A comparison of HYSPLIT backward trajectories generated from two GDAS datasets

Lin Su a, Zibing Yuan b,⁎, Jimmy C.H. Fung b,c,d, Alexis K.H. Lau b,c

a Environmental Science Programs, School of Science, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China
b Atmospheric Research Center, HKUST Fok Ying Tung Graduate School, Nansha IT Park, Guangzhou 511458, China
c Division of Environment, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China
d Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China

HIGHLIGHTS

• Vertical velocity is critical for calculating backward trajectories.
• Trajectory difference is higher in winter and smaller in summer.
• GDAS datasets are verified by comparison with observations and air pollution data.
• GDAS1 dataset is recommended to be used in the Pearl River Delta region.

ABSTRACT

The Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model is widely used to generate backward trajectories in given starting locations. However, differences exist between trajectories generated from the model with different input datasets. In this study, backward trajectories in Hong Kong in the entire year of 2011 are derived by HYSPLIT model. Two sets of Global Data Assimilation System (GDAS) output data associated with different horizontal and vertical resolutions (GDAS1 and GDAS0P5) are used as drivers in an attempt to quantify the differences between the results and discover the underlying reasons responsible for discrepancy. The results reveal that the significant differences between back trajectories generated from the two GDAS datasets can be mainly attributed to different vertical velocity calculation methods due to the absence of vertical velocity in GDAS0P5 dataset. The HYSPLIT trajectories are also sensitive to the horizontal and vertical resolutions of the input meteorological data, but to lesser extents. Results of cluster analysis indicate that when the air mass is from the north, northeast, or west with a long-to-medium range, the HYSPLIT backward trajectories are sensitive to the vertical advection calculation method and data resolution, whereas when the air mass is from the south or southwest with a long range, the trajectories are more likely to remain unchanged with the shifting of vertical velocity or data resolution. By comparing the vertical velocities with the observations and the performance in retrieving PM contributions from different directions, we conclude that GDAS1 dataset is more plausible in backward trajectory analysis in the Pearl River Delta.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model is widely used to generate air mass backward trajectories in given starting locations. HYSPLIT model is often driven by meteorological data output from the Global Data Assimilation System (GDAS) that is used by the Global Forecast System (GFS) model to place observations into a gridded model space for starting or initializing weather forecast with observational data. Traditionally, the GDAS data is with a horizontal resolution of 1° which corresponds to ~100 km × 100 km and 23 vertical layers (GDAS1, available at ftp://arlftp.arlhq.noaa.gov/pub/archives/gdas1). Starting from 2011, the Air Resources Laboratory (ARL) of the National Oceanic and Atmospheric Administration (NOAA) starts to provide GDAS data with 0.5° horizontal resolution and 55 vertical layers (GDAS0P5, available at ftp://arlftp.arlhq.noaa.gov/pub/archives/gdas0p5).

Although GDAS1 and GDAS0P5 assimilate the same observations, they have differences in various aspects, including horizontal resolution and vertical velocity field. The vertical distributions of GDAS1 and GDAS0P5

⁎ Corresponding author.
E-mail address: zibing@ust.hk (Z. Yuan).

http://dx.doi.org/10.1016/j.scitotenv.2014.11.072
0048-9697/© 2014 Elsevier B.V. All rights reserved.
are shown in Table 1. As GDAS0P5 is in hybrid sigma-pressure coordinates, the pressure level of the GDAS0P5 data is defined as

\[ P = A + B \times \text{PRSS}, \] (1)

where PRSS is the surface pressure, \( A \) is the integral fraction, and \( B \) is the decimal fraction of the level (Stunder, 2004; Stunder, 2012). Therefore, the vertical classification of GDAS0P5 in Table 1 is available when the surface pressure is assumed to be 1000 hPa. As most of the trajectories reside in the low to middle troposphere, the vertical distributions from ground level to 5000 m above ground level (AGL) are more important. There are 25 layers in GDAS0P5 but only 13 layers in GDAS1 between the surface and 5000 m AGL.

