



Master's thesis  
Master's Degree Programme in Atmospheric Sciences

## **Development of the MEGAN3 BVOC Emission Model for Use with the SILAM Chemical Transport Model**

Benjamin Foreback  
2018

Supervisor(s): Dr. Michael Boy  
Dr. Alexander Mahura

Examiner(s): Prof. Markku Kulmala  
Dr. Michael Boy  
Dr. Alexander Mahura

INSTITUTE FOR ATMOSPHERIC AND EARTH SYSTEM RESEARCH / PHYSICS  
FACULTY OF SCIENCE  
UNIVERSITY OF HELSINKI  
P.O.Box 64 (Gustaf Hållströmin katu 2)  
00014 University of Helsinki

# Contents:

Abstract.....	5
1. Introduction.....	7
1.1 Background.....	7
1.1.1 Scientific problem.....	7
1.1.2 Importance of boreal forests and BVOC emissions in models .....	8
1.1.3 Previous studies in BVOC modelling .....	9
1.2 Project Goals and Objectives .....	9
2. Methodology .....	11
2.1 The SILAM Model .....	11
2.1.1 History of SILAM.....	11
2.1.2 Model Structure .....	13
2.1.3 SILAM model domain .....	14
2.1.4 Current version of SILAM.....	15
2.1.5 Unreleased Updates.....	15
2.1.6 Required Inputs.....	15
2.1.7 Running SILAM in the CSC Computing Environment .....	16
2.1.8 Example of SILAM Run .....	17
2.2 Enviro-HIRLAM.....	18
2.2.1 Overview of Enviro-HIRLAM.....	19
2.2.2 Use of Enviro-HIRLAM output in SILAM Runs .....	19
2.3 Current Implementation in SILAM – Guenther Modified.....	20
2.3.1 USGS Eurasia Land Cover Map .....	20
2.3.2 Results from Guenther Modified in SILAM.....	21
2.3.3 Results Based on Land Classification .....	24
2.4 MEGAN as Proposed Replacement.....	26
2.4.1 Overview of MEGAN.....	27
2.4.1.1 History and Current Version of MEGAN .....	27
2.4.1.2 MEGAN Model Components .....	28
2.4.2 MODIS Input to MEGAN.....	29
2.4.3 J-Rating .....	29

3. Results and Discussion .....	30
3.1 Findings in the Current SILAM Version .....	30
3.1.1 Comparison of Enviro-HIRLAM modes in SILAM .....	30
3.1.2 Comparison of current SILAM implementation with GM algorithm to measured BVOC flux data .....	33
3.2 Initial Steps in Integration of MEGAN3 with SILAM .....	38
4. Conclusions .....	40
4.1 Primary Findings .....	40
4.2 Current Status .....	41
4.3 Future Steps .....	41
4.3.1 Integration of MEGAN3 into SILAM .....	42
4.3.2 Recommendations .....	42
5. References .....	44
6. Acknowledgements .....	48
7. Appendix .....	50
Appendix A: Glossary of Terms and Acronyms .....	50
Appendix B: Description of MODIS Data & Calculation of LAI <sub>v</sub> from MODIS .....	52
Appendix C: SILAM Meteorology for MEGAN .....	53
Appendix D: Control File and BVOC Emission File for SILAM Run .....	58
Appendix E: Detailed Steps of Installing, Compiling and Running MEGAN on CSC .....	62
E1. Overview of Steps Performed .....	62
E2. Modules, Packages and Installations .....	63
E3. Compiling and running the MEGAN3 pre-processor .....	63
E4. MEGAN3 EFP .....	64
E5. MCIP .....	64
E6. MEGAN3 Main Code .....	64
E6.1 Overview .....	64
E6.2 Compilation .....	65
E6.3 Setup of MEGAN .....	66
E6.4 Steps to initiate a MEGAN3 run .....	66
E6.5 Challenges with not running MEGAN on MPI .....	67
E7. Summary of Required Inputs .....	67

E8. Additional Notes.....	68
Appendix F: List of Updates to SILAM between Latest Public Release and Version Used for this Project .....	70

# Abstract

This project has aimed to investigate and propose improvements to the methods used in the System for Integrated ModeLing of Atmospheric coMposition (SILAM) model for simulating biogenic volatile organic compound (BVOC) emissions. The goal is to study an option in SILAM to use the Model for Emission of Gases and Aerosols in Nature, Version 3 (MEGAN3) as an alternative to SILAM's existing BVOC calculation algorithm, which is a more simplified approach.

SILAM is an atmospheric chemical transport, dispersion, and deposition modelling system owned and continuously developed by the Finnish Meteorological Institute (FMI). The model's most well-known use is in forecasting air quality in Europe and southeast Asia. Although traffic and other urban emissions are important when modelling air quality, accurate modelling of biogenic emissions is also very important when developing a comprehensive, high-quality regional and sub-regional scale model.

One of the motivations of this project is that if BVOC emission simulation in SILAM were improved, the improvements would be passed into subsequent atmospheric chemistry algorithms which form the molecules responsible to produce secondary organic aerosols (SOA). SOA have significant impacts on local and regional weather, climate, and air quality. The development in this project will therefore offer the potential for future improvement of air quality forecasting in the SILAM model.

