Characterizing Temperature and Water Vapor of the Environment using the Standardized Atmosphere Generator (SAG) Empirical Model

Raphael Panfili, Hoang Dothe, John Gruninger, and James Duff

Spectral Sciences, Inc., 4 Fourth Avenue, Burlington, Massachusetts

ABSTRACT
Optimal interpretation of remote sensing imagery requires characterizing the atmospheric composition between a sensor and the area it is observing. Timely estimates of atmospheric temperature, water vapor, and other constituents from the ground to the edge of the space environment are not always readily available. In those cases, we must supplement our knowledge of the atmosphere's composition to fill in any gaps in knowledge and empirical models of the atmosphere are useful tools for this purpose. The Standardized Atmosphere Generator (SAG) was constructed is one such empirical. It has been designed to allow all the major known, systematic variability in the atmosphere and may be used to generate atmospheric profile from the ground to 300 km consistent with user-specified temporal, geophysical, and geographical information. Output provides reasonable estimates for temperature, pressure, and densities of atmospheric constituents and can be directly incorporated into radiative transfer forward models or retrieval algorithms. SAG draws upon a number of existing empirical atmospheric models and ensures consistency of output between them. It can be used either as a stand-alone interactive program or scripted for batch execution and assist in determining atmospheric attenuation, refraction, scattering, chemical kinetic temperature profiles, and a host of other naturally occurring processes. Here, we will discuss the capabilities and performance of the SAG model for a variety of applications including its interactive and batch processing use. We will also demonstrate the physical realism of SAG through a small number of relevant use cases.

Keywords: Environment, Empirical, SAG

1. INTRODUCTION
Interpreting remote sensing imagery is improved when the atmosphere present between a sensor and observation is well understood. Much of the behavior of the atmosphere along a line-of-sight can be predicted deterministically. For example, at the solar terminator, rapid variations in the atmosphere's composition occur due to photochemical processes. This occurs over a small (several-degree) range of solar zeniths angles and can be to a large degree captured within an empirical model. Superimposed on these effects are short-term and small-scale variations associated with random processes that can be characterized and predicted statistically.

The Standardized Atmosphere Generator (SAG) is a climatological environmental model that provides location (latitude and longitude) and temporal (time of day and time of year) dependent values for environmental constituents including temperature, pressure, and concentrations of atoms, molecules, electrons, and ions needed to evaluate missile defense systems performance. SAG has been designed to allow the major known, systematic variabilities in the atmosphere, including terminator and other diurnal effects, to be practically incorporated into a number of strategic models for the defense community. SAG can be utilized as a stand-alone model, as a scriptable tool, or through an application programming interface (API). SAG is now part of the American Institute of Aeronautics and Astronautics Guide to Reference and Standard Atmosphere Models.1

Here, we'll summarize the various models that make up SAG's current capabilities in Section 2. Next, in Section 3.1 we'll discuss recent updates that allow users to assimilate selected measurements into the empirical model. We'll compare the results of the SAG model to some selected measurements in Section 4. Finally, we'll provide concluding remarks in Section 5.

1 Further author information: (Send correspondence to Raphael Panfili)
Raphael Panfili: E-mail: rpanfili@spectral.com
2. THE SAG MODEL

SAG consists of a collection of models designed to inter-operate. Each model, which we’ll refer to as a “component” of the overall model, consists of a prescriptive algorithm, any static data required by that algorithm, and an internal interface.

2.1 MSIS

The Mass Spectrometer and Incoherent Scatter (MSIS) algorithms were developed between 1977 and 2000. The name derives from two groups of instruments used to collect data that then informed the algorithm. Five space-based mass spectrometers (San Marco 3, Aeros A, and Atmospheric Explorer A, B, and C) and four ground-based incoherent scatter radars (Arecibo, Jicamarca, Millstone Hill, and St. Santin) were used to provide neutral densities of the atmosphere. The basic mathematical formalism in MSIS\(^2,3\) was originally developed to parameterize measurements taken by the Orbiting Geophysical Observatory satellites. Successive iterations built upon this formalism, first to include longitude variations and more exotic effects such as geomagnetic storms and high solar flux data\(^4\) and then later to extend MSIS from the upper atmosphere (above about 100 km) down to the ground\(^5,6\).

