A Novel Approach to the Source Region Impact Factor Using Outputs from the Hybrid Single Particle Lagrangian Integrated Trajectory Model

---

A Thesis
Presented to
The Division of Mathematics and Natural Sciences
Reed College

In Partial Fulfillment
of the Requirements for the Degree
Bachelor of Arts

Jonathan Tamsut

May 2015

Approved for the Division
(Chemistry)

Christopher Walsh

Acknowledgements

First and foremost, I would like to thank my parents for all their support. I would also like to thank Frances, Grant, Liat, Jasmin and Isabella for playing such a pivotal role in my life.

I would like to thank my advisor, Christopher Walsh, for all his input and for introducing me to this project and meteorology in general. To all who have taught me while at Reed, I say thank you. I wouldn't have grown nearly as much as I did without your help.

I would like to thank Larisa Owechko for proofreading early (and late) drafts of this work, as well as Julie Fry for her helpful input.

Table of Contents

Abstract

Identifying the location of an emission source contributing to the ambient air composition of some downwind receptor is critical information for air quality control. Source region impact factors are a metric of this relationship. An improved method for computing a source region impact factor using the outputs of an atmospheric trajectory model – the Hybrid Single-Particle Lagrangian Integrated Trajectory Model (HYSPLIT) – was developed. This was done computationally and resulted in a program called Trajpy v. 2.0[1]. The improvements made allow one to apportion more generally and to take into account factors that are a function of altitude, namely wet deposition due to rain scavenging and air parcel altitude. Trajpy v. 2.0 also has the capability to remove the "inherent peak bias", which is due to the fact that all backwards trajectories converge at the receptor. This was done by creating a conditional probability that enables one to relate the altitude of an air parcel to its composition at the receptor.

[1]The source code for this program can be found in Appendix B or [23].

Introduction

The aim of this project is to create an improved instantiation of the Residence Time Analysis (RTA) model. The output of a Residence Time Analysis. A backwards trajectory is an approximation of the trajectory a particle took to reach its current position. A backwards trajectory must be supplied by a atmospheric backwards trajectory model. The atmospheric backwards trajectory model used to compute source region impact factors in this work is the Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) model [3].

This chapter will expound upon the aforementioned topics in more depth. This will be done by first providing motivation for using the Residence Time Analysis model. Next, the reader will be accorded with the appropriate context for using and understanding Residence Time Analysis and atmospheric models, in general. This will be done by providing the reader with a primer on the composition and dynamics of Earth's atmosphere, paying special attention to the means by which particles are transported through the atmosphere. Relevant aspects of meteorology[2] will be discussed. Next general features of an atmospheric chemical transport model will be discussed with the HYSPLIT model discussed in particular. Lastly, receptor modeling will be introduced.

0.1 Motivation for Using Residence Time Analysis to Compute a Source Region Impact Factor

Residence Time Analysis is used as a means of quantifying the outputs of an atmospheric backwards trajectory model. In order to motivate the usage of Residence Time Analysis one must first explain what an atmospheric backwards trajectory model is. An atmospheric backwards trajectory model takes a geographic location of interest as input (called a *sink* or *receptor*), then traces the motion of a box of air (called an *air parcel*) backwards in time and therefore approximates the trajectory an air parcel took to reach the aforementioned geographic location (Figure 1). The atmospheric backwards trajectory model used in this work is the Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) model created and maintained by the Air Resources Laboratory (ARL) of the National Oceanic and Atmospheric Association (NOAA) [3].

Air parcels are useful constructs in atmospheric modeling and will therefore be

[2]Meteorology is the scientific study of the atmosphere.

mentioned throughout this work. They are merely imaginary boxes drawn around atmospheric gases. Air parcels can have gaseous material flow in and out of them.

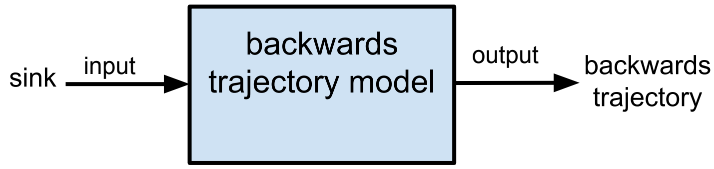

Figure 1: A schematic of an atmospheric backwards trajectory model like HYSPLIT. The parameter specified as input is the longitude and latitude of the sink or receptor. One wants to know how much a potentially far off pollution source affects the receptor. The output is the modeled backwards trajectory of an air parcel.

In order to compute a SRIF a series of grids called *source regions*[3] must be drawn around the region of interest. This region includes the source, the receptor and the area where the particle traveled while being transported from the pollution source to the receptor site. A backwards trajectory can be thought of as being made up of a series of *endpoints*. Each endpoint is connected to its subsequent and previous endpoint. A source region impact factor is then the percentage of trajectory endpoints that lay inside a particular grid and is thereby a quantity that represents the effects of a pollution source on a region up to thousands of kilometers away. A more thorough exposition of how SRIFs are computed will be given in the next chapter.

The outputs of Residence Time Analysis – source region impact factors – can be used as a quantitative tool to guide those developing strategies to manage air quality. In fact, this was the original use of SRIFs.[4] Understanding the effects of a pollutant source on another region some distance away can be critical information when conducting an air quality study. Common questions proffered to those conducting air quality studies are: What is the contribution of source A to the concentration of pollutants at site B? Where should one put a future source of pollution to minimize the amount of damage that pollution has on the surrounding area? What will be the air quality at some time in the future? Source region impact factors can help answer these questions.

0.2 Chemical Prerequisites

The Earth's atmosphere is an essential component for life on Earth. It is the source of carbon dioxide for plant photosynthesis and oxygen for respiration and absorbs

[3]These technically don't have to be grids. I am referring to "grids" to make this explanation more digestible. Read on and you will see you can use any two-dimensional shape to make source-region boxes.

[4]A more in depth exposition of the humble beginnings of the source region impact factor can be found in Chapter 1.

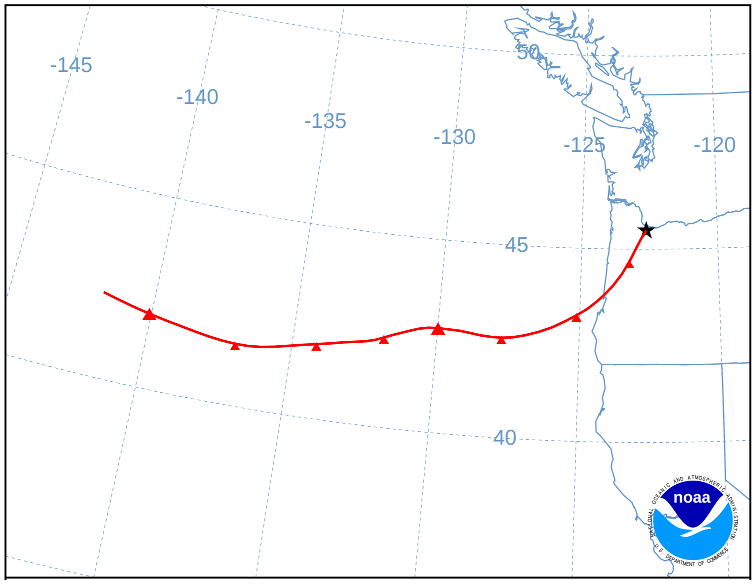

Figure 2: This is the output of a backwards trajectory model. This backwards trajectory was computed by the National Oceanic and Atmospheric Association's Hybrid Single-Particle Lagrangian Integrated Trajectory Model [3]. The ★ denotes the region of interest specified as model input. The red line is the trajectory of an air parcel up until the ★. The trajectory begins in the Pacific Ocean and ends at the Arthur F. Scott Laboratory of Chemistry.

high energy electromagnetic radiation with wavelengths below 300 nanometers that would otherwise be harmful to living organisms. Because the Earth's atmosphere is highly dynamic and multifaceted, the field of meteorology is a highly interdisciplinary venture. From the investigation of individual chemical reactions occurring in each layer of the atmosphere, to the mathematical modeling of air flow, to the study of how aerosols scatter incoming solar radiation, atmospheric chemistry (and meteorology in general) includes elements from many fields. The fundamental goal of atmospheric chemistry, though, is to understand the factors that dictate the concentration of species in the atmosphere.

In this work in particular, we will be concerned with how particles in Earth's atmosphere travel in both the horizontal[5] and vertical (changes in altitude) directions. Incoming solar radiation creates differences in pressure and temperature in Earth's atmosphere. To understand how temperature and pressure can play a role in atmospheric transport one must first have an understanding of some underlying physical theory.

0.2.1 The Chemistry of Gases

The atmosphere is a collection of gases that surround terrestrial Earth. Gas is a state of matter. Particles in a gas have weak intermolecular forces between them

[5]By horizontal I mean trajectories that can be thought of as in parallel to the surface of the Earth.

and due to this gaseous particles usually have much greater kinetic energy than molecules in the liquid or solid state do. The pressure of a gas is the force per unit area resulting from the collisions of a gas molecule with its surrounding surfaces. The ideal gas law describes the properties of a low concentration sample of gas (Equation 1). This law which relates pressure, volume and temperature for a given gas is given by,

$$PV = nRT, \tag{1}$$

where P is the pressure of the gas, V is the volume the gas occupies, n is the number of moles of the gas, R is the ideal gas constant and T is the absolute temperature of the gas. Temperature can be described in terms of kinetic energy, whereby warmer gases consist of molecules with a higher mean kinetic energy and cooler gases are made up of molecules with a lower mean kinetic energy. An ideal gas has properties that can be accurately modeled by the ideal gas law. The partial pressure of an ideal gas, P_x, in a mixture of gases of total pressure, P_{total}, is defined as the pressure that would be exerted by the molecules of gas, x, if no other gases were present and is given by,

$$P_x = C_x P_{total}, \tag{2}$$

where C_x is the mixing ratio of a gas. The mixing ratio of a gas is the ratio of the number of moles of gas x to the total number of moles in a gaseous sample. The density of a gas is its mass per unit volume. In an ideally gaseous mixture made up of gases A, B and C the total pressure of the gaseous mixture is given by the sum of partial pressures for each gas. This is known as Dalton's law of partial pressures and is given by,

$$P_{total} = P_A + P_B + P_C, \tag{3}$$

where P_A, P_B and P_C are the partial pressure of gas A, B and C, respectively, and P_{total} is the total pressure of the gaseous sample. It can often be assumed that gases in Earth's atmosphere are sufficiently dilute enough to be modeled accurately by the ideal gas equations.

Another useful metric in atmospheric chemistry is vapor pressure. In a closed system a liquid will have some if its molecules or atoms have enough kinetic energy to escape to the gas phase. Vapor pressure is the pressure of this gas over its respective liquid. When a liquid is in thermodynamic equilibrium[6] with its gas it is said that the vapor pressure of the gas has reached *equilibrium vapor pressure*.

0.2.2 Thermodynamics

The first law of thermodynamics states that the change in internal energy of a system is equal to the heat added to the system from its surroundings plus the work done on the system by its surroundings and is given by,

[6] A system in thermal equilibrium with its surroundings can be thought of as having constant internal energy. See the next section for a discussion of thermodynamics.

$$\Delta H = q + w, \tag{4}$$

where heat, q, is energy transfer between a system and its surroundings due to a temperature difference between the system and its surroundings and work, w, is energy transfer between the system and its surroundings due to a nonzero net force between the system and its surroundings. Heat input into a system from its surroundings corresponds to a positive value (+q), while heat released or removed from the system is considered to be a negative quantity (-q). Moreover, if the energy of the system is increased by work then we say work is done on the system by its surroundings which yields positive work (+w). Alternatively, when work is done on the surroundings by the system work is negative (-w).

It will be useful if we further develop the notion of thermodynamic *work*. We can quantify work by imagining a cylinder filled with gas and a piston that can increase and decrease the volume of the cylinder (Figure 3). The piston will move either upwards when the gas inside the cylinder exerts a force on it, or downwards when external pressure causes the piston to depress.

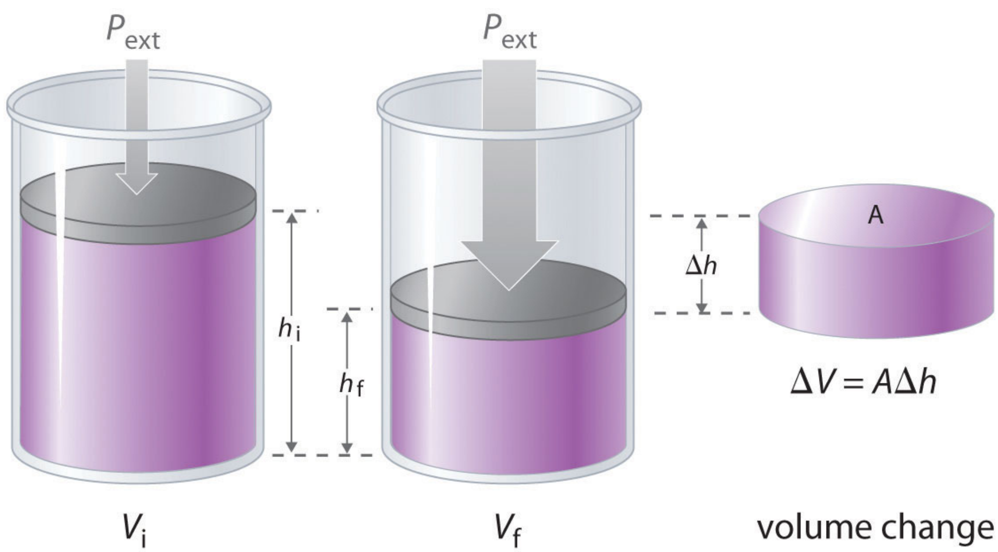

Figure 3: Illustration of thermodynamic work. The cylinder's face has an area, A. The external pressure, $P_{external}$ can be thought of as being the result of placing a massive object on the piston. Used without permission from *Principles of General Chemistry* by Bruce Averill and Patricia Eldredge [17].