Because SILAM requires meteorological forecast as input boundary conditions, this study used output generated by the Environment-High Resolution Limited Area Model (Enviro-HIRLAM), developed by the HIRLAM Consortium in collaboration with universities in Denmark, Finland, the Baltic States, Ukraine, Russia, Turkey, Kazakhstan, and Spain.

Enviro-HIRLAM includes multiple aerosol modes, which account for the effects of aerosols in the meteorological forecast. Running SILAM with and without the aerosol effects included in the Enviro-HIRLAM meteorological output showed that aerosols likely caused a minor decrease in BVOC emission rate.

This project has focused on the boreal forest of Hyytiälä, southern Finland, the site of the Station for Measuring Ecosystem-Atmosphere Relations - II (SMEAR-II, 61.847°N, 24.294°E) during a one day trial on July 14, 2010.

After performing a test run over the Hyytiälä region in July 2010 for analysis, it was found that SILAM significantly underestimates BVOC emission rates of both isoprene and monoterpene,

likely because of an oversimplified approach used in the model. The current approach in SILAM, called ‘Guenther Modified’, uses only a few equations from MEGAN and can be classified as a strongly simplified MEGAN version, with selected assumptions. It references a land cover classification map and lookup table, taking into account only three parameters (air temperature, month, and solar radiation) when performing the calculations. It does not take into account several other important parameters, which affect the BVOC emission rates. Based on qualitative analysis, this appears to be a simplified but limited approach.

Therefore, based on these findings, the next step to improve SILAM simulations is to propose a full implementation of MEGAN as a replacement to the current logic in SILAM, which is to use land classification and a lookup table for BVOC emission estimates. MEGAN, which is a much more comprehensive model for simulating BVOC emissions from terrestrial ecosystems. MEGAN includes additional input parameters, such as Leaf Area Index (LAI), relative humidity, CO<sub>2</sub> concentration, land cover, soil moisture, soil type, and canopy height.

Furthermore, this study found that in the future, simulations involving BVOCs could also potentially be improved in SILAM by adding modern schemes for chemical reactions and SOA formation in future development of SILAM.

After gaining in-depth understanding of the strengths and limitations of BVOC in the SILAM model, as practical result, some recommendations for improvements to the model are proposed.

# 1. Introduction

The goal of this project is to investigate and propose potential improvements to the Biogenic Volatile Organic Compound (BVOC) emission simulation in the System for Integrated modeling of Atmospheric composition (SILAM), an atmospheric chemical transport, dispersion, and deposition model developed by the Finnish Meteorological Institute (FMI).

This project focuses in particular on the boreal forest region in central Finland but the proposed approach and technology can be applied to any other model domain for which SILAM is used.

## 1.1 Background

The SILAM model, originally developed in the 1980s as an atmospheric transport model for emergency management purposes, has been expanded and eventually evolved into a multi-purpose model. Its primary use as of late is for air quality forecasting. One of the factors that influences aerosol formation, which subsequently affects air quality is BVOC emissions from forested regions.

### 1.1.1 Scientific problem

Currently, the algorithm in SILAM for simulating BVOC emission is very simplified and does not consider all parameters and processes involved. Because aerosols contribute to air quality and forested areas cover a significant portion of the Earth's surface, it is important to gain better understanding of how the model forecasts BVOC emissions that lead to organic aerosol formation over these large forested regions, with the goal to develop an improvement to the model.

Because the algorithm for Secondary Organic Aerosol (SOA) formation in SILAM is based on BVOC emission, it is important to simulate the BVOC emissions as accurately as possible. Without being able to pass accurate prediction of emission rates of gases into the SILAM's SOA algorithms, the predictions of SOA will consequently not be accurate. Therefore, better calculation of BVOC emission rates is the starting point for accurate prediction of aerosol formation.

### 1.1.2 Importance of boreal forests and BVOC emissions in models

When thinking about air quality and atmospheric aerosols, the first thing that is likely to come to mind is a polluted city, e.g. Beijing, Shanghai, Los Angeles, Warsaw, London, etc. Specifically, emissions from vehicles and heavy industry can easily be blamed for the formation of aerosols and elevated levels of pollution. However, biogenic emissions, including from forests, also play a very important role in global aerosol formation (Tunved et al., 2006). In fact, cities cover only about 3% of the world's land cover, whereas forests cover 31% (WorldBank, 2018).

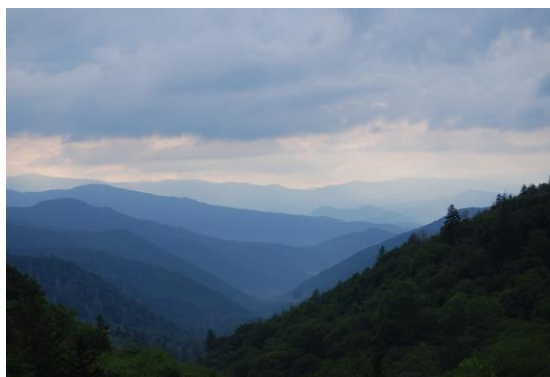
Furthermore, boreal forests represent one third of all forest land cover (Tunved et al., 2006).

Therefore, in terms of surface area covered, studying the formation of aerosols in boreal forest regions is critical for understanding global aerosol processes.

