



Supplement of

Identifying source regions of air masses sampled at the tropical high-altitude site of Chacaltaya using WRF-FLEXPART and cluster analysis

Diego Aliaga et al.

Correspondence to: Victoria A. Sinclair (victoria.sinclair@helsinki.fi)

The copyright of individual parts of the supplement might differ from the article licence.

910 Model verification

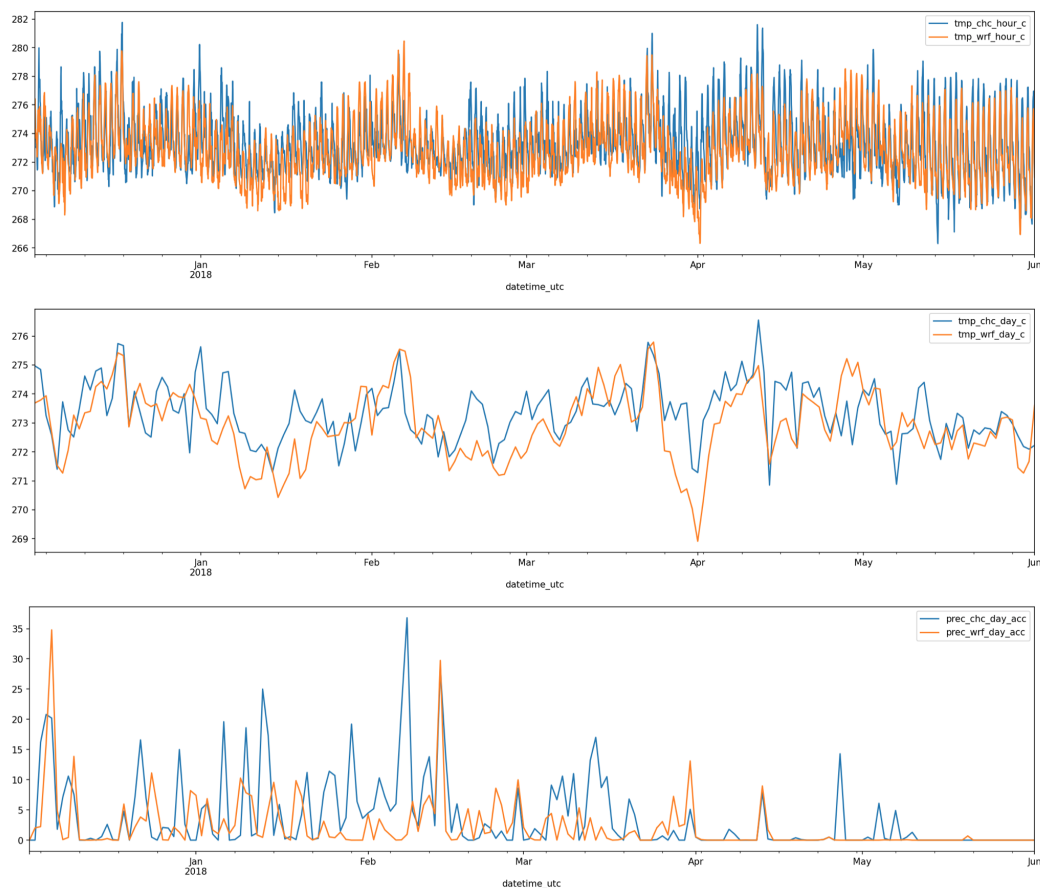


Figure S1. Observations from CHC quantitatively compared to the WRF model output from the closest grid point for the mean hourly and daily temperature and the accumulated daily precipitation. Using the hourly temperature data and the daily accumulated precipitation

An in-depth verification is challenging due to the limited observations in this area. However, we have observations from CHC which we have quantitatively compared to the WRF model output from the closest grid point (which in the model has a surface elevation of 5049 m, which is 111 m lower from the observation height). Using the hourly temperature data and the daily accumulated precipitation data (Fig. S1) we have computed the mean bias (MB), the mean absolute error (MAE) and the root mean square error (RMSE) following equations 5.1, 5.2 and 5.3 respectively in Chapter 3 of “Forecast Verification: A Practitioner’s Guide in Atmospheric Science, Second Edition. Edited by Ian T. Jolliffe and David B. Stephenson. C 2012 John Wiley & Sons, Ltd. Published 2012 by John Wiley & Sons, Ltd. ” For temperature the MB=-0.42°C indicating that WRF predicts slightly cooler temperatures than observed, the MAE = 1.35°C and the RMSE= 1.73°C. When these values are

