INPUT DATA REQUIREMENTS FOR LAGRANGIAN TRAJECTORY MODELS

by Kenneth P. Bowman, John C. Lin, Andreas Stohl, Roland Draxler, Paul Konopka, Arlyn Andrews, and Dominik Brunner

The Lagrangian approach to fluid motion follows a fluid parcel as it moves with the flow. Lagrangian methods have proven to be very useful for understanding the properties of atmospheric flows, particularly for problems related to transport, dispersion, and mixing of trace constituents or other atmospheric properties. In particular, Lagrangian methods can be used to identify pathways for atmospheric transport by computing the trajectories of air parcels as they move in the atmosphere.

Atmospheric applications of Lagrangian methods range from micrometeorological to global scales and have a long history in the meteorological literature (Petterssen and Namias 1940; Welander 1955; Wiin-Nielsen 1959; Djurić 1961; Kida 1977; Hsu 1980; Kida 1983). Recent examples of the use of Lagrangian methods include predicting the transport and dispersion of radioactive materials released following the accidents at the Fukushima nuclear power station in Japan (Stohl et al. 2012), tracking of ash clouds produced by volcanic eruptions (Kristiansen et al. 2012; Webster et al. 2012), modeling the atmospheric component of the global carbon cycle (Lin et al. 2004; Trusilova et al. 2010), the exchange of water vapor and ozone between the troposphere and stratosphere (Homeyer et al. 2011), and dehydration of the stratosphere (Schoeberl and Dessler 2011). At a recent conference on Lagrangian methods (Lin et al. 2011, 2013), the participants discussed the evolving requirements for input data for Lagrangian kinematic models. This paper provides a brief review of Lagrangian methods and presents proposals for improving the accuracy of atmospheric Lagrangian models through better access to meteorological analysis and forecast products.
LAGRANGIAN METHODS. Model equations.
Dynamical models of fluid motion can be developed using a Lagrangian framework, but in this discussion we restrict our attention to kinematic Lagrangian models, which compute the trajectories of air particles given the Eulerian velocity field. A Lagrangian trajectory model solves the kinematic equations of motion:

\[
\frac{dx}{dt} = v(x,t), \quad x(0) = x_0
\]

(1)

where \( x(t) \) is the position of a hypothetical massless air particle as a function of time \( t \); \( v(x,t) \) is the velocity as a function of position and time; and \( x_0 \) is the initial position of the particle. The solution to (1) is the path of the air particle through the atmosphere (the trajectory). Given appropriate velocity data, trajectories can be computed either forward or backward in time. Trajectories can be computed diagnostically, using archived velocity data, or prognostically, using model forecasts of winds.

The kinematic equation of motion (1) comprises a set of coupled, first-order, ordinary differential equations for each component of motion. If \( v \) can be written analytically, it may be possible to find a closed-form analytical solution to (1). Even for relatively simple velocity fields, however, the motion of an air particle can be extremely complex, or even chaotic, and it is necessary to find approximate numerical solutions to (1). For real atmospheric flows, where velocity fields are obtained either directly from observations or from data assimilation systems, a numerical solution is required. A numerical solver for (1) typically consists of two parts: a method to interpolate a discrete, gridded velocity field \( v(x_0, t) \) to an arbitrary point \( x \) and time \( t \) within the domain of interest, and a numerical scheme for integrating (1) forward or backward in time given the interpolated velocities and a set of initial conditions.

Lagrangian dispersion models augment pure trajectory models by including parameterizations of the effects of unresolved scales of motion on the path of a particle. Dispersion model applications range from the microscale, such as the emission from an automobile tailpipe, to the global scale, such as the transport of a cloud of volcanic ash. In order to estimate the properties of the unresolved flow, dispersion models require additional information, such as wind shear, stability, and turbulence parameters.

