A Data Resource for Cloud Cover Simulations
-MSc Dissertation Proposal-

Graham Nicholas Sortino
University of Edinburgh
s0561240@sms.ed.ac.uk

March 24, 2006

Abstract:

This Document outlines my MSc Dissertation Project Proposal, the purpose of which is to build a climate data resource that will assist scientists John Latham and Stephen Salter in their work to decrease the temperature increase caused by global warming. The data resource, once constructed will be used to answer a number of questions related to their work such as where are the most optimal locations to use their technology and what are the short term after effects of its use.

I. Introduction

The Purpose of this MSc Dissertation project is to build a climate data resource that is global in size. This will then be used by Atmospheric Physicist John Latham with the assistance of Engineer Stephen Salter to answer a series of questions related to their work on a proposal to counteract the effects of global warming. However, before discussing the details of this project further it is necessary to understand why it is important and how it fits into Latham & Salter's work.

In 1977 Climatologist Sean Twomey published his influential paper on cloud physics entitled "Influence of Pollution on the Short-Wave Albedo of Clouds." In it he showed that a cloud's albedo (ie. its reflectivity) has a direct correlation to the number of its cloud condensation nuclei (or CCN) which are small particles that clouds form around [20].

In order to comprehend the significance of these findings it is important to understand that clouds have a net cooling effect on the atmosphere. Or to put it another way: without clouds the Earth would be warmer. This is because overall clouds block a higher percentage of incoming energy from the sun (known as shortwave radiation) then the long-wave radiation emitted from the Earth, which they trap [10]. This is an over simplification of an immensely complex system [12], which taken all together is known as the Earth Radiation Budget. Furthermore, the factor which most directly contributes to a clouds ability to block shortwave radiation and thus cool the Earth is its albedo. Therefore, the significance of Twomey's research is that by artificially increasing the amount of cloud condensation nuclei (henceforth referred to as CCN) in the atmosphere scientists theoretically have a mechanism to reduce the temperature increase caused by global warming. This phenomenon is known as the "Twomey Effect."

Since Twomey's discovery climatologists such as [5, 6, 7, 8, 9, 12, 13, 14, 4, 11] have either been trying to find practical methods of inducing CCN into the atmosphere [4, 11] or attempting to produce climate models that could be used to predict the aftereffects of such an experiment [5, 6, 7, 8, 9, 12, 13, 14]. To the best of my knowledge no one has yet to be wholly successful in either endeavor. The prior is a challenge because of ethical and technological reason while those faced by the latter are mostly technical.

Recently, however, National Center for Atmospheric Research, Atmospheric
Physicist John Latham along with the assistance of University of Edinburgh Engineer Stephen Salter have begun putting a proposal together [4, 11] to address this issue. Their plan is to develop an ocean vessel that is capable of seeding the marine boundary layer, which is the lowest part of the atmosphere, with CCN. Their vessel works by spraying the tiny salt particles found in sea water, which are an ideal CCN into the lower atmosphere thus increasing cloud albedo.

The plan faces many of the same ethical and technological challenges as previous proposals. One such technological challenge is where are the "most ideal" places to use these spray vessels? "Most ideal" can be broadly defined as locations where there is low initial CCN, albedo, and a large amount of incoming short-wave radiation. However, as will be seen there are many other factors that must be taken into consideration as well. Thus the task of answering this question is highly non-trivial.

Given this context, the objective of my MSc Dissertation Project is to build a data resource that will help Salter, Latham, and possibly others answer these types of questions.

The rest of this proposal is broken down as follows. Section 2, further defines the purpose of this project, while section 3 discusses previous work in the field. Sections 4 and 5 address the methods and techniques respectively that I propose to answer these questions. Outputs of the project are laid forth in section 6 and finally in section 7 the proposal concludes with my anticipated work plan.

II. Purpose

As was mentioned in the introduction, the purpose of this project is to build a global climate data resource encompassing a number of relevant variables. Once it is completed it will be made available to answer climate questions such as: "What are the ideal locations to get the most useful Twomey effect? And, how much spraying of cloud condensation nuclei would achieve the greatest Twomey effect at the identified sites?" These questions are important because they address critical issues in Latham & Salter's proposal. It should also be noted that the utility of this project does not end once these issues are resolved. This is because scientists can also use this resource as a tool to answer a number of analogous questions.