915

920 compared to the standard deviation (s.d.) of the temperature from WRF (s.d. = 2.11°C) and the observations (2.61°C), we can
conclude that the average magnitude of the WRF forecast error is small relative to the variation. The corresponding values
for daily accumulated precipitation are MB=-1.6 mm (WRF is slightly too dry on average), MAE= 3.5 mm and RMSE= 6.3
mm. However, as precipitation is very variable from day-to-day (s.d. = 5.0 mm in the observations compared to the mean
observed value of 2.45 mm), we also compute the occurrence of hits, false alarms, misses, and correct negative values by a
925 given precipitation threshold (i.e. a contingency table) and the corresponding accuracy defined as the number of hits + the
number of correct negatives divided by the total number of days. Hit is when both the model and observation have precipitation
exceeding the threshold, correct negative is when the model and observation both have precipitation below the threshold. Miss
is when there is precipitation exceeding the threshold in the observations but not in the model and false alarm is when the
model has precipitation exceeding the threshold but the observations do not. The resulting values are shown in Table S2.

Diurnal analysis

930 When analysing the diurnal variation of the SRR for each of the 18 clusters we first objectively identify the clusters that present
a diurnal cycle by transforming their time series into the frequency domain by applying a fast Fourier transform and plotting the
resulting frequency spectrum (Fig. S4). In this domain, we visually identify those clusters that present a local peak around the
frequency 1 day⁻¹. Manual inspection of the clusters' timeseries indicated that not every day included in each cluster identified
by the fast Fourier transform process to have a peak at 1 day⁻¹ had a pronounced diurnal cycle. Therefore, the clusters are then
935 subjected to a dynamic time warp (DTW) grouping procedure (Tavenard et al., 2020) that classifies the days with and without
a diurnal peak i.e. influence from the boundary layer. The results of this classification are shown in Figs. S5 to S8.

domain	lower left [°]		upper right [°]		p.g.r.*	res.**	number of cells		parent start	
	lat.	lon.	lat.	lon.			w.e.†	s.n.‡	i	j
D01	-32.2	-89.4	-0.5	-43.2	1	38.0	118	86	1	1
D02	-26.3	-78.7	-7.2	-53.9	4	9.5	253	205	28	18
D03	-20.5	-70.9	-13.9	-62.0	3	3.2	274	214	80	65
D04	-17.2	-69.0	-15.6	-67.3	3	1.0	154	151	61	110

Table S1. Description of the domains used for the WRF simulation. * parent grid ratio, ** resolution (km), † west east, ‡ south north

threshold	hit	Corr. neg	miss	False alarm	accuracy
0 mm	99	38	11	34	0.75
1 mm	42	77	35	28	0.65
5 mm	12	118	36	16	0.71

Table S2. Occurrence of hits, false alarms, misses, and correct negative values by a given precipitation thresholds (i.e a contingency table) and the corresponding accuracy defined as the number of hits + the number of correct negatives divided by the total number of days. Hit is when both the model and observation have precipitation, correct negative is when the model and observation both have no precipitation. Miss is when there is precipitation in the observations but not in the model and false alarm is when the model has precipitation but the observations do not.

Grid	Dimension	Numberof levels	Resolution	N. ofgrid points	Disksize [GB]	Compresseddisk size [GB]
low resolution rectangular	x	468	10 km	468		
	y	340	10 km	159,120		
	z	30	0.5 km	4,773,600		
	t	4248	1 hour	20,278,252,800	81.1	7.4
	τ	96	1 hour	1,946,712,268,800	7786.8	707.9
high resolution rectangular	x	459	1 km	459		
	y	450	1 km	206,550		
	z	30	0.5 km	6,196,500		
	t	4248	1 hour	26,322,732,000	105.3	9.6
	τ	96	1	2,526,982,272,000	10107.9	918.9
log-polar	r	31	* $a = 0.18$	31		
	θ	36	10 degree	1,116		
	z	30	0.5 km	33,480		
	t	4248	1 hour	142,223,040	0.6	0.1
	τ	96	1 hour	13,653,411,840	54.6	5.0

Table S3. Relation between the rectangular grids and the log-polar grid.