Depressing the piston, and consequently reducing the volume in the cylinder, can be done by placing an object with sufficient weight on the piston. If the mass placed on this piston is denoted by M and the piston moves some distance, Δh, then we can express the work done on the system, by the surroundings, in terms of the weight of a mass placed on the piston,

$$w = -Mg\Delta h \tag{5}$$

where g is gravitational acceleration[7]. If we divide Mg by the area of the face of the piston (expressed as A) and multiply Δh by A we have,

$$w = -\frac{Mg}{A} \times A\Delta h. \tag{6}$$

One can now see that $\frac{Mg}{A}$ is merely the external force per unit area, or *external pressure*, exerted on the gas by its surroundings and AΔh is the change in volume of gas in the cylinder so work is then,

$$w = -P_{external}\Delta V, \tag{7}$$

where ΔV is the change in volume in the cylinder. This expression expresses thermodynamic work in terms of the volume and pressure of a gas.

0.3 Meteorology on the Local and Global Scale

The aim of this section is to provide the context in which the Residence Time Analysis model operates. The atmosphere is an extremely complex system so scientists studying the atmosphere must create conceptual divisions in order to make atmospheric processes easier to understand. Using the basic physical theory developed in the last section one can get a rather intuitive understanding of the trends and processes in the atmosphere, but also create mathematical relations that enable one to construct a mathematical model of these atmospheric processes.

0.3.1 Structure of Earth's Atmosphere

The pressure, density and temperature of gases in the atmosphere are highly spatially variable. All three of these variables are functions of altitude (Figure 4). Atmospheric pressure is defined as the weight of atmospheric air above a horizontal plane divided by the area of that plane. A useful relation called the barometric law quantifies the relationship between pressure and altitude. This law describes an exponential relationship between altitude and pressure. More specifically, barometric law says pressure decreases exponentially with altitude.

The main gaseous constituents of Earth's atmosphere are nitrogen which comprises roughly 78% of the gaseous material in Earth's atmosphere, oxygen which comprises about 21% and argon which comprises about 1% of the gaseous material in Earth's atmosphere [2]. Water vapor is the next most common gaseous constituent in the atmosphere. The Earth's atmosphere also contains many trace gases such as carbon dioxide, methane and ozone.

Atmospheric scientists have divided the layers of Earth's atmosphere into layers with a constant temperature gradient[8] (Figure 4). The lowest four layers of

[7]Note we are merely using the definition of work here. Work is defined as force times distance. Here Δh is a distance and Mg is the gravitational force (i.e., weight).

[8]Within a single temperature gradient the temperature is either constantly decreasing or increasing in an approximately linear fashion.

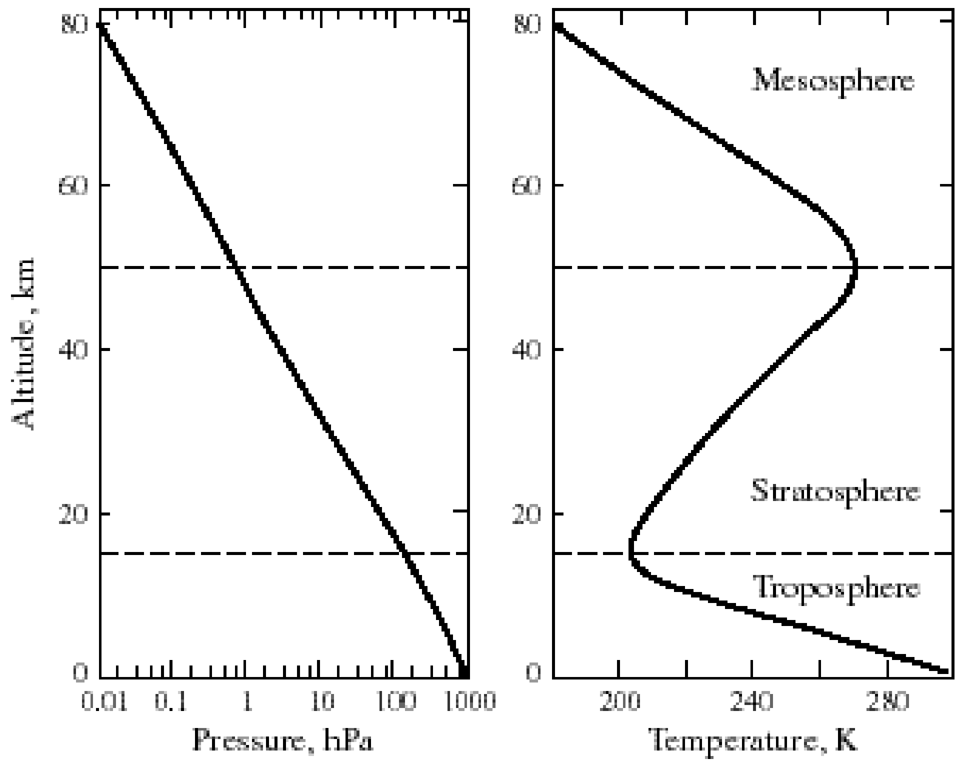

Figure 4: Mean pressure and temperature versus altitude at 30°N in March. Image used without permission from *Introduction to Atmospheric Chemistry* by Daniel J. Jacob [1].

the atmosphere (from lowest to highest altitude) are the troposphere, stratosphere, mesosphere and thermosphere. The troposphere – the lowest layer of Earth's atmosphere – extends from sea level to roughly 10 to 16 km in altitude. This layer has a relatively homogenous mixture of gases due to circulating air. The troposphere is divided into the boundary layer which is the region from about 500 meters to 3000 meters in altitude and the free troposphere which is the rest of the atmosphere. The boundary layer is the portion of the troposphere that is influenced by the Earth's surface and responds to surface conditions on a relatively short time-scale. Gas and vapor emissions from Earth enter the atmosphere from the lower levels of the troposphere. The amount of mixing in the first kilometer of the atmosphere is largely dependent on the temperature profile of the air and wind speeds. The *mixed layer depth* is the altitude at which the mixed layer exists. The mixed layer is the layer of the atmosphere where their is a high level of turbulent motion that causes air at different layers to mix together. The mixed layer is caused by atmospheric instability[9]. The mixed layer can dilute an air parcel that is highly concentrated with some pollutant.

Relative humidity is a metric often used in weather reports. It is also given as an output of HYSPLIT. It is a measure of atmospheric water concentration and is the ratio of partial pressure of water vapor in a gas sample to the equilibrium

[9]See Section 0.3.2 for explanation of atmospheric instability.

vapor pressure of water in a gas sample. In other words, it is simply the ratio of the partial pressure of water vapor in an air sample to the equilibrium vapor pressure of an air sample.

0.3.2 Local and Global Atmospheric Transport

The dynamics of gases and particles in the atmosphere are dictated mainly by gravity, the pressure-gradient, the Coriolis effect and differences in pressure due to differential heating of the Earth's surface by incoming solar radiation. In the horizontal direction (i.e, the direction of atmospheric flow that is parallel to Earth's surface) the Coriolis effect and differences in pressure work together to cause atmospheric particles to move thousands of kilometers across the surface of the Earth. The vertical profile of mass in Earth's atmosphere is determined mostly by the pressure-gradient and gravitational pull of Earth. Gravity acts to accelerate massive objects towards Earth.

Coriolis effect and Differential Heating of Earth's Surface

The Coriolis effect can be most simply illustrated by imagining the trajectory of a particle traveling from the North Pole, due South, towards a target on the equator (Figure 4). From the reference frame of an observer in outer space, at a fixed distance from Earth and rotating *with* Earth, the particle travels in a straight line from the North Pole to the equator (Figure 4, top). Now imagine what the trajectory appears like for an observer fixed on Earth's surface. The tangential velocity near the North Pole is much less than the tangential velocity at the equator. A particle that starts at the North Pole and travels to the equator takes some amount of time to reach a point on the equator. During this time Earth continues to spin on its axis. Therefore, to an observer on Earth (i.e., rotating with Earth) it seems as if the trajectory of the particle has been deflected to the right of the particle's original target (Figure 5, bottom). The Coriolis effect is sometimes called the *Coriolis force* because it appears to be the product of some imaginary force, although its not in fact a force. The Coriolis effect is due to observations being made in a rotating reference frame: Earth.

The Sun heats parts of Earth to different extents. This causes differences in air pressure at different latitudes and longitudes. The differences in temperature create horizontal pressure gradients. Since pressure is defined as force per unit area, a difference in pressure across some volume results in a net force being exerted on any particle in that given volume. This is why gases tend to move from areas of higher pressure to areas of lower pressure.

As air moves from high to low pressure it is deflected by the Coriolis force. When the pressure-gradient force and the Coriolis effect cancel one another out resulting in a stable, horizontal flow of air the resulting flow is called the *geostrophic flow*. At altitudes below 1 kilometer the path of a particle is impeded by terrestrial obstacles such as mountains, trees and buildings. This fricative force opposes the motion of a particle. The combination of the aforementioned factors combine to

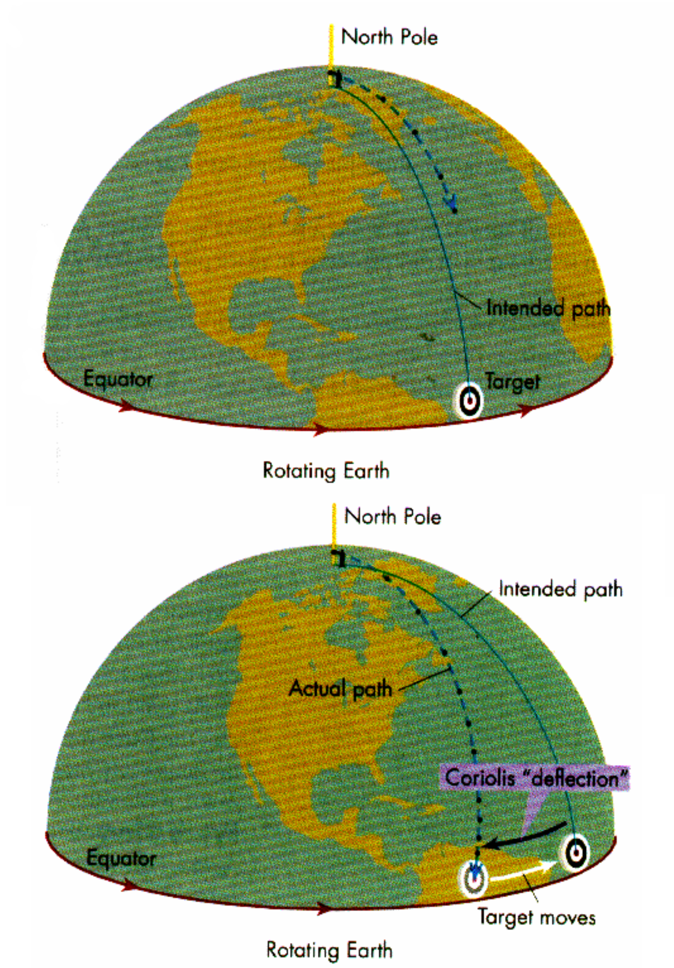

Figure 5: Diagram of the Coriolis effect. A particle begins at the North Pole and travels to a point on the equator. The trajectory of the particle appears different from the reference from of an observer at a fixed distance from Earth in outer space (top) and from an observer on Earth (bottom). Image used without permission [30].

transport particles across the globe (Figure 6).

Vertical Transport, Atmospheric Stability and Potential Temperature

Above we established the main driving forces of horizontal transport on the synoptic scale: the Coriolis effect, differences in pressure across the surface of the Earth due to differential exposure to incoming solar radiation and fricative forces with Earth's terrain. In the vertical direction both buoyancy brought on by a pressure-gradient force and gravity cause particles and gases in the atmosphere to change altitude.

The concept of the pressure-gradient force is an important one. To derive an expression for the pressure-gradient lets consider a stationary air parcel with a thickness of dz and horizontal area of A (Figure 7) at an altitude of z. The air in the

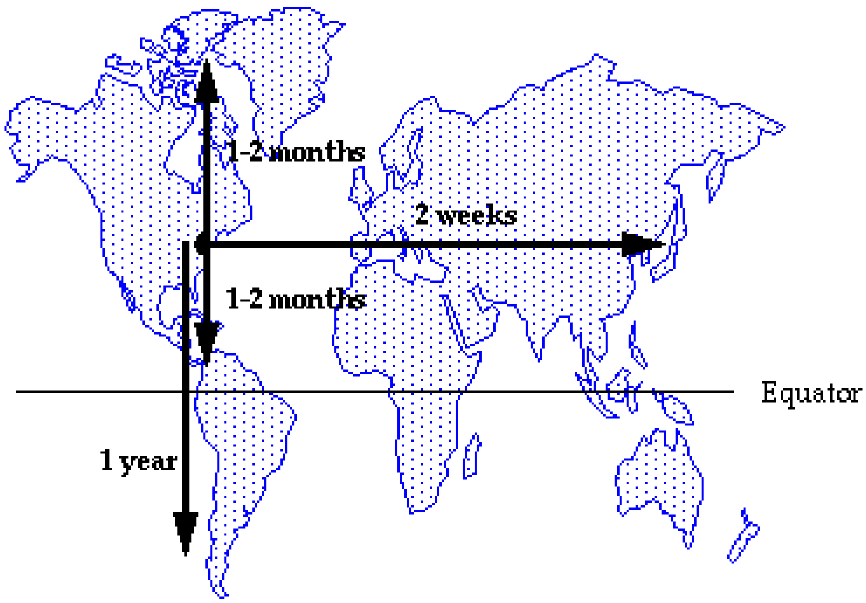

Figure 6: Typical time scales for horizontal transport of atmospheric species in the troposphere. Image used without permission from *Introduction to Atmospheric Chemistry* by Daniel J. Jacob [1].

atmosphere exerts an upwards force from the bottom of the the air parcel[10] – we will call this *P(z)A* – and a downward force from above the air parcel – lets call this *P(z + dz)A*. The net force from the differences in pressure above and below the air parcel is given by Equation 8,

$$(P(z) - P(z + dz))A \tag{8}$$

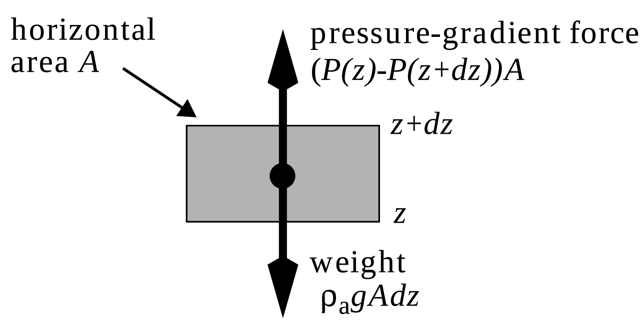

Figure 7: Diagram of the pressure-gradient force and weight acting on air parcel which results in buoyant motion. Image used without permission from *Introduction to Atmospheric Chemistry* by Daniel J. Jacob [1].

and is called the *pressure-gradient force*. For an air parcel to be in equilibrium[11] and

[10]This force is due to gas molecules bumping into the air parcel's molecules.