MSIS continues to be used extensively by the scientific community with source code freely available. Its widespread adoption has led to extensive testing against experimental data by the international scientific community. Despite being constructed using data that is now nearly two decades old, it continues to perform well during inter-comparisons of satellite drag predictions relative to other high-altitude atmospheric models.\(^7\) We make extensive use of the output generated MSIS component within SAG, modifying some output data to account for diurnal variability. Example output from the MSIS component within SAG is shown in Figure 1.

![Figure 1. MSIS-00 temperature at 150 km.](image)

2.2 Zonally Averaged Climatology

The Naval Research Laboratory (NRL) Zonally Averaged Climatology (ZAC) database\(^8,9\) contains a self-consistent set of atmospheric trace constituents. It contains the first seven HITRAN radiators (H\(_2\)O, CO\(_2\), O\(_3\), N\(_2\)O, CO, CH\(_4\), and O\(_2\)) as well as concentrations for NO\(_2\), HNO\(_3\), N\(_2\), and atomic oxygen. Monthly mean data is provided on a vertical grid of 1 km resolution up to 25 km and 5 km resolution above 25 km and up to 120 km. Data is presented on a 10\(^\circ\) latitude grid with monthly means provided. In addition, the database captures day/night ozone variability. The database was created through a combination of observational data and simulations of
the dynamics of these trace constituents in the atmosphere. It made use of atmospheric modeling performed at NASA\textsuperscript{10} including the Middle Atmosphere Project (MAP).\textsuperscript{11} Data from seven NRL ZAC radiators are used within SAG. Profiles for N\textsubscript{2}O, CO, CH\textsubscript{4}, NO\textsubscript{2}, and HNO\textsubscript{3} are used in their entirety while atomic oxygen and ozone profiles are used in conjunction with a separate component to capture diurnal variability. An example of the monthly mean variability used within the ZAC component is shown in Figure 2.

![Methane Mixing Ratio (PPMV) for September](image)

**Figure 2.** Example of water vapor concentrations from URAP data in SAG.

### 2.3 Trace Atmospheric Constituents

The Air Force Geophysics Laboratory (AFGL) Atmospheric Constituents Profiles (ACP)\textsuperscript{12} is a collection of physically reasonable atmospheric profiles of temperature and molecular concentrations for 28 species. The first seven molecular species (H\textsubscript{2}O, CO\textsubscript{2}, O\textsubscript{3}, N\textsubscript{2}O, CO, CH\textsubscript{4}, O\textsubscript{2}) have six separate profiles. Those profiles are consistent with the U.S. Standard Atmosphere, 1976 Edition\textsuperscript{13} and, alternatively, with the zonal profiles specified in the U.S. Standard Atmosphere, 1962 Edition.\textsuperscript{14–16} A single atmospheric profile is specified for the remaining 21 species. Data from five AFGL ACP radiators are used within SAG. Profiles for SO\textsubscript{2} and NH\textsubscript{3} are used in their entirety. Separately, profiles for H\textsubscript{2}O, NO, and OH are used to fill gaps in the available data record or to smoothly transition from between outputs from other SAG components.

### 2.4 Nitric Oxide

The Student Nitric Oxide Explorer (SNOE) was a satellite launched to measure nitric oxide (NO) variability in the thermosphere.\textsuperscript{17} SNOE scientific objectives were to monitor soft x-rays\textsuperscript{18} and auroral activity\textsuperscript{19} to determine their impact on NO production. We refer to the SNOE data along with the algorithm we use with that data as the SNOE component in SAG. It provides NO densities within an altitude range limited to between 97 km and 150 km. Below and above this range, we make use of the AFGL ACP component and smoothly interpolate in between. Diurnal variation in densities are further superimposed upon the measured data. The magnitude of the diurnal variation is used to scale the SNOE data. This is accomplished by computing separate daytime and nighttime profiles.\textsuperscript{20} An example of monthly mean SNOE data is shown in Figure 3.