Apart from the horizontal and vertical resolutions, the main difference between GDAS1 and GDAS0P5 is that GDAS1 includes vertical velocity while GDAS0P5 does not. As a result, the vertical motion of the trajectory generated from GDAS1 can be calculated directly from the model output vertical wind velocity, whereas the vertical velocity of GDAS0P5 has to be calculated from the vertical integration of the horizontal velocity divergence:

\[ w = \int \text{DIV} dz, \] (2)

where \( w \) is the vertical velocity, and DIV is the horizontal velocity divergence calculated by the formula

\[ \text{DIV} = \frac{\text{DELU}}{DX} + \frac{\text{DELV}}{DY}. \] (3)

\( DX \) and \( DY \) are grid distance in x- and y-directions, while \( \text{DELU} \) and \( \text{DELV} \) are the u-component and v-component of the horizontal velocity divergence calculated from formulae (4) and (5), respectively:

\[ \text{DELU} = 0.5 \times (U(i + 1, j, k) - U(i - 1, j, k)) \] (4)

\[ \text{DELV} = 0.5 \times (V(i, j + 1, k) - V(i, j - 1, k)), \] (5)

where \( U \) and \( V \) are the u-component and v-component of the horizontal velocity, respectively (Draxler et al., 2012a, 2012b).

Studies have shown that HYSPLIT backward trajectories tend to differ significantly with different input meteorological data (e.g. Cabello et al., 2008; Gebhart et al., 2005) or with different vertical transport methods (Harris et al., 2005). Differences in HYSPLIT trajectories with different data inputs tend to increase when air mass is passing through areas with complicated topography, diversified land-use, and meteorological patterns that are often in a scale much less than the data resolution. The Pearl River Delta (PRD) in southern China is such an area with frequent occurrences of micro-scale circulations, such as land-sea breeze, mountain-valley wind and urban heat island effect (Ding et al., 2004; Zheng et al., 2010; Li et al., 2012; Zhong et al., 2013). It is therefore essential to quantify the differences and examine the reliability and robustness of the backward trajectories generated from different input datasets when adopting the backward trajectory analysis in this region. Such an examination is unfortunately absent.

Current backward trajectory analysis in the PRD region is mostly driven by either data converted from the final reanalysis (FNL) data of National Center for Environmental Prediction (NCEP) operational model runs with horizontal resolution of 190.5 km or GDAS1, both of which are too coarse to depict the micro-scale meteorology that prevails in the PRD, therefore pose uncertainties in trajectories (e.g. Cheng et al., 2006; Lee et al., 2006; Li et al., 2006; Wang et al., 2006; Guo et al., 2009; Li et al., 2009; Ding et al., 2011; Ho et al., 2011). With finer horizontal and vertical resolutions, GDAS0P5 dataset is assumed to be potentially capable of better reproducing the actual meteorological conditions and providing more reliable backward trajectories. This possible improvement needs to be corroborated and documented by sophisticated analyses and comparisons with GDAS1 results and the observations.

The main objectives of this study are, first, to quantify the differences between HYSPLIT backward trajectories generated from two different GDAS datasets over 72-hour periods in the PRD region; and second, to conduct sensitivity experiments and quantify the contributions of the main factors to the differences between the trajectories generated from the two datasets; and last but not least, to provide suggestions for choosing input data in backward trajectory analysis in the Pearl River Delta. The model configuration and statistical methods are described in Section 2. The comparison results and discussion are presented in Section 3, and the summary and concluding remarks are presented in Section 4.

### 2. Methodology

In this study, 72-hour backward trajectories are calculated four times a day (02UTC, 08UTC, 14UTC and 20UTC) from January 1 to December 31, 2011, using HYSPLIT version 4.8 (Draexler and Hess, 1997; Draxler et al., 2012a, 2012b). The starting location for the trajectories is the Central/Western (CW) air quality monitoring station (latitude: 22.29°N, longitude: 114.14°E) at an elevation of 500 m above ground level (AGL). CW is located in a suburban area in the north of Hong Kong Island.