[11]By equilibrium I mean the magnitude of the upwards and downwards forces are equal and

not gaining or losing altitude the pressure-gradient force must equal the downward gravitational force of the air parcel itself. Using the density of the air parcel, ρ_a, we can express this as,

$$\rho_a g A dz = (P(z) - P(z + dz))A. \tag{9}$$

Rearranging the above expression and canceling out A yields,

$$\frac{P(z) - P(z + dz)}{dz} = -\rho_a g. \tag{10}$$

Employing the concept of a derivative[12] Equation 10 becomes,

$$\frac{dP}{dz} = -\rho_a g. \tag{11}$$

The resulting net force exerted on an air parcel is a product of the vector sum of the pressure-gradient force and the weight of the air parcel and is called the air parcel's *buoyancy*.

Atmospheric Stability Buoyant motion in the atmosphere is generated by a vertical temperature gradient. Buoyant motion causes atmospheric gas to change altitude and consequently be surrounded by gases of variable pressures and temperatures. This leads to what is called *atmospheric stability*. What follows is a derivation of atmospheric stability.

Consider an atmosphere that has a horizontally homogenous temperature profile denoted by T_{ATM}(z)[13] (see Figure 8). We will denote an air parcel with some altitude, z, with an A (once more, see Figure 8). Assume that an external force pushes the air parcel upwards from its current altitude, z, to a higher altitude, z + dz (Figure 9). The pressure at this higher altitude is less than the pressure at the lower altitude. This causes the air parcel to expand and to perform work on its surroundings. If we also assume that the air parcel does not exchange heat with its surroundings and therefore rises adiabatically (dQ = 0), then the change in internal energy of the air parcel is entirely due to the work the air parcel performs on its surroundings. Therefore, the air parcel cools as it rises. As the gaseous molecules of an air parcel cool they become more dense. If an air parcel is more dense than the gas that surrounds it, the air parcel drops to a lower altitude. The question now is: How quickly does the air parcel cool relative to how quickly the surrounding air cools?

The change in temperature with respect to altitude of an air parcel, assuming it changes altitude adiabatically, is known as an air parcels *adiabatic cooling rate*[14]. The *adiabatic profile*, T_A, of an air parcel is shown in Figure 8. If an air parcels cools at a slower rate than the surrounding air the air parcel will continue to rise due

opposite.

[12]Here, I am taking the limit $\Delta z \to 0$.

[13]By *temperature profile* I mean that temperature changes as a function of altitude.

[14]The adiabatic cooling rate can also be expressed as $\frac{dT_A}{dz}$.

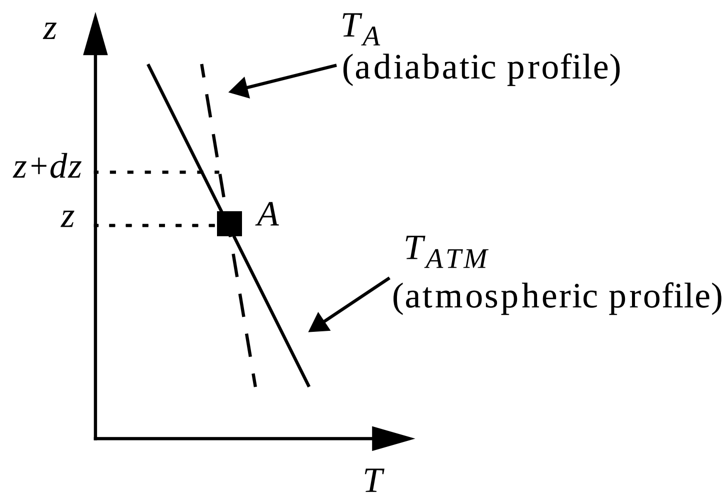

Figure 8: A plot of temperature of atmospheric gas versus the altitude of atmospheric gas representing atmospheric instability. The rate of change in temperature with respect to altitude of the adiabatic air parcel is greater than the rate of change in temperature with respect to altitude of the surrounding atmospheric gas. Image used without permission from *Introduction to Atmospheric Chemistry* by Daniel J. Jacob [1].

to buoyant motion. If the air parcel cools at a faster rate than the surrounding air the air parcel will return to its initial, lower altitude after a slight increase in altitude. Whether the air parcel keeps on rising depends on the relative temperature gradients of the air parcel, T_A, and its surrounding atmosphere, T_{ATM}(z). An air parcels that continues to rise when subject to small initial pushes is said to be *unstable*. On the contrary, if the vertical motion of an air parcel is suppressed due to surrounding air being relatively colder than the air parcel, this atmosphere is said to be *stable*. An unstable atmosphere is one where gases are well mixed and is conducive to turbulent flow.

Turbulent Flow and Potential Temperature The vertical pressure profile of the atmosphere (and therefore the vertical buoyancy profile) is determined by the vertical gradient of temperature in the atmosphere. As gas rises in the atmosphere it becomes cooler and therefore becomes more dense, counteracting buoyant forces that caused this gas to rise. On a larger scale this causes vertical flow in the atmosphere. Irregularities arise in this flow. These irregularities are called *turbulence*.

In a flowing fluid there are two flow regimes: laminar and turbulent. A laminar flow is one in which the flow is smooth and regular and a turbulent flow is unpredictable and irregular. To determine whether a flow is laminar or turbulent one can use an empirically derived formula to calculate a dimensionless quantity called Reynolds number (Re),

$$Re = \frac{UL}{\vartheta} \tag{12}$$

where U is the mean speed of the flow, L is the length of the flow and ϑ is the viscosity of the fluid. The transition from laminar to turbulent flow generally takes place at Reynolds numbers in between 1000 and 10,000.

Now lets develop the notion of *potential temperature*. Assume an air parcel somewhere in the atmosphere has a temperature T and a pressure p and is initially in thermodynamic equilibrium with its surrounding atmospheric gas. Now, if this air parcel is brought down, adiabatically, to the surface of the Earth with a pressure of p_0 = 1000 mbar then this air parcel will have a temperate of θ called the potential temperature. The potential temperature of air is related to the initial temperature of the air parcel, T, and the initial pressure, p, of the air parcel by,

$$\theta = T\left(\frac{p_0}{p}\right)^{0.286}. \tag{13}$$

0.3.3 Deposition Wet and Dry

The removal of atmospheric species occurs by two general processes: *dry deposition* and *wet deposition*. Dry deposition refers to the process by which material from the atmosphere is deposited onto the surface of the Earth without the aid of precipitation. Wet deposition is the transfer of airborne particulate matter from the atmosphere in aqueous form.

Dry Depostion The main factors that determine if a particle undergoes dry deposition are the levels of atmospheric turbulence, the chemical properties of the particle and the chemical properties of the surface itself. If one attempts to model dry deposition it quickly becomes clear that one cannot track the microphysical pathways of a particle as it travels from the atmosphere to the ground. Therefore we must abstract away from this tedious interpretation and have a theory that is more concerned with bulk properties.

A common model used to understand dry deposition assumes that dry deposition flux is directly proportional to the local concentration, *C*, of the depositing species at some height above the surface multiplied by a proportionality constant, v_d, called the deposition velocity. Deposition velocity has units of length per unit time. The expression gives the vertical dry deposition flux,

$$F = -v_d C. \tag{14}$$

The use of the above expression allows one to combine the complexities of dry deposition into a single value that is a function of altitude, turbulence and the specific chemical properties of the species and the surface: the deposition velocity. As a particle undergoes dry deposition it travels through different layers of the atmosphere with varying degrees of turbulence and eventually reaches the surface which may or may not have a chemical affinity for the particle undergoing deposition. The factors that deposition velocity, v_d, takes into account are the wind resistance a particle experiences while traveling towards the surface of the Earth and how much of a chemical affinity the surface of the Earth has for that particle.

Wet Deposition In this work we will be mainly concerned with wet deposition due to a raining cloud. This is called *precipitation scavenging*. As a raindrop falls through the air it collides with airborne particles and collects these particles, consequently transporting them to the surface of the Earth. This is a complex process with particulars that are out of the scope of this work. Broadly speaking, precipitation scavenging is proportional to the average size and density of rain drops in a precipitation event. Generally speaking, wet deposition is positively correlated with rainfall levels.

0.4 Modeling Earth's Atmosphere

This work deals with two different classes of atmospheric models: chemical transport models and receptor models. A backwards trajectory model is a type of chemical transport model. Chemical transport models, in general, take into account the microphysical processes[15] of an air parcel to approximate the paths particles take when traveling through the atmosphere. Chemical transport models may also take into account the chemical reactivity of a particle traveling through the atmosphere. Receptor models are statistical models that attempt to relate measured concentrations at a receptor to their sources, without explicitly reconstructing the paths particles take when traveling through the atmosphere. HYSPLIT is a chemical transport model.

0.4.1 Chemical Transport Models

Chemical transport models are used to understand the features that dictate the concentration of chemical species in the Earth's atmosphere. The concentration of a chemical species in the atmosphere is determined, mainly, by three factors: emissions of species from Earth, the chemistry of a species[16] and how a species is transported through the atmosphere. The two most basic flavors of chemical transport models are the box model and the puff model.

In a box model an imaginary box is constructed around some volume of the atmosphere. This box is fixed in space and what varies is the composition of gases inside the box. One can then compute the concentration of some arbitrary species – called species X – inside a box (Figure 9).

The amount of particle X in this box will increase due to both anthropogenic and natural emissions of particle X from Earth – denoted with an E in Figure 6 – and from chemical reactions that produce species X – denoted with a P in Figure 6.

[15]For a detailed exposition of these *microphysical processes* I recommend *Atmospheric Chemistry and Physics* by Seinfeld and Pandis [2].

[16]This revolves around the question of whether a species reacts to make more of itself or reacts to cause less of the species exist in the atmosphere?

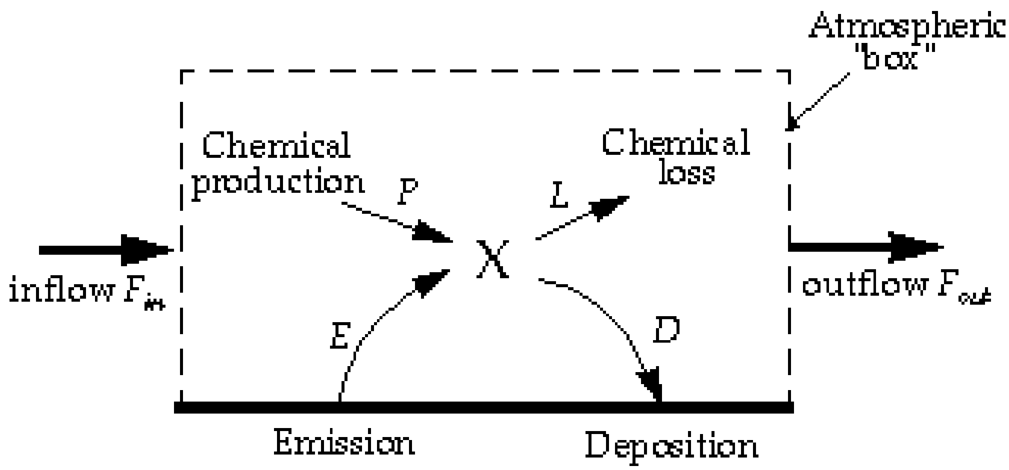

Figure 9: A representation of a one-box model representation of the atmosphere. The box is an air parcel. Image used without permission from *Introduction to Atmospheric Chemistry* by Daniel J. Jacob [1].

The amount of particle X may decrease due to chemical loss – denoted with an L – or by deposition – denoted with a D (Figure 6). If we assume mass conservation, the rate of change in the mass of species X in the box (with respect to time) is the difference between the rate at which sources add species X to the box minus the rate at which sinks remove species X from the box. This is given by,

$$\frac{dm}{dt} = \Sigma sources - \Sigma sinks = F_{in} + E + P - F_{out} - L - D, \tag{15}$$

which uses the variables defined above as well as the notions of the transport of species X into the box, F_{in}, and the transport of species out of the box, F_{out}[17].

Intimately related to the change in mass of some species in an air parcel with respect to time (Equation 15) is the *residence time* of a species. Residence time is simply the average of the time every particle of some species spends in the atmosphere.

In a puff model, the movement of a particle in a fluid flow is modeled. This differs from a box model because the domain of the model moves in space. The "box" in a box model is fixed in space. The puff in a puff model is a fluid element that is sufficiently small that all points of the puff move with the same velocity relative to one another, but is big enough to contain a statistically significant representative ensemble of a given particle.

