Supplementary Material

A holistic framework to estimate the origins of atmospheric moisture and heat using a Lagrangian model

Jessica Keune[1], Dominik L. Schumacher[1], Diego G. Miralles[1]

[1]Hydro-Climate Extremes Lab (H-CEL), Ghent University, Ghent, 9000, Belgium

Correspondence to: Jessica Keune (jessica.keune@ugent.be), Dominik Schumacher (dominik.schumacher@ugent.be)

1. Analysis with the software framework *HAMSTER v.1.0.0*

The Lagrangian analysis in the main manuscript was run with the Heat And MoiSture Tracking framEwoRk ***HAMSTER v1.0.0***, as published on https://github.com/h-cel/hamster.

The commands for the analysis were as described below. For simplicity, we omit the years (set by --ayyyy and --ryyyy for the analysis and run years, respectively) and the analysis months (set by --am), which covered all months between January 1980 and December 2016 for this analysis. Furthermore, commands are only illustrated for one experiment (--expid "ALL-ABL") and one city (Denver, --maskval 1001). The analysis is split in two parts:

1. The global detection and quantification of fluxes based on two-step trajectories
2. The attribution of source regions for heat and precipitation, and their bias-correction using data from the previous step.

A short description of the commands used for both parts is given below.

1.1. Global detection and quantification of fluxes based on two step trajectories

For the global analysis as presented in this paper, we extract two-step trajectories globally (using --ctraj_len 0) and without any masking netCDF file described in the paths.txt

```
python main.py --steps 0 --ctraj_len 0 --pathfile paths.txt
```

and we detect precipitation globally using --cprec_dqv 0 --cprec_rh 80, evaporation using --cevap_dqv 0 –-fevap_drh False, and sensible heat using --cheat_dtemp 0 --fheat_drh False –-fevap_drh False, the latter conditioned on at least one occurrence (--cpbl_strict 1) within the maximum ABL (--cpbl_method "max") between the two time steps — which refer to the "ALL-ABL" experiment described in the main manuscript:

```
python main.py --steps 1 --expid "ALL-ABL" --cheat_dtemp 0 --fheat_drh False –-fevap_drh False --fheat_rdq False --cevap_dqv 0 --cpbl_method "max" --cpbl_strict 1 --cpbl_factor 1 --cprec_dqv 0 --cprec_rh 80 --pathfile paths.txt
```

Both hamster steps 0 and 1 were run with a paths.txt file that looks as follows:

```
# MASK
maskfile  = None

# location of original flexpart files (untarred)
path_orig = "./flexpart/simulations/eraint_global"

# path and base name for global t2 diag data
base_f2t_diag = "global"
path_f2t_diag = "./flexpart/hamster_analysis/eraint_global/flex2traj_t2"

# paths for processed data
path_diag = "./flexpart/hamster_analysis/eraint_global/diagnosis"
```

The output of the first step are 6-hourly h5-files which contain parcel positions and properties for a specific date and the prior time step. The output of the second step is a monthly netCDF file which contains three variables, i.e., P, E and H, gridded onto a regular 1x1° global grid, for all 6-hourly time steps of the month.

1.2. Attribution of source regions for heat and precipitation and their bias-correction.

To filter for parcels residing over the city of Denver, and to estimate the locations where these parcels were moistened or warmed in the last 15 days, we construct a netCDF file mask_cities3x3.nc that contains the values 1001 over the 3x3° surrounding of Denver, and adjust paths.txt as follows:

```
# MASK
maskfile  = "mask_cities3x3.nc"
```

```
# location of original flexpart files (untarred)
path_orig = "./flexpart/simulations/eraint_global"

# location of the reference data used for bias correction (e.g., ERA-Interim)
path_ref  = "./eraint/by_var_nc/1x1/12-hourly"

# path and base name for global t2 diag data
base_f2t_diag = "global"
path_f2t_diag = "./flexpart/hamster_analysis/eraint_global/flex2traj_t2"

# path and base name for trajectory data
base_f2t_traj = "denver"
path_f2t_traj = "./flexpart/hamster_analysis/eraint_global/denver/00_f2t"

# paths for processed data
path_diag = "./flexpart/hamster_analysis/eraint_global/diagnosis"
path_attr = "./flexpart/hamster_analysis/eraint_global/denver/02_attr"
path_bias = "./flexpart/hamster_analysis/eraint_global/denver/03_biascorr"
```

To extract 16-day trajectories, we run

```
python main.py --steps 0 --ctraj_len 16 --maskval 1001 --pathfile paths.txt
```

To get (biased) estimates of the source regions of precipitation and heat for Denver, we then employ the same settings as for the global "ALL-ABL" analysis, and we evaluate the source contributions over 15 days into the past using the linear discounting and attribution.

```
python main.py --steps 2--ctraj_len 15 --expid "ALL-ABL" --cheat_dtemp 0 --fheat_drh False –-fevap_drh False
$fheat_rdq False --cevap_dqv 0 --cpbl_method "max" --cpbl_strict 1 --cpbl_factor 1 --cprec_dqv 0 --cprec_rh 80 --
mattribution "linear" --maskval 1001 --pathfile paths.txt
```

