# Appendix

## Appendix 1.1 Pollution and meteorological station details for London

The table below depicts metadata for air quality stations in London.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Import function (Openair)** | **Location (name)** | **Latitude** | **Longitude** | **Background type** | **Station code** | **Data retrieved** |
| importKCL | Lambeth - Brixton Road | 51.46411 | -0 .114581 | Kerbside | MY1 | NO2, NOx and O3 hourly mean mass concentration (µg/m³) |
| importAURN | Westminster - Marylebone Road | 51.522530 | - 0.154611 | Kerbside | LB4 | Wind speed (m/s), wind direction (°), ambient air temperature (°C) |

## Appendix 1.3 Pollution and meteorological station details for Beijing

The table below depicts the information of meteorological data and air quality data for Beijing.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Station name** | **Latitude°** | **Longitude°** | **Station type** | **Data retrieved** |
| Aotizhongxin | 116.397 | 39.982 | Air quality station | PM10 mean hourly mass concentration (µg/m³) |
| Hadian | 116.2905556 | 39.98694444 | Meteorological station | Wind speed (m/s) and wind direction (°) |

## Appendix 1.4

The table below reports the Software and version employed in this study.

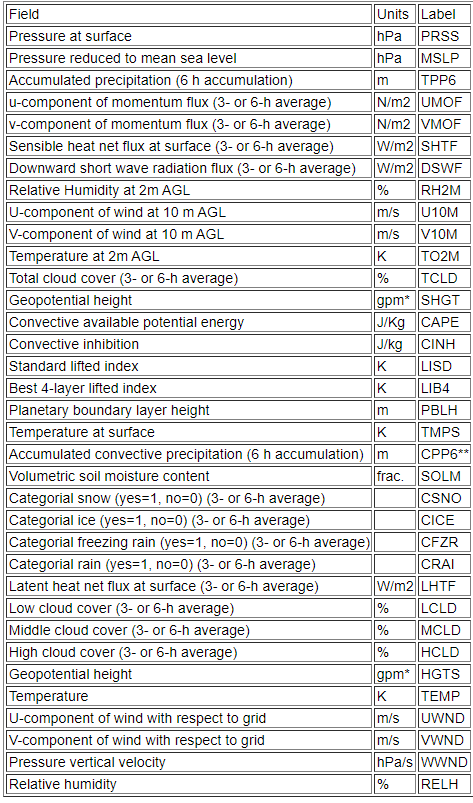
|  |  |  |  |
| --- | --- | --- | --- |
| **Software name** | **Developer** | **Version** | **Release date** |
| R software | The R Foundation for Statistical Computing | 3.6.3 | 29th February 2020 |
| Igor Pro | Wavemetrics | 6.3.8.1 |  |
| Python | Python Software Foundation | 3.7.3 | 25th March 2019 |
| HYSPLIT | NOAA and Australia's Bureau of Meteorology. | 4.2.0 | September 2019 |
| PySPLIT | Mellissa S.C. Warner | 0.3.5 | 1st June 2019 |
| Zefir | Jean-Eudes PETIT | 3.7 | 9th January 2020 |

A special license was requested to Wavemetrics (Academic license) as the package Zefir was supported by Igor Pro version 6.3.7 but compatibility with version 7 was not guaranteed.

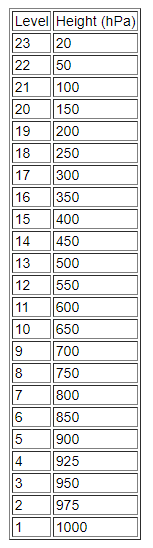
## 

## Appendix 1.5 GDAS1 meteorological files

The table below depicts the Meteorological fields in GDAS1 files for 3-h accumulation average fields.



The table below depicts a description of vertical levels in GDAS1 data files.



## Appendix 1.6

The table below reports the geographical coordinates of the BT source locations and periods.

|  |  |  |  |
| --- | --- | --- | --- |
| **BT source location** | **Latitude°** | **Longitude°** | **Periods** |
| London | 51.50 | 0.12 | January 2014, 2015, 2016, 2017 and 2018 |
| Beijing | 40.00 | 116.00 | November and December 2013, 2014, 2015, 2016 and 2017 |