In addition to air quality, aerosols from forests also affect cloud-climate interactions, which is a significant uncertainty in climate studies (Kulmala et al., 2004). It is important to understand the emission process in order to understand subsequent formation of aerosols. The composition, size, concentration, and optical properties of these aerosols all play a role in aerosol-climate interactions. Furthermore, forcing from aerosols (direct and indirect aerosols effects) varies significantly with location, season, time of day, and altitude, which illustrates the need to understand BVOC emissions and reactions not only in cities but over forested regions as well. (Seinfeld and Pandis, 2006).

Isoprene is the BVOC with the highest emission rates globally and monoterpenes are considered to have an important impact on the SOA budget. Therefore, this project focused specifically on the two chemical species isoprene and monoterpene.

Figure 1 shows a visual example of high amounts of aerosols that form in a forested area. This image shows the effects of aerosols on light scattering in a forested area, which can directly and indirectly affect local and regional weather and climate.



**Figure 1:** The Great Smokey Mountains (Tennessee/North Carolina, USA) is one of the most well-known location for being able to view SOA. The mountains get their name from the fact that heavy aerosols from the forest frequently create the appearance of smoke covering the mountain range (image source: Pixabay).

### 1.1.3 Previous studies in BVOC modelling

Previous studies on integrating BVOC emission into local and regional chemical transport models date back at least to the mid-1990s. One such study includes introducing the effects of urban trees into the U.S. Environmental Protection Agency's Urban Airshed Model (UAM); the purpose of this study was to estimate the cost-effectiveness of planting new trees in attempt to improve air quality in Sacramento, California, USA (McPherson et al 1997). Other studies include influence of BVOC emission and subsequent reactions on ozone pollution in Barcelona, Spain (Toll and Baldasano, 2001). Diem and Comrie (2001) also involved modelling BVOC emissions and the effects on ozone pollution in the American cities of Los Angeles, California and Tucson, Arizona. Numerous other studies have been performed since.

Recent studies include Zhou et al (2017), which integrated the MEGAN (version 2) model into the model to Simulate the concentrations of Organic vapours, Sulphuric Acid and Aerosols (SOSAA), which estimated emissions of BVOCs over the boreal forest of Hyytiälä, Finland. Their study found that emission rates modelled by MEGAN compared reasonably well to flux measurements at SMEAR-II. The Zhou et al (2017) study was used as a comparison to this study.

## 1.2 Project Goals and Objectives

The ultimate goal of this project is to develop ideas and propose recommendations to improve SILAM so that it can better simulate BVOC emissions, which would result in the model doing a better job at forecasting aerosols and air quality. Once the strengths and weaknesses of the current algorithm in SILAM and MEGAN3 are compared, a plan can be made for integration of

MEGAN3 into a future version of SILAM. Due to the complexity of both models, this project does not do the full technical implementation, rather it serves as a starting point for a larger-scale project of integrating the two models together.

## 2. Methodology

The methodology for this project includes investigating SILAM's current BVOC emission algorithm, exploring its strengths and weaknesses, and determining if there is possibly a way to improve the emission estimates in SILAM.

### 2.1 The SILAM Model

SILAM is a versatile atmospheric chemical and particle transport, dispersion, and deposition modelling system used for a wide variety of real-world applications, including but not limited to forecasting of air quality, smoke dispersion, radionuclide dispersion from nuclear bombs or accidents, and releases from chemical or biological terrorism. At the present day, SILAM is probably best known as an air quality forecasting model, and daily public forecasts are available on SILAM's website at <http://silam.fmi.fi>

SILAM can be operated in both modes: forward, which forecasts dispersion from a given source, and inverse, which traces a measured compound back to its source. Depending on the desired use case, the model can be used as either a Lagrangian model, which follows an air parcel through space and time, or as a Eulerian model, where chemical species are transported through fixed grid cells.

The model dates back to the late 1980s and is developed and managed by the Finnish Meteorological Institute (FMI). As written in Sofiev et al. (2015), the SILAM code from the most recent release is publicly available under the request from the FMI.

#### 2.1.1 History of SILAM

The first SILAM code was developed in the late 1980s following the nuclear meltdown and explosion at the Chernobyl nuclear power plant in present-day Ukraine in April 1986. This disaster exposed the need to develop both forward and inverse modelling of dispersion of radionuclides for emergency response and decision-making purposes (Sofiev and Siljamo, 2014). Forward modelling, that is forecasting the transport of fallout resulting from a major accidental release, was seen as important as fears of another incident rose. Additionally, inverse modelling, which is tracing back a compound or gas to its source, was also seen as important. The need for backtracing came from the fact that the Soviet Union did not immediately report situation with the Chernobyl incident to international nuclear authorities. In fact, it wasn't until alarms activated at a nuclear power plant in Sweden that authorities in western Europe were made aware

of the disaster (Bengtsson, 1986). In the years following, the model was developed with focus on nuclear emergency response and preparedness.

SILAM was first developed as a Lagrangian model. In the early 2000s, the Eulerian scheme was added to SILAM so that it could be run in either Lagrangian or Eulerian mode (Sofiev et al., 2015).