Based on data from organizations such as NOAA's Climate Diagnostics Center [23] and models created by the Earth Radiation Budget Experiment (ERBE) [33], it can be proved that during the summer months the majority of incoming short-wave radiation is focused in the northern hemisphere, while in the winter the opposite is true. Therefore I predict that the most ideal locations will follow this formula (ie. summer in the north and winter in the south). Additionally, I expect that areas closest to the poles would be most ideal because incoming shortwave radiation is highest there with values reaching up-to 400 watts/meter^2 [33]. These are both examples of the types of hypotheses I would like this project to support.

It is important to stress that this project will not be used to determine what the long-term effects of increasing cloud condensation nuclei on the environment would be. Answering this requires the expert knowledge of climate scientists and the use of extremely complex predictive models that require massive amounts of computing power [15]. Instead it tries only to answer the more technical questions of what are the most suitable spray locations to use the vessels, what are the short term after effects on cloud albedo & CCN, and other related questions.

III. Background

In John Latham's 2002 paper [4], he discusses a number of challenges facing his proposal foremost of which are the technological considerations of how to artificially introduce CCN into the atmosphere and predicting potential changes in the Earth's climate. The prior is being addressed by Salter [11] and the latter is being investigated using climate models. Models such as [21] indicate that rainfall is sensitive to anthropogenic aerosols (ie. those introduced by humans) into the atmosphere. Additionally, Latham predicts that increasing CCN over the sea would skew the temperature contrast between the land and ocean.

Perhaps, due to these challenges much climate research has focused either on understanding small climate subsystems such as cloud microphysics [7] or large-scale predictive models. This project fits somewhere in the middle because it is not looking at the medium to long-term after effects on the climate and similarly it is not investigating only one sub-system such as cloud microphysics. Because of this not much work has actually been put into creating a resource to identify potential spray locations.

While not much has been done in this specific area there are a number of organizations dedicated to gathering and studying climate data [23, 24, 25, 26, 27, 28, 29, 33, 34]. These groups normalize and transform the data into a format that can then be used by other scientists. This is very helpful as their datasets will be utilized by this project. It should be noted, however, that all the variables required for it cannot be found within one single organization. For example, ISCCP [26] has information on cloud properties, while PODAC [24] is interested in ocean surface data. Thus while much work has been done it appears that no single group/scientist has tried to answer the types of questions Latham and Salter are interested in. This is further evidenced by the fact that no single organization possesses datasets encompassing all of the variables required for this project.

IV. Methods

As alluded to in the introduction the "most ideal" spray locations are defined by many factors including:

- Cloud Condensation Nuclei (CCN)
- Albedo (Cloud Reflectivity)
- Cloud Droplet Number Concentration
- Incoming Shortwave Solar Radiation
- Outgoing Longwave Radiation
- Height's of cloud base and top
- Amplitude and spectral properties of the wind in the bottom 25 meters.
- Cirrus (high) clouds *(Note: They have a net-warming effect and their interference should be minimized if possible)*
- Wind speed & direction at sea level
- Thickness of the marine boundary layer

Each factor must be properly weighted for importance. For example, initial cloud albedo should have a higher precedence than heights of cloud base and top; however, if a cloud base is too high it may make seeding impossible. Thus careful consideration must be taken in assigning the correct weights to each variable.

Additionally, because there are so many variables that must be taken into account it will most likely be impossible to gather all of them from one data source. Thus a major challenge is data integration [17].

Currently I am evaluating a number of different data sets for suitability in this project including but not limited to those provided by:

- NOAA-CIRES Climate Diagnostics Center
- Physical Oceanography DAAC
- The Greenhouse Effect Detection Experiment (NASA-GEDEX)
- The British Atmospheric Data Center
- International Satellite Cloud Climatology Project (ISSCP)
- ARM - Atmospheric Radiation Measurement
- National Environmental Satellite, Data, and Information Service (NESDIS)
- Climate & Global Dynamics
- Atmospheric Data Center
- Earth Observing System (EOS)
- Television Infrared Observation Satellite (TIROS)
- Upper Atmosphere Research Satellite
- Earth Radiation Budget Experiment (ERBE)

Each source is stored in its own domain specific (and oftentimes arcane) data format. For example, data sets provided by the International Satellite Cloud and Climatology Project (ISCCP) [26] are stored in a binary format that are accessed using a FORTRAN program

The vast majority of data to be used comes from satellites with global coverage. Each satellite may have taken measurements calibrated in a slightly different manner, at different times, and locations. Thus another challenge is that of data cleaning/transformation to convert the disparate sources into a uniform format. In order to accomplish this I am investigating the use of data cleaning approaches such as those proposed by [18, 19, 22] each of which handles a different class of problems. Finally, it is important that scientists accessing this resource know the provenance or lineage [16] of the results they obtain in order to properly verify its accuracy.