Finally, to bias-correct these source regions, we use the diagnosed fluxes from the previous step and a references data set — here ERA-Interim — and adjust the source-receptor relationships day by day (--bc_time "daily"):

```
python main.py --steps 3 --expid "ALL-ABL" --maskval 1001 --bc_useattp True --bc_aggbwtime False --bc_time "daily" --pathfile paths.txt
```

The output of the first step are 6-hourly h5-files which contain parcel positions and properties for a specific date and the 16-days prior to that date. The output of the second step is a monthly netCDF file which contains the source regions of precipitation ("E2P") and the source region of heat ("Had") for the city of Denver. These variables are of dimension time x level x latitude x longitude, where time refers to the 'arrival date', level refers to the backward days (ranging from 0 to -15 with 0 being the 'arrival date') and latitude x longitude being again a global 1x1° grid. By setting --bc_aggbwtime False in the bias-correction step, we retain the 4D dimension for this analysis; the netCDF output of this step contains 6 variables: "E2P" as the biased estimates of daily source regions for Denver, "E2P_Es" as the source-corrected estimates of precipitation origins, "E2P_Ps" as the sink-corrected estimates of precipitation origins, and "E2P_EPs", which is the source- and sink-corrected estimate of these source regions. For heat advection, the final output file "Had", which is a copy of the biased estimates, and "Had_Hs", which represents the source-corrected source regions of heat.

2. Analysis performed with *hamster*.

The settings for all experiments contained in the main manuscript are as listed in Table S1 (listing detection criteria only). To assess the impact of the attribution methodology for the estimation of precipitation source regions, all experiments were repeated with --mattribution random in addition to the linear discounting / attribution (--mattribution linear).

	ALL-ABL	**SOD08**	**FAS19**	**SCH19**	**SCH20**	**RH-15**
P criteria						
--cprec_dqv	0	0	0	0	0	0
--cprec_rh	80	80	80	80	80	80
ABL criteria (for *E* and *H*)						
--cpbl_method	max	max	max *	max	max	max
--cpbl_factor	1	1	1 *	1	1	1
--cpbl_strict	1	1	0	1	1	1
E criteria						
--cevap_dqv	0	0.0002	0.0001	-	-	0
--fevap_drh	False	False	False	False	False	True
--cevap_drh	15 **	15 **	15 **	15 **	15 **	15
H criteria						
--cheat_dtemp	0	-	-	1	1	0
--fheat_drh	False	False	False	False	False	True
--cheat_drh	15 ***	15 ***	15 ***	15 ***	15 ***	15
--fheat_rdq	False	False	False	True	False	False
--cheat_rdq	15 ****	15 ****	15 ****	10	15 ****	15 ****

Table S 1. Overview of hamster flags used for the detection of *P*, *E* and *H* in this study. * not used due to --cpbl_strict 0. ** not used due to --fevap_drh False. *** not used to due --cheat_drh False. **** not used due to --fheat_rdq False.

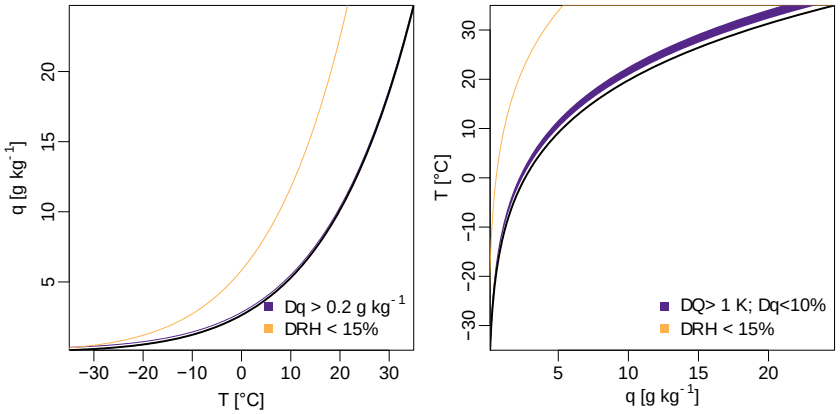

Figure S 1. Visualization of the detection of *E* based on changes in specific humidity above a threshold as in Sodemann et al., 2008 (purple) and relative humidity changes (yellow) in dependency of the temperature; and visualization of *H* based on potential temperature changes and dependent on the specific humidity as in Schumacher et al., 2019 (purple) and based on relative humidity (yellow) in dependency of the specific humidity content.

3. Validation statistics

Additional results for the validation of the detection criteria are shown below.