Two meteorological datasets archived on the ARL server are used to drive the HYSPLIT model: (a) GDAS1 with 1-degree horizontal resolution, 23 vertical levels of pressure coordinates, and a temporal resolution of 3 h; and (b) GDAS0P5 with 0.5° horizontal resolution, 55 vertical levels of hybrid sigma-pressure coordinates, and a temporal resolution of 3 h. Trajectories with complete endpoints over 72 h for both datasets are retained for comparison, yielding 1425 pairs of valid trajectories in 2011.

#### 2.1. Sensitivity experiments

GDAS1 and GDAS0P5 datasets show differences in the horizontal resolution, vertical resolution and vertical advection calculation method. To evaluate the individual effect on the differences in backward trajectories, four sensitivity experiments with two intermediate scenarios are designed, as illustrated in Table 2.

In the first experiment (EXP_CTRL), HYSPLIT is driven by the two datasets with their original settings: (a) GDAS1 with vertical velocity from model output and (b) GDAS0P5 with vertical velocity calculated from horizontal divergence. The EXP_CTRL results illustrate the grand improvement needs to be corroborated and documented by sophisticated analyses and comparisons with GDAS1 results and the observations.

In the second experiment (EXP_W), HYSPLIT is driven by (a) GDAS1 with vertical velocity from model output and (b) GDAS0P5 with vertical velocity calculated from horizontal divergence. The EXP_W results reveal the impact of the vertical advection calculation method on the difference in EXP_CTRL.

In the third experiment (EXP_HR), HYSPLIT is driven by (c) GDAS1 with vertical velocity calculated from horizontal divergence and (d) GDAS0P5 with the same vertical velocity calculation but only with 23 vertical layers. The 23 vertical layers are extracted from

<table>
<thead>
<tr>
<th>Pressure-level</th>
<th>Physical height (m)</th>
<th>GDAS1</th>
<th>GDAS0P5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000–900</td>
<td>0–900</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>900–800</td>
<td>900–1800</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>800–700</td>
<td>1800–3000</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>700–600</td>
<td>3000–4000</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>600–500</td>
<td>4000–5000</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>500–400</td>
<td>5000–6500</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>400–300</td>
<td>6500–8500</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>300–200</td>
<td>8500–10,500</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>200–100</td>
<td>10,500–13,500</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>100–0</td>
<td>13,500–18,000</td>
<td>2</td>
<td>14</td>
</tr>
</tbody>
</table>
the original 55 layers with the pressure levels closest to the GDAS1 layers. The EXP_HR results indicate the contribution of horizontal resolution to the difference in EXP_CTRL. Note that the GDAS1 data is in the pressure coordinate while the GDAS0P5 data is in the hybrid sigma-pressure coordinate, and the contribution of horizontal resolution to difference between trajectories generated from GDAS1 and GDAS0P5 data might be overestimated in EXP_HR.

In the final experiment (EXP_VR), HYSPLIT is driven by (d) GDAS0P5 with 23 vertical layers and (b) GDAS0P5 with 55 vertical layers, with both vertical velocity calculated from horizontal divergence. The EXP_VR results identify the contribution of vertical resolution to the difference in EXP_CTRL.

2.2. Quantification of differences

To quantify the differences between the trajectories generated from GDAS1 and GDAS0P5, the included angle (IA) and the absolute height difference (AHD) are calculated. As shown in Fig. 1, the IA is the angle between the lines that link the trajectory endpoints to the starting point of CW and is used to evaluate the difference in direction. The AHD is the height difference between corresponding trajectories generated from the two datasets. For the two corresponding endpoints generated from the two datasets, \( A = (x_1, y_1, z_1) \) and \( B = (x_2, y_2, z_2) \), the IA and AHD are calculated by

\[
IA = \arccos\left(\frac{x_1x_2 + y_1y_2}{\sqrt{x_1^2 + y_1^2 + x_2^2 + y_2^2}}\right) \tag{6}
\]

\[
AHD = |z_2 - z_1| \tag{7}
\]

IAs and AHDs between the corresponding trajectories in each sensitivity experiment for the preceding 24, 48, and 72 h are calculated. The number of large IAs in different seasons or transport patterns is also used as metrics for difference quantification. A large IA is defined as an angle greater than 60° at 72 h.