Once the data warehouse is fully constructed a set of experiments will be designed and conducted in order to test how well it meets the evaluation criteria listed in the following section.

Before embarking on the tasks listed above I must invest the effort to become as much of an expert as possible in climate data and measurements. To accomplish this I have been and will continue reading climate data literature and discuss my proposed methodologies with domain experts.

V. Evaluation

This project will be evaluated by how effectively it can answer queries based on the large number of variables it must take into account. To accomplish this, locations which it identifies as “most ideal” will need to be cross-checked using other climate sources such as those offered by NOAA’s CDC [23]. This will provide a general accuracy metric, however, the vast amount of domain knowledge required to fully evaluate its effectiveness is beyond my capabilities. Therefore, I will enlist the assistance of a domain expert. This is important because in order for this project to be successful the data gathered must be defendable. That is this project should not be considered a success if climate experts are able to find holes in the data sources I’ve chosen due to incompatibility or because it produces results that cannot be confirmed or are contradicted by other sources. The role of domain expert will most likely be fulfilled byLatham. However, if he is unavailable there exists the option of enlisting a member of the British Atmospheric Data Center or another University of Edinburgh colleague.

The domain expert’s role in evaluating this project is two fold. First, after choosing the data sources it is important that they be evaluated for suitability of integration by the domain expert. This will help ensure that the data used in the project is adequately defendable and it will also contribute to more accurate findings. The second role is in evaluating the validity of the outputs produced by the data warehouse. This involves taking a sample of results and seeing if the most suitable locations really do exhibit the types of conditions predicted. If both of these questions can be answered in the affirmative this would indicate that the project has met its defined objectives. However, if either or both cannot then they must be re-addressed as required. This could involve choosing a different set of data sources and or changing the weights of each of the variables/functions.