During the rise of Al Qaeda and other international terrorist organizations in the 1990s and early 2000s, chemical and biological terrorism became a global concern. A 2003 report from multiple United States agencies stated that Al Qaeda had interest in and was capable of deploying chemical weapons on the large scale (Parachini, 2003). This new threat brought forward the need for emergency modelling beyond just nuclear accidents, but for quick decision making and emergency response for a major release of any nuclear, chemical, biological or particle release. This posed new challenges for atmospheric chemical transport modelling, and it led to further development of SILAM to be able to model dispersion of any generic compound, not just radionuclides. Moreover, it became necessary to minimize computing time to run models for emergency response, especially in cases where the release may be close to densely populated areas. The SILAM group at FMI focused heavily on developing solutions to these challenges, for instance parallelization of the model code (Sofiev and Siljamo, 2014).

In 2005, air quality forecasting was added to SILAM, which greatly expanded the model beyond its original use (Sofiev, 2010). This brought the model into public visibility and everyday operational use. Since 2006, daily model forecasts for a number of air quality parameters (currently NO<sub>2</sub>, SO<sub>2</sub>, NO<sub>x</sub>, particulate matter, ozone, and carbon monoxide) is publicly available for Europe, Southeast Asia, and the whole globe on SILAM's website, <http://silam.fmi.fi> (Sofiev, 2011).

In 2010, the eruption of the Eyjafjallajökull volcano in Iceland caused widespread disruption to air traffic in Europe and across the Atlantic Ocean. SILAM has worked in a partnership with MACC (Monitoring Atmospheric Composition and Climate; <http://www.gmes-atmosphere.eu/>), an EU-funded programme that provides support to governments and private companies in monitoring and forecasting atmospheric conditions. During and after the Eyjafjallajökull eruption, the model was used by MACC consortium as one of three models for forecasting dispersion of the volcanic ash plume (MACC, 2010)

The following year, the Grimsvotn volcano in Iceland erupted, also causing disruption to air traffic. As with the previous year, the FMI provided forecasts of ash plume spread using the SILAM model (FMI, 2011a).

Also in 2011, the meltdowns of three reactors at the Fukushima Daiichi Nuclear Power Plant in Japan, which were caused by a powerful earthquake and subsequent tsunami, released extreme amounts of radionuclides into the atmosphere and ocean. The Nuclear Safety Commission (NSC) of Japan estimated that in total 150 petabecquerels (PBq) of iodine-131 and 12 PBq of caesium-137 were released as a result of this disaster. This makes it the second worst nuclear accident in world history, following Chernobyl, in terms of amount of released radioactive material (Bannai, 2011). The FMI provided extensive support to Japanese authorities during and after the disaster, deploying the SILAM model for forecasts of radionuclide dispersion (FMI, 2011b).

An example of using SILAM in inverse mode to trace measurements back to a source is when unusual concentrations of caesium-137 were measured in Helsinki (Finland) and Stockholm (Sweden) in April 2013. An inverse run of SILAM, along with NOAA's Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLOT) model, were able to trace the radionuclides to a smelter near Moscow (Russia) which was melting radioactive material, presumably by mistake (NERIS, 2015).

### 2.1.2 Model Structure

SILAM is, first of all, the Atmospheric Chemical Transport model, and therefore does not perform weather forecasts. Therefore, it requires meteorology forecast as input to drive the model's simulations. SILAM accepts weather forecast input from the High Resolution Limited Area Model (HIRLAM) or the European Centre for Medium-range Weather Forecasts (ECMWF) among other forecast models. SILAM is an offline model, which means that it accepts the meteorological input at the beginning of the model run, i.e. after the weather model run is completed. It uses the same weather forecast (i.e. 3D meteorological fields) output interpolated into the SILAM model domain throughout the run. It therefore does not couple with the NWP model, nor does it integrate real-time forecasts from the weather model (Sofiev et al., 2017; Baklanov et al., 2014). The meteorological forecast input is only read once at the beginning of the model run.

Technically, a model run of SILAM is driven by a "control file." This control file specifies all parameters for a model run, such as start time, time steps, type of run (forward or inverse), emission sources (e.g. BVOCs, smokestacks, fires, a bomb, or any other generic point or area source), type of compounds released, deposition schemes, chemical reactions, aerosols, and list of output parameters to be saved. The control file is divided into categorized sections, called "namelists". Some namelists, such as time, physical domain and output parameters, are required. Others, such as emission parameters, are optional, depending on the purpose of the model run.

All necessary input files for the model run must be declared in the control file. Pointers to supplementary definition files, such as detailed descriptors of BVOCs, fires, chemical releases, etc., can also be defined in the same control file.

Supplementing the main control file, there are additional model descriptors called “cocktail” files. These cocktails contain further information for the model to run but are considered somewhat more internal to the model. The cocktails are typically not modified between multiple runs. Examples of parameters contained in the cocktails are units, whether a species is in particle or gas phase, component fraction, and other chemical and physical properties of the chemical species in the model.

Some “standard cocktails” are core to the model and define internal parameters and structure. These should only be modified by the SILAM code developers or expert users who are very familiar with the model.

