Contents

A) Yeti fundamentals

This document serves as a basic guide for assembling input data and configuration for conducting a Yeti run. The version of Yeti described herein has been developed using Python version 3.7.6. Both development and testing of Yeti have been conducted in a Unix-compatible environment. However, Yeti should also function in other operating systems (such as Windows) when a compatible Python environment is available. It is assumed that the reader has a rudimentary understanding of the basic Python environment and data structures, such as lists and dictionaries, as well as basic operation and organization of HBEFA. The reader should note that all variables are written in `camelCase`.

A1) System and data prerequisites

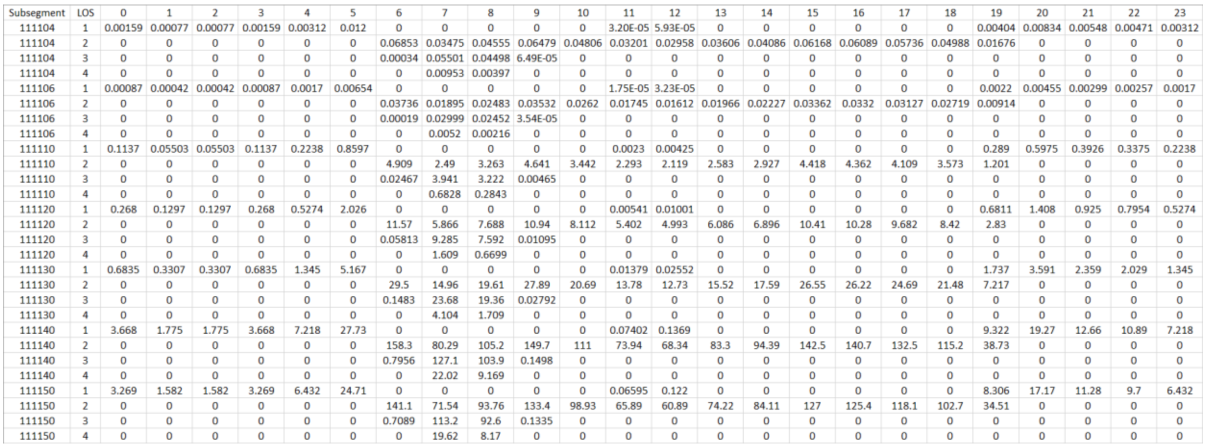
The following Python packages are used in Yeti, which should be part of the standard installation:

`concurrent.futures`, `io`, `itertools`, `math`, `os`, `shutil`, `sys`, `timeit`, `yaml`

Any missing dependencies can be rectified using the `pip3` utility. Yeti has been tested with the pubic edition of HBEFA versions 3.3 and 4.1, which can be purchased on-line from INFRAS (Zürich, Switzerland) through the HBEFA home page (`https://hbefa.net/`). Further, traffic activity data in Yeti format must be tabulated according to available subsegments, and LOS corresponding to the version of HBEFA used. The onus is therefore on the user to procure HBEFA and generate activity data in a suitable format.

A2) Input and output file format

In order to streamline workflow and reduce computational effort, Yeti was conceived with a high level of symmetry in data representation and processing. Text files are used as the sole data exchange medium between the user and other systems. Configuration data are written in YAML format, as indicated by the dependency to the same-named Python package shown previously. On the other hand, input and output data are stored in tab-delimited tabulated text data with column header labels. This facilitates manual inspection of these data using a variety of software packages from text editors to spreadsheet programs. Figure S1 shows a sample of such a tabulated file viewed in the latter.