It should be noted that individual particle trajectories ultimately become unpredictable due to the underlying chaotic nature of atmospheric flows, even in the absence of stochastic turbulence parameterizations. The time scale on which this occurs depends on the type of flow being studied (turbulent planetary boundary layer, large-scale stratospheric flow, etc.). Despite this eventual loss of predictability, like Eulerian models, Lagrangian methods have proven to be quite powerful, particularly for understanding dispersion, stirring, and mixing in the atmosphere (e.g., Bowman 1993; Sutton et al. 1994; Legras et al. 2005).

For real atmospheric flows, trajectory calculations will have errors that arise from three sources:

1) errors in the gridded winds themselves, which could result from measurement error, or from Eulerian model approximations, such as subgrid-scale parameterizations, that enter the analyzed fields through the data assimilation process;

2) sampling errors that follow from the fact that velocity fields are available at finite spatial and temporal resolution and must be interpolated to particle locations; and

3) truncation errors that come from the use of an approximate numerical scheme to integrate (1) in time.

Current computer hardware, an appropriate choice of numerical methods, and optimized numerical codes generally allow trajectory calculations to be carried out quickly, efficiently, and accurately, even for large numbers of particles. High numerical accuracy can be achieved by varying the time step size and observing the convergence of the numerical solutions. Input-output considerations aside, trajectories for multiple particles are independent of each other and thus are trivially parallelizable. For these reasons, type 3 errors are generally small relative to the other error sources.

The dominant errors in trajectory calculations typically come from errors in the winds themselves (type 1 errors) or from the limited spatial and temporal resolution of the gridded wind fields (type 2 errors) (Stohl et al. 2001). It is those two sources of error that we concentrate on in this paper, and we focus on trajectory models that use velocity fields produced by operational global or regional data assimilation and weather forecast models or by reanalysis systems. These models are generally operated by government agencies, such as the U.S. National Centers for Environmental Prediction (NCEP), the European Centre for Medium-Range Weather Forecasts (ECMWF), the United Kingdom Met Office (UKMO), or the Japan Meteorological Agency (JMA). In addition to large-scale computational resources and forecast models, the major modeling centers have the necessary data ingest and assimilation capabilities to create high-quality meteorological analyses. At present, operational forecast models and reanalysis...
systems are virtually all hydrostatic, and thus include parameterizations for convection, as well as for unresolved turbulent mixing.

**Types of trajectory models.** Many different types of trajectory models have been developed over the years (Stohl 1998), and a variety of different models are currently in use around the world for research and operational applications. Characteristics of selected models are listed in Table 1. Model trajectories have been verified at different scales using balloon tracking (Reisinger and Mueller 1983; Knudsen et al. 1996; Baumann and Stohl 1997), tracer releases (Haagenson et al. 1987; Draxler 1991; Stohl et al. 1998), satellite observations of volcanic SO$_2$ clouds in the stratosphere (Schoeberl et al. 1993), satellite ozone data (Bowman 1993), in situ observations of natural tracers (Bourqui et al. 2012), and comparisons with conserved meteorological parameters (Stohl and Seibert 1998).

Early models often incorporated only the horizontal components of the wind, neglecting vertical motion (Petterssen and Namias 1940; Djurić 1961). These models typically assumed the flow to be isobaric or isentropic. While these approximations may be sufficient in some circumstances, newer trajectory models are fully three dimensional. Trajectory models have been developed that use pressure coordinates, isentropic coordinates, or the native vertical coordinate system of the Eulerian model from which winds are taken (typically terrain-following sigma coordinates or more general hybrid sigma-pressure coordinates). In the lower troposphere, terrain-following coordinates have also been adopted by many trajectory models (see Table 1).

Some trajectory models use standard operational analysis and forecast output data, which provide atmospheric parameters in pressure coordinates from 1,000 hPa to the top of the Eulerian model atmosphere, $p_{\text{top}}$, which is typically in the midstratosphere or higher (10 hPa or less). These output fields are produced by interpolating the source model’s terrain-following coordinates to fixed pressure levels with resolutions on the order of 25 or 50 hPa. In locations where the surface pressure is less than 1,000 hPa, atmospheric variables on pressure surfaces that lie below the local surface are found by extrapolation. Because these pressure surfaces do not actually exist, this extrapolation produces fictitious values. Unrealistic trajectories can be generated anywhere that the 1,000-hPa surface intersects the ground.