Results from the model runs are stored in Network Common Data Format (NetCDF; developed by the National Center for Atmospheric Research (NCAR), USA), and all outputs from a run are stored in the same file. Details about the NetCDF format can be found at NCAR’s website at <https://www.unidata.ucar.edu/software/netcdf/>

Further practical details about the structure, setup and run of the SILAM modelling system can be found in the SILAM user guide: [http://silam.fmi.fi/doc/SILAM\\_v5\\_userGuide\\_general.pdf](http://silam.fmi.fi/doc/SILAM_v5_userGuide_general.pdf)

### 2.1.3 SILAM model domain

SILAM is setup by defining a domain in the control file (see Section 2.1.6 (Required Inputs) below). The horizontal domain is defined using latitude and longitude coordinates for the four corners of the model, along with the number of x and y grid cells. There is no upper or lower limit to the horizontal size of the domain or number of grid cells; the model can be run for a very small area (e.g. less than a kilometre) or for the entire globe if desired. However, computational resources and run time should be taken into consideration.

As an example, the FMI’s daily operational runs for SILAM are performed at the Finnish scale using a 0.05 x 0.05 degree grid, over Europe and Southeast Asia at a resolution of 0.1 x 0.1 degree, at globally at 0.5 x 0.5 degrees. These are the forecasts that are available on SILAM’s website.

When using SILAM for research purposes, it is possible to use a domain with different x and y grid sizes, e.g. 0.01 x 0.02 degrees.

This study used a 2 x 2 degree horizontal domain centered over Hyytiälä, Finland, with 200 x 200 grid cells (i.e. the resolution was 0.01 x 0.01 degree), which equates to approximately 200 x 200 km.

The vertical domain for SILAM can also be as small or large as desired. Additionally, the vertical layers do not have to be equal. This is especially useful for this project, where the first layer is of most interest. Therefore, vertical levels near the top of the model, which were less important, were larger than near the surface. For this project, 9 vertical layers were defined with the following thicknesses (in meters): 25, 50, 100, 200, 400, 750, 1200, 2000, 2000 (i.e. the model top is at 6725 meters above the ground). It is assumed that all of the BVOC emissions are in the first layer, as the trees at Hyytiälä's location are not taller than 25 meters.

### **2.1.4 Current version of SILAM**

The most recent version of SILAM is 5.5.1 and it is freely available. This is the public version used to generate operational forecasts which can be found on the SILAM website, <http://silam.fmi.fi>.

### **2.1.5 Unreleased Updates**

This project is using a version of SILAM provided directly by the FMI's SILAM development group, which includes some bug fixes and modifications not currently in the public release of the model. This version was last modified on 25 April 2018. For a full list of changes between the latest public release and the one used in this project, see Appendix F. This list was provided courtesy of the FMI's SILAM development group.

### **2.1.6 Required Inputs**

In order to begin a SILAM run, the following need to be provided:

- Control file to define the model run
- Output configuration file
- Cocktail file(s)
- Meteorology input, e.g. Enviro-HIRLAM
- (optional) Descriptor files for emission sources
- (optional) Any datasets to define initial conditions, e.g. gas concentrations that are known at the beginning of the model run

The control file is used to specify the following:

- The model start time, end time, and time step
- The spatial domain, including the location in latitude/longitude coordinates, the number of horizontal grid boxes, and the vertical layers
- Specification of the path (either relative or absolute) for the meteorological forecast data
- MPI (Message Passing Interface) parallel parameters, if desired to run in MPI mode; to run without the MPI, the value 1 should be specified here
- Dispersion parameters, which define the algorithm used when determining the dispersion of gases and aerosols
- Specification of whether SILAM is to be run in Eulerian or Lagrangian mode
- Output parameters, including where to put output files and how to save the data, which includes the following, which may be different than during the run:
  - output time step
  - output layers
  - Map information, either as lat/lon or specific projection information
  - Resolution
- (optional) Definition of any emissions desired; in this project, this is where the BVOC emission algorithm defined
- (optional) Definition of initial conditions; if specified, the datasets containing initial conditions mentioned above need to be specified here
- (optional) Specification of deposition algorithms
- (optional) Specification of optical properties in the atmosphere

Additionally, the NetCDF libraries and include files defined when SILAM was compiled must be available at the time of running the model. This project uses the NetCDF4 software modules, which is the latest version of NetCDF. However, NetCDF3 would be suitable as well.

### **2.1.7 Running SILAM in the CSC Computing Environment**

In this project, SILAM has been compiled and run on the Centre for Scientific Computing (CSC) supercomputers. CSC is closely working with the University of Helsinki researchers. CSC provides scientific computing services and access to their supercomputers to many universities and research organizations in Finland.

