**Richard C. Alt II**
**Submission for E= mc$^2$ Journal**
**For Dr. Mazziotti**

# Section I

Every winter, hundreds of people die needlessly in snowstorms, often unwarned and unable to comprehend the dangers they will face; at the same time, millions of dollars in government and private resources are lost when unforecasted storms suddenly develop or when anticipated major events fail to materialize. Seasonal snowfall forecasting algorithms with short and long term accuracy can eliminate or minimize each of these societal concerns. With sufficient accuracy and scientific rigor, such forecasts can be used by state-funded and private agencies for advanced planning in this age of changing climate. By giving scientists and policymakers a tool to uncover anomalous winters well in advance, my research, by extension, serves to benefit all elements of regional society. Given the universality of such methods, algorithms could be developed from climatology with the potential to serve the needs of any region of the world with access to sufficient climate data and weather modeling programs. I saw potential to apply my knowledge as a young leader in this scientific field that has traditionally been overlooked.

My experiment in seasonal snowfall forecasting aimed to combine three schools of meteorological thinking and break new ground in accurate long term winter forecasting. The experiment I have laid out was designed precisely so that it can be repeated at any location on earth with sufficient climate data. I consider my Intel project to be the ultimate expression of my lifelong passion for weather, beginning as early as my 6th birthday when the Blizzard of 1996 set new snowfall records across the East Coast. In March of 2001 came the next milestone of my still young career: another approaching snowstorm promised to equal or exceed the astonishing accumulations of January, 1996. Two feet of snowfall was forecasted, up to three feet even in elevated locations. Had this storm occurred as predicted, it would have exceeded and even doubled in a single event the historical snowfall averages for an entire season in my part of Virginia. However regrettably, this storm decided instead to move 300 miles to the northwest, leaving my town with a mere two minutes of snow showers, rather than two days. The promises cascading out of my television speakers were drowned in a torrential rainstorm in place of the expected blizzard. Naturally this was quite disappointing for me as a child, especially coupled with an entirely unexpected need to attend school the next morning. In spite of the short term disappointment, I was confronted with a much more profound inspiration. Never mind that the forecasts had gone wrong; I wanted to know why, and I wanted to offer my community a new and innovative type of forecast that would provide greater accuracy the next time around.

My research did not require access to any advanced laboratory facilities or tools, and a similar project could be pursued by any teenager with sufficient scientific prowess, mathematical knowledge, and most importantly, dedicated effort. My research was conducted primarily using a home computer and graphing calculator. The verification for my methods took place in the field. I measured the yearly snowfall on my family's property using standardized methods to record my actual results. I then compared these results against the forecasts I had generated mathematically. My research is remarkable both in its scope and its simplicity. I have taken advanced mathematics courses throughout high school, but my project did not require any techniques that could not be learned within a typical advanced high school curriculum. Indeed, I began the project during my freshman year, and only minor research into statistical methods was needed to compile the final product over the next three years. I took higher level courses, including Calculus and Statistics at the Advanced Placement level, but I needed only arithmetic, algebra, and an understanding of statistical analysis and modeling to complete my project.

While working on my project, I did develop a much greater appreciation for the amazing feats that can be accomplished when the scientific method is combined with detailed mathematical methods. It was far more rewarding to be personally involved in new and innovative research in atmospheric science than merely to be taught science and mathematics in a classroom setting. Modern meteorology owes much of its existence to the development of atmospheric modeling algorithms, and I am proud to say that I have done my own small part to expand human knowledge about the workings of our planet's atmosphere. During certain stages of my project, a specific problem would appear so difficult to resolve that it might compromise the success of the entire forecasting system, but I was careful not to let this happen. Support for scientific work is easy to find if you know where to look.