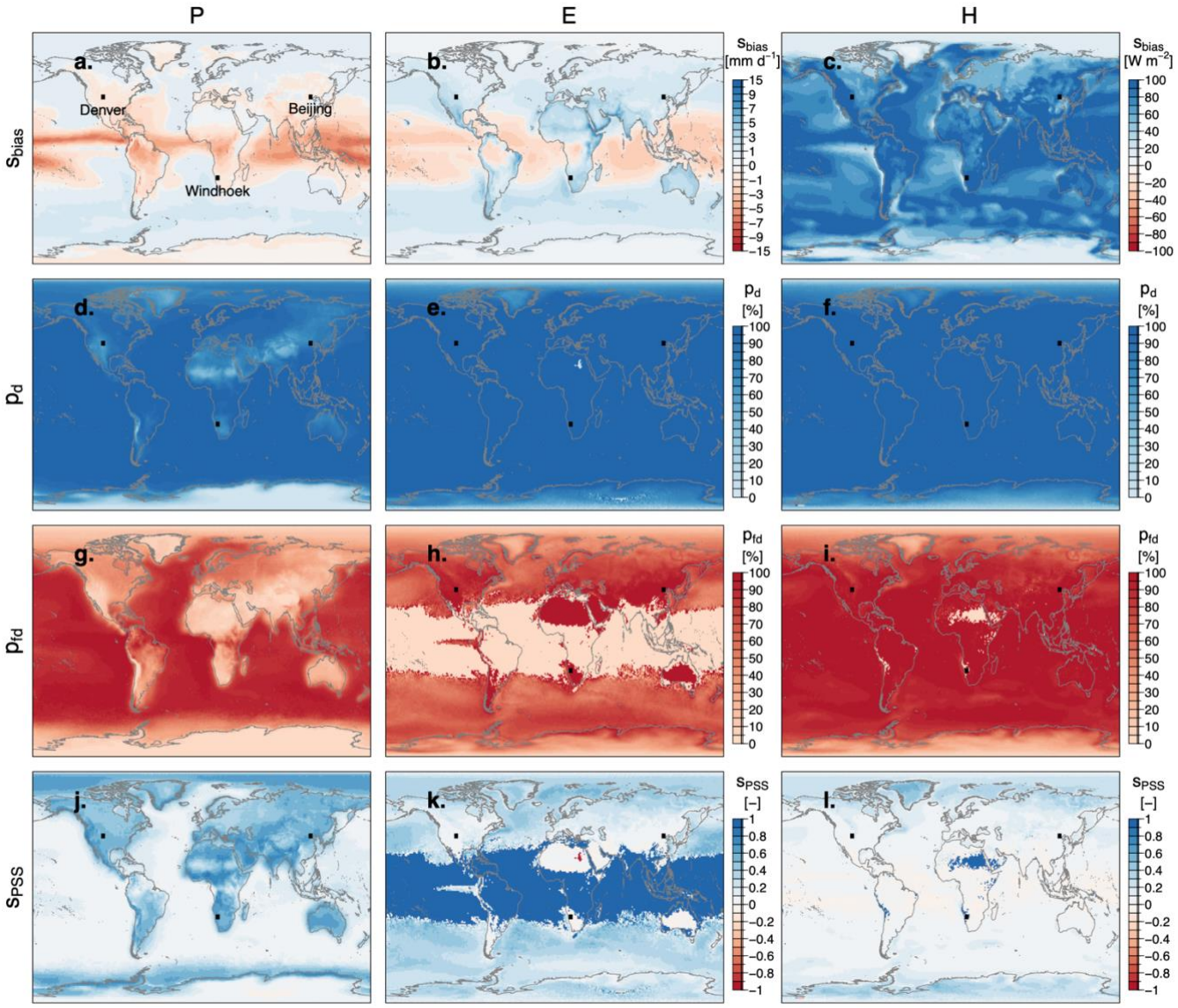

Figure S 2. Same as Fig. 2 of the main manuscript, but considering all ABL changes (ALL-ABL) for the detection of E and H.