2.3. Cluster analysis

To compare the predominant patterns of backward trajectories generated from the two datasets, backward trajectories are clustered into groups with similar transport patterns. The groups, called clusters hereafter, are represented by their mean trajectory. This method minimizes the intra-cluster differences among trajectories while maximizes the

<table>
<thead>
<tr>
<th>Scenario (a) — GDAS1</th>
<th>Scenario (c)</th>
<th>Scenario (d)</th>
<th>Scenario (b) — GDAS0P5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizontal resolution</td>
<td>1°</td>
<td>1°</td>
<td>0.5°</td>
</tr>
<tr>
<td>Vertical resolution</td>
<td>23 layers</td>
<td>23 layers</td>
<td>23 layers</td>
</tr>
<tr>
<td>Vertical velocity</td>
<td>Model output</td>
<td>Calculated from horizontal divergence</td>
<td>Calculated from horizontal divergence</td>
</tr>
<tr>
<td>Sensitivity experiments</td>
<td>EXP_W</td>
<td>EXP_HR</td>
<td>EXP_VR</td>
</tr>
</tbody>
</table>

Fig. 1. Conceptual illustration of (a) included angle (IA) and (b) absolute height difference (AHD) between trajectories generated from GDAS1 and GDAS0P5 datasets, and (c) the location of Central Western (CW) station at Hong Kong.
The clustering of trajectories is based on the total spatial variance (TSV) method (Draxler et al., 2012a, 2012b).

In the clustering process, the spatial variance ($SV$, the squared distances between the endpoints of the cluster’s component trajectories and the mean of the trajectories in that cluster), the cluster spatial variance ($CSV$, the sum of the $SV$s of all trajectories within the cluster), and the total spatial variance (TSV, the sum of the CSVs for all clusters) are calculated. The $SV$ is determined between each endpoint ($k$) along trajectory ($j$) within its cluster ($i$):

$$SV_{ij} = \sum_k \left( P_{jk} - M_{ik} \right)^2,$$

where the sum is taken over the number of endpoints along the trajectory, and $P$ and $M$ are the position vectors for the individual trajectory and the mean trajectory of cluster $i$, respectively. The $CSV$ of cluster $i$ is the sum of the $SV$s of all trajectories within that cluster:

$$CSV_i = \sum_j SV_{ij}.$$

The $TSV$ is the sum of the $CSV$s over all clusters:

$$TSV = \sum_i CSV_i.$$

The clustering process starts by assigning each trajectory to its own ‘cluster’, so that there are $i$ ‘clusters’ with one trajectory in each cluster. Initial $CSV$ and $TSV$ are both zero. The two clusters with the lowest increase in $TSV$ are then combined to form a pair. In the first iteration, every combination is tried and the two clusters that result in the lowest increase in $TSV$ are combined. After the first combination, the number of clusters reduced to $i - 1$, with one cluster consisting of two trajectories and the others consisting of one. The clustered two trajectories are

Fig. 2. Seasonal variations in the average IA, AHD, and number of cases with large IAs for trajectories generated from (a) EXP_CTRL, (b) EXP_W, (c) EXP_HR and (d) EXP_VR. Large IA means an IA larger than $60\degree$ at 72 h.
replaced by their mean trajectory for further calculation. The process continues until the last two clusters are merged, resulting in trajectories in one cluster.

In the first few iterations, the TSV rises dramatically, then gradually levels off and lifts again by the end of calculation. The final rise happens when disparate clusters start to be merged, indicating that the paired clusters are no longer similar. The ideal final number of clusters is just before the inflection point where the final rise occurs (Draxler et al., 2012a, 2012b).