Central to modeling flow in the atmosphere is the *continuity equation* for fluid dynamics. This equation describes the transport of a flowing fluid. The gaseous material that comprises the Earth's atmosphere can be modeled as a fluid. The continuity equation provides a relation for the concentration of a species in some

[17]See *Introduction to Atmospheric Chemistry* by Daniel J. Jacob for a more rigorous explanation of how this equation was conceived.

volume over time. There are two main forms of the continuity equation employed in modeling the flow of atmospheric gases: the Lagrangian form and the Eulerian form. While the technical details of how these two models differ require the use of advanced mathematics – and are thus out of the scope of this work – one can say that these two paradigms differ in that in a Lagrangian equation is derived relative to a frame of reference moving with local fluid flow (analogous to a puff model) and an Eulerian continuity equation is derived from employing the idea of fluid flowing through a volume element fixed in space (analogous to box model).

The ideas above are employed in HYSPLIT to compute backwards trajectories. Broadly speaking there are three main classes of models that are used to ascertain the previous position of an air parcel from current atmospheric conditions: trajectory models, dispersion models and chemical transport models. All three of these models utilize the mathematical concept of a vector field. A vector field (or field) is a subset of space in which each point in space has a vector associated with it; therefore each point in a vector field has a force associated with it.

Chemical Transport Models (CTM)[2] synthesize data from meteorological fields, emissions and physicochemical atmospheric processes with data about reactivity (kinetics and specific reactions) to describe the chemical and physical transformations within air masses. CTMs include characteristics from trajectory and dispersion models (wind fields and turbulence), but also consider chemistry and emissions.

The Hybrid Single-Particle Lagrangian Integrated Trajectory Model

The Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model created by the Air Resource Laboratory is a model that is capable of air parcel trajectory simulations as wells as modeling atmospheric deposition [3]. The model uses a hybrid Lagrangian-Eulerian approach to simulate the advection of an air parcel. HYSPLIT is able to compute both backwards and forwards trajectories.

Besides computing backwards trajectories HYSPLIT outputs meteorological data and can superimpose a trajectory on a map (Figure 2). One can use HYSPLIT to create multiple backwards trajectories from the same starting point and with up to three different starting heights. A starting height is the altitude an air parcel has at the receptor. This is the initial height from which a backwards trajectory created by HYSPLIT is generated.

0.4.2 Receptor Modeling in Environmental Chemistry

In order to regulate air quality it is important to understand how species released into the atmosphere travel. This includes knowing how species undergoes chemical transformations while traveling from source to sink and where sources and sinks are located. One approach to modeling the aforementioned phenomena is to model the emission and dispersion of materials from a point after making assumptions about atmospheric processes. A fundamental assumption receptor models make is based on the concept of conservation of mass. Conservation of mass is the

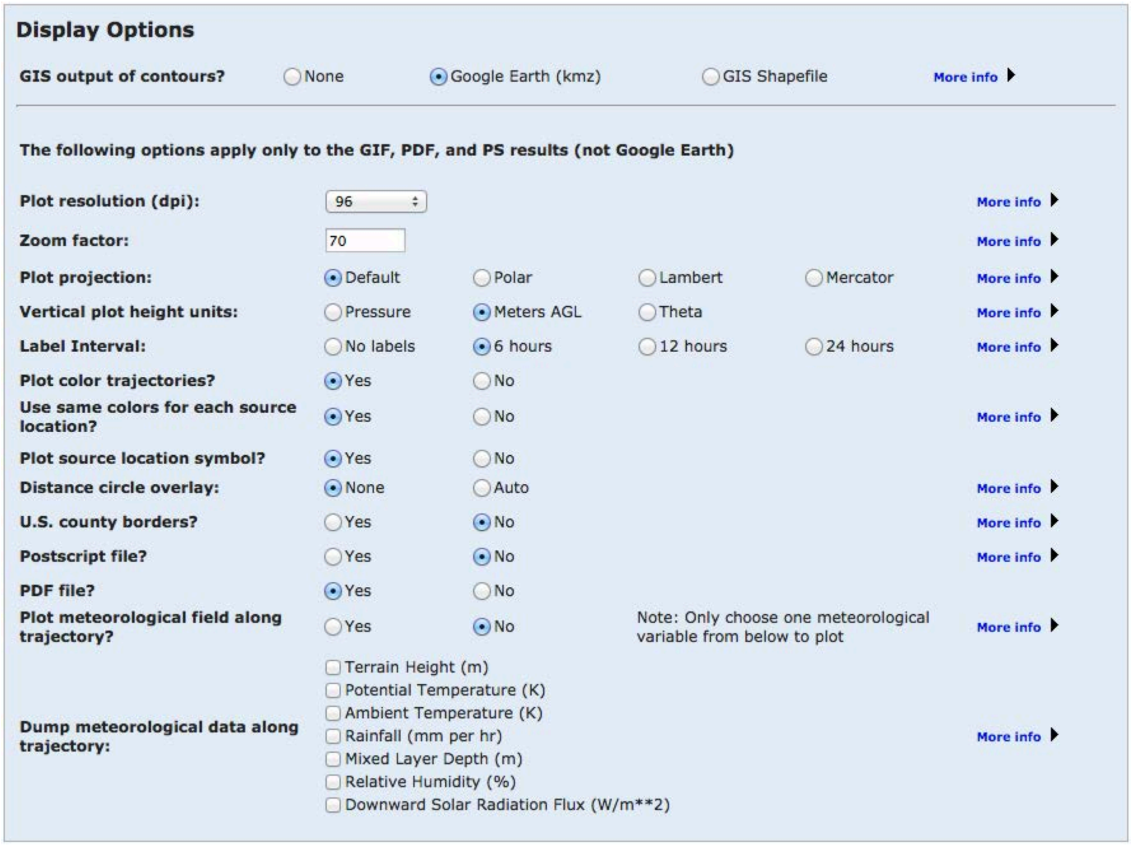

Figure 10: Screenshot of the possible inputs to HYSPLIT's trajectory model. Used without permission [3].

idea that in a closed system the mass remains constant. Only when mass enters or leaves a system does the mass change in the conservation of mass regime.

A consequence of assuming conservation of mass is ability to employ mass balance analysis. To illustrate the idea of mass balance lets imagine we want to measure the airborne particulate matter of some species, X, that is emitted from three sources. Lets say that the only sources of this species are a smokestack, automobiles and livestock. Using the concept of conservation of mass we can then say that the total concentration of a species, X, in the air is equal to the sum of the concentrations from each source. This is given by,

$$X_{total} = X_{smokestack} + X_{automobiles} + X_{livestock}. \tag{16}$$

The basic premise of receptor modeling is that the properties of material collected in the environment can be used to infer their origins. This means that receptor models are reliant on accurate and sensitive sampling methods. While the analytical techniques used to elucidate the composition of air samples are not the concern of this work, it should be mentioned that obtaining and analyzing a sample of air that is representative of ambient air quality is a difficult, yet integral component in utilizing the full power of a receptor model.

0.5 Summary

Source region impact factors have utility because they are a simple metric of source-receptor relationships and work with robust backwards trajectory models such as HYSPLIT. The necessity for understanding the relationship between a pollutant source and a region some distance away is a fundamental concern for atmospheric scientists. This concern is motivated by the understanding of long range chemical transport in the atmosphere. This work is concerned with creating a robust instantiation of the Residence Time Analysis model which is used to compute a source region impact factor.

Chapter 1

How Does One Compute a Source Region Impact Factor?

The Residence Time Analysis (RTA) model is a statistical model that is used to determine the locations of dominant pollutant *sources* affecting a far off, downwind receptor and can also give insight into major *pathways* polluted air parcels take to reach a given receptor. The output of a Residence Time Analysis is a source region impact factor (SRIF). A source region impact factor is a probability and therefore a number between 0 and 1[1]. An SRIF is the probability that a given air parcel spent time in a given *source region* and reflects the amount of time that an air parcel spent in the source region. A source region is simply a two-dimensional apportionment on Earth's surface specified by longitudes and latitudes. A source region is artificially[2] created. Residence Time Analysis utilizes a modeled backwards trajectory of an air parcel as well as information about the chemical composition of an air parcel at a given receptor to identify major pollutant sources affecting a receptor. This chapter expounds the nature of Residence Time Analysis and its output – the source region impact factor – by providing the context in which this model was formulated.

1.1 Origins of Residence Time Analysis

Residence Time Analysis was first proposed in 1985 by Ashbaugh, Malm and Sadeh [11]. This model was created in response to amendments to the Clean Air Act made by Congress in 1977 that mandated that there be a national goal to prevent and remedy air visibility in what Congress called Class I Federal areas. Class I Federal areas included national parks and wilderness reserves in the western United States [24]. There was increasing concern that growing energy needs were contributing to a decrease in air quality due to the burning of coal and petroleum. Source region impact factors were an attempt to create a relatively simple yet accurate tool that policy makers could use to identify the sources and potential dis-

[1] See Appendix A for refresher on probability theory.
[2] By "artificial" I mean the modeler creates this apportionment.

tribution of pollutants so as to draw up regulations to mitigate the adverse effects that increasing industrialization had on the air quality of Class I Federal areas.

1.2 Deriving an Expression for the Source Region Impact Factor

Residence Time Analysis usually begins with the collection and analysis of ambient air samples at a receptor. The analysis of ambient air samples yields the air's chemical composition[3]. Once the chemical composition of air at a receptor is known one may want to know the origins of these species[4]. To understand and predict the locations of the sources of a given species in an ambient air sample it is important to know the trajectory an air parcel took to reach a given receptor. We assume that air parcels which spent a large part of their trajectories flowing through areas with high pollutant concentrations will have higher pollution content once they reach a receptor than air samples that flowed through regions with little pollution emission. One way to quantify relative impacts of source regions on a receptor is to determine the spatial probability distribution of the previous positions of air parcels at given time intervals that arrived at a particular receptor over a specified time period. This is the central idea behind Residence Time Analysis.

Given a backwards trajectory that defines the set[5] of previous longitudes and latitudes of an air parcel, one can compute the spatial probability distribution of the previous longitude and latitudes an air parcel has had. To do this one must apportion the region of interest. Nuances of apportionment will be discussed later, but the basic aim of apportionment is to discretize the possible locations of an air parcel into one or more apportionments. This treats the theoretical continuum of possible positions of an air parcel as a set of discrete points and thus makes the calculation of a source region impact factor computationally tractable. The spatial probability distribution that arises from this analysis is reflective of the residence time of some particle in a given apportionment.

We will now construct a probability model that uses the residence time of an air parcel released into the atmosphere to create a spatial probability distribution[6]. Construction of the Residence Time Analysis model from first principles provides the reader with a richer and deeper understanding of the model. We begin this derivation by noting that to perform a Residence Time Analysis one must be supplied with a backwards trajectory and the longitude and latitude of a receptor. This allows us to define the sets ω and Ω which are the set of the coordinates – expressed as (longitude, latitude) – an air parcel took (i.e., all coordinates that lie on the air parcels backwards trajectory) and the set of all possible coordinates on

[3]The analytic techniques employed to understand the composition of ambient air samples are out of the scope of this thesis. For an example and discussion of such techniques see [12].

[4]By species I mean the chemical components of an ambient air sample.

[5]HYSPLIT provides us with this back trajectories.

[6]A similar derivation to the one that follows can be found in [11].

Earth's surface an air parcel could traverse – also expressed as (longitude, latitude), respectively. These are given by,

$$\Omega = \{\text{all possible coordinates on Earth's surface specified as (longitude, latitude)}\}. \tag{1.1}$$

$$\omega = \{\text{all coordinates on a backwards trajectory specified as (longitude, latitude)}\}. \tag{1.2}$$

It should be noted that ω is a subset of Ω (i.e., $\omega \subset \Omega$). It should also be noted that Ω[7] defines the sample space of this probabilistic model.

Now, one should ask: What constitutes an event in this probability model? An event is defined as a particle having a specific longitude and latitude at one instance in time. One can visualize this by imagining the position of a particle represented by a vector on a 2-D plane – with axes of longitude and latitude – at some time t (Figure 1.1).

Since we would rather deal with discrete positions we will define an event as a particle being bound by some box. If we let x (and X) denote longitude and y (and Y) denote latitudes, then we can express this box as bound by $x_i < X \leq x_{i+1}$ and $y_i < Y \leq y_{i+1}$ at some time t' (Figure 1.1). This allows us to define an *event* as,

$$(A_{ij}) = \{(X, Y) : (x_i < X \leq x_{i+1}, y_i < Y \leq y_{i+1})\}. \tag{1.3}$$

This demonstrates the need to discretize the sample space in this probability model. While the locations of a particle in this space are technically on a continuum, we will treat the particle as having discrete motion. This means the particle is either in a given apportionment or not in a given apportionment. Moreover, it should be noted that this model operates on a fixed time scale.This implies that the set of possible coordinates a particle could have (i.e., the sample space) is restricted. In a given time interval the probability of an event A_{ij} occurring is the probability that the previous position of a randomly selected air parcel was found in the ij_{th} cell at any point in the time interval.

Now that we have thoroughly defined what constitutes an event we must determine how to assign a probability to a given event. One way to do this is to obtain the joint probability distribution for the spatial distribution of particles under consideration. We can define the the probability that an event A_{ij}, defined above, will occur using the joint probability distribution $f_{X,Y}(x, y)$,

[7]If we assume a particle has some constant velocity then the sample space will actually be the set of all physically realizable points on Earth's surface. This will be a circle around the particles initial starting point with a radius of the product of the velocity of the particle and the time the particle traveled for.

$$\mathcal{P}(A_{ij}) = \int_{y_j}^{y_{j+1}} \int_{x_i}^{x_{i+1}} f_{X,Y}(x,y)dxdy. \tag{1.4}$$

The joint probability distribution can then be plotted (Figure 1.3). In practice, however, it is difficult to guess the joint probability density, $f_{X,Y}(x,y)$, to obtain the probability that an air parcel's previous position was within some grid, $\mathcal{P}(A_{ij})$. Therefore, what is done in practice is to estimate $\mathcal{P}(A_{ij})$ using a modeled backwards trajectory.