short name	SRR [%] $n_c = 6$	distance from CHC [km]	height above ground [km]	height above sea level [km]	$\frac{SRR_{<1.5km}}{SRR_{total}}$ [%]	age [h]
03_PW	29.0	518	5.0	5.9	18	51
05_PW	11.4	428	3.0	6.3	31	45
07_PW	13.9	721	6.5	7.7	10	54
08_PW	17.7	238	2.5	5.7	43	36
11_PW	19.6	465	4.5	6.6	22	53
12_PW	8.4	76	1.8	4.3	62	27

Figure S2. Describes the properties of the main 6 pathways (similar to Fig. 13). The main pathway's digits refer to the clockwise direction of the centre of mass of the air masses. The colours are unique for each PW and the same scheme is used on Fig. 13 and Fig. 6a and b. The SRR [%] columns describe the average contribution of each cluster. We also show the distance from CHC, height above ground and height above sea level of each cluster's centre of mass. Furthermore, $\frac{SRR_{<1.5km}}{SRR_{total}}$ [%] shows the ratio between the SRR below 1.5 km and the SRR summed over the full vertical column (SRR_{total}).

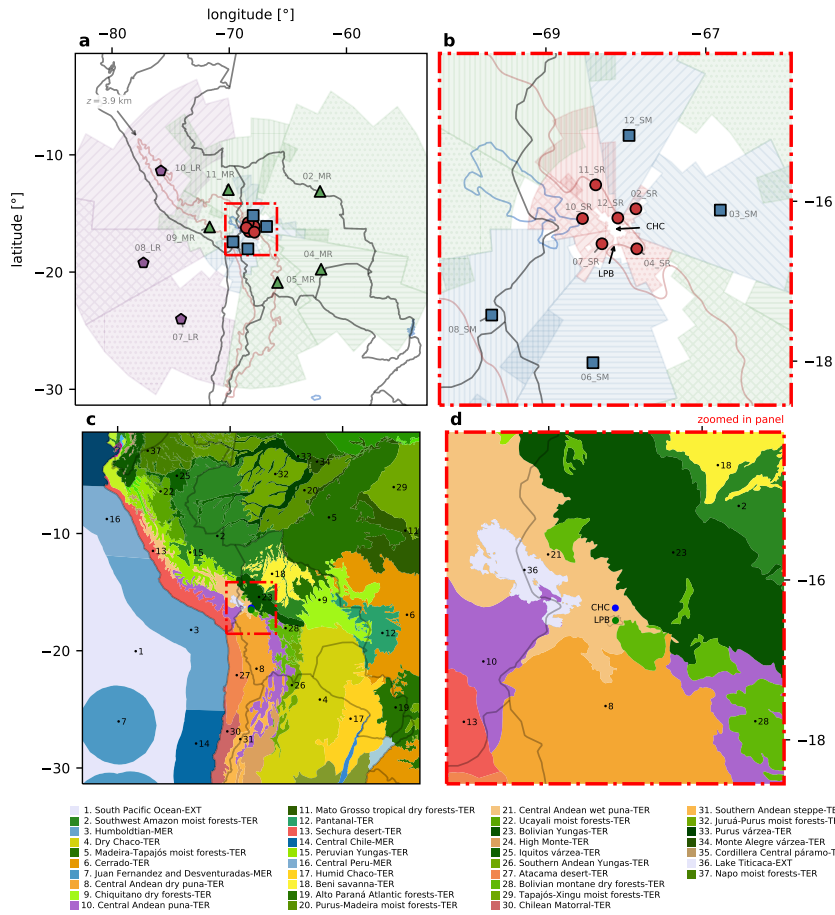


Figure S3. Panels a) and b) show the horizontal location of the 18 clusters. The colour is related to their distance range from CHC and the hatch distinguishes each cluster independently. For easier visualization, we only show 80% of the grid cells where the SRR is highest. We also show the centroid markers for each of the clusters. Each cluster centroid is marked with a disk (short range), square (short-medium range), triangle (medium range), or pentagon (long range) locator. The brown line corresponds to a height of 3.9 km a.s.l. and encircles the Altiplano plateau. Panel b) corresponds to the region inside the red rectangle in panel a). Panels c) and d) show the eco-regions as described in section 4 and are shown for comparing the area of the pathways to the under-lying biome.