Isentropic coordinate trajectory models have advantages for trajectory studies of the upper troposphere and stratosphere, where vertical velocities in $\theta$ coordinates (diabatic heating rates) are usually small (Ploeger et al. 2010, 2011). Isentropic coordinates are rarely used for trajectory studies in the lower atmosphere because of conceptual and technical difficulties near Earth’s surface, although hybrid-coordinate models have been developed (Mahowald et al. 2002). Models that use isentropic or terrain-following

---

**Table 1. Characteristics of selected trajectory models.**

<table>
<thead>
<tr>
<th>Model</th>
<th>Institution</th>
<th>Domain</th>
<th>Vertical coordinate</th>
<th>Interpolation</th>
<th>Numerical Scheme</th>
<th>Stochastic turbulence</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLEXPART</td>
<td>NILU</td>
<td>Global</td>
<td>terrain-following $z$</td>
<td>4-D linear</td>
<td>Pettersen 1940</td>
<td>Y</td>
<td>Fortran</td>
</tr>
<tr>
<td>HYSPLIT</td>
<td>NOAA ARL</td>
<td>Regional</td>
<td>$\sigma - z$</td>
<td>4-D linear</td>
<td>Pettersen 1940</td>
<td>Y</td>
<td>Fortran</td>
</tr>
<tr>
<td>LAGRANTO</td>
<td>ETH Zurich</td>
<td>Regional</td>
<td>hybrid-$p$</td>
<td>4-D linear</td>
<td>Pettersen 1940</td>
<td>N</td>
<td>Fortran</td>
</tr>
<tr>
<td>NAME</td>
<td>UK Met Office</td>
<td>Regional</td>
<td>flexible</td>
<td>4-D linear</td>
<td>Euler-Maruyama</td>
<td>Y</td>
<td>Fortran</td>
</tr>
<tr>
<td>STILT</td>
<td>multiple</td>
<td>Regional</td>
<td>$\sigma - z$</td>
<td>4-D linear</td>
<td>Pettersen 1940</td>
<td>Y</td>
<td>Fortran</td>
</tr>
<tr>
<td>TRAJ3D</td>
<td>Texas A&amp;M</td>
<td>Regional</td>
<td>flexible</td>
<td>4-D linear</td>
<td>4th order R-K</td>
<td>optional</td>
<td>IDL</td>
</tr>
</tbody>
</table>

*a Stohl et al. (2005)  
*b between wind field data simple Euler forward step  
*c Draxler and Hess (1997, 1998); Draxler (1999)  
*d Wernli and Davies (1997)  
*e optional cubic interpolation in vertical  
*f Jones et al. (2007)  
*g Lin et al. (2003)  
*h MPI-Jena, University of Utah, University of Waterloo, Harvard University, and AER  
*i Bowman (1993); Bowman and Carrie (2002)
coordinates require variables that are not generally present in standard output files, such as $\dot{\theta}$, $\dot{\sigma}$, and $\eta$.

To represent unresolved scales of motions and to ensure numerical stability and model robustness, data assimilation and forecast systems include parameterizations for convection and for horizontal and vertical mixing. These parameterizations act to smooth model variables at small space and time scales. In pressure coordinates, for example, the vertical velocity $\omega$ is the grid-scale velocity computed diagnostically from the continuity equation. Similarly, $\dot{\theta}$ is the gridbox-averaged heating rate. The resolved grid-scale velocity fields do not directly include information about vertical velocities due to convective updrafts and downdrafts or information about the stability and vertical mixing in the planetary boundary layer. Information about subgrid-scale mixing processes would be of considerable value in trajectory calculations, particularly in the turbulent planetary boundary layer, but the relevant variables are normally not included in standard, publicly available, model output files.