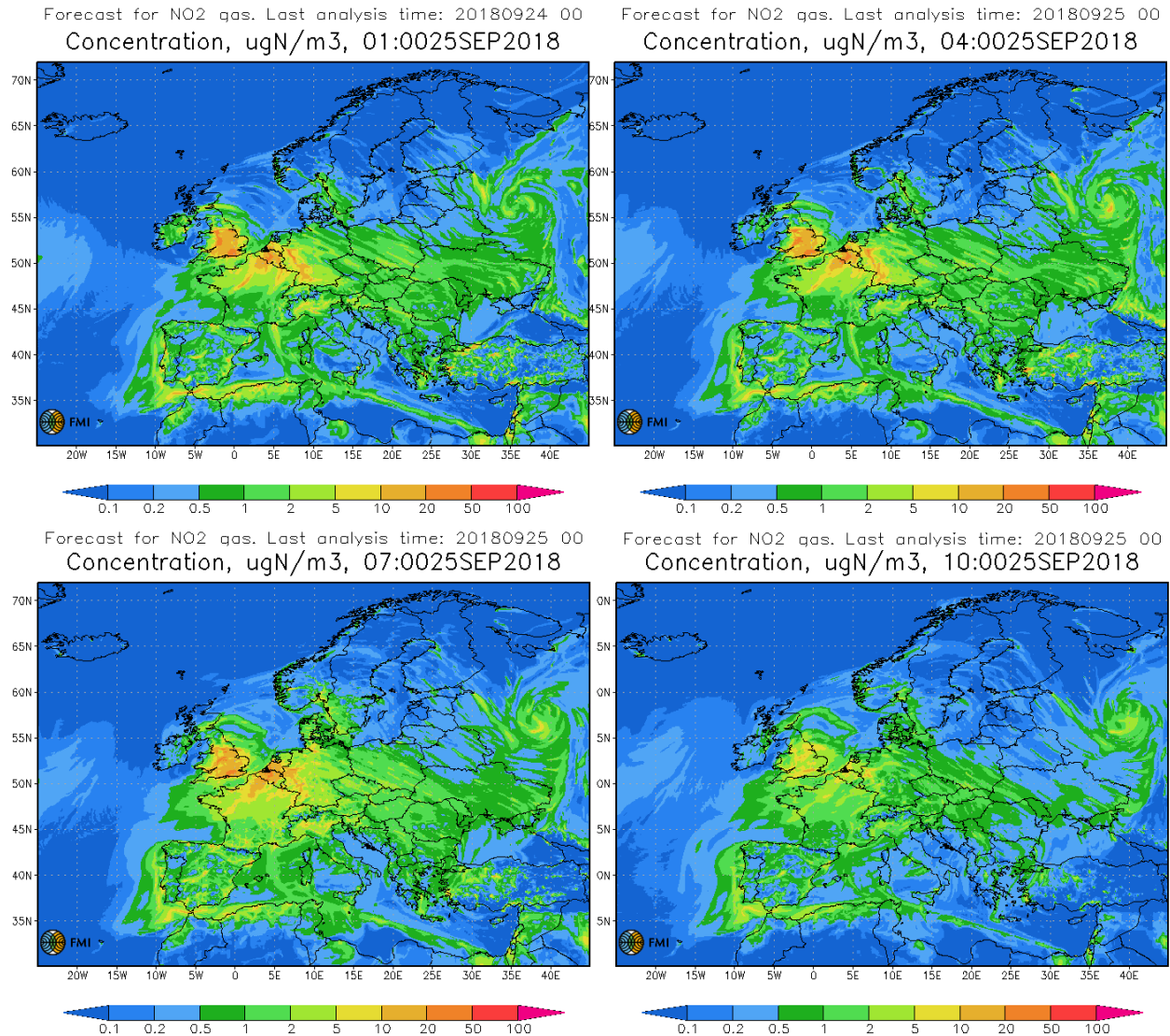
CSC has two high performance computing (HPC) supercomputers, Sisu and Taito. Both of these supercomputers are suitable for installation and running the SILAM model. They have slightly different compile and build environments from each other, but SILAM has been built on both machines. The build files and source paths had to be modified for the specific computers, rather

than for FMI's environment and paths. Note that both FMI and CSC are using the CRAY-XC40 supercomputers which helped simplify the transition of the model code between the FMI's and CSC's environments.

### **2.1.8 Example of SILAM Run**

Public forecasts over Europe and Southeast Asia are available on SILAM's website ([silam.fmi.fi](http://silam.fmi.fi)) for seven pollutants: CO, NO, NO<sub>2</sub>, O<sub>3</sub>, SO<sub>2</sub>, PM<sub>10</sub> and PM<sub>2.5</sub>.

An example of a public forecast from SILAM, showing NO<sub>2</sub> concentrations over Europe is shown in Figure 2 below.



**Figure 2:** An example of a public SILAM forecast. This particular example shows NO<sub>2</sub> concentrations over Europe simulated at 00:00 UTC on September 25, 2018, with forecasts for 01:00 (top-left), 04:00 (top-right), 07:00 (bottom-left), and 10:00 (bottom-right).

## 2.2 Enviro-HIRLAM

For this project, the High Resolution Limited Area Model (Enviro-HIRLAM) model was used to generate the meteorological input to the SILAM model runs, which is one of SILAM's options of inputs. Enviro-HIRLAM is based on the HIRLAM meteorological model but additionally can include, for example, the effects of aerosols when performing the weather forecasts.

### 2.2.1 Overview of Enviro-HIRLAM

Enviro-HIRLAM is fully online integrated numerical weather prediction (NWP) and atmospheric chemical transport (ACT) modelling system for research and forecasting of joint meteorological, chemical and biological weather (Baklanov et al., 2017). The most recent version (v.7.2) of the model is used. It is developed by the HIRLAM Consortium in collaboration with universities in Denmark, Finland, the Baltic States, Ukraine, Russia, Turkey, Kazakhstan, and Spain.