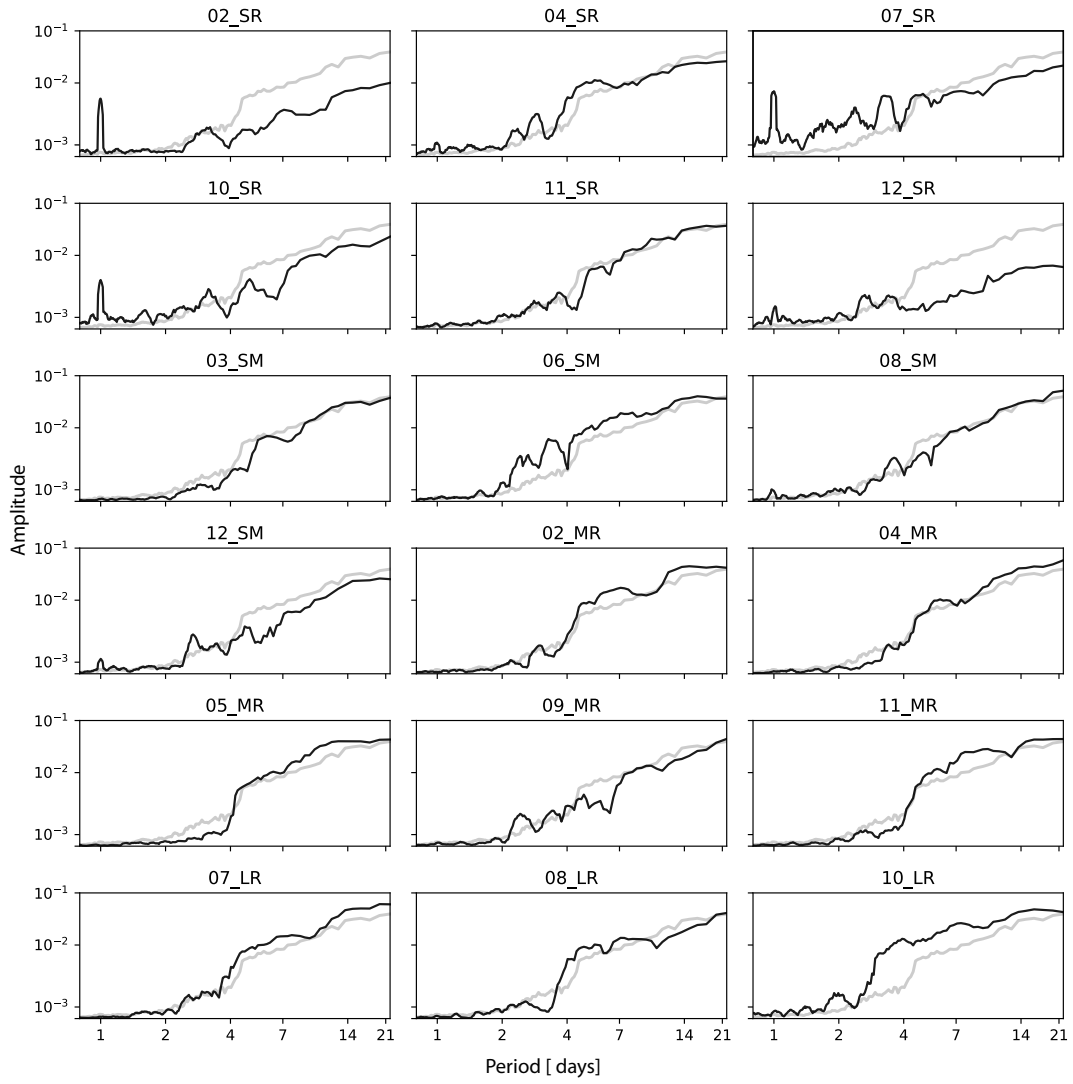


Figure S4. Power spectra of the clusters' timeseries obtained via the Fourier transform. In light gray is the median of all the spectra (shown for reference). Clusters 02_SR, 07_SR, 03_SR and 12_SR show a local peak at a daily frequency and therefore are analysed in detail in Figs. S5 S8