3. Results and discussion

3.1. Seasonal variations in the differences between trajectories

Fig. 2(a) shows the increase of the differences in the IAs, AHDs and number of large IAs in EXP_CTRL over time. The annual variations of the three parameters are similar, with the largest values in winter and the smallest in summer. In winter and spring, the average IA and number of large IAs increase significantly from the 24th to the 48th hour, then level off. In comparison, the average IA and number of large IAs do not show significant variation during the 72 h. Thus, the directional separation of the trajectories generated from GDAS1 and GDAS0P5 occurs mainly in the second 24 h. In contrast, the average AHD rises steadily with time throughout the year.

The EXP_W results in Fig. 2(b) show the contribution of different measures of vertical velocity to the grand difference between the trajectories generated from the two datasets. Indeed, Fig. 2(a) and (b) is very much alike, which illustrates that different measures of vertical velocity contribute dominantly to the grand differences between the trajectories. Vertical velocity calculated from horizontal divergence tends to differ notably from the model output vertical velocity, resulting large errors in the backward trajectories.

The average IAs, AHDs and frequency of large IAs in Fig. 2(c) and (d) show similar seasonal variations as those in Fig. 2(a) and (b), but in significantly lower values. This suggests that the effects of horizontal and vertical resolution of the input dataset on the differences in backward trajectories are much smaller.

Fig. 3 shows the seasonal variation in the average height of the trajectories generated from the two datasets. Generally, the average heights are insensitive to the horizontal and vertical resolutions of the input datasets across the 72 h. Significant differences exist over trajectories generated from GDAS1 with model output vertical velocity. In general, the vertical velocity calculated from horizontal divergence is lower in autumn and winter but higher in spring and summer than the model output vertical velocity, and the differences increase with time.

In summary, the difference between the HYSPLIT trajectories generated from GDAS1 and GDAS0P5 datasets can be mainly attributed to the deviation between model output and vertical velocity calculated from horizontal divergence. HYSPLIT trajectories are also dependent on the horizontal and vertical resolutions of the driven data, but to lesser extents. The vertical velocity calculated from vertical integration of horizontal divergence is lower in autumn and winter while higher in spring and summer than the model output vertical velocity, which leads to differences in the height of trajectories generated from GDAS1 and GDAS0P5 datasets.

3.2. Cluster analysis

Cluster analysis is conducted to identify the general patterns of air mass backward trajectories as well as the frequency of each pattern. For each sensitivity experiment, one scenario is selected as base scenario — scenario (a) in EXP_CTRL, scenario (c) in EXP_W, scenario (d) in EXP_HR and scenario (b) in EXP_VR, and the other scenario as the contrast scenario. As shown in Table 3, trajectories in the base scenarios are categorized into 6–8 clusters according to their general directions with respect to the starting location and the total transport distance of the mean trajectories. The general transport directions are northern, northeastern, eastern, southeastern, southern, southwestern, western, and northwestern. The clusters are defined as long-, medium-, and short-range transport...
patterns when the transport distance of the mean trajectory is larger than 1000 km, between 500 and 1000 km, and less than 500 km, respectively.

The most common transport pattern in the trajectories generated from scenario (a) is the short-range southwestern pattern, with a frequency of 28%. The most common pattern from the other three scenarios is the medium-range northern pattern, with frequencies of 26%, 29%, and 31%, respectively. It is worth noting that the vertical velocity of scenario (a) is from model output, while the vertical velocity of the other three scenarios is derived from horizontal divergence.

The trajectory generated from the contrast scenario is then mapped onto the cluster where its corresponding trajectory in the base scenario resides. The mean trajectories of the base scenarios and the contrast scenarios for all four experiments are shown in Fig. 4. For EXP_CTRL, the differences between mean trajectories generated from two datasets are large, especially when the transport pattern of base scenario is long-range western, long-range eastern, short-range southwestern and long-range northern; for EXP_W, the differences between mean trajectories are also large when the transport pattern is long-range northern, long-range northeastern and short-range western; for EXP_HR, the differences between mean trajectories are significantly smaller, the corresponding mean trajectories are more likely consistent with each other, except the large difference for medium-range western transport pattern; the difference between mean trajectories is the smallest for EXP_VR; the corresponding mean trajectories are quite similar with each other in all transport patterns.