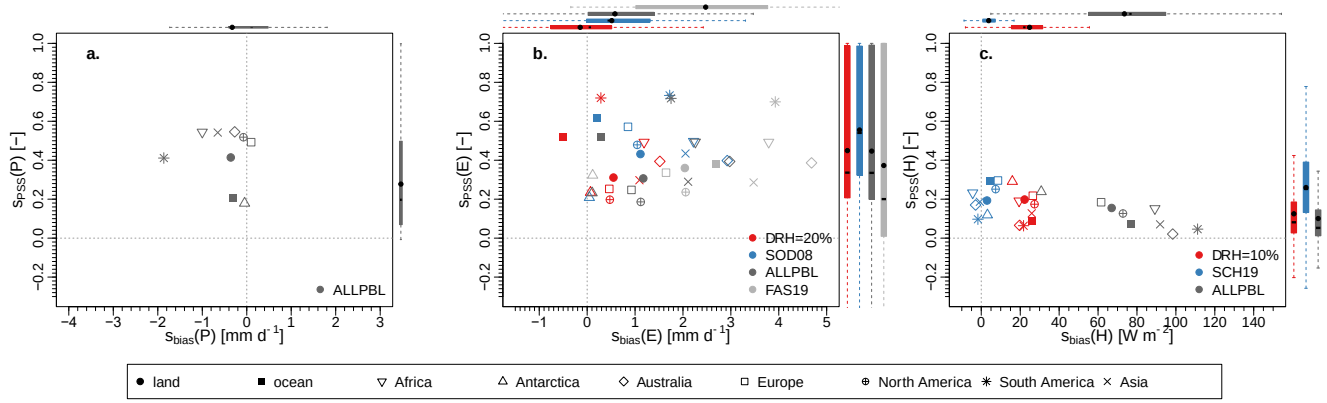

Figure S 3. Similar to Fig. 2 of the main manuscript but showing the PSS on the y-axis (threshold for detection: 1 W m^{-2} and 0.1 mm d^{-1}); and excluding SCH20 for *H*.

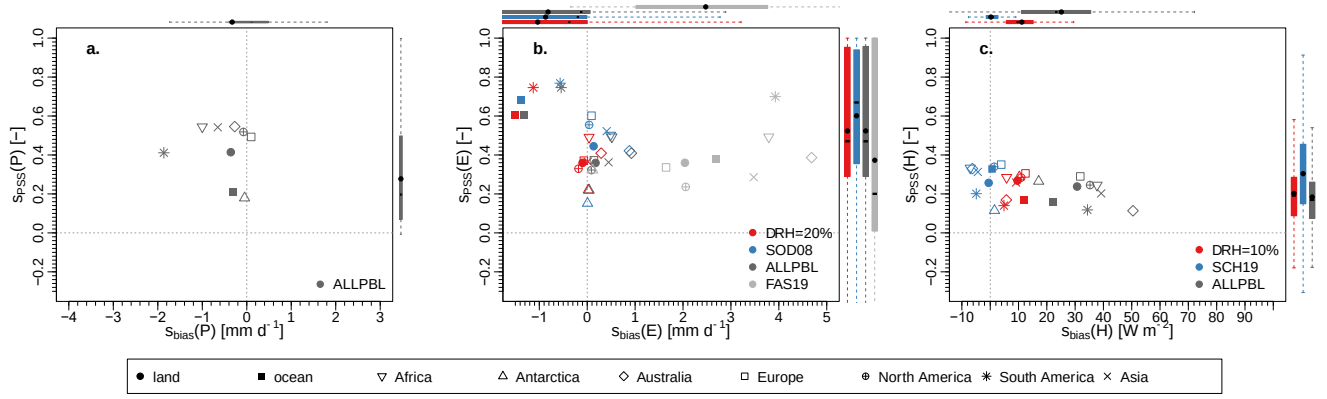

Figure S 4. Same as Fig. S2 but requiring both parcel locations to be within the maximum ABL (threshold for detection: 1 W m^{-2} and 0.1 mm d^{-1}).

4. Attribution

4.1 Attributed fraction

The specific humidity loss associated with precipitation cannot always be fully attributed to the source locations identified along the trajectory:

$$\Delta q_{t_0} \geq \sum_t f_t * \Delta q_{t_0} \quad (15)$$

over all source contributions t along a trajectory; i.e. the attributed precipitation can be (significantly) smaller than the diagnosed loss representing precipitation. This discrepancy occurs because (1) the attribution of source contributions follows a linear algorithm (see Eq. 12 of the main manuscript) and (2) the process-based selection of source locations can prohibit a full attribution. As the linear discounting assumes that all source locations contribute to rain *en route*, this automatically leads to a deficit in the amount of precipitation that can be attributed to all source locations along a trajectory. Further, the fewer source locations are detected, e.g., by restricting the source locations to positive Δq within the ABL, the lower the attributed

fraction. While one could argue that this is an argument against process-based detection criteria and the restriction of source locations within the ABL, we argue that above-ABL moisture increases are likely a result of mixing processes and that the within-ABL locations are (biased) representatives of surface processes. Contrary to Sodemann (2020), we argue that these above-ABL "source locations" do not reflect surface processes, even if the moisture mixed into these parcels originated from surface evaporation in the first place — which in turn may have taken place prior to the mixing and at a different location. These above-ABL moisture increases are, however, assumed to contribute to the specific humidity of parcels *en route* and are thus indirectly considered indirectly in the discounting procedure. In addition, expecting Δq to be biased as it reflects only the net flux $(e - p)$, one may assume that the corresponding source region contributions reflect a reliable detection of source region contributions, and the corresponding weights can be *upscaled* to 100% of the desired quantity using

$$f_{t,upscaled} = \frac{f_t}{\Sigma_t \, f_t}.$$

It is noted here, however, that bias-correcting for precipitation yields the same effect.