**REQUIREMENTS FOR TRAJECTORY MODEL INPUTS.** As discussed in the previous section, large-scale trajectory models usually obtain their input wind fields from operational weather forecast and analysis models. For many trajectory applications, the operational models and reanalyses have significant strengths. Forecast centers expend great efforts to ingest and quality control a large volume of data. Data assimilation systems provide very effective mechanisms for combining these data with model forecasts to produce high-quality analyses of the global three-dimensional state of the atmosphere. Finally, the analyses and forecasts are made available to researchers in a timely manner and archived for later retrospective studies. Despite these advantages, trajectory calculations are increasingly being constrained by the availability of specific forecast model output variables and the temporal resolution at which output is made available.

The accuracy of numerical solutions to the trajectory equations, and the potential sources of error, have been investigated by many authors for a wide range of meteorological situations (Walmsley and Mailhot 1983; Kuo et al. 1985; Kahl and Samson 1986; Merrill et al. 1986; Rolph and Draxler 1990; Schoeberl et al. 1992; Bowman 1993; Doty and Perkey 1993; Seibert 1993; Stohl et al. 1995; Stohl and Seibert 1998; Bourqui 2006; Davis and Dacre 2009). Errors due to spatial and temporal interpolation of wind fields can be quantified by carrying out model simulations at high resolution and using the meteorological data at the original and at degraded resolutions for trajectory calculations. An important finding of these and other studies [see Stohl (1998) for a review] is that the temporal and spatial resolutions of the wind fields must be in balance in order to limit the trajectory errors. For instance, an increase in spatial resolution alone yields only marginal improvements in trajectory accuracy when the temporal resolution of the wind data is low. This has recently become a problem for Lagrangian models, as most operational weather forecast centers have increased the horizontal resolution of their models over time and are now approaching 10-km resolution. The spatial resolution at which the model output is made available to the public is relatively high (e.g., 0.5°), and privileged users can retrieve data at the model’s native resolution; however, most weather centers have not increased the output frequency of their data, which is typically 3 or 6 h. The coarse temporal resolution of the model output files undersamples the wind field in time relative to space and prevents the Lagrangian modeling community from benefitting from the high spatial resolution of contemporary weather forecast models. For instance, Pisso et al. (2010) have produced special output from the ECMWF model with hourly resolution, and they have shown that, at 0.5°–1° horizontal resolution, reconstructions of stratospheric ozone profiles are substantially improved when increasing the wind field frequency from 6 to 3 h, and smaller but still significant improvements were obtained with 1-hourly resolution. Using input from a mesoscale model with 4- and 12-km resolution, Brioude et al. (2012) recently demonstrated large improvements in correlation of model results obtained from backward and forward integrations of a Lagrangian particle dispersion model when increasing the temporal resolution from 2 h to 1 h and 30 min, as interpolation errors are reduced. G. D. Carrie and K. P. Bowman (1999, unpublished manuscript) used general circulation model (GCM) winds archived at every model time step (20 min) to estimate errors due to time sampling of the wind field. For 1–5-day forward trajectories computed using wind sampling intervals $\Delta t$ between 1 and 12 h, they found that globally averaged errors in the horizontal position varied approximately as $(\Delta t)^b$, where $b$ is between approximately 1.8 and 2 (Fig. 1). With 6-h wind sampling, 5-day trajectory errors are quite large (~350 km), while changing from 3- to 1-h wind sampling reduces the error by approximately an order of magnitude.
Another way to improve trajectory accuracy is to use time-averaged winds instead of instantaneous winds for driving the Lagrangian model, as has been shown by Nehrkorn et al. (2010) and Brioude et al. (2012). Pawson et al. (2007) demonstrated similar results for an Eulerian transport model. Weather forecast centers currently store instantaneous model fields. Adding time-averaged data would double the data storage requirements. For some users, doubling the model output frequency would be more attractive than storing time-averaged fields; and it would also improve trajectory accuracy. But if data handling by the Lagrangian modeler and not storage at the weather forecast centers is the limiting factor (and it often is), it would certainly be better to use time-averaged wind fields instead of instantaneous wind fields.