To compute the spatial probability distribution without the joint probability distribution requires that backwards trajectories of air parcels be computed at specific time increments over a prescribed time interval. This means that points at regular intervals must be specified on a given trajectory. These points are called backwards trajectory *endpoints* (see Figure 1.3 for a visualization of end points). A backwards trajectory is made up of multiple endpoints separated by a fixed time. If n_{ij} is the number of trajectory segment endpoints that fell in the ij_{th} cell during a time interval T and N is the total number of endpoints in all trajectories, then the probability that an air parcel's previous position was within some grid can be approximately given by,

$$\mathcal{P}(A_{ij}) = \frac{n_{ij}}{N}. \tag{1.5}$$

We have now derived the basic instantiation of the Residence Time Analysis model and have seen how it is used to compute a source region impact factor.

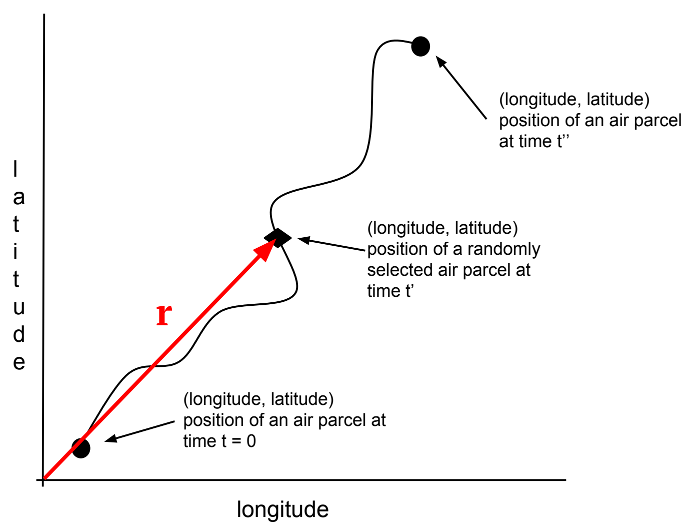

Figure 1.1: A representation of the sample space of this probability model. The trajectory of an air parcel is represented by the black curve. The circles represent the receptor and source locations. The position of a randomly selected air parcel at some time, t', is given by the vector $\vec{r}$. The head of the vector $\vec{r}$ can be thought of as an event A_{ij} if the head is located in the ij_{th} apportionment.

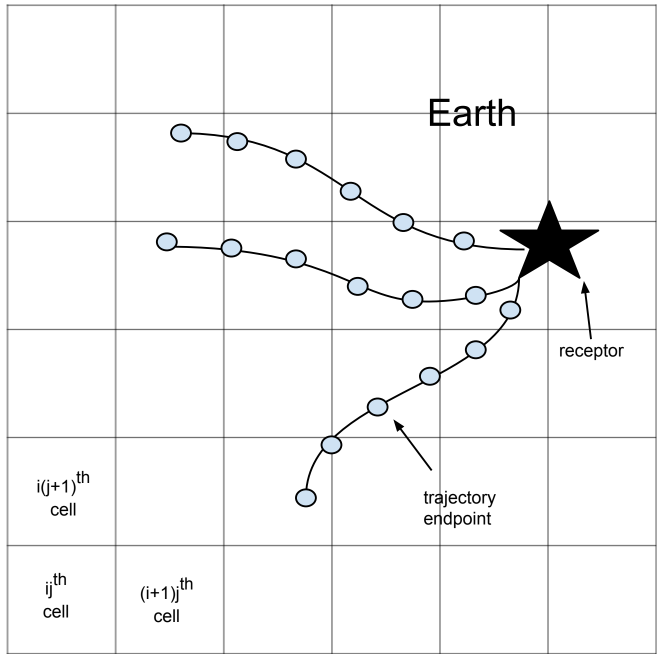

Figure 1.2: A schematic of the creation of a source region impact factor. The light blue dots represent the *endpoints* of a modeled backwards trajectory, while the black line represents the trajectory itself. The ★ symbolizes a receptor. The trajectories and grid are overlaid over the surface of the Earth. The abscissa can be thought of representing longitudinal coordinates and the ordinate can be thought of representing the latitudinal coordinates. The sample space is apportioned into multiple grids denoted by *ij*.

A source region impact factor is simply the ratio of endpoints in a given apportionment over the total number of endpoints and if multiplied by 100% gives the percentage of endpoints that were located in some source region. This is reflective of the time a particle spent in a given grid during the time it took to travel from source to receptor. A higher SRIF corresponds to a greater amount of time a particle spent in a given source region.

1.3 Previous Uses of the Model: Residence Time Analysis In Practice

Residence Time Analysis has been used widely, in part due to its simplicity, but also due to improved analytical chemical techniques and increasingly accurate modeled backwards trajectories [6, 7, 8, 9, 12]. In 2005 Begum et. al. [12] tested the efficacy of the model. This was done by comparing the location of a major airborne pollutant source – a large forest fire in Quebec – identified using Residence Time Analysis to the known location of this event. HYSPLIT was used and the receptor site was located in Northeastern United States. Begum et. al. found an agreement with the results of a Residence Time Analysis and the actual location of this forest

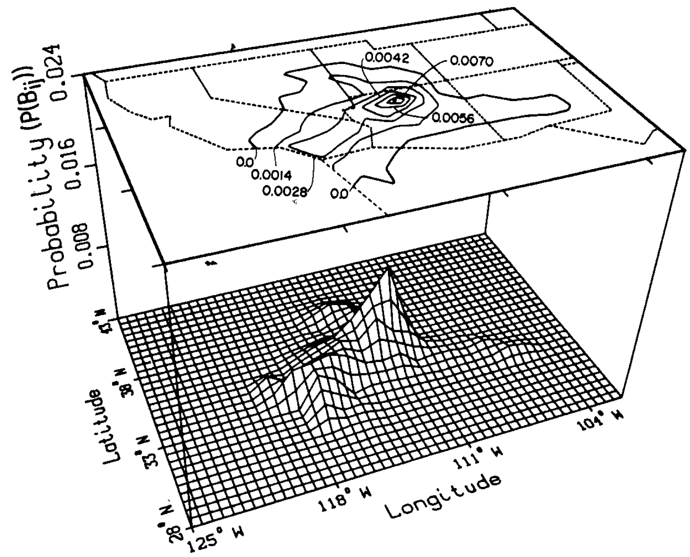

Figure 1.3: The spatial joint probability distribution of a particle. The horizontal axes are longitude and latitude. The vertical axis is the probability of that set of coordinates represented by a grid being populated by a backwards trajectory. Used without permission from [11].

fire in Quebec. Moreover, source region impact factors have previously been used to identify source regions of semivolatile organic compounds (SOCs) deposited in Central Oregon [6] and polycyclic aromatic hydrocarbons (PAHs) deposited in the Great Lakes [8]. A formula for calculating source region impact factors has been used before [7] and is given by,

$$SRIF\% = \left\{ \left[\sum_{n=1\ hour}^{k\ hours} (n(T_{SF} = 1; otherwise = 0) \right] hours/m\ hours \right\} \times 100 \quad (1.6)$$

where T_{SF} represents a binary response variable for whether a particle spent time in a single source region or not ('1' if a trajectory passed through a given source region and '0' if a trajectory did not pass through a given source region), k is the number of hours a backwards trajectory went (e.g., for a 4-day backwards trajectory k would be 4 $\times$ 24 hours = 96 hours), m is the total number of hours a backwards trajectory goes multiplied by the number of backwards trajectories calculated for a sample of air. A trajectory endpoint is given every hour so Equation 1.6 is equivalent to Equation 1.5. A higher SRIF percentage corresponds to a greater amount of time a trajectory spent in a given source region. Source region impact factors informs one of the effects a source region has on an air sample.

Residence Time Analysis is a probabilistic model. Due to this the validity of a SRIF is increased with increasing number of samples. Using HYSPLIT [3] one can compute a backwards trajectory every hour for 72 hours. HYSPLIT can also compute a backwards trajectory for three different starting points and three different altitudes. Moreover, HYSPLIT outputs data such as the height of terrain, ambient temperature, amount of rainfall, mixed layer depth and radiative flux. These variables can give insight into the deposition of pollutants, but are currently not factored into the calculation of a SRIF.

1.4 Summary

The Residence Time Analysis model can be constructed with the use of some basic probability theory. From this derivation one can see what a source region impact factor is explicitly: the probability that a given air parcel that arrived at some receptor spent time in a some apportionment. This allows one to locate the origins of measured ambient air samples in a reasonably quantitative manner. The source region impact factor has limitation in the accuracy of its conclusions. This limitation and more will be discussed in the next chapter.

Chapter 2

Limitations and Flaws of the Model

Residence Time Analysis is a relatively simple tool. With comparatively[1] simple mathematics Residence Time Analysis enables one to quantify the relationship between a receptor and a pollutant source thousands of miles away. The simplicity of this model requires that it relies on potentially inaccurate modeled backwards trajectories. Moreover this model takes a two-dimensional view. It doesn't take into account any meteorological factors that are a function of altitude, such as deposition or surface emissions which may more prominently affect air parcels traveling through low altitudes. This is problematic because one might want to know the magnitude to which a far off, downwind pollution source is contributing to the ground pollutant levels at a given receptor. Also, the computation of a source region impact factor has an inherent bias called the *inherent peak bias* due to the fact that all backwards trajectories converge at the receptor. Lastly, the source region impact factor model does not tell one where and how to apportion. This chapter discusses the aforementioned limitations and flaws of Residence Time Analysis.

2.1 Reliance on the Accuracy of a Backwards Trajectory

A source region impact factor is only as good as the backwards trajectory used to compute it. The resolution of a backwards trajectory can be thought of as the error in position associated with each trajectory endpoint. If the error in a trajectory endpoint is large – for example, larger than the source region that contains it – then the source region impact factor is no longer a reliable metric of that source-receptor relationship. Moreover, associated with each trajectory endpoint is a series of meteorological data such as rainfall and solar radiative flux. These parameters also have error associated with them.

The backwards trajectories used in this work were generated by HYSPLIT. While it is not possible to give an absolute error for all backwards trajectory end-

[1]Compared to other receptor modeling techniques which make use of more sophisticated statistics. See [14] for an in depth exposition of other receptor modeling techniques.

points, there is a methodology for approximating these errors [26]. Modeled backwards trajectories created by HYSPLIT are semi-empirical in that they rely on meteorological data at various sites around the globe [26]. This meteorological data inherently has random error associated with it (as all physical measurements do). Trajectory endpoints also contain errors due to the fact that HYSPLIT is a computational – and thus discrete – representation of a physical system. HYSPLIT models the atmospheric flow field (Figure 2.1), which is a vector field that represents the flow of air through the atmosphere [26]. These vectors can be thought of as plotted on a discrete coordinate system by HYSPLIT. Depending on the resolution of this coordinate system – how far adjacent grid points in this coordinate system are from one another – the *discrete* representation of a *continuous* physical system by HYSPLIT will have some error. The resolution error can be estimated by starting several trajectories about an initial point, but at different altitudes [29]. The divergence of these trajectories from one another will give an estimate of the uncertainty due to divergence in the flow field [26].

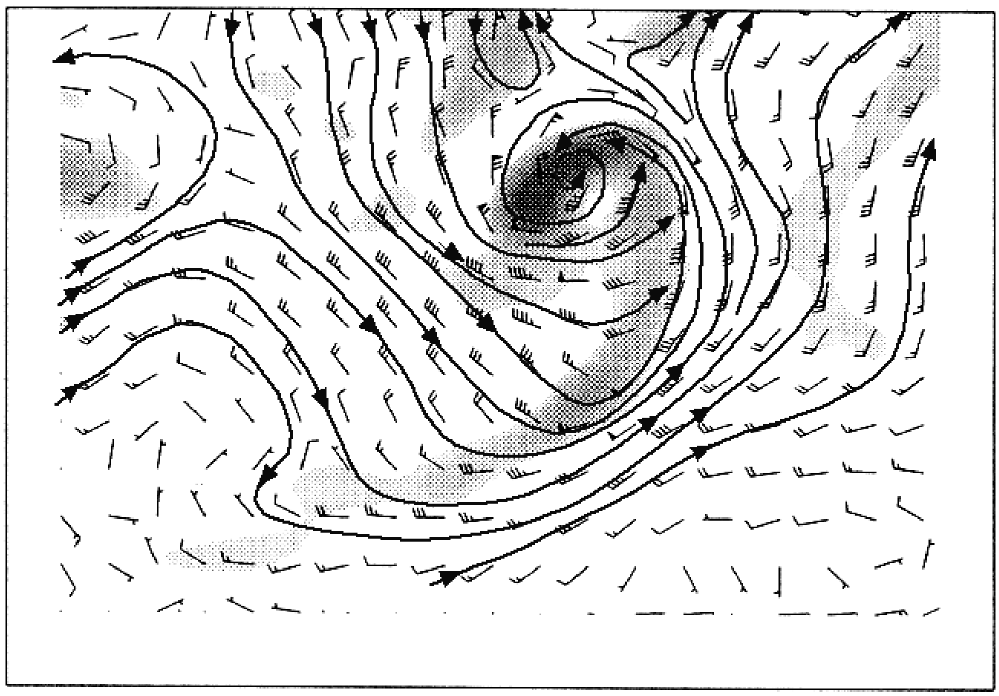

Figure 2.1: Representation of a meteorological flow field. Used without permission from [28].

When computing a backwards trajectory using HYSPLIT the accuracy of the trajectory endpoints decreases the farther away the endpoint is from the receptor and the farther back in time the endpoint is computed [22]. This is a result of errors in previous trajectory endpoints compounding one another as well as uncertainty in meteorological data [26].