The SNOE satellite has completed its mission and is no longer collecting data. What data was collected has been archived and is freely available from NASA. The SNOE component within SAG utilizes most, but not all of the data set. Separately, the NO Empirical Model (NOEM) has been created\textsuperscript{21} using the complete data set and the performance of that model has been recently compared against measurements\textsuperscript{22} taken over a larger swath of the solar cycle. We are currently evaluating the use of NOEM as a new optional component within SAG.
2.5 Water Vapor

Data from the Upper Atmosphere Research Satellite (UARS) is used to model water vapor in SAG. The data used is contained within the UARS Reference Atmosphere Profile (URAP) collection. The UARS component in SAG is not self-contained as its altitude range is limited to between (approximately) 15 km and 85 km. Below and above this range, we make use of the AFGL ACP component and smoothly interpolate in between. An example of water vapor concentrations from the UARS URAP component is shown in Figure 4.

There now exists a more comprehensive data set containing the historical record of water vapor observations.
This data set has been assembled by the National Centers for Environmental Prediction (NCEP) and National Center for Atmospheric Research (NCAR). URAP water vapor profiles are one of data sets used to develop The NCEP-NCAR Reanalysis database. We are evaluating how best to incorporate the results of this research into a future revision of the SAG model.

3. DATA ASSIMILATION

The empirical portion SAG model provides a physically consistent description of the environment based on historical observations. This is a reasonable starting point for predictive modeling but may be insufficient to characterize a specific event. In that case, a user may have a handful of measurements giving a partial picture of the environment (typically temperature, humidity, pressure, and wind speed and direction). For example, one or more ground-based weather stations may be available. Or, perhaps, a radiosonde was released near a test site. Or, a remote sensing satellite captured a unique atmospheric event at a specific location in space. For these cases, the SAG model maintains a limited data assimilation capability. When user-supplied atmospheric data is provided, the SAG model assumes that information should take precedence over a model based on historical data.

It is extremely likely that for instances in which a user supplies data to SAG, the data being supplied will not provide a complete picture of the state of the environment. For example, a radiosonde may supply only temperature, pressure, relative humidity, wind data, and nothing more. For these cases, SAG offers the ability to overly physically consistent values that are missing from the data record while preserving user-supplied measurements. The results are presented to the user as a typical SAG atmospheric profile.

3.1 Radiondes

SAG allows a user to incorporate radiosonde measurements within an atmosphere profile. Most radiosonde files (and all of the radiosonde formats currently supported within SAG) contain water density, temperature, and pressure as a function of altitude. Most radiosonde data also contains wind speed and direction information, but this is not currently used by SAG at this time. Valid data within the radiosonde profile is treated as truth and SAG consistently incorporates all other atmospheric information to complete a profile.

SAG currently accepts four common formats of radiosonde data files. Figure 5 illustrates one supported format. The NOAA Earth Science Research Laboratory utilizes the RAdiosonde OBservation (RAOB) format and data in this format is available directly from NOAA. Unfortunately, there is no single standard radiosonde data format. It is left to the user to understand the nature of the radiosonde format at hand and to supply SAG with the necessary information to process that data.

<table>
<thead>
<tr>
<th>358_74_52</th>
<th>-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.00000</td>
<td>292.7000</td>
</tr>
<tr>
<td>16.00000</td>
<td>245.5000</td>
</tr>
<tr>
<td>18.00000</td>
<td>199.8000</td>
</tr>
<tr>
<td>20.00000</td>
<td>200.2000</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5. Example of RAdiosonde OBservation (RAOB) data format.

The example shown in Figure 5 displays a common feature of radiosonde data. Most radiosonde files contain “bad” data which must be excluded from processing. In our example, those data are flagged with a fixed value
of 99999. SAG can identify and process bad data points within the four different formats of radiosonde data it currently understands.