To quantify the differences between the transport patterns for each sensitivity experiments, IAs and AHDs for corresponding trajectories in each transport pattern for all sensitivity experiments are calculated. As shown in Fig. 5, the proportion of cases with large IAs at 72 h is the highest for EXP_CTRL and EXP_W, lower for EXP_HR, and the lowest for EXP_VR. This is consistent with the results in Section 3.1, which highlight that the vertical velocity is more crucial to the HYSPLIT trajectories than the horizontal and vertical resolutions of the driven data. In general, the proportion of cases with large IAs is significantly higher when transport direction is northern or western, thus the difference between the trajectories generated from the different datasets is larger when the transport pattern is from these directions. In Hong Kong, the long- to medium-range northern, northeastern and long- to medium-range western transport patterns generally occur in winter, which explains why the largest differences are found in winter for all three

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Scenario (a) — GDAS1</th>
<th>Scenario (c)</th>
<th>Scenario (d)</th>
<th>Scenario (b) — GDAS0P5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transport patterns</td>
<td>SW/S (28%)</td>
<td>N/M (20%)</td>
<td>N/M (29%)</td>
<td>N/M (31%)</td>
</tr>
<tr>
<td></td>
<td>E/L (22%)</td>
<td>NE/L (15%)</td>
<td>W/M (20%)</td>
<td>E/L (21%)</td>
</tr>
<tr>
<td></td>
<td>N/M (21%)</td>
<td>E/L (12%)</td>
<td>NE/L (18%)</td>
<td>S/M (21%)</td>
</tr>
<tr>
<td></td>
<td>SW/L (11)</td>
<td>S/L (11%)</td>
<td>SW/L (18%)</td>
<td>S/L (10%)</td>
</tr>
<tr>
<td></td>
<td>N/L (10%)</td>
<td>SW/L (9%)</td>
<td>SE/M (9%)</td>
<td>SW/L (8%)</td>
</tr>
<tr>
<td></td>
<td>W/L (8%)</td>
<td>N/L (7%)</td>
<td>NW/L (6%)</td>
<td>N/L (7%)</td>
</tr>
<tr>
<td></td>
<td>W/S (4%)</td>
<td>W/L (6%)</td>
<td>W/N (2%)</td>
<td>NN/L (2%)</td>
</tr>
</tbody>
</table>

L: long-range, M: medium-range, S: short-range.
In summary, the results of the cluster analysis indicate that when the transport direction is northern, northeastern, or western, the HYSPLIT trajectories are sensitive to the vertical velocity or data resolution, and when the transport pattern is long- to medium-range southern or long-range southwestern, the HYSPLIT trajectories are more likely to remain unchanged with the shifting of vertical velocity or data resolution.

3.3. Validation of GDAS1 and GDAS0P5 datasets

3.3.1. Validation of vertical velocity

In previous sections, we have identified that the vertical velocity plays a significant role in the discrepancies between trajectories generated from GDAS1 and GDAS0P5 datasets. In this section, we examine the representativeness of vertical velocity in GDAS1 data in reproducing the actual condition in the atmosphere, and try to show the deviation of the vertical velocity in GDAS0P5 data which is calculated from horizontal divergence.

Firstly, a comparison between vertical velocity from GDAS1 and observational data in Hong Kong is conducted. The hourly wind profiler data with 26 vertical layers below 1500 m AGL at Sha Lo Wan station, a rural site in the northwestern part of Hong Kong (latitude: 22.29°N, longitude: 113.90°E), are used to evaluate the performance of GDAS1 data. This station is selected to avoid the impact of urban land-use on the winds. The vertical velocity from GDAS1 is interpolated into the observational vertical layers at Sha Lo Wan station every 3 h for comparison.