If we think of the computation of a HYSPLIT backwards trajectory as an experiment with random error associated with it, then we must compute many backwards trajectories and obtain the geographic coordinates of many trajectory endpoints if we are to have an accurate source region impact factor. In other words, a large number of trajectory endpoints is required to give the source region impact factor statistical significance.

2.2 Problems with Apportioning

How one apportions determines the source region impact factor garnered from Residence Time Analysis. For example, apportioning by a 1° longitude by 1° latitude grid creates what is known as the "pole problem" [22]. The problem is due to the fact that by apportioning in this manner the grids nearest the equator are smaller than the grids farther away from the equator so the number of backwards trajectory endpoints found in multiple grids cannot be compared on a per unit basis. The areas are different and therefore grids differ in their occupation of the overall sample space.

Additionally, if one apportions very small source regions then a given source region may not have enough trajectory endpoints located within the source region to garner a statistically significant conclusion. Grids that have a width and/or length that is much less than the average distance between trajectory endpoints or grids that are smaller then the resolution of a backwards trajectory will have this problem.

2.3 A Two-Dimensional Model

One can see from the derivation of the Residence Time Analysis model that it takes a two-dimensional view. Only longitude and latitude are considered in the model. This obviously is an unrealistic view of Earth's atmosphere. The introduction of this work expounded the many ways altitude shapes the properties of air. From pressure and temperature to the stability of an air parcel, many properties of the Earth's atmosphere are functions of altitude. While the computation of a backwards trajectory from HYSPLIT takes into account factors that are a function of altitude the SRIF is not reflective of any of these factors [29].

The outputs of a HYSPLIT backwards trajectory computation include elements that play a role in deposition such as rainfall. Rainfall is a means by which wet deposition occurs [2]. Therefore, if one is concerned with how a far off pollutant source is affecting a pollutant receptor then deposition might be included.

Each trajectory endpoint, resulting from a HYSPLIT backwards trajectory calculation, has an altitude associated with it. This enables one to track the changes in altitude an air parcel has as it travels towards the receptor. Air pollutant levels may be a function of altitude so air parcels that are within the mixed layer may be polluted more than air parcels that are above the mixed layer.

2.4 Inherent Peak Bias

The *inherent peak bias* is a product of backwards trajectories all converging at the receptor. This means that the apportionment containing the receptor contains at least as many trajectory endpoints as there are backwards trajectories. Therefore the maximum density in the joint probability distribution created from plotting the longitudes and latitudes of trajectory endpoints against their respective probabilities of occurrence tends to occur at the receptor (Figure 1.3). This phenomena gives higher SRIF values to the apportionments near the receptor and is due to the increasing density of trajectory endpoints in apportionments near the receptor relative to apportionments that are farther away.

This leads someone who uses Residence Time Analysis to conclude that the source is where it might not be. That is, the source is nearer to the receptor than it actually is. If this bias could be removed then one would get a more accurate SRIF. Removing this bias would normalize[2] the peak in the aforementioned joint probability distribution.

2.5 The Need for Automation

In practice, computing a SRIF involves generating many backwards trajectories using HYSPLIT. This creates a large amount of data. The output of a single HYSPLIT back trajectory calculation is a text file called a *data dump* (Figure 2.2).

```
  2     1
GDAS    15     4     8     0     0
GDAS    15     4    15     0     0
  1 BACKWARD OMEGA
15     4    15    17    45.030 -122.312    500.0
  8 PRESSURE THETA     AIR_TEMP RAINFALL MIXDEPTH RELHUMID TERR_MSL SUN_FLUX
  1     1    15     4    15    17     0     1     0.0   45.030 -122.312      0.0    951.3    282.7    278.6      0.0    572.6     72.9    697.8    209.4
  1     1    15     4    15    16     0     2    -1.0   45.050 -122.341      9.9    954.1    280.6    276.8      0.0    365.2     76.6    671.8    133.6
  1     1    15     4    15    15     0     3    -2.0   45.060 -122.369     34.2    954.7    278.5    274.8      0.0    144.2     80.3    646.8     57.2
  1     1    15     4    15    14     0     2    -3.0   45.074 -122.415     91.7    950.5    277.8    273.7      0.0    111.1     82.2    606.6     37.6
  1     1    15     4    15    13     0     1    -4.0   45.101 -122.488    156.4    948.6    278.1    273.8      0.0     78.6     82.6    542.4     18.3
  1     1    15     4    15    12     0     0    -5.0   45.138 -122.583    221.2    950.6    278.3    274.2      0.0     48.1     82.2    460.3      0.0
  1     1    15     4    15    11     0     1    -6.0   45.177 -122.677    293.0    952.2    278.9    274.9      0.0     38.9     79.6    381.0      0.0
  1     1    15     4    15    10     0     2    -7.0   45.219 -122.753    372.6    949.2    279.3    275.1      0.0     31.8     77.3    319.4      0.0
  1     1    15     4    15     9     0     3    -8.0   45.274 -122.815    455.7    944.5    279.6    275.1      0.0     26.3     75.8    271.7      0.0
  1     1    15     4    15     8     0     2    -9.0   45.358 -122.894    550.0    940.7    280.0    275.1      0.0     39.3     75.0    215.0     33.0
  1     1    15     4    15     7     0     1   -10.0   45.484 -123.011    635.9    936.6    280.3    275.1      0.0     27.2     73.4    145.0     71.9
  1     1    15     4    15     6     0     0   -11.0   45.645 -123.172    738.6    928.3    280.4    274.5      0.0    128.8     73.2    150.5    109.6
  1     1    15     4    15     5     0     1   -12.0   45.819 -123.377    900.6    910.8    280.5    273.1      0.0    320.9     77.1    141.3    144.7
  1     1    15     4    15     4     0     2   -13.0   45.985 -123.618   1102.7    888.7    280.9    271.5      0.0    524.2     77.9    112.7    182.4
  1     1    15     4    15     3     0     3   -14.0   46.145 -123.900   1303.0    868.7    281.4    270.2      0.0    655.1     71.3     62.6    222.4
  1     1    15     4    15     2     0     2   -15.0   46.304 -124.225   1447.2    853.7    281.7    269.3      0.5    912.7     64.7     35.6    369.0
  1     1    15     4    15     1     0     1   -16.0   46.454 -124.579   1542.1    847.7    281.9    268.8      0.0   1103.3     66.2     21.9    527.7
  1     1    15     4    15     0     0     0   -17.0   46.598 -124.961   1649.5    840.8    281.9    268.3      0.0   1143.4     64.1      8.9    703.1
  1     1    15     4    14    23     0     1   -18.0   46.731 -125.353   1756.9    824.7    282.3    267.1      0.1   1208.3     65.5      3.4    727.5
  1     1    15     4    14    22     0     2   -19.0   46.847 -125.728   1815.6    818.3    282.4    266.6      0.0   1257.5     65.9      0.8    760.1
  1     1    15     4    14    21     0     3   -20.0   46.956 -126.080   1833.3    816.8    282.3    266.4      0.0   1278.2     66.2      0.0    792.7
  1     1    15     4    14    20     0     2   -21.0   47.068 -126.421   1822.1    818.1    282.2    266.4      0.0   1296.0     65.2      0.0    603.3
  1     1    15     4    14    19     0     1   -22.0   47.191 -126.759   1803.0    821.0    282.0    266.6      0.0   1317.8     64.3      0.0    412.8
  1     1    15     4    14    18     0     0   -23.0   47.327 -127.096   1799.7    821.8    281.9    266.5      0.0   1335.2     62.8      0.0    223.7
  1     1    15     4    14    17     0     1   -24.0   47.468 -127.444   1832.2    818.7    281.9    266.3      0.0   1343.5     60.7      0.0    157.0
```

Figure 2.2: A HYSPLIT data dump generated from running a HYSPLIT back trajectory calculation for a Normal type of trajectory with 1 starting location and a total run time of 24 hours. This calculation had one starting height of 500 meters.

The HYSPLIT data dump contains information such as potential temperature (Kelvin), relative humidity (%), terrain height above sea level (m), rainfall (mm per hour) and mixed layer depth (m). In order to process this amount of data in a

[2]By normalize I mean adjust for effects converging trajectories have on a SRIF.

reasonable manner the process of computing a SRIF must be automated with the aid of a computer.

2.6 Summary

Residence Time Analysis is a simple yet potentially robust way to determine the locations of pollutant sources. A model of this simplicity has its flaws though. From a lack of physical dimensionality to complications of apportioning, the limitations of Residence Time Analysis incite the need for simple improvements to the model. These improvements should not detract from the usability of the model, but should rather make this model more reflective of reality. The aim of this work is to resolve these limitations by providing a practical and theoretical framework with which to reason through these limitations.

Chapter 3

A Novel Method to Compute a Source Region Impact Factor

The output of a HYSPLIT backwards trajectory calculation is a text file specifying the latitude and longitude of each trajectory endpoint. Each trajectory endpoint is separated temporally by one hour. HYSPLIT also has the capability to report other variables of meteorological interest such as rainfall (in units of mm per hour) for each trajectory endpoint. In order to increase the statistical validity of a SRIF many backwards trajectories (and subsequently many trajectory endpoints) must be used. A problem then arises: How do you deal with such a large data set? Consequently, it is both appropriate and useful to wrangle this data using the aid of a computer and to automate the simple yet tedious task of using all the HYSPLIT backwards trajectory output files to compute a source region impact factor.

Christopher Walsh created a program to do so called Trajpy v. 1.0, which computed a source region impact factor using the outputs of HYSPLIT. The work of this thesis was to create Trajpy v. 2.0 [23]. Trajpy v. 2.0 is an improvement over Trajpy v. 1.0 and addresses some of the limitations and flaws of the source region impact factor model that Trajpy v. 1.0 does not address. Trajpy v. 2.0 has an updated algorithm to create source region apportionments of any polygonal shape using the ray casting algorithm. Moreover the "inherent peak bias" was addressed in Trajpy v. 2.0. Lastly the effects of wet deposition on a far off downwind receptor were considered by implementing a schema that uses empirical weighting factors to make a SRIF contain information about deposition. The aim of this chapter to elaborate on the aforementioned improvements.

3.1 n-sided Polygonal Apportioning

The computation of a source region impact factor involves apportioning a given region of interest. In Trajpy v. 1.0 a source region could only be a 4-sided polygon. Trajpy v. 2.0 allows the user to specify an n-sided polygon as a source region. This was done by implementing the ray casting algorithm based on the odd-even rule [21]. As stated before the outputs of a backwards trajectory computed by HYSPLIT are trajectory endpoints specified by longitude and latitude. To compute a SRIF one must determine if a trajectory endpoint is either inside or not inside a given source region. This was done using the ray casting algorithm (Figure 3.1).

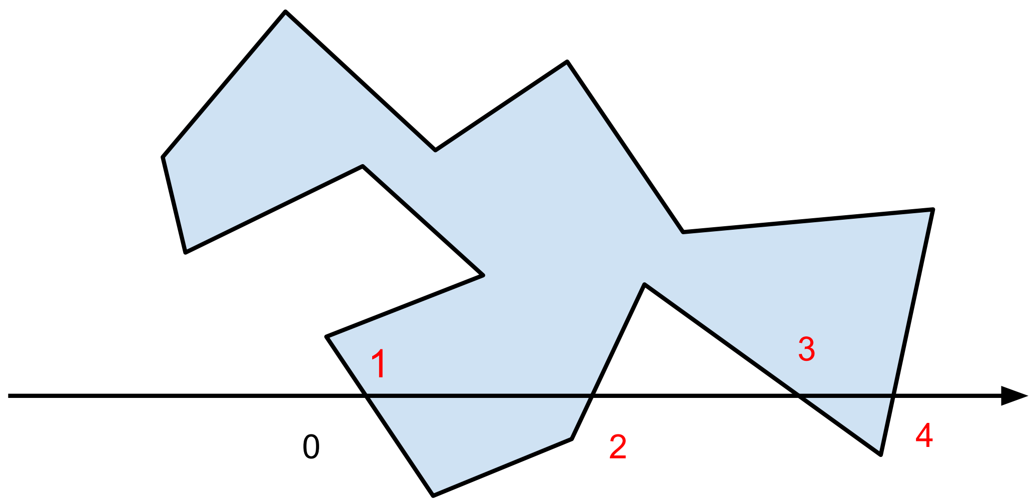

Figure 3.1: Graphical representation of the ray casting algorithm. The light blue polygon can be thought of as a source region. A ray starting at a trajectory endpoint and pointing in an arbitrary direction crosses the border of the polygon a certain number of times. If the ray crosses the border of the polygon an *even* number of times the trajectory endpoint is outside the source region. The opposite is true if the the ray crosses the border of the source region an *odd* number of times.

This is a simple algorithm that works by first drawing a ray starting at a given point – in this case the point in question is a backwards trajectory endpoint from a HYSPLIT output – that travels in any fixed direction and ends outside a given polygon. If the ray is not on the edge of the polygon and intersects the edges of the polygon an even number of times then the point is outside the polygon and if the ray intersects the polygon an odd number of times it is inside the polygon. The vertices of the polygon must be given to Trajpy v. 2.0 by the user in the form of latitude and longitude in decimal degrees (Figure 3.2). This allows one to specify source regions of any polygonal shape even those of relatively complex geometry.

Trajpy v. 2.0 also extends the apportionment capabilities to apportioning with a circle. If the radius and center of a circular source region are specified the distance formula can be used to determine if a trajectory endpoint is within a circular apportionment or not. The distance formula is simply given by,

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, \tag{3.1}$$

where x_2 and x_1 are longitudes of the center of the source region and the trajectory endpoint, respectively, and y_2 and y_1 are latitudes of the center of the source region and the trajectory endpoint.

