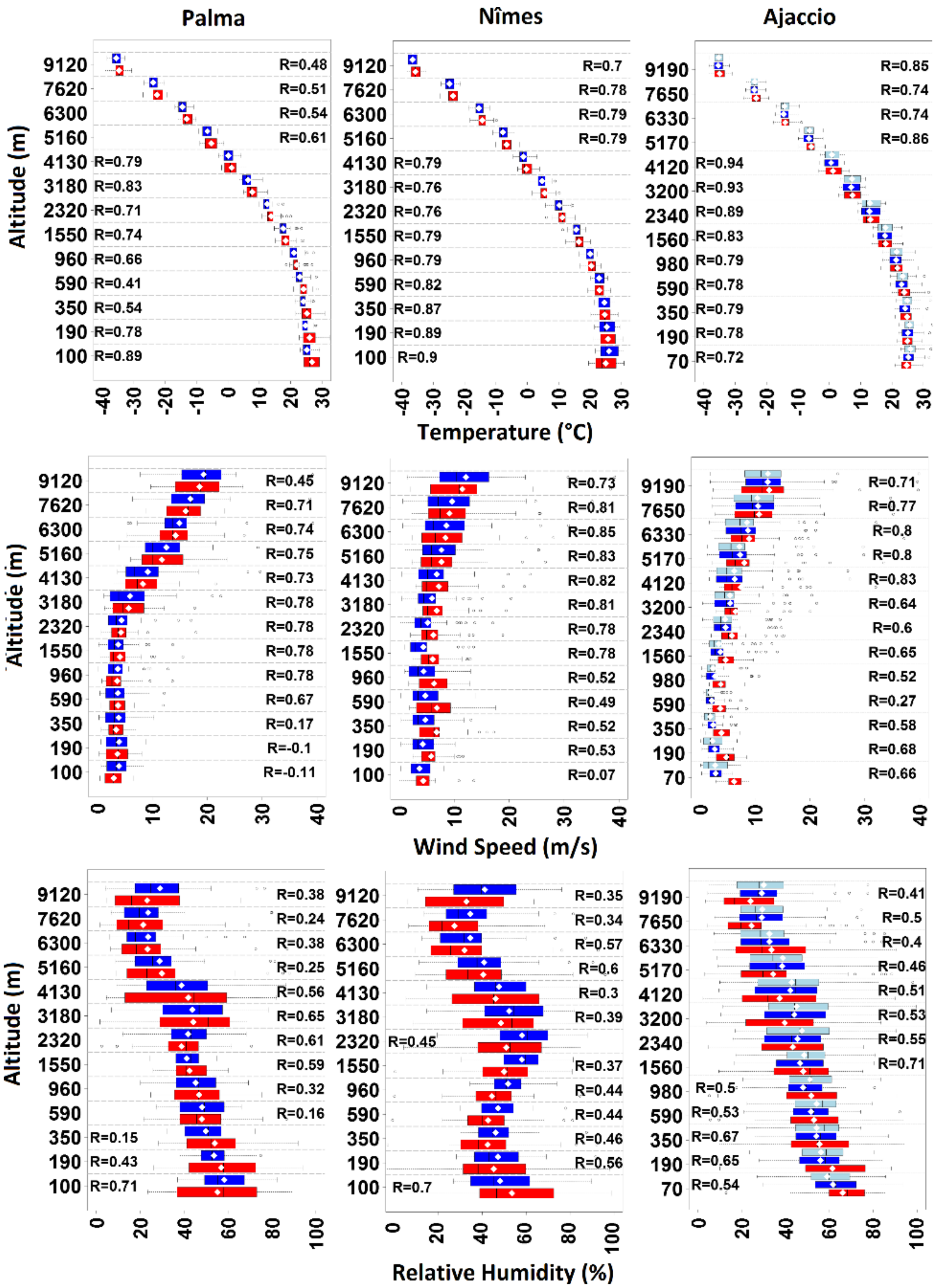

Figure SI-3. Boxplots showing the comparison of all radiosounding launches available (total of 32 balloons for each site) in the period of 10 to 30 july. Each row of figure represents a parameter, each column a station. The Y axis shows the altitude and the X axis shows Temperature (°C), wind speed (m.s-1) and relative humidity (%). The blue box shows the simulations and the red box the observations. If possible, both D10 (dark blue) and D3 (light blue) simulations are displayed, else only the D10 one. On each box the white dot shows the average over 32 soundings, the bars the variability (minimum and maximum with outliers shown with dots). R values represent the correlation coefficient between measurements and observation for each level.