Subsegment	LOS	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
111104	1	0.00159	0.00077	0.00077	0.00159	0.00312	0.012	0	0	0	0	0	3.20E-05	5.93E-05	0	0	0	0	0	0	0.00404	0.00834	0.00548	0.00471	0.00312
111104	2	0	0	0	0	0	0	0.06853	0.03475	0.04555	0.06479	0.04806	0.03201	0.02958	0.03606	0.04086	0.06168	0.06089	0.05736	0.04988	0.01676	0	0	0	0
111104	3	0	0	0	0	0	0	0.00034	0.05501	0.04498	6.49E-05	0	0	0	0	0	0	0	0	0	0	0	0	0	0
111104	4	0	0	0	0	0	0	0	0.00953	0.00397	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
111106	1	0.00087	0.00042	0.00042	0.00087	0.0017	0.00654	0	0	0	0	0	1.75E-05	3.23E-05	0	0	0	0	0	0	0.0022	0.00455	0.00299	0.00257	0.0017
111106	2	0	0	0	0	0	0	0.03736	0.01895	0.02483	0.03532	0.0262	0.01745	0.01612	0.01966	0.02227	0.03362	0.0332	0.03127	0.02719	0.00914	0	0	0	0
111106	3	0	0	0	0	0	0	0.00019	0.02999	0.02452	3.54E-05	0	0	0	0	0	0	0	0	0	0	0	0	0	0
111106	4	0	0	0	0	0	0	0	0.0052	0.00216	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
111110	1	0.1137	0.05503	0.05503	0.1137	0.2238	0.8597	0	0	0	0	0	0.0023	0.00425	0	0	0	0	0	0	0.289	0.5975	0.3926	0.3375	0.2238
111110	2	0	0	0	0	0	0	4.909	2.49	3.263	4.641	3.442	2.293	2.119	2.583	2.927	4.418	4.362	4.109	3.573	1.201	0	0	0	0
111110	3	0	0	0	0	0	0	0.02467	3.941	3.222	0.00465	0	0	0	0	0	0	0	0	0	0	0	0	0	0
111110	4	0	0	0	0	0	0	0	0.6828	0.2843	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
111120	1	0.268	0.1297	0.1297	0.268	0.5274	2.026	0	0	0	0	0	0.00541	0.01001	0	0	0	0	0	0	0.6811	1.408	0.925	0.7954	0.5274
111120	2	0	0	0	0	0	0	11.57	5.866	7.688	10.94	8.112	5.402	4.993	6.086	6.896	10.41	10.28	9.682	8.42	2.83	0	0	0	0
111120	3	0	0	0	0	0	0	0.05813	9.285	7.592	0.01095	0	0	0	0	0	0	0	0	0	0	0	0	0	0
111120	4	0	0	0	0	0	0	0	1.609	0.6699	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
111130	1	0.6835	0.3307	0.3307	0.6835	1.345	5.167	0	0	0	0	0	0.01379	0.02552	0	0	0	0	0	0	1.737	3.591	2.359	2.029	1.345
111130	2	0	0	0	0	0	0	29.5	14.96	19.61	27.89	20.69	13.78	12.73	15.52	17.59	26.55	26.22	24.69	21.48	7.217	0	0	0	0
111130	3	0	0	0	0	0	0	0.1483	23.68	19.36	0.02792	0	0	0	0	0	0	0	0	0	0	0	0	0	0
111130	4	0	0	0	0	0	0	0	4.104	1.709	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
111140	1	3.668	1.775	1.775	3.668	7.218	27.73	0	0	0	0	0	0.07402	0.1369	0	0	0	0	0	0	9.322	19.27	12.66	10.89	7.218
111140	2	0	0	0	0	0	0	158.3	80.29	105.2	149.7	111	73.94	68.34	83.3	94.39	142.5	140.7	132.5	115.2	38.73	0	0	0	0
111140	3	0	0	0	0	0	0	0.7956	127.1	103.9	0.1498	0	0	0	0	0	0	0	0	0	0	0	0	0	0
111140	4	0	0	0	0	0	0	0	22.02	9.169	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
111150	1	3.269	1.582	1.582	3.269	6.432	24.71	0	0	0	0	0	0.06595	0.122	0	0	0	0	0	0	8.306	17.17	11.28	9.7	6.432
111150	2	0	0	0	0	0	0	141.1	71.54	93.76	133.4	98.93	65.89	60.89	74.22	84.11	127	125.4	118.1	102.7	34.51	0	0	0	0
111150	3	0	0	0	0	0	0	0.7089	113.2	92.6	0.1335	0	0	0	0	0	0	0	0	0	0	0	0	0	0
111150	4	0	0	0	0	0	0	0	19.62	8.17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure S1: Exemplary tab-delimited tabulated data file used in Yeti viewed in a spreadsheet program.

B) Yeti directory structure

As part of the symmetric design concept, Yeti also manifests a high degree of parallelism in the way the input and output data are organized on the file system, so that they could be processed and retrieved quickly and easily. Each set of Yeti emissions is known as a run case, where the input parameters and output data are stored under the same directory, aptly called the case directory. HBEFA emission factors and traffic data can also be stored there for reference, but due to their size they should be introduced as symbolic links (in a Unix-like environment). Alternatively, their location can be specified directly in the user configuration.

B1) Source code

Figure S2 shows the default source and executable structure of Yeti. For better source and data organization, the source code and output data should be stored under a home directory, indicated in Fig. S1 as `yeti/`. The shell script `yeti` is a wrapper to the executable script `main.py`. The location of each Yeti run case, where the case configuration and output files are located, can be specified as a command line argument for `yeti`. If no argument is supplied, Yeti will look for case configuration from the path `yeti/runs/default/`, where emission data are also written out.

B2) Yeti run case

The file organization for a Yeti run case is shown in Fig. S3. As mentioned previously, Yeti uses the path `yeti/runs/default/` by default, while input and output data for multiple cases can be independently stored under the `runs/` directory or elsewhere. All input and output file names and directory structures under each case path are identical, so for illustration, descriptions made for the default Yeti run case (that is, `yeti/runs/default/`) can be applied directly to all other run cases without losing generality.

All execution options specific to the case are contained in the file `config.yaml` in YAML format. This configuration file contains locations of other input data (that is, emission factor, traffic, and meteorological data), basic operation parameters, and emission strategies. Further specification for `config.yaml` will be detailed in Section C1. During execution, Yeti output emission data in the `emissions/` directory. An overview of roadlinks and vehicle subsegments used in the case can be found in the files `keyTable_roadlink.txt` and `keyTable_Subsegment.txt` respectively, in tab-delimited text format. Should the case be re-executed, the `emissions/` directory will be backed-up with a sequentially applied five-digit identifier. A duplicate of the input `config.yaml` contains the exact default user-defined parameters used for the case in question, while the accompanying `logfile` provides basic diagnostics and statistics of the run.

The processed emission data are saved in the sub-directory `hourly/`. The data are first organized in directories based on road link ID and corresponding traffic direction, such that road links with two-way traffic will be identified in two directories. In turn, the emission data are located under each road link / direction pair. Each file contains hourly emissions from all vehicle subsegments for each pollutant, emission strategy, traffic activity and meteorological profiles.