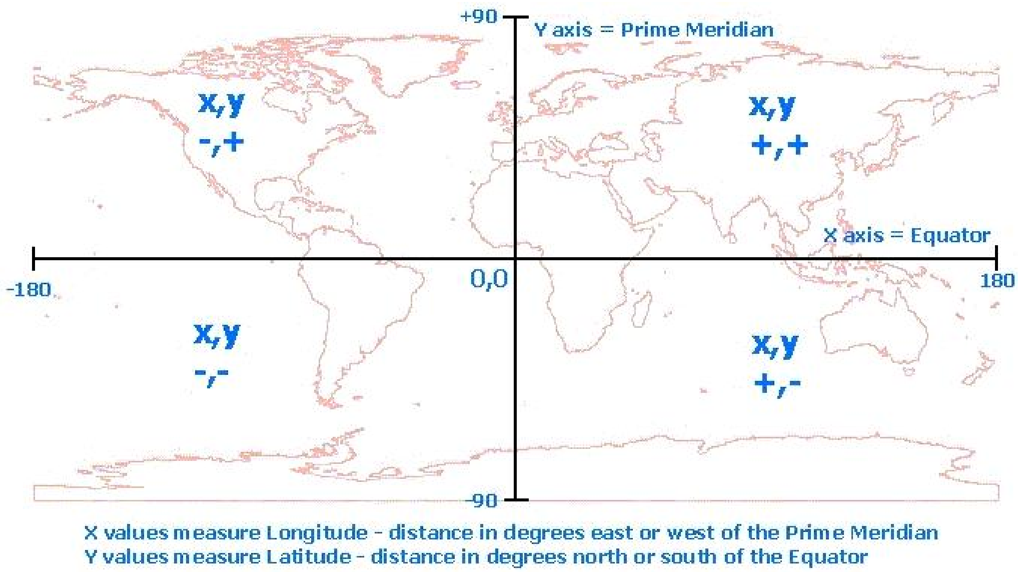

Figure 3.2: Visualization of decimal degree representation of latitude and longitude. Used without permission from [27].

3.2 Removing the Inherent Peak Bias

The problem of the *inherent peak bias* was introduced last chapter. Using a method similar to the one proposed by Ashbaugh et. al. [11], Trajpy v. 2.0 resolves the inherent peak bias. The inherent peak problem can be canceled out of the joint probability distribution that results from Residence Time Analysis (Figure 1.3) by dividing a source region impact factor by another source region impact factor. As long as these two source region impact factors are based on the same apportionment scheme, what gets divided out is the inherent peak bias. The source region impact factors both need to be computed from the same set of trajectory endpoints. In other words for a given use of RTA one can use two different sets of backwards trajectories to cancel out the inherent peak bias.

Some pollutant events only affect air parcels that travel at low altitudes, below the mixed layer depth. Trajpy v. 2.0 was originally[1] intended to quantify source-receptor relationships of pollutants that do no travel at high altitudes, therefore a scheme do so was developed. This scheme is similar to one implemented in [11]. This is done in Trajpy v. 2.0 by defining a conditional probability denoted by $\mathcal{P}(H_{ij}|A_{ij})$ which can be stated as: if a randomly selected air parcel is located in a given apportionment – say the ij_{th} apportionment – what is probability the air parcel passes through that apportionment with a low altitude? This would be useful if an air parcel would only have measurable pollutant levels at the receptor if the air parcel passed through an apportionment at a low altitude. This is defined as,

[1] Trajpy v. 2.0 is a tool that Christopher Walsh plans to use to study source-receptor relationships of species that travel at low altitudes.

$$\mathcal{P}(H_{ij}|A_{ij}) = \frac{\mathcal{P}(H_{ij} \cap A_{ij})}{\mathcal{P}(A_{ij})}. \tag{3.2}$$

$P(A_{ij})$ is still the probability that an air parcel resides in a given source region. $\mathcal{P}(H_{ij})$ is the probability that the air parcel resides in this air parcel at a low altitude. Since H_{ij} is a subset of A_{ij} then,

$$H_{ij} \cap A_{ij} = H_{ij}. \tag{3.3}$$

This results in Equation 3.2 becoming,

$$\mathcal{P}(H_{ij}|A_{ij}) = \frac{\mathcal{P}(H_{ij})}{\mathcal{P}(A_{ij})}. \tag{3.4}$$

The probability that air parcels resides in a given source region is given by,

$$\mathcal{P}(A_{ij}) = \frac{n_{ij}}{N}, \tag{3.5}$$

where n_{ij} is the number of trajectory endpoints that reside in the ij^{th} cell and N is the total number of trajectory endpoints. We can define $\mathcal{P}(H_{ij})$ similarly,

$$\mathcal{P}(H_{ij}) = \frac{k_{ij}}{N}, \tag{3.6}$$

where k_{ij} is number of trajectory endpoints in a cell below some cutoff altitude and N is the total number of trajectory endpoints. Therefore, Expression 3.4 can be restated as,

$$\mathcal{P}(H_{ij}|A_{ij}) = \frac{k_{ij}}{n_{ij}}. \tag{3.7}$$

Trajpy v. 2.0 computes Expression 3.7 when a cutoff altitude is given as input. This is an altitude in which if an air parcel travels under, this air parcel will most likely become polluted. The inherent peak bias is normalized because both k_{ij} and n_{ij} are from the same source region and thus scale the same in regards to trajectory endpoint density. This effectively cancels the contribution to the ij^{th} SRIF due to the inherent peak bias. This conditional probability should be used with the original expression for a SRIF (Equation 1.5) because the ambient air concentration at the receptor may be dependent on the altitude an air parcel has as it travels through a source region as well as the total number of trajectory endpoints in a source region.

3.3 Factoring in Wet Deposition

A source region impact factor is the probability that an air parcel traveled through a given source region. If we assume that air parcels assimilate components of the surrounding air as they travel, then a source region impact factor can also be viewed as the likelihood that a measured ambient air sample originated from a given source region. Air parcels arrive at receptors at a variety of altitudes in variable weather conditions. Given that HYSPLIT provides an approximate value for the amount of rainfall at each trajectory endpoint, one may want to predict the amount of a species an air parcel will deposit to the Earth's surface in aqueous form. While the process of wet deposition is a complex one the amount of rainfall is generally positively correlated with the mass of species deposited to Earth's surface in aqueous form [2].

Therefore, we can modify the expression for Residence Time Analysis to give us the probability that a pollutant is from a given source and is deposited in aqueous form at the receptor site. This will be done by assuming that the probability that wet deposition will occur at receptor site and the probability that a given pollutant is from a given source are independent probabilities.

Since these probabilities are independent of one another they can be multiplied to get the probability that both these events occur. If we let $\mathcal{P}(A_{ij})$ denote the probability that a given pollutant is from the ij_{th} grid and $\mathcal{P}(D)$ be the probability that the particle undergoes deposition at the receptor site then the probability that a pollutant from a given source will be deposited in aqueous form at a receptor site is simply the product of these two probabilities,

$$\mathcal{P}(DAR) = \mathcal{P}(A_{ij})\mathcal{P}(D), \tag{3.8}$$

where $\mathcal{P}(DAR)$ is the probability that a particle from a given source undergoes deposition at the receptor (DAR).

$\mathcal{P}(DAR)$ can be empirically derived by taking ground water samples and mapping those to modeled rainfall. One can create a set of probabilities that stand in for $\mathcal{P}(DAR)$ that correspond to a range of measured water samples at the receptor.

3.4 Summary

An instantiation of the Residence Time Analysis model – Trajpy v. 2.0 – was created. This instantiation was based of Trajpy v. 1.0 created by Christopher Walsh. Trajpy v. 2.0 has improved ability to create source regions of various shapes. Moreover, Trajpy v. 2.0 resolves the inherent peak bias and includes factors that are altitude dependent. The next chapter discusses further work that can be done to Trajpy v. 2.0.

Conclusion

Residence Time Analysis has proven to be a useful model [6, 7, 9]. For this reason, Trajpy v. 2.0 was created in order aid those wanting to quantify source-receptor relationships. Trajpy v. 2.0 is an open source project. This means it highly likely that others (like myself) will add features to it in the future. The main areas of Trajpy v. 2.0 requiring further work are of two categories: interface improvements and validation of the model itself. Interface improvements are those that make Trajpy v. 2.0 easier to use. Model validation will make the source region impact factor computed by Trajpy v. 2.0 more accurate and informative. This chapter will elaborate upon these potential improvements.

4.1 Further Work

4.1.1 Model Validation

Trajpy v. 2.0 factored in deposition by creating a scheme to relate rainfall levels at the receptor – approximated by HYSPLIT – to measured ground level pollution. This requires empirical data. The collection of groundwater or snow samples and subsequent chemical analysis of these samples will allow one to come up with values for $\mathcal{P}(D)$ (see Chapter 3) that may allow for Trajpy v. 2.0 to have predictive capability for deposition levels of a system. Moreover, by reporting approximated rainfall levels Trajpy v. 2.0 makes it easier for the user to use this tool to make connections between rainfall and groundwater concentration of some species.

4.1.2 Improvements to Trajpy v. 2.0's Interface

A Web Interface HYSPLIT is a model that can be rendered by a web browser. The calculation of a backwards trajectory is done on servers run by the Air Resources Laboratory. In order to run the model one has to input various data. This is time consuming. It would be useful to develop a tool that could request multiple backwards trajectories at a time from HYSPLIT, thus decreasing the amount of tedious data entry required by the user.

A tool to do this is currently in development (by me), but is not yet complete at the time of writing. This tool would essentially mimic the process a human would go through when typing in the URL of HYSPLIT into a browser, choosing what input parameters one wants for their backwards trajectory and requesting a backwards trajectory from HYSPLIT.

Automating The Creation of a Series of Source Regions In Trajpy v. 2.0's current form one has to specify the vertex of each source region. If one wants to construct a grid of many source regions this becomes a tedious and time consuming task. It would be useful to automate this. This would be done by specifying the latitude and longitude of one source region vertex and the relationship one wants the other source regions to have with that point. One would also have to specify the shape and size (for example, width and height) they want all source regions to have. This tool would then essentially "draw" the source regions for the user. In practice this would be done by reporting a list of longitudes and latitudes that specify the vertex of each source region box.

4.2 Final Thoughts on Residence Time Analysis

The Residence Time Analysis model is a simple one. It does not take into account the chemical reactivities of the particles in question and relies on modeled backwards trajectories. If a chemical species is highly reactive it is likely that the species will be a chemically distinct one when it reaches the receptor. This may make tracing the source of that species more difficult. This also suggests that one has to have an understanding of what the potential sources are. Therefore, RTA should be used with chemical data and knowledge of individual chemical reactions to get a more complete picture of atmospheric processes.

Moreover while the simplicity of Residence Time Analysis greatly limits the conclusions one can make with just a source region impact factor the Residence Time Analysis model does provide one with the ability to see trends and relative impacts. Improvements to Residence Time Analysis which are made at the expense of the simplicity of the model should therefore be avoided.

Furthermore, for each trajectory endpoint HYSPLIT outputs approximations of six meteorological features:

1. The height of the terrain a trajectory endpoint is over.
2. The potential temperature of that column of air.
3. Temperature at that trajectory endpoint.
4. Rainfall levels at trajectory endpoint.
5. Relative humidity.
6. Down solar radiative flux which is a measure of the intensity of sunlight that a trajectory endpoint experiences.

All of these features play a role in the complex dynamical process that occur in Earth's atmosphere [2]. It is difficult to incorporate these into Residence Time Analysis due to the complex nature of atmospheric process. It is therefore foolish to attempt to make highly accurate conclusions by incorporating one or more of the aforementioned meteorological variables into the model to explain a complex atmospheric process. Lastly, it should be noted that backwards trajectory endpoints may not be accurate, but in practice source regions are made to be sufficiently large to account for errors in trajectory endpoints [6, 8, 13].

Appendix A

Mathematical Appendix

A.1 Set Theory

The notation of set theory is used in this work. An explanation of the notation used ensues.

A set is a collection of objects. Here I mean *objects* abstractly. These objects could be anything. The objects that make up the set are called its *elements*. For example the set S with n elements is given by,

$$S = \{x_1, ..., x_n\}. \tag{A.1}$$

Moreover, to express that x_n is in set S we say (i.e., x_n is an element of set S) ,

$$S \in x_n. \tag{A.2}$$

If we are given another set T and we observe that every element in set S is also in the set T we say S is a *subset*of T and we write,

$$S \subset x_n. \tag{A.3}$$

Moreover if we want to define all elements of a set, V, such that all elements of V are the elements of S that are greater than or equal to 3 we would denote this as,

$$V = \{x \in S : x \geq 3\}. \tag{A.4}$$

A.2 Elements of a Probabilistic Model

Source region impact factors are probabilistic models. For this reason a brief review of the theory of probability is appropriate here. Generally speaking, a probabilistic model is a mathematical description of an uncertain situation. For example, one can construct a probabilistic model for the rolling of die. The notion of set theory proves to be useful for describing probabilistic phenomenon.