For relative humidity, there does not seem to be an apparent relation between the correlation and the vertical levels. For this parameter, general correlation of 0.42, 0.46 and 0.45 is seen at Palma, Nimes and Ajaccio respectively, for a general relative bias of -0.1%, 9% and -0.4%. Contrary to wind speed, the bias becomes stronger at altitudes higher than 3000m (2%, 10% and 11% respectively).

Available comparisons of meteorological parameters at the two supersites for chemical measurements are presented in table 4 (for the three domains D10, D3, and D1 at Ersa and for D10 at Es Pinar). In general, the results for these comparisons become more accurate when the domain has a higher resolution (Table 4). Figure 5 shows the simulation-observation comparisons for these two stations for four parameters including temperature (T), relative humidity (RH), wind speed (WS) and wind direction (WD). For Ersa, the correlation and the bias become better for domains with higher resolution. As illustrated in figure 5, the comparison for temperature and relative humidity (figures 5-a1 5-b1) for this station show excellent results for the D1 domain. For wind speed (figure 5-c1), while longer-lasting synoptic scale events such as the strong wind conditions between 22 to 26 July and also 28 to 30 July are well-represented by the model, some smaller wind changes are disregarded. It is interesting to notice that in both of these two strong wind periods, the wind direction shows little variation which is also seen by the model. For Es Pinar, the comparisons become less satisfying. An overestimation in the simulation of daily minima of the temperature is seen for this station, which could be explained by the fact that this is a coastal station and the corresponding cell in the simulations is half covered by land and half by sea. The slight overestimation in relative humidity daily maxima could also be attributed to the same reason. Wind speed is overestimated in the simulations, here probably again due to local features not depicted by the 10-km resolved fields are determining.

		Ersa				Es Pinar			
		R	RMSE	Bias	Mean_obs	R	RMSE	Bias	Mean_obs
D10	T (°C)	0.79	2.30	1.48	23.17	0.55	2.72	-2.01	28.49
D10	RH (%)	0.48	17.29	3.91	70.22	0.48	10.24	-1.03	70.10
D10	WS (m/s)	0.60	1.88	-0.39	3.71	0.60	2.27	1.97	1.13
D10	WD (°)	0.13	117.5	46.9	194.2	0.13	124.1	-35.8	165.6
D3	T (°C)	0.81	2.52	1.99	23.17				
D3	RH (%)	0.48	17.59	-4.56	70.22				
D3	WS (m/s)	0.69	1.63	0.05	3.71				
D3	WD (°)	0.27	101.3	36.10	194.2				
D1	T (°C)	0.86	1.54	0.70	23.17				
D1	RH (%)	0.54	16.51	-1.67	70.22				
D1	WS (m/s)	0.66	1.90	0.61	3.71				
D1	WD (°)	0.37	92.0	25.0	194.2				

Table SI-3. Comparison of hourly meteorological parameters at the two ChArMEx supersites of Ersa and Es Pinar from 10 July to 5 August, 2013. R is the correlation coefficient, RMSE, the root mean square error and Mean_obs the average of observations.

The model succeeds in catching the different regimes and the general changes of the wind speed while overestimating its intensity. This fact is also seen in wind rose diagrams shown in figures SI-4-d1 and SI-4-d2. The prevalent wind direction is correctly simulated in both sites, however, wind directions with less occurrences are sometimes not seen by the model.