5. Source regions of heat and precipitation

5.1 Source regions of heat

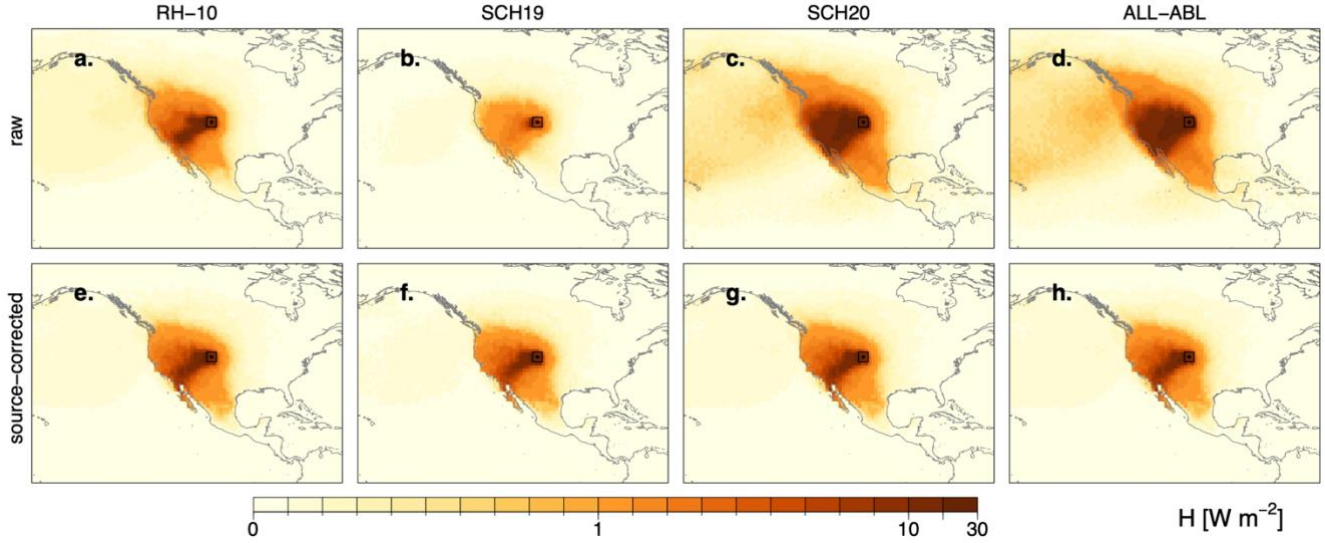

Figure S 5. Same as Fig. 4 but for Denver**.**

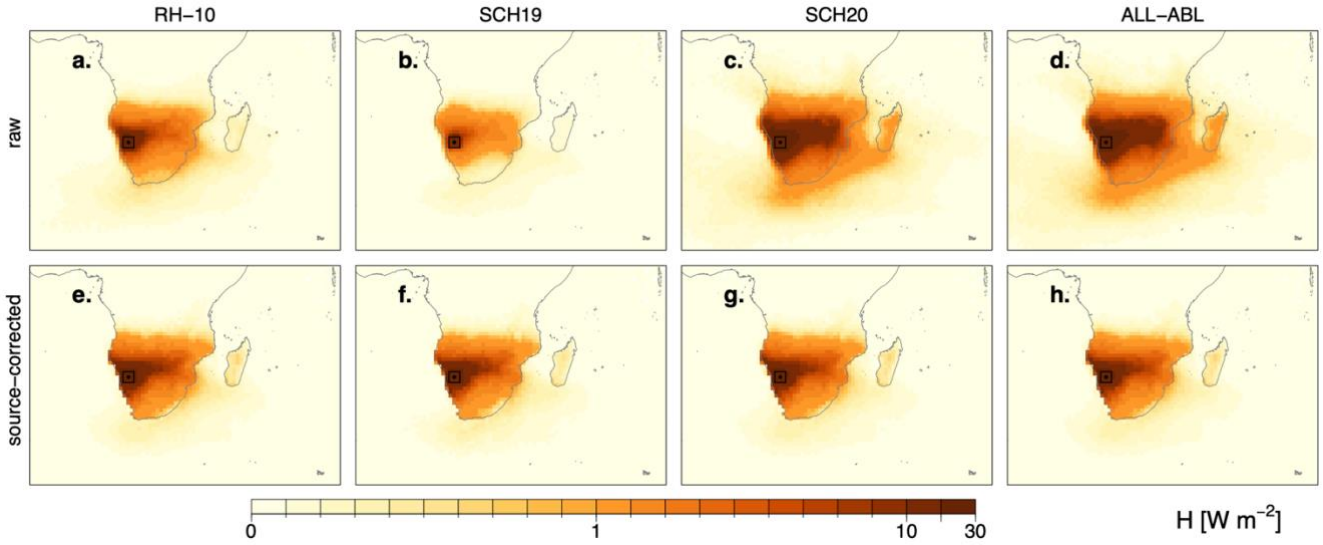

Figure S 6. Same as Fig. 4 but for Windhoek.