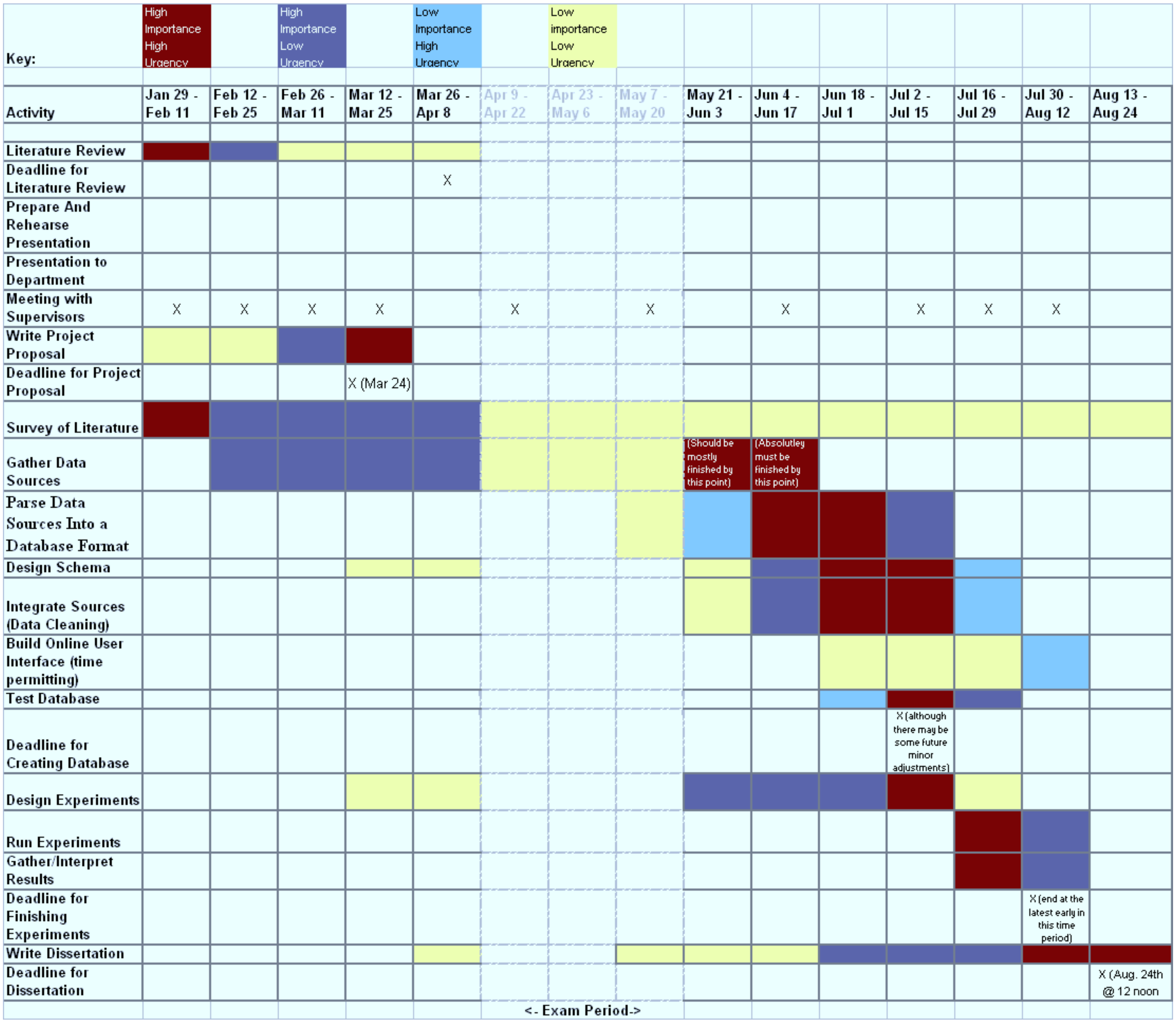
VI. Outputs

The output of this project will be a data resource which will be used by Latham, Salter, and possibly others as they continue working on their project to build a spray vessel capable of increasing the reflectivity of clouds. Additionally, time permitting I will: build a web interface to the data resource so that other scientists could use it for their research and begin trying to answer some of the questions posed by Latham and Salter.

VII. Work Plan

The final section of this proposal identifies the anticipated workflow of my project along with expected completion dates. Each phase will have some overlap from other phases. For example, writing will take place during all phases although for illustrative purposes they are each listed as disjoint. I have given extensive consideration to the amount of work required to complete this project and have determined that although difficult this project can be completed in the allotted time. This can be seen in the summary table directly below or in the more comprehensive Gantt chart listed below that.

Milestone	**Estimated Time of Completion**
-Analyze the project and do literary review	-March 1st
-Define the problem and write research proposal	-March 24th
-Gather data sources	-June 10th
-Design schema	-July 15th
-Integrate sources	-July 15th
-Test database	-July 20th
-Design experiment	-July 20th
-Run experiments	-August 1st
-Gather results	-August 7th
-Complete dissertation paper	-August 24th



Gannt Chart modeling my predictive work flow

VIII. Bibliography

[1] J. Caron. NetCDF-Java (version 2.2) User's Manual. UNIDATA, Aug. 2004. http://www.unidata.ucar.edu/software/netcdf/docs/.

[2] R. Rew, G. Davis, S. Emmerson, and H. Davies. The NetCDF Users' Guide. UNIDATA, Dec. 2004. http://www.unidata.ucar.edu/software/netcdf/docs/.

[3] P. Daum and Y. Liu. Anthropogenic Aerosols: A Clearer Understanding. Brookhaven National Laboratory (BNL), Pub Date Unknown. http://www.arm.gov/science/research/show.php?id=R00042.

[4] J. Latham. Amelioration of Global Warming by Controlled Enhancement of the Albedo and Longevity of Low-Level Maritime Clouds. Royal Meteorological Society, 2002.

[5] Q. Han, W. B. Rossow, J. Chou, and R. M. Welch. Global Survey of the Relationships of Cloud Albedo and Liquid Water Path with Droplet size Using ISCCP

[6] Q. Han and R. Welch. The Effect of Ecosystems on Cloud Microphysics and Aerosol Distribution (Annual Progress Report FY 97/98). National Institute for Global Environmental Change, 1998. http://nigec.ucdavis.edu/publications/ar/annual97/greatplains/project83.html.

[7] Q. Han, W. B. Rossow and A. A. Lacis. Near-Global Survey of Effective Droplet Radii in Liquid Water Clouds Using ISCCP Data. Journal of Climate, April 1994.