Over the course of my project, various difficulties arose, which I overcame with the aid of a wealth of local resources. My first problem was inherent in the research process: locating the data. The National Climatic Data Center online records were exemplary in detail, but several key storms were missing from the records. On my parents' advice, I turned to the newspaper archives at my local library and, sure enough, my missing storms could be identified on the front pages. A second obstacle materialized during 2004-05, the first winter surveyed, when virtually no snowfall occurred during the first half of the winter. By the end of January, no single storm had amounted to more than two inches total, and not one of my forecasts had indicated significantly below normal snowfall for the season as a whole. I did not let this discourage me; rather, I returned to my climate records and found just how anomalous the lack of snow was, in contrast to the statistically very snowy El Nino climatology I had used in developing two of the three sub-forecasts for that year. My earth science teacher suggested that such a remarkable dearth of snow was notable on its own, and perhaps an underlying symptom of global climate change. When a five inch storm finally struck in late February of 2005, my concerns were assuaged, as the relative accuracy of at least one of my methods was now ensured for its first winter of testing. My psychological determination to continue with the long term applications of my project was quickly renewed.

Determination is key to the success of a research project, but it is also prudent to focus on improving real world conditions. This can be a critical factor in arranging the support you will need if a portion of your project does not go as planned. I did not set out to do science for its own sake, but rather to use my knowledge of science and mathematics to actively make a difference in my local community. I felt it as my duty to designate the beneficiaries of my research project, and I found many locals teeming with interest in what I set out to accomplish. The best science ought to be in the public interest, even in a non-traditional way. I learned many things from my research experience, but this is one of the most significant. Though these goals may seem very farfetched even for the most talented of high school students, they are nonetheless worthy aspirations to pursue.

Finally, in order for a research project to be successful, it must also be engaging to the individual. If you are not having fun, then you need to consider doing something entirely different. Of course there will be downturns along the way—a late night computer failure that destroys 75% of your research paper-in-progress, an anomaly that threatens to discredit your entire experiment, or even an accidental lab fire that destroys valuable equipment. Despite these and other difficulties, research should be on the whole, an enjoyable experience. To find a solution, you must first find the motivation to succeed. This can only come from a genuine interest in your chosen topic, and only you can make the firm decision to continue forward.

## Section II

My project which I submitted to the Intel Science Talent Search, titled Developing and Testing a New Seasonal Snowfall Forecasting Method, uses the scientific method to compare and analyze seasonal forecasting methods to determine their effectiveness in predicting total winter snowfall in my hometown of Fredericksburg, Virginia. In this project, three different methods of seasonal forecasting are constructed and tested to find their predictive ability in forecasting total seasonal snowfall during the winter seasons of 2004-05, 2005-06, and 2006-07. Using the results of these trials, a composite forecast was then assembled to determine how best to combine the three sub-methods to produce a forecast with the highest overall accuracy. The three forecasting methods chosen were:

1. Historical analog years taken from a historical database of summer and fall temperature and precipitation trends
2. Historical analog years based on four global climate indices

3. Human analysis of temperature and precipitation forecasts from three major computer climate ensembles for the December-February three month period of 2005 and the November-January, December-February, and January-March periods of 2006 and 2007.

The results from each of the methods were tabulated (numerically, for the first two methods, and in the context of normal conditions for the model analysis method) and compared against actual snowfall observations I recorded during each winter season. The standard for comparison was the average total seasonal snowfall in inches for Fredericksburg, Virginia. Two averages were selected, one using the standard 30 year base period from 1971-2000, and the second using a 50 year base period from 1955 to 2004. The 50 year mean snowfall for a single winter season in Fredericksburg, Virginia is approximately 17.4 inches, and the 30 year mean is 15.7 inches. A 30 year record is the standard used in the meteorological community to establish climate for a given region. The 50 year average was also included to increase the sample size and account for variation in seasonal snowfall due to chance factors, rather than legitimate climate influences. All three seasons studied were near the long term 30 and 50 year averages in total seasonal snowfall (totaling 15.0, 17.5, and 11.1 inches respectively for 2004-05, 2005-06, and 2006-07). I hypothesized that my statistical methods would be most accurate in predicting snowfall during seasons that did not deviate strongly from the historical pattern of the data, so the results I observed during these years should provide an effective representation of the three methods' overall effectiveness.