5.2 Source regions of precipitation

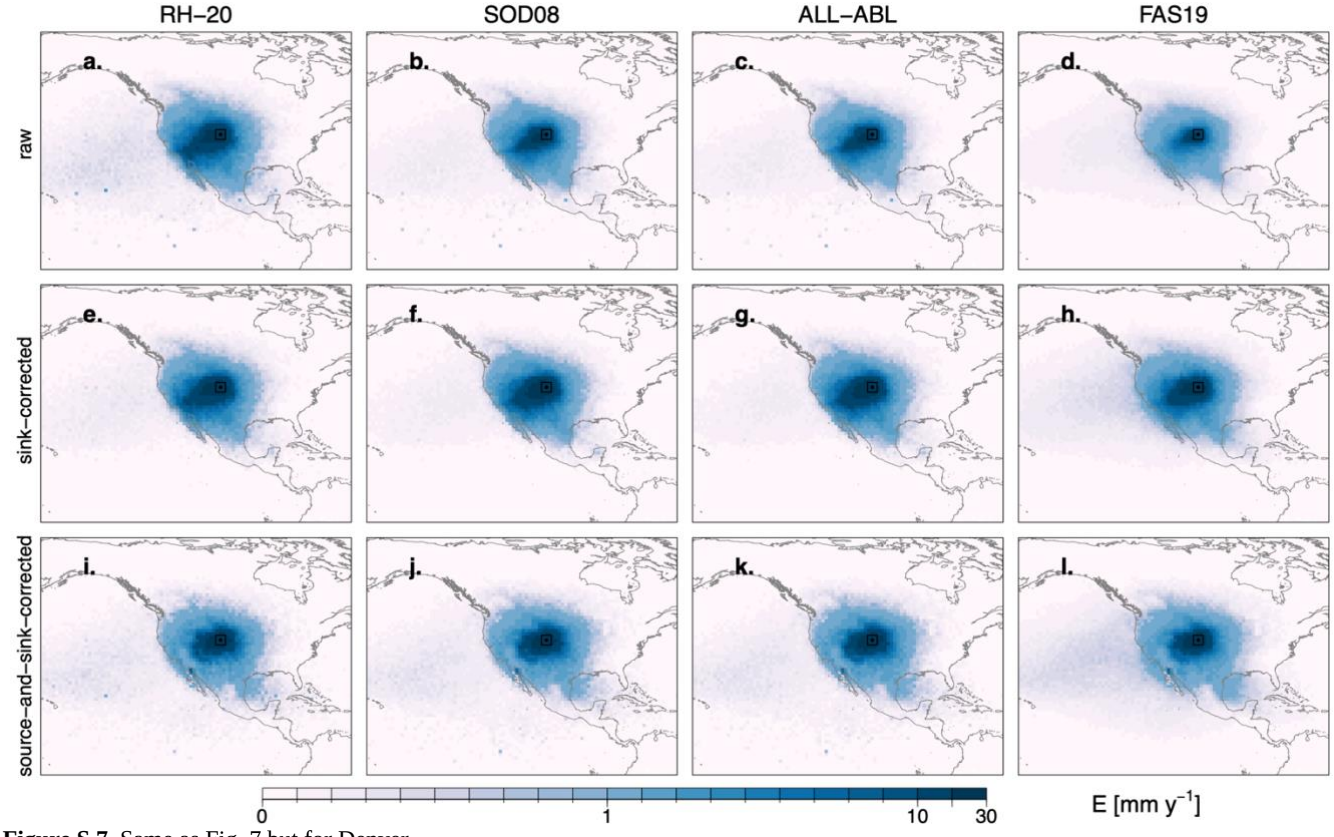

Figure S 7. Same as Fig. 7 but for Denver.