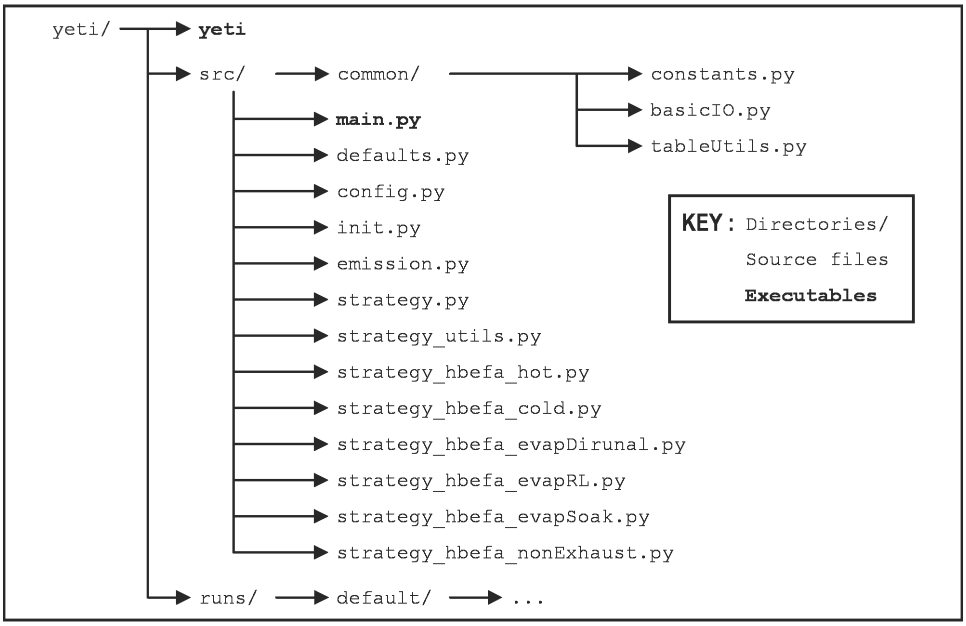


Figure S2: Yeti source code structure.

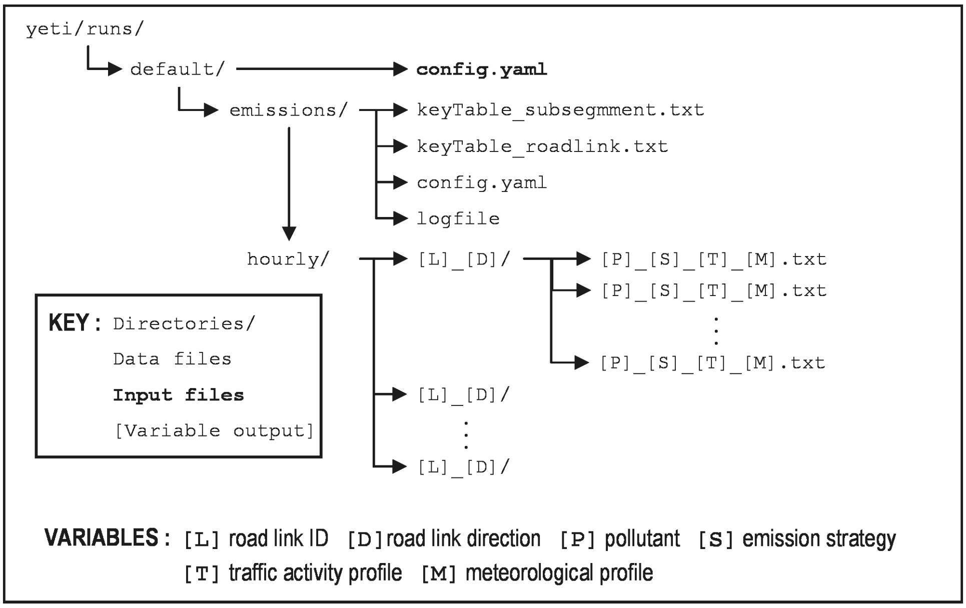


Figure S3: Yeti case directory structure and organization.

B3) HBEFA emission factor data

The layout of the HBEFA data tables used in Yeti is illustrated in Fig. S4. By default, Yeti looks for the HBEFA data under the directory `hbefa/` under the run case directory, but it can be customized to a different user-defined location. All HBEFA data of the same version are stored as tabulated text files in the same directory, and therefore it is advisable to distinguish each set of HBEFA data tables should multiple versions exist. Yeti can be configured to automatically identify the version of HBEFA when the version number is present as the suffix of the folder (e.g., `hbefa_3.3` and `hbefa_4.1`).

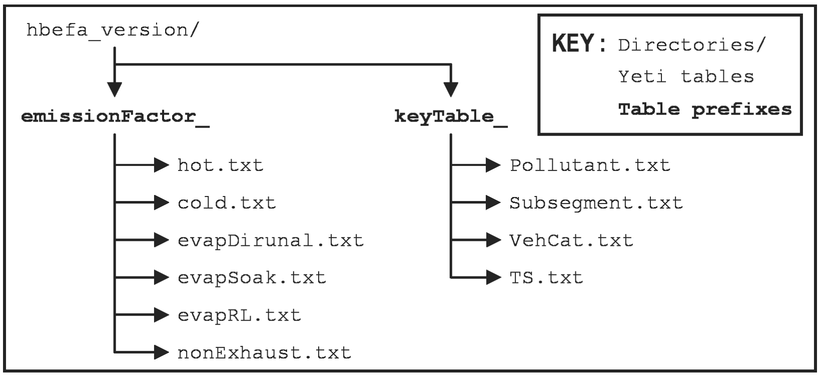


Figure S4: Yeti HBEFA data directory structure and organization.

There are two types of HBEFA data tables that are used by Yeti, which are accordingly indicated on the file naming convention. The first are the emission factor files – identified with the prefix `emissionFactor_` – which contain emission factors from HBEFA. The second are known as key tables – recognizable by the prefix `keyTable_` – containing basic information on pollutants, vehicle subsegments, traffic situations (TS), etc. The two sets of data work in combination in Yeti to serve as a light-weight relational database, and can be extracted from the public edition (i.e., Microsoft Access runtime application) of HBEFA. Table S1 shows the header required by each of the data tables.