## Appendix 1.7 BT arrival altitude and surface sensitivity analysis

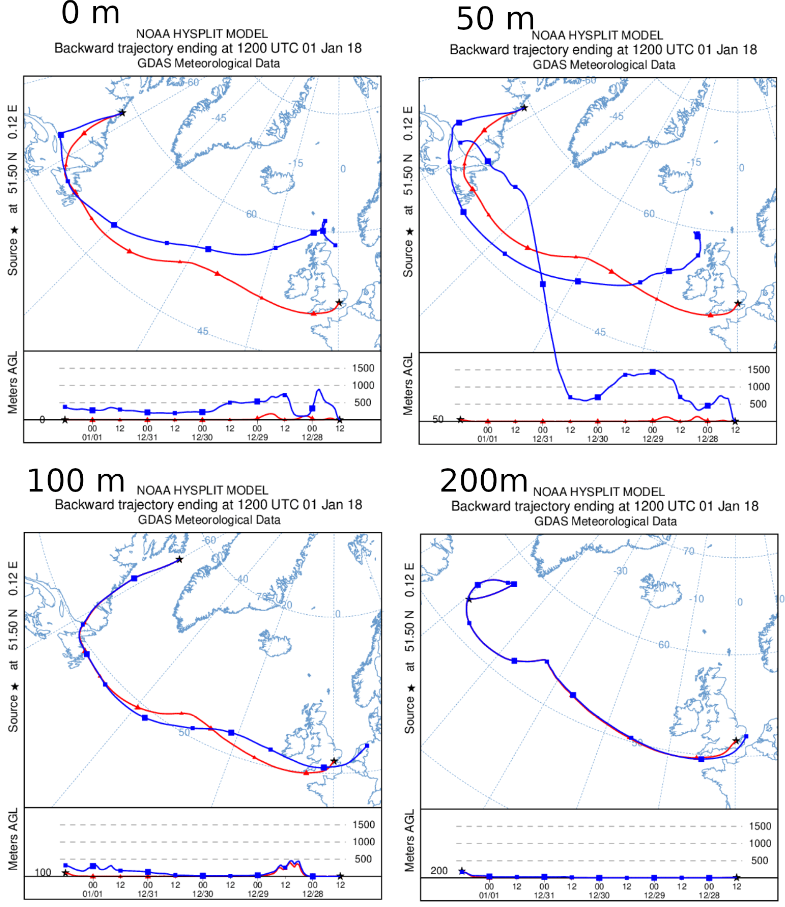


Figure 1: Panel depicting BTs (in red) and correspondent reverse trajectories (in blue) source location altitude. BT source location coordinates and arrival time are displayed on the figure.

The surface sensitivity test revealed the effect of the ground surface on trajectory paths. It is evident from figure 1 that BT generated at 0, 50 and 100 m above ground have poor agreement with the paths of trajectories generated with source location at the endpoint of original BT (reverse trajectories). This suggests that for altitudes of 0, 50, 100 and 200 m AGL, the ground surface has a significant effect on the trajectory path, hence leading to errors. BT with source location at 200 meters have good agreement with their corresponding reverse trajectory, which signifies the effects of the ground surface is less severe (figure 1). However, the ground has still a significant effect, as the trajectory path appears to entirely at an altitude of 0m altitude, which is unexpected. The altitude at which minimal influence from the ground surface was identified at 400m (visualized in figure 4).

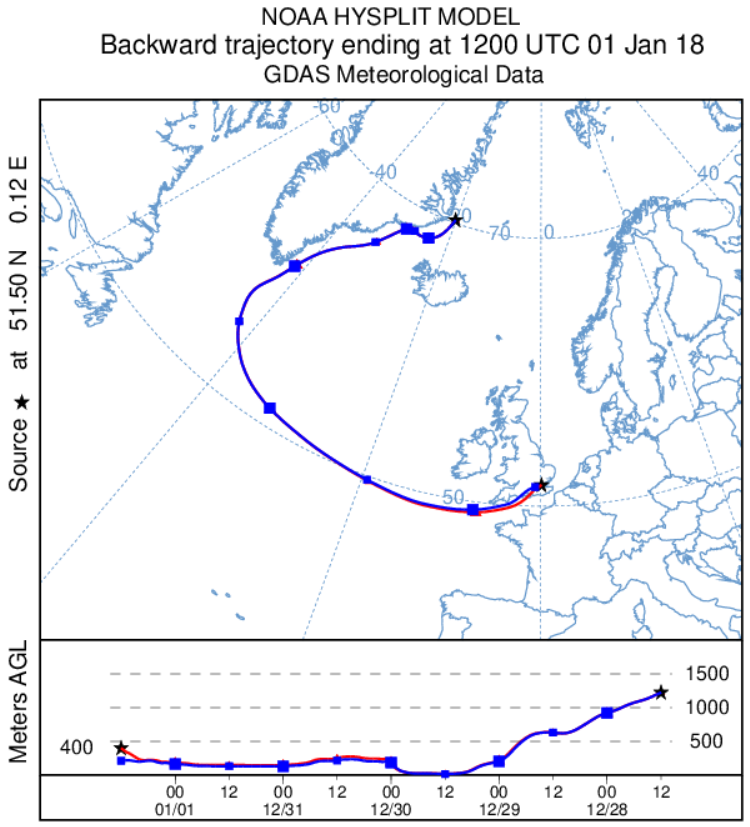


Figure 2: diagram depicting a BT (in red) and correspondent reverse trajectory (in blue) with source location altitude and endpoint altitude respectively at 400 m. The source location time, and coordinates of the BT source location and endpoint of reverse trajectory respectively are displayed on the figure.