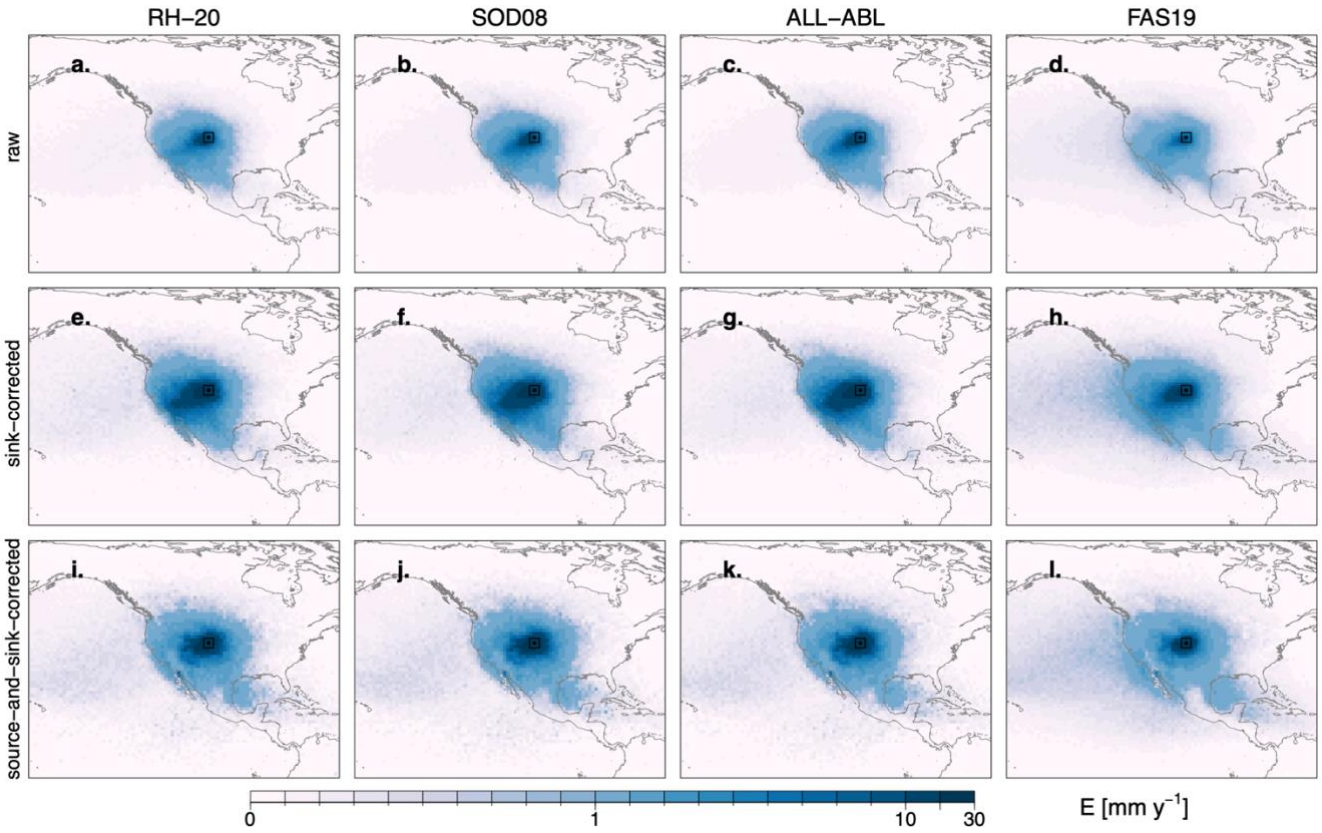

Figure S 8. Same as Fig. S7 but employing the random attribution.