Finally, a radiosonde will stop collecting data at some altitude in the stratosphere. At that point, it becomes necessary to seamlessly transition to the empirical modeling in SAG to create a complete atmospheric profile. For that we use a collection of canonical SAG atmospheres (CSAs) that span the physical space of the problem. We show an example of a radiosonde dataset with a superimposed best-fit CSA in Figure 6. The CSAs are designed to reduce fluctuations of values in the dataset, leading to a relatively smooth atmospheric profile. It excludes the possibility of capturing events such as temperature inversions. Future updates to SAG may allow users greater flexibility in choosing the degree to which certain radiosonde features are retained.

![Figure 6. Comparison of radiosonde temperature profile with the best fit Canonical SAG Atmosphere.](image)

### 3.2 Single Profiles

SAG provides the user with the ability to assimilate a limited set of external profiles as input data sets. This option is currently available for temperature, molecular water and ozone profiles, or any combination of the three. External profile data is typically input into SAG using a simple two-column ASCII file which summarizes the desired profiles. The first line of this file provides a user-selected name (currently unused), which can also be used to hold user-defined parameters. Following the header line will be a profile in two-column format. Each row contains one altitude in kilometers followed by value (temperature or molecular concentration). An example external profile data set is shown in Figure 7.

There are a few limitations to the external profiles a user can supply. The first is that external profile data is assumed to be contiguous with the smallest and largest provided altitudes defining the range the data spans. The order is immaterial as SAG will ensure profiles are monotonically increasing in altitude, but there can be no gaps in the profile. Next, external profiles in temperature are limited in range from the ground to 72.5 km. This allows the user to model phenomena such as sudden stratospheric warming. Molecular profiles for water and ozone follow a similar pattern with the exception that the altitude restriction is removed.

### 4. DATA COMPARISONS

Here, we will compare water vapor profile predicted by SAG with ongoing satellite measurements. Water is among the most variable component of the atmosphere and will be the most stressing element for an empirical
environmental model. The Meteorological Operational satellite program (MetOp) is a joint collaboration between the European Space Agency (ESA) and European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT). Additional assistance is provided from the French Centre National d’Etudes Spatiales (CNES) and the U.S. National Oceanic and Atmospheric Administration (NOAA). MetOp is the first in a series of European satellites dedicated to near-real-time operational meteorology observations. The first polar orbiting satellite (MetOp-A) was launched on October 19, 2006. On 17 September 2012, the follow-on MetOp-B was successfully launched into polar orbit, which is now in a co-planar orbit and nearly half an orbit out of phase with MetOp-A.

The METOP-B data set includes Temperature, H\(_2\)O Mass Mixing Ratio, and other quantities yielding cloud, snow, ice info. Data are specified by layers defined by constant Pressure. There are data for pressure levels from 1013 hPa (ground level) to 0.005 hPa (80km). The METOP-B dataset is at high resolution, each orbit with 1150 horizontal scans and 90 cells per scanline with scanlines approximately 18 km apart. We utilize the hypsometric equation in a piecewise manner to determine altitudes given METOP-B pressure and temperature measurements. With a translation from MIRS pressures to altitudes, we have all the information we need to perform a direct comparison against SAG. Inputs to SAG from METOP-B include date, time, latitude, longitude, and altitude. From that, we extract the molecular concentrations predicted by SAG. SAG outputs are converted to the native units of density \( \rho \) (molecules per cubic centimeter) to that of the METOP-B data (grams of water per kg of atmosphere) using:

\[
\rho_{H_2O}(g/m^3) = \rho_{H_2O}(molec/cm^3) \cdot \left( \frac{10^6}{N_A} \right) \cdot w_{H_2O} \tag{1}
\]

Where \( N_A \) is Avogadros number and \( w_{H_2O} \) is the molecular weight of water. With this conversion, METOP-B data can be directly compared by creating the ratio:

We acquired all METOP-B data taken for four days in a calendar year: March, June, September, and December 6th, 2016. This, in total, provided us with 30 GB of data containing slightly more than 10 billion individual water measurements. We next winnowed this collection down by considering only the central scanline in each set. With this data reduction step, we are left with 14,253,900 individual water measurements spanning 116 NetCDF files.