Errors in computed trajectories are often most obvious in the vertical coordinate. As with horizontal errors, vertical errors can be reduced by improving the temporal sampling of the high-frequency components of the flow, such as the portion of the gravity wave spectrum that is resolved by the model but not captured in 3- or 6-hourly output. Although verification is difficult in the upper-troposphere and higher layers, using time-averaged diabatic heating rates and $\theta$ coordinates appears to give better results than purely kinematic calculations in $p$ coordinates, where high-frequency components would be badly aliased (Ploeger et al. 2010; Diallo et al. 2012).

Other useful diagnostics of trajectory errors include deviations from mass conservation (Stohl and Seibert 1998) and comparisons between forward-time and backward-time simulations (Lin et al. 2003). Mass nonconservation and time asymmetry could be introduced from imbalances due to data assimilation (Byun 1999), interpolation to pressure levels in the NWP output (Trenberth 1991), and conversions between vertical coordinate systems (Hoerling and Sanford 1993).

**SUMMARY AND RECOMMENDATIONS.**
The standard output products from current operational forecast models and reanalyses have characteristics that limit the accuracy of atmospheric trajectory and dispersion models. Because much of the information needed by trajectory models is, in fact, computed as part of the data assimilation and forecast process, these limits are largely artificial. Specifically,

1) the spatial resolution of global models has increased substantially over the last decade, while the output frequency has not increased from the standard 3- or 6-h intervals;
2) model output is usually provided only in pressure coordinates;
3) diabatic heating rates are generally not available; and
4) information about subgrid-scale processes, such as vertical motion from the convective or boundary layer parameterizations, is not included in model output files.

The trajectory model developers and users at the workshop discussed which changes in the data streams from the operational modeling centers would be most beneficial for trajectory models. In order of priority, the greatest benefits to trajectory model accuracy would come from:

1) increasing model output frequency to match the continuing increases in spatial resolution (at least hourly for 0.5° grids, increasing to 30 min as model resolution approaches 10 km);
2) providing model output, including vertical velocities, in native coordinates so that input fields can be tailored by the users for specific meteorological situations and trajectory models;
3) including diabatic heating rates for use in $\theta$-coordinate models;
4) including convective mass fluxes or vertical velocities and subgrid-scale mixing information in the output files; and
5) time averaging of winds between the output intervals to improve trajectory accuracy.
These changes will not eliminate errors in the winds themselves (type 1 errors), but they will dramatically reduce errors from interpolating the discretized wind fields, converting from model to pressure coordinates, and other factors that can be controlled by the end user. Providing higher temporal resolution and additional model output variables may require greater data storage capacity and transmission bandwidth, but both storage and bandwidth continue to increase in capacity and decline in cost. Better input data for Lagrangian models will lead to advances in basic scientific research and faster, more accurate responses to urgent situations such as fires, chemical or radiation releases, or volcanic eruptions. While the focus of this paper is on Lagrangian models, the suggested changes in model output and archiving would also benefit many other users, such as off-line Eulerian chemical transport models.

We envision that the operational centers would benefit in return through the use of their products by a broader research community. Different analyses and reanalyses are known to produce significantly different vertical velocities—in the tropical tropopause layer, for example. Passive tracers are well known to help in diagnosing transport errors (e.g., Boering et al. 1996; Xueref et al. 2004). Similarly, comparisons of Lagrangian transport properties of different analyses will provide additional useful diagnostics.

Toward this end, our specific recommendations are designed to more closely connect the operational forecast and analysis models with offline Lagrangian models. The closer coupling will reveal problems more readily. Currently, errors in trajectory models are difficult to trace back to specific components of the operational model, as they can originate from the myriad issues already mentioned: coordinate transformation, coarse resolution, or unavailable variables. Adopting the recommendations would help to close the loop and enable the Lagrangian modeling community to provide feedback to further improve the operational systems.

ACKNOWLEDGMENTS. The authors wish to thank the AGU and the organizers of the Chapman conference for providing a stimulating venue to discuss these issues.

REFERENCES