Although the Enviro-HIRLAM model can simultaneously simulate meteorological and atmospheric composition fields, this project has only considered using Enviro-HIRLAM generated 3D meteorological fields output for purposes of required meteorological input for SILAM.

The meteorological part of the Enviro-HIRLAM model is based on the original HIRLAM NWP model (Undén et al., 2002). The Enviro- part of the model includes components typical for the atmospheric chemical transport modelling. In particular, in the Enviro-HIRLAM model there are modules for pollution transport-dispersion-deposition, chemistry and aerosols, and atmospheric composition feedbacks. The aerosol multi-compound approach, aerosol feedbacks on radiation and cloud microphysics are included (see more details and references in Baklanov et al., 2017).

Compared to any independent NWP model run and output, the Enviro-HIRLAM model has the ability to include the effects of aerosols in its algorithms. In this study, there are three different modes available:

- Control/Reference mode – no aerosol effects;
- Direct aerosol effects – includes direct radiative effects of aerosols in weather forecasts;
- Indirect aerosol effects – includes indirect effects of aerosols on weather forecasts;

### 2.2.2 Use of Enviro-HIRLAM output in SILAM Runs

One component of this study was comparing the weather forecast output from Enviro-HIRLAM, used in SILAM, with aerosol effects included when performing the numerical weather forecast.

The purpose of using Enviro-HIRLAM rather other numerical weather model output is that Enviro-HIRLAM includes the option of aerosol effects, as described above. This gave us the ability to understand how the different modelled aerosol effects can influence BVOC emissions in SILAM.

## 2.3 Current Implementation in SILAM – Guenther Modified

First, the only two BVOCs currently included in SILAM are isoprene and monoterpene. The calculation is based on the algorithm in Guenther et al, 1995, hereafter referred to as “Guenther Modified” (thereafter, GM). GM is a simplistic calculation which uses a land classification map, an emissions lookup table, and a few simple equations. In the current implementation, the only parameters taken into account in the equations are temperature, month of year, and solar radiation.

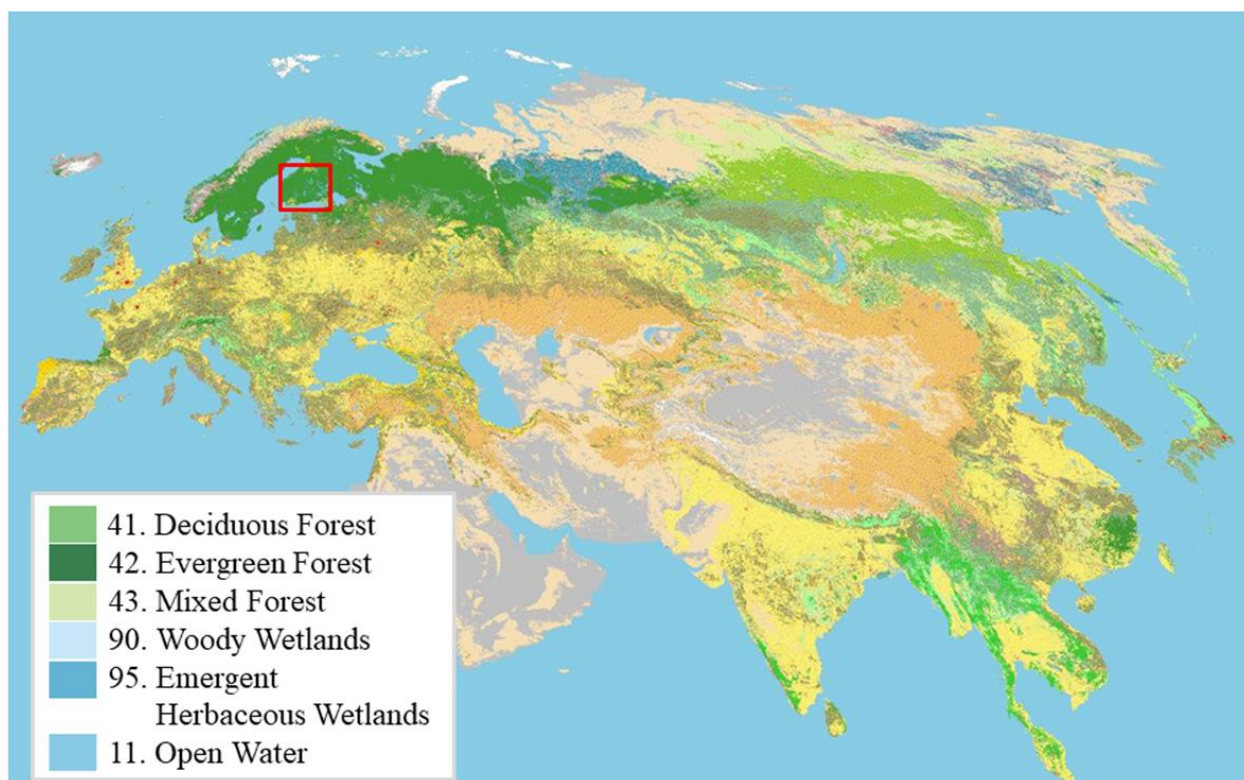
The current code in SILAM includes a block of code that serves as a “placeholder” for selecting other options of BVOC emission. The intent is that the type can be specified in the SILAM control file. However, as of the current version, the placeholder is empty and not yet implemented. Currently, GM is the only valid option in the control file.