B4) Traffic activity data

The organization of traffic activity data, integrated with subsegment and LOS data from a specific version of HBEFA, is presented in Fig. S5. As with the HBEFA data tables, Yeti will by default look for the traffic data in the directory `trafficData/` in the case directory, although the user could specify an alternate location in the configuration. Additionally, the Yeti can be optionally configured to distinguish the HBEFA version with which of the traffic data is generated in the same manner as HBEFA data tables.

The key table containing information for each road link present in the traffic data (`keyTable_roadlink.txt`) is required by Yeti. Optionally, a second key table, `keyTable_fleetSubsegment.txt`, contains subset of the HBEFA key table `keyTable_Subsegment.txt`, for the subsegments appearing only in the vehicle fleet appearing in the road network of interest, can be used by Yeti. The hourly count of each vehicle subsegment passing through each road link in each direction are collected in the directory `hourlyData/`, with each set of data stored in individual directories.

The count data for each traffic activity profile are stored in a table file in a format shown in Fig. S1. A header row indicates the subsegment ID, LOS, and the hourly counts starting at midnight (00h) local time. Typically, the traffic profiles are specific to the type of day throughout the week, and in Yeti, they are classified as Mondays to Thursdays (`workday`), Fridays (`friday`), Saturdays (`saturday`), as well as Sundays and holidays (`holiday`). The file `linkInfo.yaml` contains information of the link and road direction for the user.

B5) Meteorological data

The meteorological data used in Yeti are the diurnal temperature profiles and the seasonal RVPs, which are stored in separate files in table format in a common directory. By default, Yeti will look for the traffic data in the directory `meteorology/` in the case directory, although it can be configured to look for the files in a user-specified location. The temperature data file is `keyTable_temperature.txt`, which contains the hourly diurnal temperature profile for each meteorological profile. The header for the tab-delimited temperature table must be presented as follows:

```
profile   00   01   …   23
```

Table S1: Header information for all HBEFA data tables used in Yeti.

Emission factor file	HBEFA data fields
`emissionFactor_hot.txt`	IDVehCat IDSubsegment IDTS IDPollutant Grad V EFA
`emissionFactor_cold.txt`	IDVehCat IDSubsegment IDPollutant Condition EFA
`emissionFactor_evapDiurnal.txt`	IDVehCat IDSubsegment IDPollutant Condition EFA
`emissionFactor_evapSoak.txt`	IDVehCat IDSubsegment IDPollutant Condition EFA
`emissionFactor_evapRL.txt`	IDVehCat IDSubsegment IDPollutant RoadCat Condition EFA
`emissionFactor_nonExhaust.txt`	IDVehCat IDSubsegment IDTS IDPollutant Grad V EFA
`keyTable_Pollutant.txt`	IDPollutant Pollutant
`keyTable_Subsegment.txt`	IDSubsegment Subsegment IDVehCat Segment SizeClass Technology
`keyTable_VehCat.txt`	IDVehCat VehCat
`keyTable_TS.txt`	IDTS TS

Where the field `profile` refers to the labels denoting each meteorological profile, and the fields `00` to `23` contain the hourly temperature (in °C) for the corresponding profile. On the other hand, the seasonal RVP data can be found in the file `keyTable_temperature.txt`, and the contents of the file is defined by its header as shown below:

```
profile  season  RVP
```

As with the meteorological data, the field `profile` refers to the same meteorological profile labels. The field season corresponds to the `season` for which the RVP is valid (i.e., spring, summer, autumn, and winter). The values of the RVP, expressed in kPa, are located under the field `RVP`. The user should be aware that the meteorological profile labels will be used to distinguish the output of the emission data in Yeti and thus they must be identical in the two files.

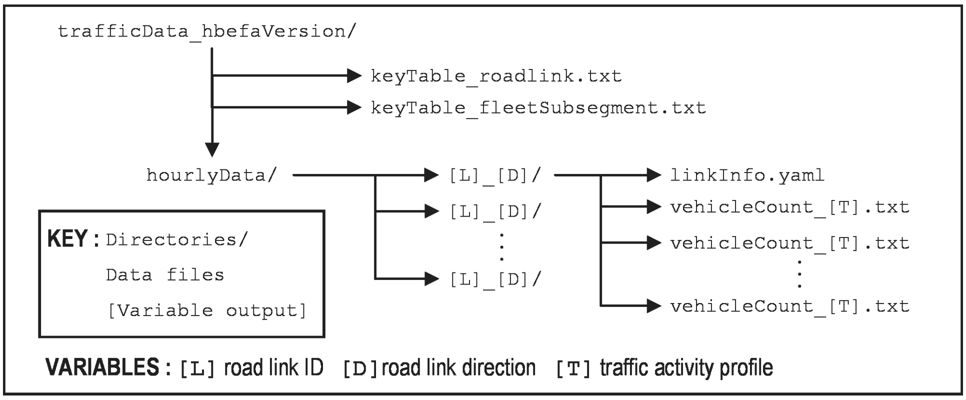


Figure S5: Yeti Traffic data directory structure and organization.

C) Using Yeti

C1) Essential configuration parameters

The run configuration for Yeti is written in YAML format. Most options are presented with two indentation levels, with further levels being represented explicitly by dictionary or list structures. Table S2 shows the minimum required user parameters, their default values (if available), as well as a brief description of their function. Identifiers in boldface are mandatory.