[8] J. Haywood and O. Boucher. Estimates of the Direct and Indirect Radiative Forcing Due to Tropospheric Aerosols: A Review. Review of Geophysics, 38, 4: 513-543 Nov. 2000.

[9] W. Cotton. Development of a Radiative Cloud Parameterization Scheme of Stratocumulus and Stratus Clouds Which Includes the Impact of CCN on Cloud Albedo *(Annual Report)*. National Office Progress Reports, Jul. 1993 - Jul 1994

[10] NASA Facts On Line. Clouds and the Energy Cycle. http://eospso.gsfc.nasa.gov/ftp_docs/Clouds.pdf.

[11] S. Salter. Sea-Going Hardware for the Implementation of the Cloud Albedo Control Method for the Reduction of Global Warming. 2006 (This has not been published)

[12] S. Schwartz. The Whitehouse Effect - Shortwave Radiative Forcing of Climate By Anthropogenic Aerosols: an Overview. J. Aerosol Sci. Vol. 27, No. 3: 359-382 1996

[13] S. Schwartz. Cloud Droplet Nucleation and its Connection to Aerosol Properties. Int'l Conf. Nucleation and Atmospheric Aerosols: 770-779, 1996.

[14] Q. Han, W. Rossow, J. Zeng, and R. Welch. Three Different Behaviors of Liquid Water Path of Water Clouds in Aerosol-Cloud Interactions. Journal of the Atmospheric Sciences Vol. 59 Is. 3: 726-735, 2002.

[15] G. H. Athens. IT Struggles with Climate Change. Computer World, Feb. 2006. http://computerworld.com/hardwaretopics/hardware/story/0,10801,108314,00.html.

[16] P. Buneman, S. Khanna, and W-C. Tan. Why and Where: A Characterization of Data Provenance. Lecture Notes in Computer Science Vol. 1973: 316-330, 2001.

[17] E. Rahm and P. A. Bernstein. A Survey of Approaches to Automatic Schema Matching. The VLDB Journal 10: 334-350 (2001).

[18] V. Raman and J. M. Hellerstein. Potter's Wheel: An Interactive Data Cleaning System. 27th VLDB Confrence: 2001.

[19] P. Bohannon, W. Fan, M. Flaster, and R. Rastogi. A Cost-Based Model and Effective Heuristic for Repairing Constraints by Value Modification. SIGMOD 2005: June 2005.

[20] S. Twomey. Influence of Pollution on the Short-Wave Albedo of Clouds. Journal of Atmospheric Science: 34, 1149-1152, 1977.

[21] L. Rotstayn, D. Ryan, and B. F. Penner. Precipitation Changes in a GCM Resulting From the Indirect Effects of Anthropogenic Aerosols. Geophys. Res. Lett., 27, 3045-3048, 2000.

[22] H. Gaallhardas, D. Florescu, D. Shasha, E. Simon, and C. Saita. AJAX: An Exstensible Data Cleaning Tool. In SIGMOD, 2001

[23] NOAA-CIRES Climate Diagnostics Center. http://www.cdc.noaa.gov/.

[24] Physical Oceanography DAAC. http://podaac.jpl.nasa.gov/.

[25] The Greenhouse Effect Detection Experiment (GEDEX). Provided By: The British Atmospheric Data Center. http://badc.nerc.ac.uk/data/gedex/.

[26] International Satellite Cloud Climatology Project (ISCCP). http://isccp.giss.nasa.gov/index.html.

[27] Atmospheric Radiation Measurement (ARM). http://www.arm.gov.

[28] NOAA Satellite and Information Service - National Environmental Satellite, Data, and Information Service (NESDIS). http://www.nesdis.noaa.gov/.

[29[Climate & Global Dynamics. http://www.cgd.ucar.edu/.

[30] Earth Observing System (EOS). http://eospso.gsfc.nasa.gov/.

[31] Television Infrared Observation Satellite (TIROS). http://www.earth.nasa.gov/history/tiros/tiros.html.

[32] Upper Atmosphere Research Satellite. http://umpgal.gsfc.nasa.gov/.

[33] Earth Radiation Budget Experiment (ERBE). http://asd-www.larc.nasa.gov/erbe/ASDerbe.html.

[34] Atmospheric Science Data Center. http://eosweb.larc.nasa.gov/.