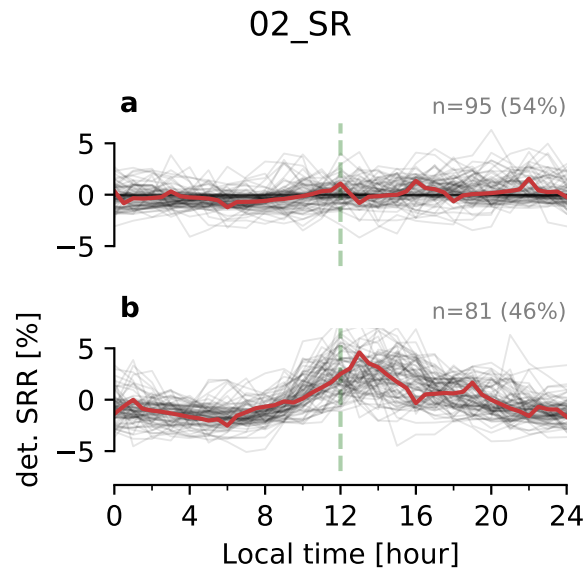


Figure S5. Detrended daily SRR timeseries of cluster 02_SR divided into a) no diurnal influence and b) diurnal influence by applying a dynamic time warping classification with 2 groups. In grey, we show the timeseries for each single modelled day and red line is the group median evolution. The detrended diurnal cycle timeseries were generated by decomposing the SRR raw signal into a trend (SRR_{trend}) and a diurnal cycle ($SRR_{diurnal\ cycle}$) in the following fashion: $SRR_{diurnal\ cycle}(t) = SRR_{raw\ signal}(t) - SRR_{trend}(t)$. The SRR_{trend} is obtained by applying a running mean with a window of 24 hours to the raw signal.

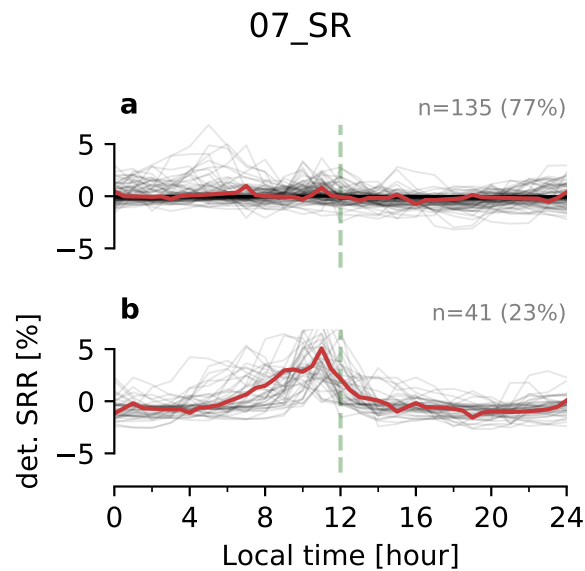


Figure S6. Similar to Fig. S5 but for cluster 07_SR.

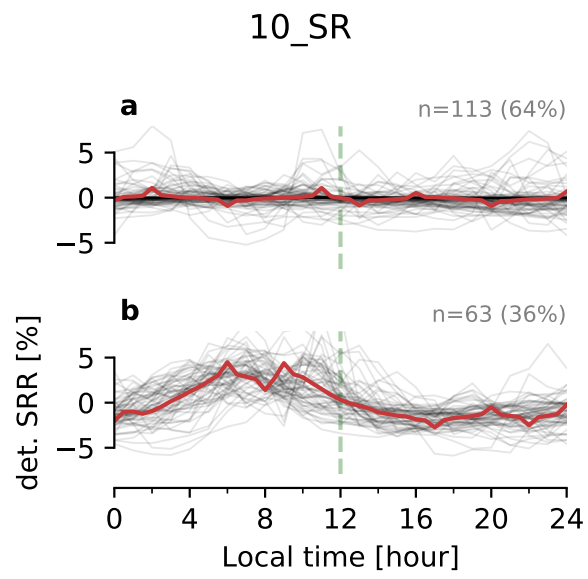


Figure S7. Similar to Fig. S5 but for cluster 10_SR.