C2) Running a case

As mentioned in Sections B1, a Yeti run can be started using the shell script `yeti` located in the `yeti/` directory. The user can either specify the path to the run case (Fig. S2) as argument to the script or, with no argument, have Yeti use the case path of `run/default`. Once the input file and configuration have been properly assembled, Yeti should start with the following notification:

```
------------------------------------------------------------------
 Yeti 1.0 - DEAGGREGATION UTILITY FOR TRAFFIC EMISSIONS INVENTORY
------------------------------------------------------------------
 *** NOTE : opening logfile runs/default/emissions/logfile
    reading table for roadlinks topology
    reading emission factors
    reading emission key tables
    reading meteorological tables
    harmonizing emission factor tables
    compacting tables
    indexing emission factor tables
    creating folder for hourly emissions under runs/default
    processing disaggregated emissions on 8 processes
 24 of 24 road links processed
    Yeti completed
```

Diagnostic information concerning status and statistics the run can be found in the log file (`runs/default/emissions/logfile`). An excerpt is presented below:

```
------------------------------------------------------------------------
--   reading table for roadlinks topology
------------------------------------------------------------------------
 loading traffic roadlink table runs/default/trafficData_3.3/keyTable_roadlink.txt
 10082 records retrieved
 time elapsed: 0.058 seconds
------------------------------------------------------------------------
--   reading emission factors
------------------------------------------------------------------------
 loading HBEFA emission factor for strategy_hbefa_hot
 985527 records retrieved
time elapsed: 3.815 seconds
------------------------------------------------------------------------
--   reading emission key tables
------------------------------------------------------------------------
```

Table S2: Essential Yeti configuration parameters as specified in `config.yaml`.

Section and Key	Type	Default	Description	Remarks
general				
hbefaVersion	string	3.3	HBEFA version	Required if appendVersion is True
appendVersion	Boolean	False	Attaches HBEFA version to all relevant input and output files and paths	
numProcesses	integer	4	Number of processes spawned for processing emissions	
output				
logfile	string	'logfile'	Name of Yeti log file	
precision	integer	6	Floating point write precision	
encoding	string	'utf-8'	Output text encoding	Uses Python 3 encoding names
hbefa				
tableRoot	string	'hbefa'	Location of HBEFA data tables	If appendVersion is True, hbefaVersion will be attached to tableRoot
encoding	string	'utf-8'	Input text encoding	Possible options are 'latin' and 'ascii'
traffic				
tableRoot	string	'trafficData'	Location of traffic activity data tables	If appendVersion is True, hbefaVersion will be attached to tableRoot
useTrafficSubsegment	Boolean	False	If Yeti vehicle subsegments are limited to those present in the traffic data	Uses complete HBEFA vehicle subsegment listing if it is set to False
lengthConversion	float	0.001	Converts output length unit to km	
subset	[string]	N/A	Defines a unique subset of street links from the traffic data to be processed to reduce runtime (e.g., for a smaller region of interest).	Subsets can be specified using full or partial road name and / or link ID.
encoding	string	'utf-8'	Input text encoding	Possible options are 'latin' and 'ascii'
meteorology				
useMeanProfiles	Boolean	False	If intra-day mean values are used	
meanSeasons	[string]	['summer']	Labels for each mean profile	Required if userMeanProfiles is True
meanTemperatures	[float]	[20.0]	Mean temperatures [°C] to be defined for each meteorological profile	Required if userMeanProfiles is True
meanRVPs	[float]	[70.0]	Seasonal RVPs [kPa] to be defined for each meteorological profile	Required if userMeanProfiles is True
tableRoot	string	'meteorology'	Location of traffic activity data tables	Required if userMeanProfiles is False
encoding	string	'utf-8'	Input text encoding	Possible options are 'latin' and 'ascii'

(*Table continued next page …*)

```
loading HBEFA emission key table runs/default/trafficData_3.3/keyTable_fleetSubsegment.txt
 223 records retrieved
time elapsed: 0.122 seconds
------------------------------------------------------------------------
--   reading meteorological tables
------------------------------------------------------------------------
 loading hourly temperature profile table runs/default/meteorology/keyTable_temperature.txt
 4 records retrieved
time elapsed: 0.001 seconds
------------------------------------------------------------------------
--   harmonizing emission factor tables
------------------------------------------------------------------------
 table runs/default/hbefa_3.3/emissionFactor_cold.txt reduced from 198744 to 8112 records
 time elapsed: 10.576 seconds
------------------------------------------------------------------------
--   compacting tables
------------------------------------------------------------------------
table runs/default/hbefa_3.3/emissionFactor_hot.txt reduced from 985527 to 246192 records
time elapsed: 46.566 seconds
```

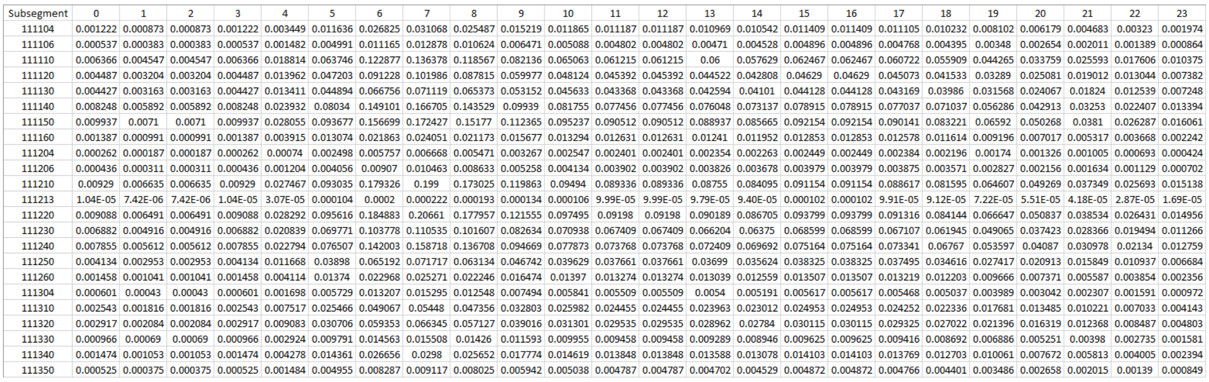
Table S2: Essential Yeti configuration parameters (continued)