Fig. 7 shows scatters of the vertical velocities from GDAS1 data (the red dots) and observational values (green dots) with altitude at Sha Lo Wan in January, April, July and October 2011. The observational velocities have a wide range of (−16, 16) cm/s and are increasingly scattered with altitude in all seasons. The model output vertical velocities from GDAS1 data are also scattered but with narrower ranges of (−4, 4) cm/s in January, (−10, 6) cm/s in April, (−15, 5) cm/s in July and (−12, 6) cm/s in October. Similar with the observational data, the model output vertical velocities become more and more scattered with altitude. In comparison, the vertical velocities calculated from horizontal divergence in GDAS0P5 data (black dots) are most concentrated in a narrow range around 0 cm/s.

In summary, the results of the cluster analysis indicate that when the transport direction is northern, northeastern, or western, the HYSPLIT trajectories are sensitive to the vertical velocity or data resolution, and when the transport pattern is long- to medium-range southern or long-range southwestern, the HYSPLIT trajectories are more likely to remain unchanged with the shifting of vertical velocity or data resolution.

Fig. 5. Proportion of cases with large IAs in different transport patterns.

Fig. 6. Average IAs (degree, a–d) and average AHDs (m, e–h) for transport patterns.
As given in Table 4, the agreement between vertical velocity from GDAS1 data and observational data is relatively low but acceptable due to the coarse grid space of GDAS1 data and numerous data points. Nevertheless, the IOAs between vertical velocities calculated from horizontal divergence in GDAS0P5 data and observational values are significantly lower.

We further examine the global distributions of monthly average vertical velocity from GDAS1 and GDAS0P5 datasets at 850 hPa, as displayed in Fig. 8. It is noted that the magnitudes of vertical velocities from GDAS1 are smaller than those from GDAS0P5, which explains why the trajectories generated from GDAS0P5 can transport to higher altitudes than those from GDAS1 data in summer when the convection is active, and transport to lower altitudes in winter when China is controlled by stable and depressed systems. More importantly, the GDAS1 distributions show significant terrain effect that the magnitudes of vertical velocity are larger along the coastline area than those inland or over the ocean. In contrast, there is no such effect in the GDAS0P5 distributions. This also indicates that the model output vertical velocity from GDAS1 is more representative to the realistic situation than those from GDAS0P5.

3.3.2. Tracking pollutant source contributions from different directions

An important application of backward trajectory analysis is to investigate the potential source direction of pollutants that impact the starting location. In this study, hourly PM\(_{10}\) and PM\(_{2.5}\) concentrations measured at the CW air monitoring stations in Hong Kong for the entire year of 2011 were applied to further validate the two datasets. Hours with backward trajectories are extracted to have 1460 PM\(_{10}\) and PM\(_{2.5}\) concentration records in this analysis. To eliminate the impact of different vertical velocity schemes in the two datasets, only backward trajectories at the preceding 24 h are examined. Backward trajectories during these 24 h are closer to the ground where most of the PM pollutants are emitted, therefore are more indicative to the pollutant source directions.

The average PM\(_{10}\) and PM\(_{2.5}\) concentrations contributed from different directions during the preceding 24 h are plotted in Fig. 9. The general patterns of the two plots are similar, with higher PM concentrations associated with northerly wind and lower PM concentrations associated with southerly wind. Such a pattern is determined by the geographical location of Hong Kong, with the rapidly urbanized PRD to the north/northwest and the South China Sea to the south. Therefore, air pollution in Hong Kong is generally worse during winds with a northerly component while cleaner during winds with a southerly component.

In spite of similar pattern, slight difference exists between the two plots with higher north/south contrast for the GDAS1 dataset (Fig. 9a).