Using the example of flipping a fair, two-sided coin one can define what the sample space is, what constitutes an event in this experiment and what the probability law will be for each event. The sample space, Ω is the set of all possible outcomes in an experiment. The sample space for the flipping of a coin is given by,

$$\Omega = \{heads, tails\}. \tag{A.5}$$

Another key feature of a probability model is a probability law which assigns to a set A (A is called an *event*) of possible outcomes a nonnegative number, $\mathcal{P}(A)$, that denotes the likelihood of event A from occurring. For example the probability that a given flip of a coin will land on heads (i.e., we will observe the event of the coin landing with its heads side facing up) is,

$$(heads) = 1/2. \tag{A.6}$$

Appendix B

Source Code

This chapter contains information on how to run Trajpy v. 2.0 as well as the source code. All of this information can be found on the internet at the associated GitHub repository [23]. Trajpy v. 2.0 runs on Python 2.7. It requires that a backwards trajectory be supplied by HYSPLIT. The trajectory output files should be saved with a .csv filename extension. These output files should then be put into the same directory as Trajpy v. 2.0. Moreover a separate text file is required that has the number of output files as its first line and the name and filename extension of the trajectory output files as each subsequent line. Once you have your trajectories configured correctly (they are all in the same directory, have their file names in a text file with the number of trajectory data dump files as the first item in this text file) you can use Trajpy v. 2.0 to compute a SRIF. The trajvarsv2.py module must also be in the same directory as trajpyv2.py.

```
#------------------------------------------------------------
#
# Trajpyv2.py
# Created by Christopher D. Walsh and Jonathan H. Tamsut
#------------------------------------------------------------

import csv
import trajvarsv2

#------------------------------------------------------------
#
# function readTraj:
#------------------------------------------------------------

def readTraj(filename):
    nump = 0 # This is the number of points in your trajectory.
    trajListReader=csv.reader(open(filename), dialect='excel')
    number_of_traj_files = trajListReader.next()[0]
    print ' '
    # Now lets process each data dump!
    num_files = int(number_of_traj_files)
    trajList=[]
    for count in range(1,num_files+1):
        current = trajvarsv2.trajectory()
```

```
        current.ID=count-1
        trajList.append(current)
    print "Trajectory␣objects,␣list␣initialized."
    print ''
    print "Reading␣trajectory␣files␣and␣appending␣to
␣␣␣␣Trajvars␣version␣2.0!"
    for i in range(0, num_files):
        with open(trajListReader.next()[0], 'r+') as csvfile:
            trajReader=csv.reader(csvfile, dialect='excel')
            for row in trajReader:
                new_row=row[0].split()
                if len(new_row) > 10:
                    nump += 1
                    current=trajvarsv2.trajectory()
                    current.ID=i
                    current.grid=eval(new_row[1])
                    trajList[current.ID].grid.append(current.grid)
                    current.year=eval(new_row[2])
                    trajList[current.ID].year.append(current.year)
                    current.month=eval(new_row[3])
                    trajList[current.ID].month.append(current.month)
                    current.date=eval(new_row[4])
                    trajList[current.ID].date.append(current.date)
                    current.hour=eval(new_row[5])
                    trajList[current.ID].hour.append(current.hour)
                    current.min=eval(new_row[6])
                    trajList[current.ID].min.append(current.min)
                    current.forecasthour=eval(new_row[7])
                    trajList[current.ID].forcasthour.
                    append(current.forcasthour)
                    current.age=eval(new_row[8])
                    trajList[current.ID].age.append(current.age)
                    current.lat=eval(new_row[9])
                    trajList[current.ID].lat.append(current.lat)
                    current.lon=eval(new_row[10])
                    trajList[current.ID].lon.append(current.lon)
                    current.hag=eval(new_row[11])
                    trajList[current.ID].hag.append(current.hag)
                    current.pressure=eval(new_row[12])
                    trajList[current.ID].pressure.append
                    (current.pressure)
                    current.theta=eval(new_row[13])
                    trajList[current.ID].theta.append(current.theta)
                    current.airtemp=eval(new_row[14])
                    trajList[current.ID].airtemp.
                    append(current.airtemp)
                    current.rain=eval(new_row[15])
                    trajList[current.ID].rain.append(current.rain)
                    current.mixdepth=eval(new_row[16])
                    trajList[current.ID].mixdepth.
                    append(current.mixdepth)
                    current.rh=eval(new_row[17])
                    trajList[current.ID].rh.append(current.rh)
                    current.msl=eval(new_row[18])
```

```
                    trajList[current.ID].msl.append(current.msl)
                    current.flux=eval(new_row[19])
                    trajList[current.ID].flux.append(current.flux)
    print '␣'
    print 'HYSPLIT␣paramters␣stored!'
    for x in range(0,num_files):
        trajList[x].numpoints=nump
        trajList[x].numtraj=num_files
    print ''
    print 'trajList␣populated!␣You␣are␣using␣' + str(nump) + '␣
       trajectory␣endpoints!'
    print ''
    return trajList

#------------------------------------------------------------------------
#
# procedure makeBox:
#------------------------------------------------------------------------

def makeBox(Tag,box_points,numtraj,polygon,radius):
    points_in_box = 0
    points_below_alt = 0
    rain_scav = 0
    for q in range(0,numtraj):
        srifTmp=srifMake(trajResult[q],box_points,polygon,radius)
        points_in_box += srifTmp[3]
        points_below_alt += srifTmp[1]
        rain_scav += srifTmp[4]
    cond_imp = points_below_alt/points_in_box
    impfactor= (points_in_box/75.)
    print 'SRIF:␣' + str(impfactor)
    print '␣'
    print "Conditional␣prob:␣" + str(cond_imp)
    print '␣'
    print "In␣" + str(rain_scav) + "␣source␣regions␣rainfall␣was␣above␣
       the␣cutoff␣value."
    print "␣"

#------------------------------------------------------------------------
#
# function srifMake:
#------------------------------------------------------------------------

def lat(point):       # helper function that gives you latitude
    return point[0]

def long(point):      # helper function that gives you latitude
    return point[1]

# check_intersect checks if a ray pointing due east crosses
# a line of a polygon. This polygon is a SRIF box. check_intersect is
# a helper function for the ray casting algorithm.

def check_cross(point, segstart, segend):
```

```
    plat = lat(point)
    plong = long(point)
    # Here we are checking to see if a ray due east will pass a
        horizontal line
    # created by the boundaries of an apportionment.
    cond1 = lat(segstart) < plat and lat(segend) > plat
    cond2 = lat(segstart) > plat and lat(segend) < plat
    # note that the when specifying (lat, long) you must specify
    # coordinates opposite of what is normally expressed in Cartesian
        coordinates
    # (y,x) => (lat, long)
    if (not cond1) and (not cond2):
        return False
    y = long(segstart) − long(segend) # distance in x direction
    x = lat(segend) − lat(segstart) # distance in the y direction
    c = x ∗ long(segstart) + y ∗ lat(segstart)
    x = (c−y∗plat)/x
    if x < plong:
        return True
    else:
        return False

# box_points should be a list of lists made up
# of a series of points specified as [lat, long]
# connected to one another.

def is_in_box(point,box_points,radius=None):
    # box_points must be a list of points (a list of lists) in (lat,long
        ) form
    count, k = 0, 0
    while k + 1 < len(box_points):
        if check_cross(point, box_points[k], box_points[k+1]):
            count += 1
        k += 1
    if count \% 2 != 1:
        return False
    else:
        return True

def is_in_circle(center,point,radius):
    plat = lat(point)
    plong = long(point)
    cen_lat=lat(center)
    cen_long=long(center)
    d = ((plat−cen_lat)∗∗2 + (plong−cen_long)∗∗2)∗∗(1/2)
    if d < radius:
        return True
    else:
        return False

def srifMake(Tc,box_points,polygon,radius):
    srifReturn=[0,0,0,0,0]
        srifMake outputs srifReturn which is a list of numbers.
```

```
        nump=len(Tc.lat) # 10-day back trajectory; must specify how many
             points
                         for y in range(0,nump):
        point = []
        lat=Tc.lat[y]
        lon=Tc.lon[y]
        alt=Tc.hag[y]
        rainfall=Tc.rain[y]
        point.insert(0, lat)
        point.insert(1, lon)
               if polygon is True:
            if is_in_box(point, box_points):
                if alt <1000:
                    srifReturn[0]+=1
                if alt <2000:
                    srifReturn[1]+=1
                if alt <3000:
                    srifReturn[2]+=1
                if alt <10000:
                    srifReturn[3]+=1
        else:
            if is_in_circle(point, box_points, radius):
                if alt <1000:
                    srifReturn[0]+=1
                if alt <2000:
                    srifReturn[1]+=1
                if alt <3000:
                    srifReturn[2]+=1
                if alt <10000:
                    srifReturn[3]+=1
        if rainfall > 0:   # set rainfall cut off
                srifReturn[4]+=1
    return srifReturn


#---------------------------------------------------------------
#
# Main: This is where you execute Trajpy v2.
#
#---------------------------------------------------------------

trajResult=readTraj("trajlist.txt")
trajsrif=[]
numtraj=trajResult[1].numtraj
for j in range(0,numtraj):
    tmpRcd=trajvarsv2.srifcount
    trajsrif.append(tmpRcd)


#-----------------------------------------------------------------------------
#
#trajvars.py
#-----------------------------------------------------------------------------
```

```
# input classes

class trajectory():
    def __init__(self):
        self.name=''
        self.ID=0
        self.numpoints=0
        self.numtraj=0
        self.grid=[]
        self.year=[]
        self.month=[]
        self.date=[]
        self.hour=[]
        self.min=[]
        self.forcasthour=[]
        self.age=[]
        self.lat=[]
        self.lon=[]
        self.hag=[]
        self.pressure=[]
        self.theta=[]
        self.airtemp=[]
        self.rain=[]
        self.mixdepth=[]
        self.rh=[]
        self.msl=[]
        self.flux=[]



# output classes

class srifcount():
    def __init__(self):
        self.ID=0
        self.count1km=0.
        self.count2km=0.
        self.count3km=0.
        self.countTotal=0.
```

References

[1] Daniel J. Jacob. *Introduction to Atmospheric Chemistry*. Princeton University Press. 1999.

[2] John Seinfeld and Spyros Pandis. *Atmospheric Chemistry and Physics*. Wiley-Interscience. 2006.

[3] "HYSPLIT - Hybrid Single Particle Lagrangian Integrated Trajectory Model." Air Resources Laboratory.

[4] Mark Z. Jacobson. *Fundamentals of Atmospheric Modeling*. Cambridge University Press. 2005.

[5] Stanley E. Manahan. *Environmental Chemistry*, 6th Edition. Lewis Publishers. 1994.

[6] Toby Primbs, et. al. Influence of Asian and Western US Urban Areas and Fires on Atmospheric Transport of PAHs, PCBs, and FTOHs in the Western US. Environ Sci. Technol. 2008, 42, 6385-6391.

[7] Toby Primbs. Trans-Pacific and Regional Atmospheric Transport of Anthropogenic Semivolatile Organic Comounds in the Western U.S. (Doctoral dissertation). 2007. Retrieved from Google Books.

[8] William D. Hafner and Ronald A. Hites. Potential Sources of Pesticides, PCBs, and PAHs to the Atmosphere of the Great Lakes. Environ. Sci. Technol. 2003, 37, 3764-3773.

[9] William D. Hafner and Ronald A. Hites. Effects of Wind and Air Trajectory Directions on Atmospheric Concentrations of Persistent Organic Pollutants near the Great Lakes, Environ. Sci. Technol. 2005, 39, 7817-7825.

[10] Zoë L. Fleming, Paul S. Monks, Alistair J. Manning. Review: Untangling the influence of air-mass history in interpreting observed atmospheric composition. Atmos. Research. 2012, 104, 1-39.

[11] Lowell L. Ashbaugh, et. al. A Residence Time Probability Analysis of Sulfur Concentrations at Grand Canyon National Park. Atmos. Env. 1985, 19, 1263-1270.

[12] Bilkis A. Begum, et. al. Evalution of the potential source contribution function using the 2002 Quebec forest fire episode. Atmos. Env. 2005, 39, 3719-3724.

[13] Philip K. Hopke. Recent developments in receptor modeling. J. Chemometrics. 2003, 17, 255-265.

[14] Philip K. Hopke. *Receptor Modeling in Environmental Chemistry*. John Wiley Sons. 1985.

[15] Dimitri P. Bertsekas and John N. Tsitsiklis. *Introduction to Probability*. Athena Scientific. 2008.

[16] Donald A. McQuarrie and John D. Simon. *Molecular Thermodynamics*. University Science Books. 1999.

[17] Bruce Averill and Patricia Eldredge. *Principles of General Chemistry*. Unnamed publisher. 2012.

[18] Richard L. Poirot, et. al. Comparative Application of Multiple Receptor Methods to Identify Aerosol Sources in Northern Vermont. Environ. Sci. Technol. 2001, 35, 4622-4636.

[19] Alexander V. Polissar, Philip K. Hopke and Richard L. Poirot. Atmospheric Aerosol over Vermont: Chemical Compostion and Sources. Environ. Sci. Technol. 2001, 35, 4604-4621.

[20] "Health". United States Environmental Protection Agency. 6 February 2014. Web. www.epa.gov. April 2015.

[21] J.D. Foley, A. van Dam, S.K. Feiner, J.F. Hughes. *Computer Graphics: Principles and Practice, 2nd Edition*. Addison-Wesley. 1990.

[22] John N. Westgate and Frank Wania. On the Construction, Comparison, and Variability of Airsheds for Interpreting Semivolatile Organic Compounds in Passively Sampled Air. Environ. Sci. Technol. 2011, 45, 8850-8857.

[23] Jonathan Tamsut. Trajpy-v2. GitHub. https://github.com/jtamsut/Trajpy-v2.

[24] "List of 156 Mandatory Class I Federal Areas". *epa.gov*. United States Environmental Protection Agency. April 2105.

[25] Air Resource Laboratories. HYSPLIT Trajectory FAQs. http://www.arl.noaa.gov/faq_d2.php. April 2015.

[26] "Trajectory Error." HYSPLIT. Air Resources Laboratory. Web. 22 Apr. 2015.http://www.arl.noaa.gov/documents/workshop/NAQC2007/HTML_Docs/trajerro.html.

[27] Decimal longitude and latitude. Digital image. HowTo: Identify the Spatial Reference, Projection, or Coordinate System of Data. Environmental Systems Research Institute, 23 Sept. 2014. Web. 9 Apr. 2015.

[28] Doswell III, Charles. "A Primer on Vorticity for Application in Supercells and Tornadoes." The University of Oklahoma, 2000. Web. 22 Apr. 2015.

[29] HYSPLIT Tutorial. Air Resources Laboratory. June 2012. http://www.arl.noaa.gov/documents/workshop/Spring2012/HYSPLIT_Tutorial.pdf.

[30] Coriolis Effect. Digital image. University of Oregon. Web. http://abyss.uoregon.edu/~js/glossary/coriolis_effect.html.