My project combines these three well known forecasting methods to generate a seasonal snowfall outlook that has the potential to aid all sectors of local and regional society. To accomplish this goal, I analyzed temperature and precipitation trends from previous years, correlated values for key atmospheric indices, and developed a third forecast method by studying climate model output and making a scientific judgment about expected snowfall with each solution. All three methods are objective, scientific, and easily repeatable using data from any region of the country and similar model reading skills. In preparing the three forecasts, I synthesized and expanded upon previous accomplishments to explore new algorithms of seasonal snowfall forecasting. During October of 2004, I derived three primary analog years for each of the first two methods, using historical climate data I had compiled from the two nearest stations owned and operated by the Southeast Regional Climate Center. The method for selecting these analog years was complex, but the choices were intended to be both mathematical and logical. I considered the similarities in observed conditions between the current summer and fall seasons and the conditions observed during previous years.

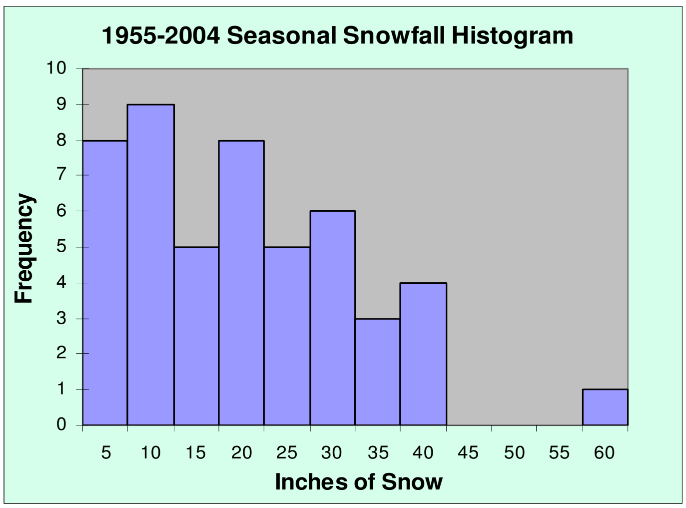
For comparison to historical climate, the first of my methods, I focused on two key variables (the same two I would later use to establish forecasts based on model analysis): temperature and precipitation. Attention was paid not only to the observed values in any given month, but also to the trend observed over multiple consecutive months leading up to a past winter. I made the decision early on in the forecasting process to categorically limit the years that could be selected as analogs using the temperature and precipitation based method to those years in which the same phase of the El Nino Southern Oscillation was observed as in the current year.

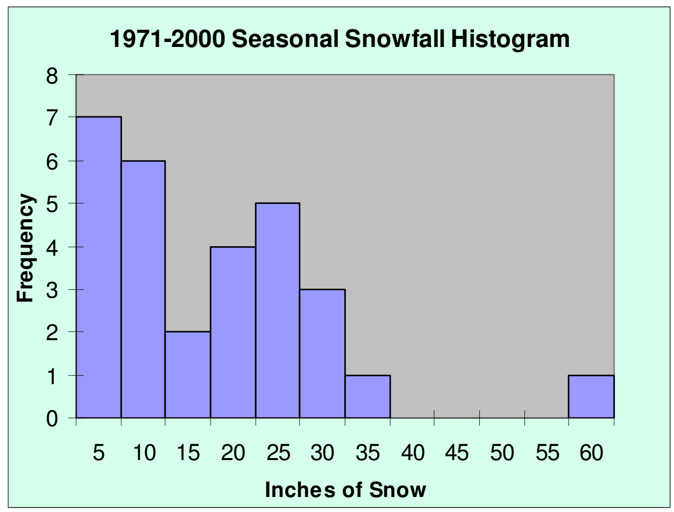
I never had the opportunity to calculate the impact that this decision had on the accuracy of my forecasts, but, looking back, I am concerned that I allowed one factor to influence two of the three forecasting methods, meaning that they were not truly independent of each other—which would have been ideal to ensure a high standard of scientific credibility. After selecting the three past years that I determined were most similar to the conditions observed during the present summer and fall, I averaged the total amount of snowfall, in inches observed during the three winters that followed those historical summer and fall seasons. During the second two winters, this method was expanded to include years prior to 1954 for which accurate climate data was available, in order to increase the sample size from which to choose analog years.