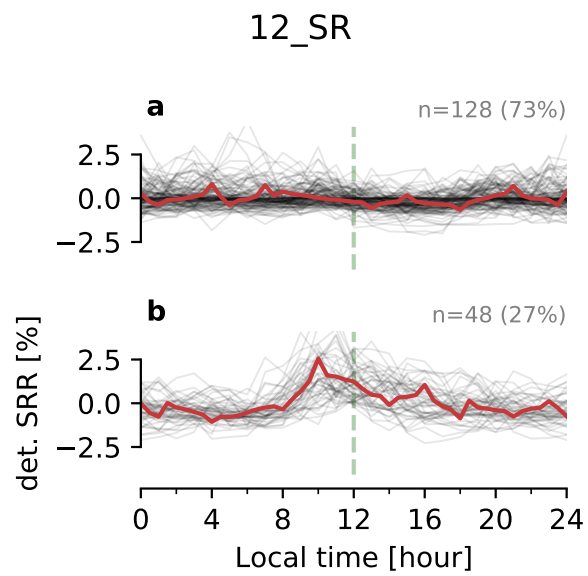


Figure S8. Similar to Fig. S5 but for cluster 12_SR.

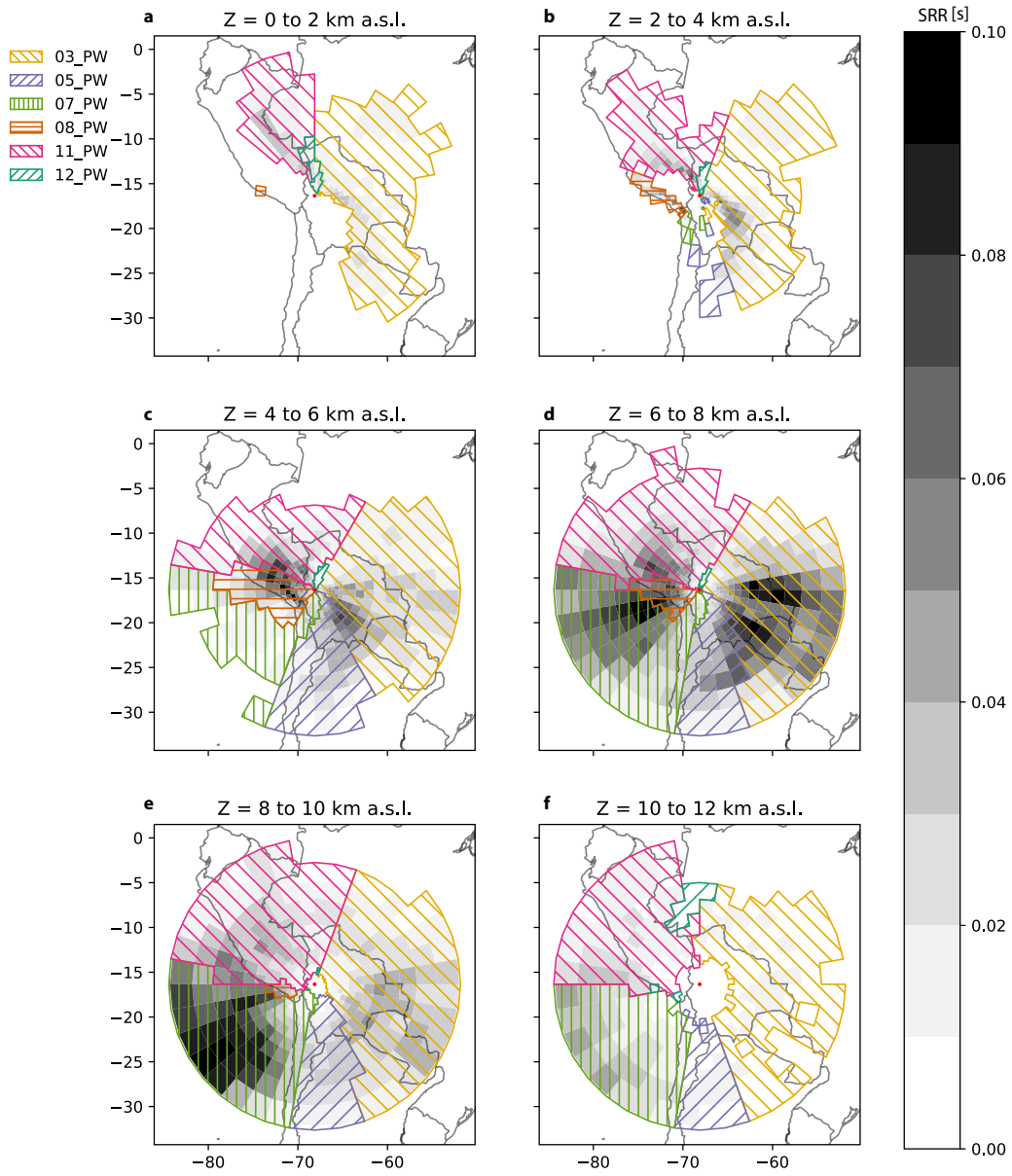


Figure S9. Hourly mean SRR (grayscale) and pathways regions (colors and etching) at different height levels for the full study period (December 2017 to May 2018).