## Appendix 1.8 HYSPLIT clustering algorithm and TSV

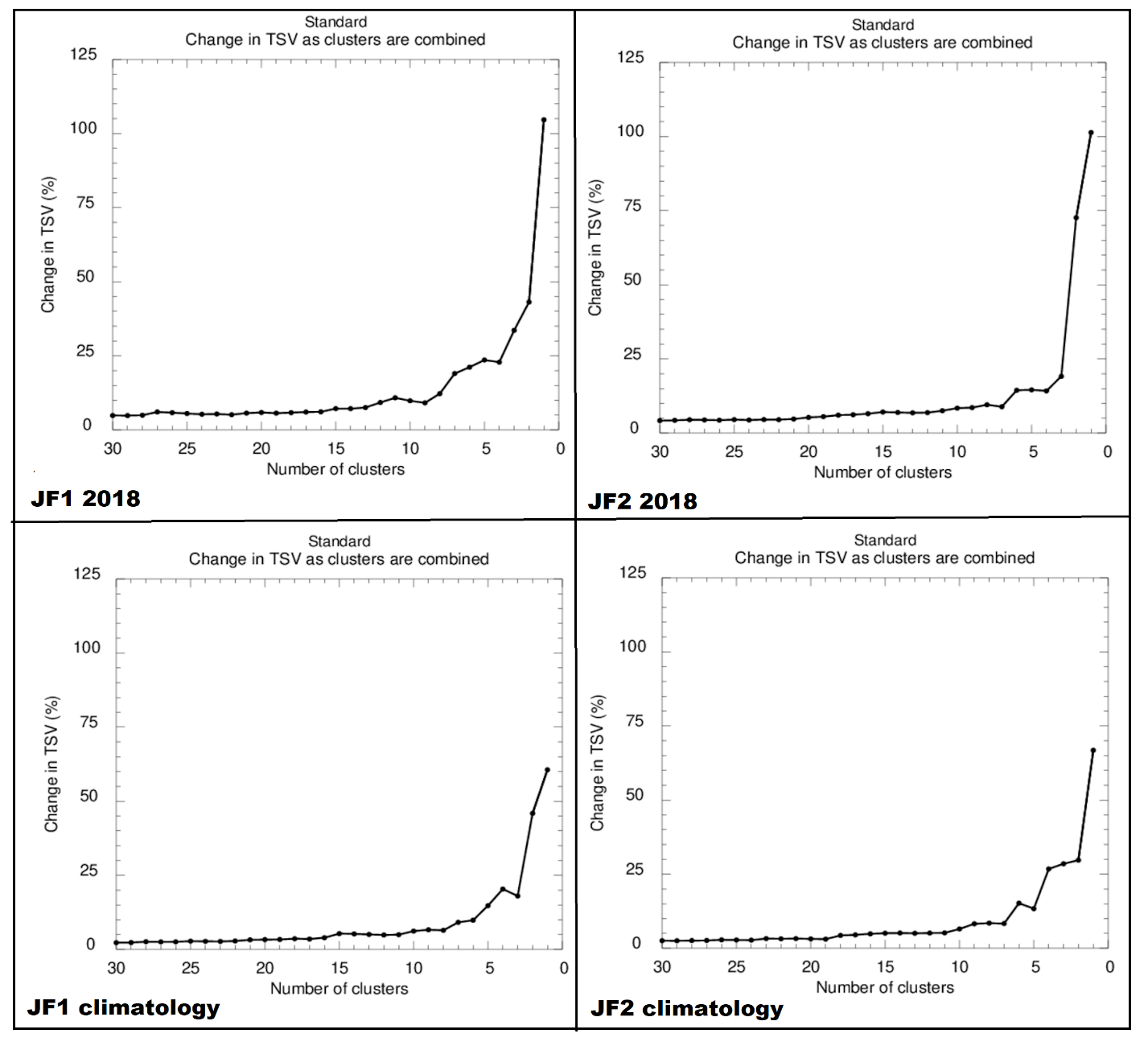
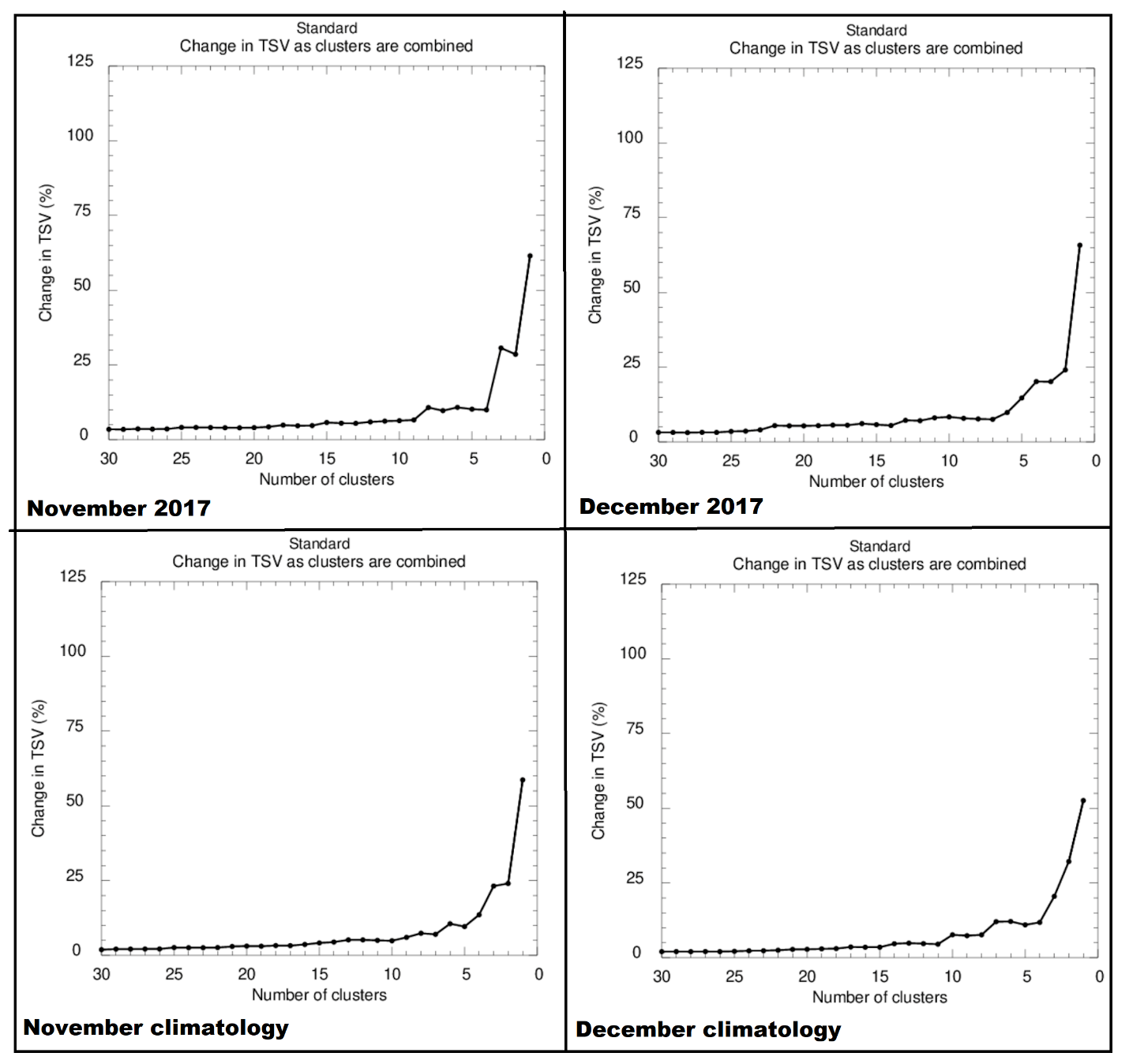
The clustering procedure minimises the spatial variability between BT of the same cluster whilst maximising the variability between clusters, thus allowing to merge the BT in different groups having similar synoptic scale transport patterns. HYSPLIT default algorithm adopts different computational steps. The first entails calculating spatial variance (SV) between an individual trajectory and the cluster mean. Then these differences are summed, representing the SV of each trajectory within the cluster, the cluster spatial variance (CSV). Lastly, the sum of all the CSVs over all clusters is summed to obtain the total spatial variance (TSV), which relevance is discussed later in this paragraph. The clustering procedure starts by assigning each trajectory to its own cluster. Then the spatial positions on each cluster are compared to all other trajectories singularly, one in each step, forming pairs and the SV is computed for each pair. The pair with minimum CSV is merged to produce a new cluster and, in each iteration, TSV is computed for each pair and pairs with minimum values are combined in clusters until only cluster is left. At the beginning of the procedure, TSV increases considerably, then plateaus for following iterations, to rise again towards the end of the computation, when the main selected clusters are merged. To adopt clusters with minimal intra-cluster variation and maximal inter-cluster variation (natural clusters), HYSPLIT developers suggest the ideal number of clusters should be just before the sudden increase in TSV. The TSV diagrams for BT clusters in this study are depicted in figure 3 for London and figure 4 for Beijing.

Figure 3: TSV diagrams for London in the study periods.

Figure 4: TSV diagrams for London in the study periods.