Section and Key	Type	Default	Description	Remarks
strategy_hbefa_hot				
active	Boolean	False	If hot exhaust emissions are calculated	At least one strategy must be active
strategy_hbefa_cold				
active	Boolean	False	If cold excess emissions are calculated	At least one strategy must be active
baseColdStartFactor	float	0.0	Fraction of hourly traffic attributed to cold starts for all HBEFA road types	
customColdStartFactor	{ int : float }	N/A	Fraction of hourly traffic attributed to cold starts for specified HBEFA road types	Overrides baseColdStartFactor
strategy_hbefa_evapDiurnal				
active	Boolean	False	If diurnal evaporative HC emissions are calculated	At least one strategy must be active
strategy_hbefa_evapSoak				
active	Boolean	False	If evaporative HC emissions from hot soak are calculated	At least one strategy must be active
baseHotSoakFactor	float	0.0	Fraction of hourly traffic attributed to hot soaks for all HBEFA road types	
customHotSoakFactor	{ int : float }	N/A	Fraction of hourly traffic attributed to hot soaks for specific HBEFA road types	Overrides baseHotSoakFactor
strategy_hbefa_evapRL				
active	Boolean	False	If evaporative HC emissions from running losses are calculated	At least one strategy must be active
strategy_hbefa_nonExhaust				
active	Boolean	False	If non-exhaust PM10, PM2.5, and BC emissions are calculated	At least one strategy must be active

```
------------------------------------------------------------------------
--   indexing emission factor tables
------------------------------------------------------------------------
 creating indices for emission factor table runs/default/hbefa_3.3/emissionFactor_hot.txt
time elapsed: 0.959 seconds
------------------------------------------------------------------------
--   creating folder for hourly emissions under runs/default
------------------------------------------------------------------------
 emissions for the following pollutants will be processed:
       HC CO NOx PM PM (non-exhaust) BC (non-exhaust) PM2.5 (non-exhaust)
 the following emission strategies have been activated:
       strategy_hbefa_hot
------------------------------------------------------------------------
--   processing disaggregated emissions on 8 processes
------------------------------------------------------------------------
 time elapsed: 27.293 seconds
------------------------------------------------------------------------
--   Yeti completed
------------------------------------------------------------------------
```

C3) Obtaining results

The `emissions/` directory contains the disaggregated emissions results of the Yeti run case, with the layout shown in Fig. S2, where the hourly emissions for each vehicle subsegment are first organized by road link ID, traffic direction in a directory, then individually saved in tabulated format according to pollutant, emission strategy, traffic activity and meteorological profiles. Figure S6 shows an example of Yeti output, as viewed in a spreadsheet program. In addition, summary data on the run case configuration can be found in the `emissions/` root directory. Information on road links and vehicle subsegments are saved in the files `keyTable_roadlink.txt` and `keyTable_subsegment.txt` respectively. A complete listing of the particular case run configuration can be inspected in the file `config.yaml`, which can be used to re-execute the Yeti case with settings identical to the original.