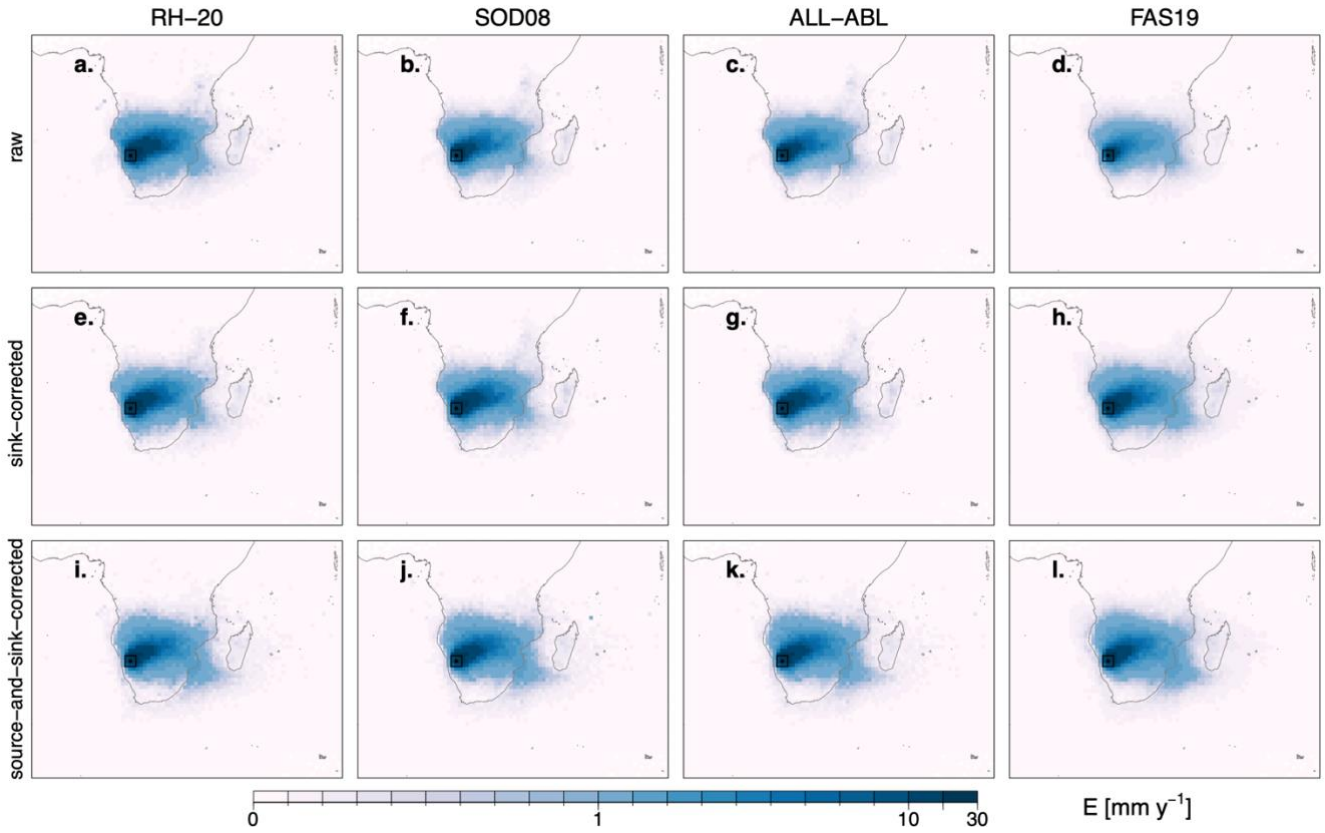

Figure S 9. Same as Fig. 7 but for Windhoek.

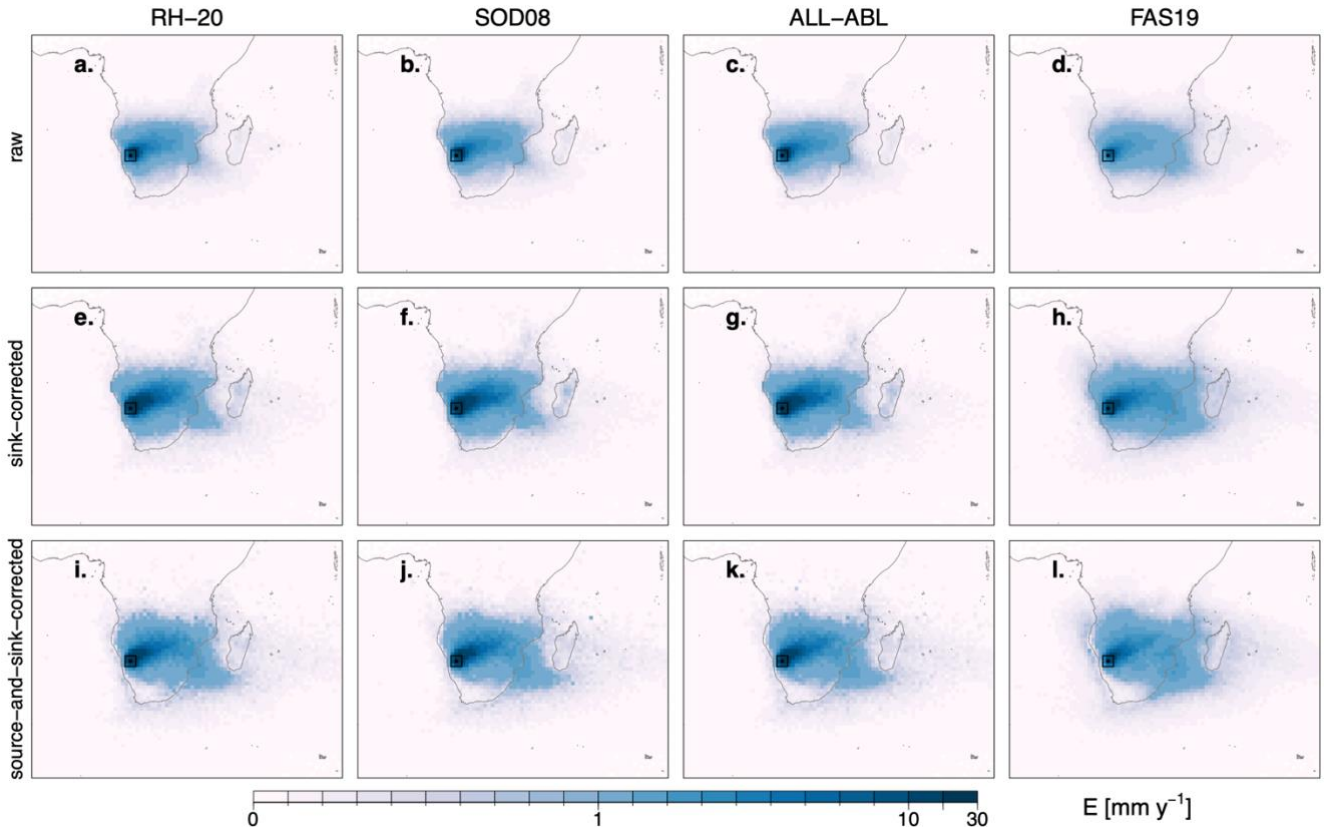

Figure S 10. Same as Fig. S9 but employing the random attribution.