We then established an automated workflow in which the NetCDF data was extracted, altitudes were computed for each scan line, and SAG executed with conditions to match the provided data set. The result was two key image representations of the output as shown in Figure 8. On the left is a map indicating the location of the METOP-B scan for a particular data set. Ground-layer water mass density is plotted but legends are omitted: the primary purpose of this map is to ground the geographic location of the data being compared. The time and date of the scan is embedded in the output with time defined as the instance in which the first element in the data set was captured. The original name of the NetCDF file, as tagged by NOAA, is also embedded in the output.
The right-hand plot in Figure 8 shows the direct comparison of SAG numerical output (x-axis) and METOP-B measurements (y-axis). Results are plotted in the native units of the METOP-B data (grams water per kilograms air) and presented on a log scale as the data spans several orders of magnitude. The results are binned and the density of points in the scatter plot is shown. Perfect correlation between measurements and model would result in a line along the diagonal while fully uncorrelated results would result in an image without any discernable pattern. We do not expect a perfect match against measurement. SAG is an empirical model and will not be able to capture day-to-day fluctuations in low-altitude weather. Nonetheless, we expect to see correlation between the empirical model and measurement.

We next examine the statistical nature of the data sets. The results, for the entire collection of data analyzed, are shown in Figure 9. The data has mean values both greater and less than the mean of the SAG model results, depending on the total water concentration in question. Standard deviations are generally lower than those from the SAG model. This implies it may be necessary to more fully sample the temporal range of the data collect to obtain a fuller picture. That is, the data used for this comparison was collected over only four days. The data appear to be more tightly bound than output provided by SAG, which was itself generated from a different data set but over finer gridding of calendar dates.

![Figure 8. Example map of METOP-B data scan (left) and SAG versus METOP-B comparison (right).](image1)

![Figure 9. Mean and standard deviation of measurements compared to SAG.](image2)
We see generally good correlation at the lowest and highest altitudes (corresponding to lowest and highest water vapor concentrations, respectively). In between, the data and SAG are still correlated, but to a lesser degree. Some of this can be explained by local weather. Specifically, the precise location of the tropopause has a large impact on the amount of water vapor the atmosphere can hold. This indicates a path towards incorporating data into SAG at a future date. Should local water concentrations be known, the SAG water vapor profiles can match tropopause altitudes to improve upon predicted output. The results also give some hint as to the natural limits of an empirical model to capture such a variable component of the atmosphere in the absence of anchoring information. Let us consider the data itself in the absence of any empirical model. We can then inquire as to the degree of correlation between subsets of the data itself. A previous investigation\textsuperscript{26} considered this in the context of comparing results from the GOME-2A instrument on the METOP-A satellite and the GOME-2B instrument on the METOP-B satellite. What was found, when performing month-over-month comparisons was that there exists very little mean bias between measurements made by the two instruments. In addition, and relevant to the above discussion, the report also found month-to-month variance of the data sets. The authors reached the conclusion that "the standard deviation for water vapor data is dominated by natural variability and is therefore quite large."

5. CONCLUSIONS

The Standardized Atmosphere Generator (SAG) endeavors to be a model of the atmospheric background suitable to meet defense needs. It strives to form a consistent, complete picture of the environment to support a variety modeling needs. We've described the key components which make up SAG and multiple methods to access the model. We've illustrated a recent effort to validate the model against data. Going forward, we anticipate adding capabilities to meet emerging needs. This includes further refinements to the existing model as well as an expansion of the range of data that can be assimilated into SAG.

ACKNOWLEDGMENTS

This unnumbered section is used to identify those who have aided the authors in understanding or accomplishing the work presented and to acknowledge sources of funding.

REFERENCES