Subsegment	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
111104	0.001222	0.000873	0.000873	0.001222	0.003449	0.011636	0.026825	0.031068	0.025487	0.015219	0.011865	0.011187	0.011187	0.010969	0.010542	0.011409	0.011409	0.011105	0.010232	0.008102	0.006179	0.004683	0.00323	0.001974
111106	0.000537	0.000383	0.000383	0.000537	0.001482	0.004991	0.011165	0.012878	0.010624	0.006471	0.005088	0.004802	0.004802	0.00471	0.004528	0.004896	0.004896	0.004768	0.004395	0.00348	0.002654	0.002011	0.001389	0.000864
111110	0.006366	0.004547	0.004547	0.006366	0.018814	0.063746	0.122877	0.136378	0.118567	0.082136	0.065063	0.061215	0.061215	0.06	0.057629	0.062467	0.062467	0.060722	0.055909	0.044265	0.033759	0.025593	0.017606	0.010375
111120	0.004487	0.003204	0.003204	0.004487	0.013962	0.047203	0.091228	0.101986	0.087815	0.059977	0.048124	0.045392	0.045392	0.044522	0.042808	0.04629	0.04629	0.045073	0.041533	0.03289	0.025081	0.019012	0.013044	0.007382
111130	0.004427	0.003163	0.003163	0.004427	0.013411	0.044894	0.066756	0.071119	0.065373	0.053152	0.045633	0.043368	0.043368	0.042594	0.04101	0.044128	0.044128	0.043169	0.03986	0.031568	0.024067	0.01824	0.012539	0.007248
111140	0.008248	0.005892	0.005892	0.008248	0.023932	0.08034	0.149101	0.166705	0.143529	0.09939	0.081755	0.077456	0.077456	0.076048	0.073137	0.078915	0.078915	0.077037	0.071037	0.056286	0.042913	0.03253	0.022407	0.013394
111150	0.009937	0.0071	0.0071	0.009937	0.028055	0.093677	0.156699	0.172427	0.15177	0.112365	0.095237	0.090512	0.090512	0.088937	0.085665	0.092154	0.092154	0.090141	0.083221	0.06592	0.050268	0.0381	0.026287	0.016061
111160	0.001387	0.000991	0.000991	0.001387	0.003915	0.013074	0.021863	0.024051	0.021173	0.015677	0.013294	0.012631	0.012631	0.01241	0.011952	0.012853	0.012853	0.012578	0.011614	0.009196	0.007017	0.005317	0.003668	0.002242
111204	0.000262	0.000187	0.000187	0.000262	0.00074	0.002498	0.005757	0.006668	0.005471	0.003267	0.002547	0.002401	0.002401	0.002354	0.002263	0.002449	0.002449	0.002384	0.002196	0.00174	0.001326	0.001005	0.000693	0.000424
111206	0.000436	0.000311	0.000311	0.000436	0.001204	0.004056	0.00907	0.010463	0.008633	0.005258	0.004134	0.003902	0.003902	0.003826	0.003678	0.003979	0.003979	0.003875	0.003571	0.002827	0.002156	0.001634	0.001129	0.000702
111210	0.00929	0.006635	0.006635	0.00929	0.027467	0.093035	0.179326	0.199	0.173025	0.119863	0.09494	0.089336	0.089336	0.08755	0.084095	0.091154	0.091154	0.088617	0.081595	0.064607	0.049269	0.037349	0.025693	0.015138
111213	1.04E-05	7.42E-06	7.42E-06	1.04E-05	3.07E-05	0.000104	0.0002	0.000222	0.000193	0.000134	0.000106	9.99E-05	9.99E-05	9.79E-05	9.40E-05	0.000102	0.000102	9.91E-05	9.12E-05	7.22E-05	5.51E-05	4.18E-05	2.87E-05	1.69E-05
111220	0.009088	0.006491	0.006491	0.009088	0.028292	0.095616	0.184883	0.20661	0.177957	0.121555	0.097495	0.09198	0.09198	0.090189	0.086705	0.093799	0.093799	0.091316	0.084144	0.066647	0.050837	0.038534	0.026431	0.014956
111230	0.006882	0.004916	0.004916	0.006882	0.020839	0.069771	0.103778	0.110535	0.101607	0.082634	0.070938	0.067409	0.067409	0.066204	0.06375	0.068599	0.068599	0.067107	0.061945	0.049065	0.037423	0.028366	0.019494	0.011266
111240	0.007855	0.005612	0.005612	0.007855	0.022794	0.076507	0.142003	0.158718	0.136708	0.094669	0.077873	0.073768	0.073768	0.072409	0.069692	0.075164	0.075164	0.073341	0.06767	0.053597	0.04087	0.030978	0.02134	0.012759
111250	0.004134	0.002953	0.002953	0.004134	0.011668	0.03898	0.065192	0.071717	0.063134	0.046742	0.039629	0.037661	0.037661	0.03699	0.035624	0.038325	0.038325	0.037495	0.034616	0.027417	0.020913	0.015849	0.010937	0.006684
111260	0.001458	0.001041	0.001041	0.001458	0.004114	0.01374	0.022968	0.025271	0.022246	0.016474	0.01397	0.013274	0.013274	0.013039	0.012559	0.013507	0.013507	0.013219	0.012203	0.009666	0.007371	0.005587	0.003854	0.002356
111304	0.000601	0.00043	0.00043	0.000601	0.001698	0.005729	0.013207	0.015295	0.012548	0.007494	0.005841	0.005509	0.005509	0.0054	0.005191	0.005617	0.005617	0.005468	0.005037	0.003989	0.003042	0.002307	0.001591	0.000972
111310	0.002543	0.001816	0.001816	0.002543	0.007517	0.025466	0.049067	0.05448	0.047356	0.032803	0.025982	0.024455	0.024455	0.023963	0.023012	0.024953	0.024953	0.024252	0.022336	0.017681	0.013485	0.010221	0.007033	0.004143
111320	0.002917	0.002084	0.002084	0.002917	0.009083	0.030706	0.059353	0.066345	0.057127	0.039016	0.031301	0.029535	0.029535	0.028962	0.02784	0.030115	0.030115	0.029325	0.027022	0.021396	0.016319	0.012368	0.008487	0.004803
111330	0.000966	0.00069	0.00069	0.000966	0.002924	0.009791	0.014563	0.015508	0.01426	0.011593	0.009955	0.009458	0.009458	0.009289	0.008946	0.009625	0.009625	0.009416	0.008692	0.006886	0.005251	0.00398	0.002735	0.001581
111340	0.001474	0.001053	0.001053	0.001474	0.004278	0.014361	0.026656	0.0298	0.025652	0.017774	0.014619	0.013848	0.013848	0.013588	0.013078	0.014103	0.014103	0.013769	0.012703	0.010061	0.007672	0.005813	0.004005	0.002394
111350	0.000525	0.000375	0.000375	0.000525	0.001484	0.004955	0.008287	0.009117	0.008025	0.005942	0.005038	0.004787	0.004787	0.004702	0.004529	0.004872	0.004872	0.004766	0.004401	0.003486	0.002658	0.002015	0.00139	0.000849

Figure S6: Exemplary hourly emission output from Yeti viewed in a spreadsheet program.

D) Remarks on data processing

D1) Subsegment mapping between HBEFA version 3.3 and 4.1

Vehicle subsegment definitions between different HBEFA versions are not unconditionally compatible. Although it is expected that the same subsegment IDs introduced in an older version would still appear unaltered in newer versions, they might not be used or assigned in vehicle fleet composition from one version to another, resulting from various technical and logistical factors. Thus, a vehicle subsegment allocated to a particular subsegment ID in one version could be reassigned to a different subsegment ID in another version. In the context of the Berlin fleet, this includes, but not limited to, by the use of segment (engine size) independent passenger vehicle subsegments, the identification of diesel vehicle subsegments equipped with the "defeat device" for manipulating standardized emission tests – and those with subsequent corrective software update, as well as hybrid and flex-fuel vehicles operating under their respective powertrain, all of which introduced in HBEFA version 4.1 under new subsegments.