Table 4

<table>
<thead>
<tr>
<th>Year-month</th>
<th>Index of agreement (IOA)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GDAS1</td>
</tr>
<tr>
<td>201101</td>
<td>0.20</td>
</tr>
<tr>
<td>201104</td>
<td>0.26</td>
</tr>
<tr>
<td>201107</td>
<td>0.32</td>
</tr>
<tr>
<td>201110</td>
<td>0.31</td>
</tr>
</tbody>
</table>

Fig. 7. Vertical velocity from observation (OBS.) at Sha Lo Wan station, from GDAS1 data, and calculated from horizontal divergence in GDAS0P5 data for January, April, July and October 2011.
Fig. 8. Global distributions of monthly average vertical velocity from GDAS1 and GDAS0P5 datasets at 850 hPa in January, April, July and October 2011.
Average PM$_{10}$ concentration during northerly wind is 69 μg/m$^3$ for GDAS1 and 64 μg/m$^3$ for GDAS0P5. In comparison, PM$_{10}$ concentration during southwestery wind is 26 μg/m$^3$ for GDAS1 and 29 μg/m$^3$ for GDAS0P5. As the pollutant emissions from north and southwest directions of the CW station are fundamentally different, greater difference from GDAS1 results indicates the plausibility of GDAS1 in quantitatively retrieving pollution source contributions. In addition, the PM$_{2.5}$ patterns in GDAS1 point to the northwest, which is also more reasonable than the north-pointing GDAS0P5 pattern considering the significant impact from the central PRD (Guangzhou-Foshan area) where most of the PM$_{2.5}$ emissions are located. In summary, backward trajectories from GDAS1 are more realistic with better performance in retrieving PM contributions from different directions.

4. Conclusions

In this study, we evaluate the performance of two GDAS datasets as the driven data of HYSPLIT model in generating backward trajectories in Hong Kong for the entire year of 2011. Factors that may lead to differences between the trajectories generated from the two datasets are identified and the differences caused by each factor are quantified by a series of sensitivity experiments. The results reveal that the significant difference between trajectories generated from GDAS1 and GDAS0P5 datasets can be mainly attributed to the difference in vertical motion calculation methods due to the absence of vertical velocity in GDAS0P5. In winter and spring when northern or northeastern air mass prevails, the vertical velocity calculated from horizontal divergence is smaller while in summer and autumn when southern or southwestern air mass prevails, the vertical velocity calculated from horizontal divergence is larger than that in GDAS model output. Both lead to uncertainties in terms of trajectory heights and directions.

Vertical velocities in GDAS1 data are validated by comparing with the observations at Sha Lo Wan in northwestern Hong Kong. As the vertical velocities of GDAS0P5 are calculated from horizontal divergence, it is expected that GDAS1 vertical velocity is closer to the measurements with higher IOA. GDAS1 also performs better in retrieving contributions from different directions to the ambient PM$_{10}$ and PM$_{2.5}$ levels in Hong Kong. Therefore, it is concluded that, although GDAS0P5 has finer horizontal and vertical resolutions, it should be unfortunately used with caution as the vertical velocity is absent and has to be calculated from horizontal divergence, therefore poses greater uncertainties to the generated backward trajectories. GDAS1 is still recommended for driving HYSPLIT model in areas with complicated topography and diversified land-use, such as the PRD region.

Acknowledgments

This work was supported by grants SRHIPO01, Environmental Conservation Funds (4/2009), NSFC/RCG Joint Funds (N_HKUST631/05), NSFC-GD Joint Funds on key project (U1033001), the Fu Tak Lam Foundation (FTFL08/09.EG01), and the Fok Ying-Tung Graduate School (NRC06/07.SC01). The authors acknowledge Hong Kong Observatory for providing the wind profiler data at Sha Lo Wan.

References


Li Y, Tau AXF, Fung JCH, Zheng YJ, Zhong L, Louie PKK. Ozone source apportionment (OSAT) to differentiate local regional and super-regional source contributions in

Stunder BJB. Global Data Assimilation System (GDAS) Archive Information. Silver Spring, Maryland, USA: NOAA Air Resources Laboratory; 2004 (Available at http://ready.arl.noaa.gov/gdas1.php).

Stunder BJB. GDAS0P5 Information. Silver Spring, Maryland, USA: NOAA Air Resources Laboratory; 2012 (Available at ftp://arlftp.arlhq.noaa.gov/archives/gdas0p5/).