In brief, the second method was based on my analysis of the values of various atmospheric indices known to impact climate on the regional and presumably, local level. I used the same format as with the first method, selecting past years that correlated well with the current conditions of several global climate phenomena during the summer and fall months. For this method, I considered four significant global climate records: The El Nino Southern Oscillation, the Quasi-Biennial Oscillation in the Central Pacific (involving the wind direction high in the atmosphere), the 10.7 cm Solar Flux (a measured number released by government agencies—from these numbers, a seasonal average can be calculated), and, finally, the total Northern Hemisphere Snow Cover in millions of square miles. Again, the three most similar seasons in the historical record were selected as analog years to model the upcoming winter. The mean snowfall of these three years was taken as the snowfall forecast for the climate indices method.

To establish a consensus forecast for the final method, based on the output of computer models, I collected and analyzed model forecasts as they became available during the fall months and used known correlations between temperature, precipitation, and snowfall to synthesize a forecast for total snowfall during the 2004-05 season. The end result in 2004-05 was near normal snowfall, with the method of model analysis appearing dramatically more accurate than the two climatology based methods, which had both somewhat over-predicted total snowfall for that season.

Because I could not generate a numerical forecast with the model analysis approach for a specific amount or range of amounts for a given season's snowfall, I needed to use a categorical approach. I could not put a number on the expected snowfall, but I could put seasons into groups based on how much snowfall occurred and use my model reading skills to predict which group the coming winter would fall into. I grouped the model analysis based forecasts into one of three categories: near normal snowfall, above normal snowfall, and below normal snowfall. I offered the quartiles of the 30 and 50 year distributions of total seasonal snowfall as an approximation of the near, above, or below normal categories. Any amount falling between the $1^{st}$ and $3^{rd}$ quartiles of the distribution would be considered "near normal snowfall." "Above normal snowfall" was defined as any winter in which the final snowfall total exceeds the $3^{rd}$ quartile in at least one of the two periods of record. Finally, any winter in which the total observed snowfall fell below the $1^{st}$ quartile of either distribution was labeled as a "below normal snowfall" season. These five number summaries of the snowfall distributions, as they are known in statistics, played a critical role in my forecasting methods and can be expressed more effectively with a graph, such as a histogram or a boxplot. The histograms and boxplots included below provide a visual picture of the snowfall data for my hometown of Fredericksburg. The histogram bars measure the number of years in the 30 or 50 year record that fell into the given range of snowfall, and the boxplot displays the data in terms of the median (middle amount of snowfall), maximum (highest amount of snowfall), 1st Quartile (the amount of snowfall that is higher than 25% of all snowfall totals in the dataset but lower than the other 75%), and 3rd Quartile (the amount of snowfall that is higher than 75% of all snowfall totals in the dataset but lower than the highest 25%)




**Figures 1 and 2: Histograms illustrating the distribution of seasonal snowfall in Fredericksburg, Virginia over the 50 and 30 year data periods, respectively.**