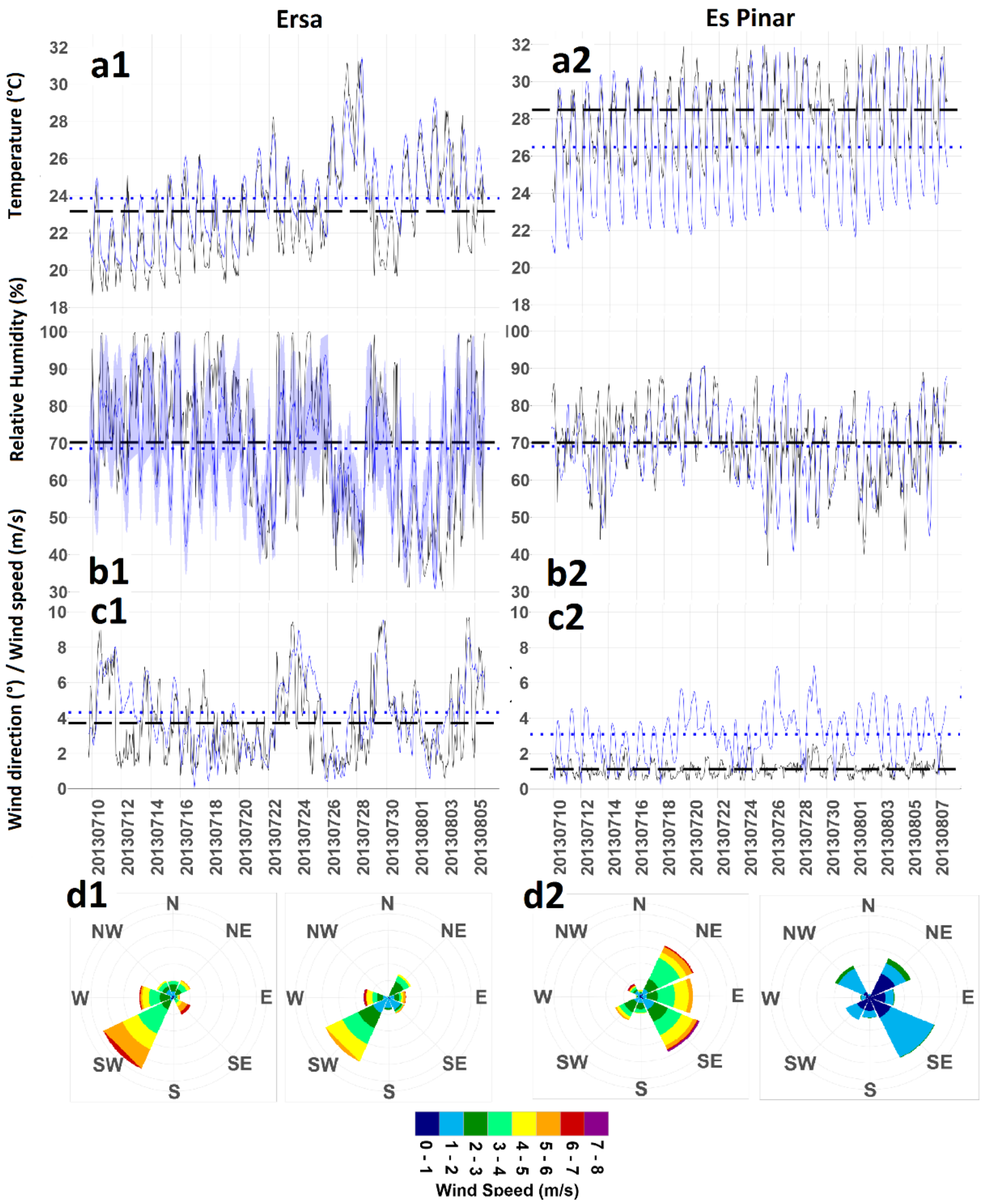

Figure SI-4. Comparison of observed and simulated coincident meteorological parameters for at Ersa (left) and Es Pinar (right) stations. For Ersa D1 simulations and for Es Pinar D10 simulations are shown. In plots a to c, simulations are in blue and observations in black. For relative humidity at Ersa, the envelope corresponding to the spatial representativeness is given. In figures d1 and d2, simulations are shown at the left side, and observations on the right.