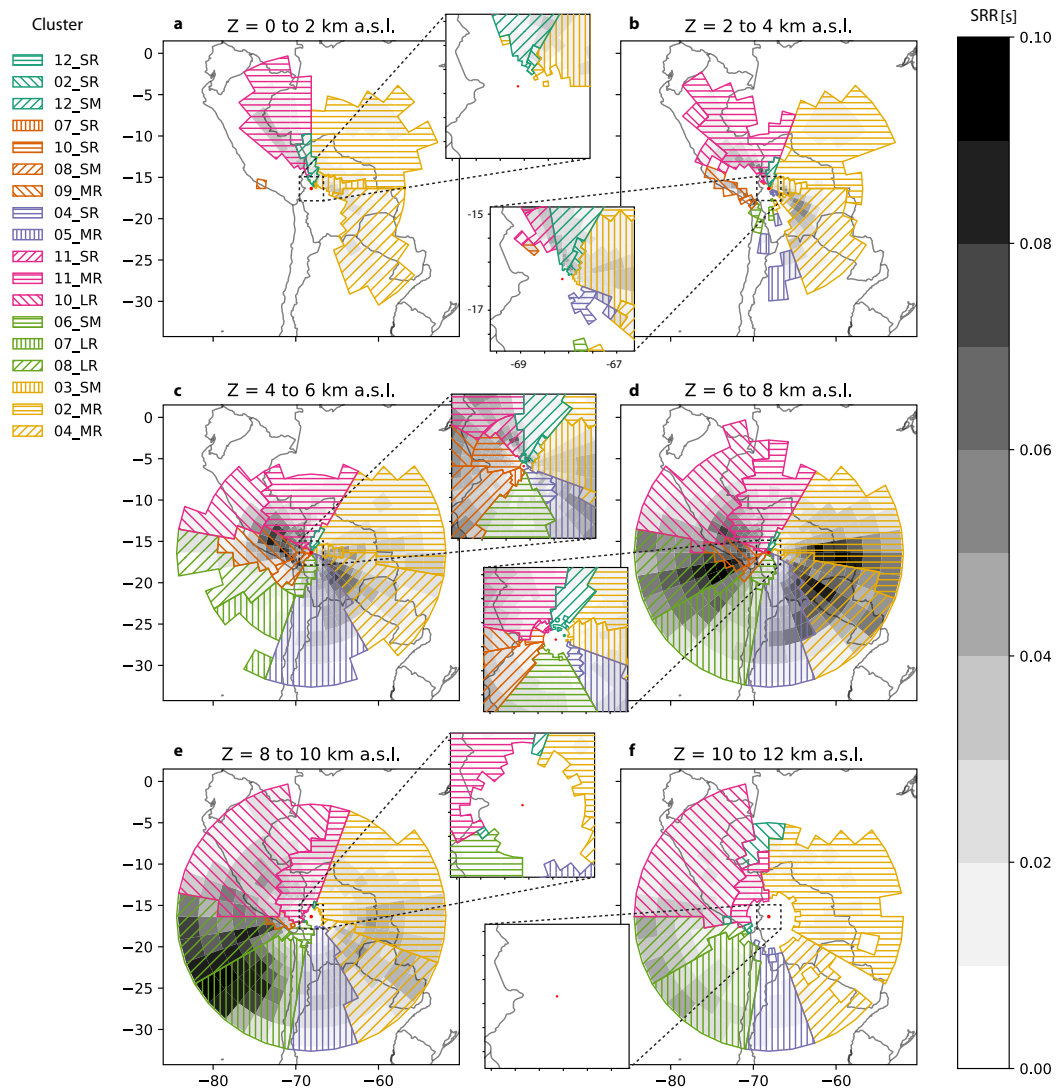


Figure S10. Hourly mean SRR (grayscale) and clusters (colors and etching) at different height levels for the full study period (December 2017 to May 2018).