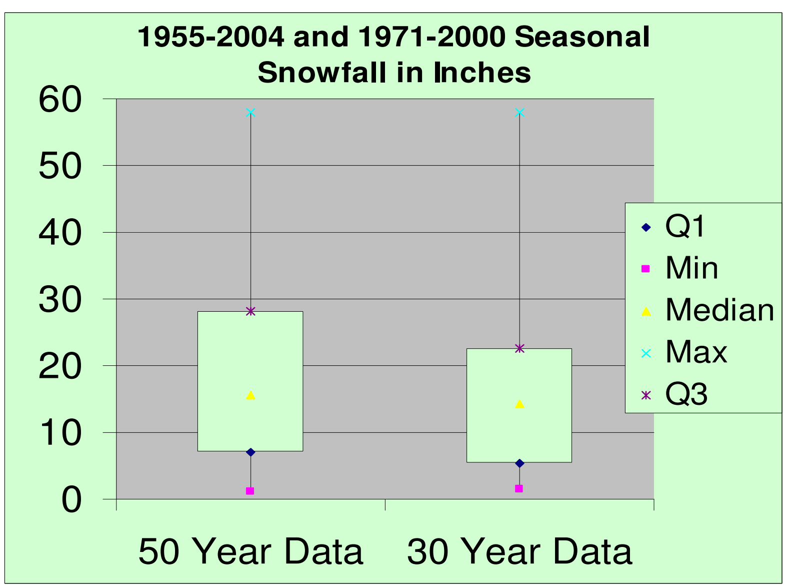


**Figure 3: Side-by-Side Box Plots of the 50 and 30 year snowfall data for Fredericksburg, Virginia**

As this categorical system offers only two degrees of freedom (meaning that there are only two other alternatives possible if one category is predicted), its accuracy can be influenced by short term random variation and its skill is limited because each category includes a wide range of possible snowfall totals. For this reason, I made several cautionary notes to be considered with the broad application of my model analysis method. Barring these limitations, I was quite satisfied with the single season performance of the computer model analysis method.

After achieving moderate success during my first year, I approached the project again during the summer of 2007 and repeated my methods in a re-analysis format for past conditions observed during the winters of 2005-06 and 2006-07. The model analysis method remained the most effective forecast system over the long term, but its utility was limited due to the required categorical, rather than numerical interpretation of the snowfall data. The two analog based methods improved significantly in their overall accuracy during the second two seasons, but there is not yet enough evidence to attribute this phenomenon to a specific cause.

I concluded my research report with an analysis of the uncertainty inherent in my forecasts, as any aspiring scientist should. I focused on the uncertainty that developed from combining categorical and numerical variables in the same forecast system, the small geographic range I covered (focusing on a specific region of Virginia), and on the possible hidden impact of an underlying upward trend in global temperature that could be influencing snowfall totals at the local level. I took various steps to reduce uncertainty, including using both a 30 and 50 year snowfall average to test for noticeably different results. Expanding the project to cover three consecutive winters reduced the role of chance in the outcome of the forecasts (although a sample of several decades' worth of winters would be ideal). The question of independence and the inclusion of the El Nino phenomenon in two separate methods also came up.

Would these methods perform equally well during an extremely snowy winter or during a winter with little or no snowfall? Was there a subconscious impact of human bias during my reanalysis of the second two years? The inherent psychological impact of "forecasting the past" raises concerns as well, because the real world results are already known to the forecaster. Interestingly, seasonal snowfall was significantly higher during single decade of the 1960's, so much higher that the addition of this decade almost entirely explains the difference between the 30 and 50 year mean snowfall figures. Was unusually high snowfall during the earlier period merely an anomaly or is there a long term downward trend in seasonal snowfall—which would be quite consistent with a warming climate? I do not have these answers. I can aspire to find them, but some solutions will always remain unknown. The goal of science, then, must be to quantify and then to minimize these chance factors and observational biases. I must first acknowledge the uncertainty of my project, then work actively to limit its long term impact.

A snowfall forecasting method with the power to distinguish between harsh and mild winters several months in advance has great impact and broad implications for all sectors of the local community, such as transportation services, emergency planning agencies, and local businesses affected by winter storms. By testing and refining forecasts for future winters, my method can be expanded and built upon in the scientific community, with increasing skill and scientific rigor for each repeated application as the sample size of winters rises with a simultaneous increase in geographic scope of the individual forecasts. My method develops a new idea in snowfall forecasting that provides a base for scientific research to build upon and has great potential to solve several of the logistical problems commonly associated with snowstorms.

Advancements in seasonal forecasting stand to benefit community resources and protect human life, on local, national, and even international levels. If the methods I have devised can be expanded and widely applied, virtually every segment of society stands to benefit. Such diverse sectors as emergency response, government financial planning, transportation industries, and utility providers can benefit greatly from the nearly limitless potential of seasonal snowfall forecasting, using easily replicable, climate-derived methods.