On the other hand, emission factors for non-exhaust PM in HBEFA 3.3 can only be obtained through its limited availability "expert" version. Importing them from HBEFA 4.1 is a plausible option, but this requires vehicle subsegments to be redefined from 4.1 to 3.3. In addition to the displacement subsegment IDs, the accommodation and attribution of subsegments representing technologies that are still unaccounted for in HBEFA 3.3 are also of primary concern. Their treatment in this case, however, could be less rigorous, as non-exhaust PM is independent of powertrain systems and thus subsegment characteristics such as fuel type and exhaust treatment technology have a diminished influence over non-exhaust PM.

Therefore, some kind of methodology is required to ascertain a systematic bi-directional alignment, or mapping, of vehicle subsegment IDs between HBEFA 3.3 and 4.1, or other relevant versions. While manual mapping is theoretically most flexible, it is labor intensive, having to account for about over 900 subsegment definitions, and it does not always completely incorporate displaced subsegments. Thus, the emission factors for the mapped subsegments may not be representative of those of its original subsegment, as in the case of vehicles belonging to different Euro VI classes, or multi-fuel vehicles. The method presented below is indicative of one such possible way with which an effective subsegment mapping can be made, using a combination of classification, lexicon analysis, and heuristics.

The process begins by selecting active subsegments for each HBEFA version, that is, subsegments that have non-zero emission factors in the corresponding version. This is accomplished by looking up each subsegment in the emission factors tables. The two sets of subsegments are then separated into match pools belonging to the same combination of HBEFA categories and, if available, the Euro emission standard. This strategy can significantly improve computational effort by restricting the matching operation into smaller pools, since the computational effort, that is, the number of look-ups required to find a match, scales quadratically with the size of the pool.

Eligible map candidates for each subsegment are located lexically by selecting those subsegment descriptors with the minimum string edit distance defined by Levenshtein (1966), an algorithm which scales logarithmically with the length of both strings. Degenerate cases, that is, a subsegment having more than one match with identical minimum Levenshtein distance, are eliminated using a heuristic match to the subsegment descriptors. The listing of heuristic matching used is presented in Table S3.

Table S3. Heuristic used in HBEFA subsegment mapping.

> 12 t	↔	> 7.5-12 t
151 cc	↔	≤ 250 cc
251 cc	↔	> 250 cc
BEV	↔	EE
EA189	↔	< 2.0 L
Euro-6c	↔	Euro-6
Euro-6d-temp	↔	Euro-6d1
Fuel cell	↔	EE
LPG	↔	CNG
Rigid Truck	↔	RT

The lexical heuristics can still be determined manually by inspecting mappings that still remain degenerate following the lexical analysis. A subsegment map is established between the versions when a unique eligible match is found. Degenerate subsegments (still having non-unique eligible matches) as well as invalid subsegments (having no matches) will not be provided a matching subsegment. Such cases typically represent subsegments whose powertrain technologies or Euro emission classes are unavailable in the target HBEFA version.

Subsegment maps from HBEFA 3.3 to 4.1, and vice versa, are produced using the aforementioned methodology. A total of 606 matches out of 970 subsegments are mapped from version 3.3 to 4.1, while, in the other direction, 704 maps have been identified out of a total of 1840 subsegments. These mapping are then applied to the fleet composition data, so that vehicle subsegment IDs from the base HBEFA version can be substituted into those for the target version, as part of the traffic data pre-processing.

D2) Discretizing Yeti data for chemical transport models

One of the objectives for developing Yeti is to produce high-resolution traffic emissions data for chemical transport modelling studies that are consistent with the meteorological conditions for the region and period of interest. An algorithm is currently being undertaken using a k-dimensional tree approach (Bentley, 1975), to reconcile structure and topological heterogeneities between the road network and corresponding road surfaces in the model grid in a computationally efficient manner. It has been tested on some modelling platforms such as WRF-Chem (Grell et al, 2005) and OpenFOAM (Weller et al, 1998; Chan and Butler, 2021), as shown in the example in Figure S7.

References

Bentley (1975) Commun ACM 18(9):509-517.
Chan & Butler (2021) Geosci Model Dev 14:4555-4572.
Grell et al (2005) Atmos Environ 39(17):6957-6975.
Levenshtein (1966) Sov Phys Dokl 10(9):707-710.
Weller et al (1998) Comput Phys 12:620-631.

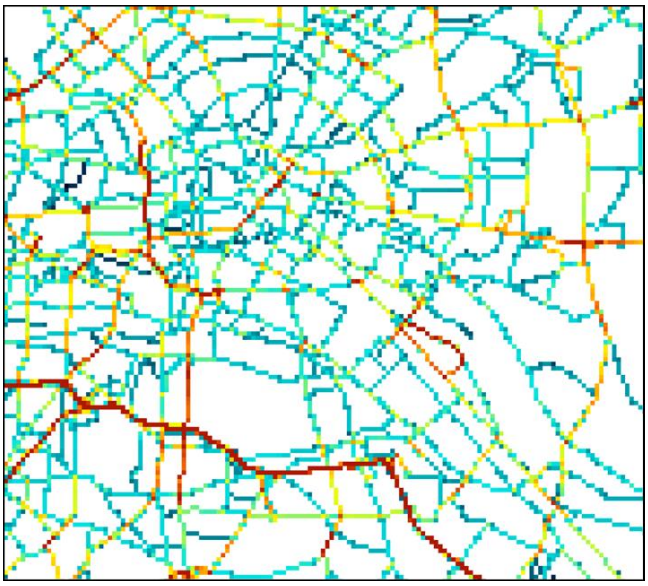


Figure S7. An illustrative example of gridded emissions output over a WRF-Chem domain at a horizontal resolution of 100 m using the k-dimensional tree algorithm (Bentley, 1975).