Based on this placeholder, it is evident that the developers of SILAM have the intention of implementing a newer advanced algorithm for BVOC emission.

### 2.3.1 USGS Eurasia Land Cover Map

The land classification in SILAM is based on the United States Geological Survey (USGS) Eurasia Land Cover Characteristics (ELCC) Data Base Version 2.0, which is a map with 254 different potential classifications with a resolution of 1km x 1km (USGS ELCC, 2018). For each grid point, SILAM references this map to determine the land classification identifier. From there, it takes the classification and references a lookup table, using three parameters (temperature, solar radiation, and relative humidity), and determines the isoprene emission.

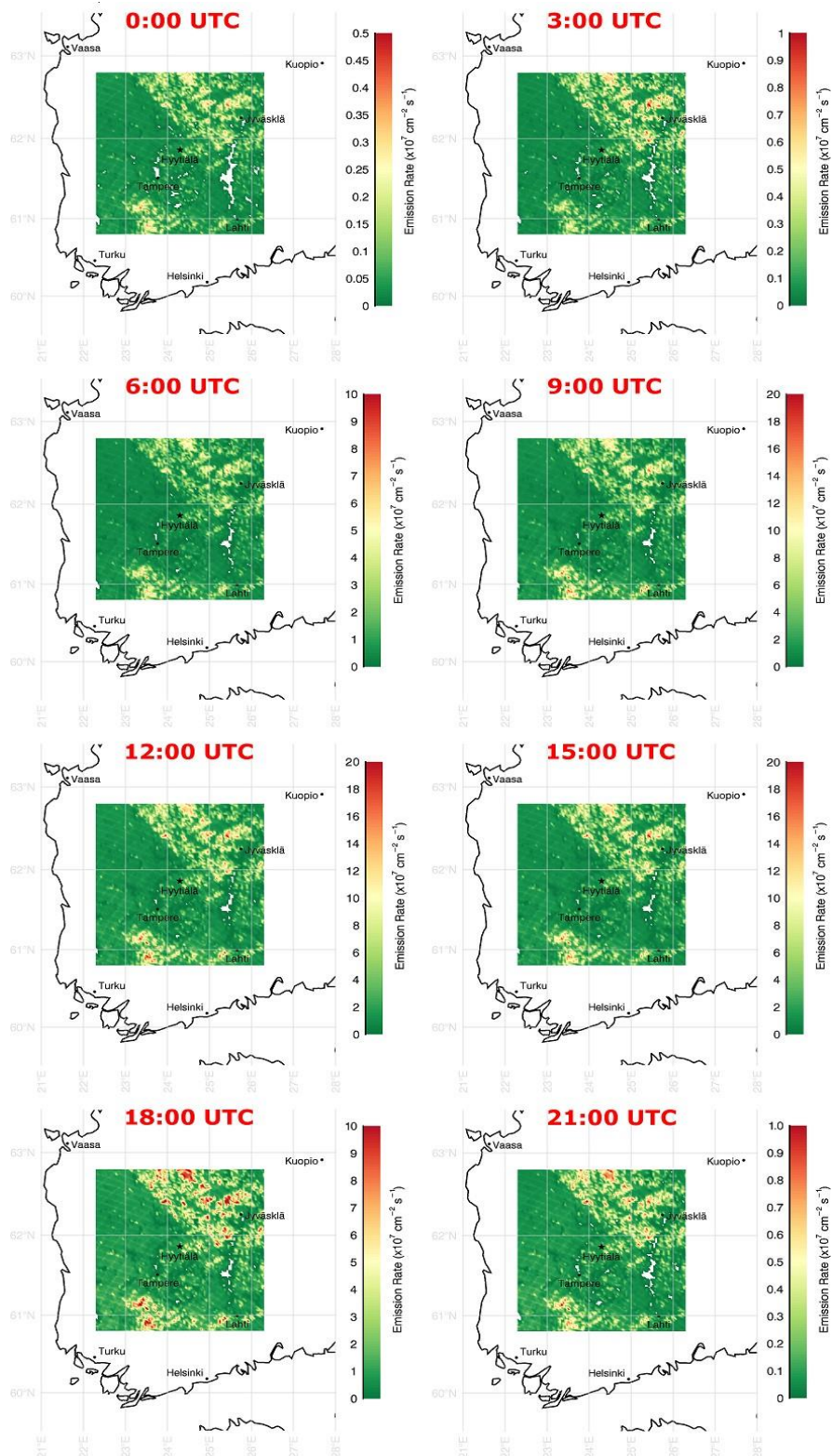
This map is shown in Figure 3, with a few of the classifications relevant to Finland shown in the legend. The southern Finland region is outlined in red (note that this is simply a reference and is not the same as the grid run for SILAM in this study). The map shown here is colorized to RGB for ease of visually interpreting the file.



**Figure 3:** US Geological Survey (USGS) Eurasia Land Cover Database Version 2.0. The database includes 254 land cover classifications, and the legend shows a few of the classifications applicable to Finland.

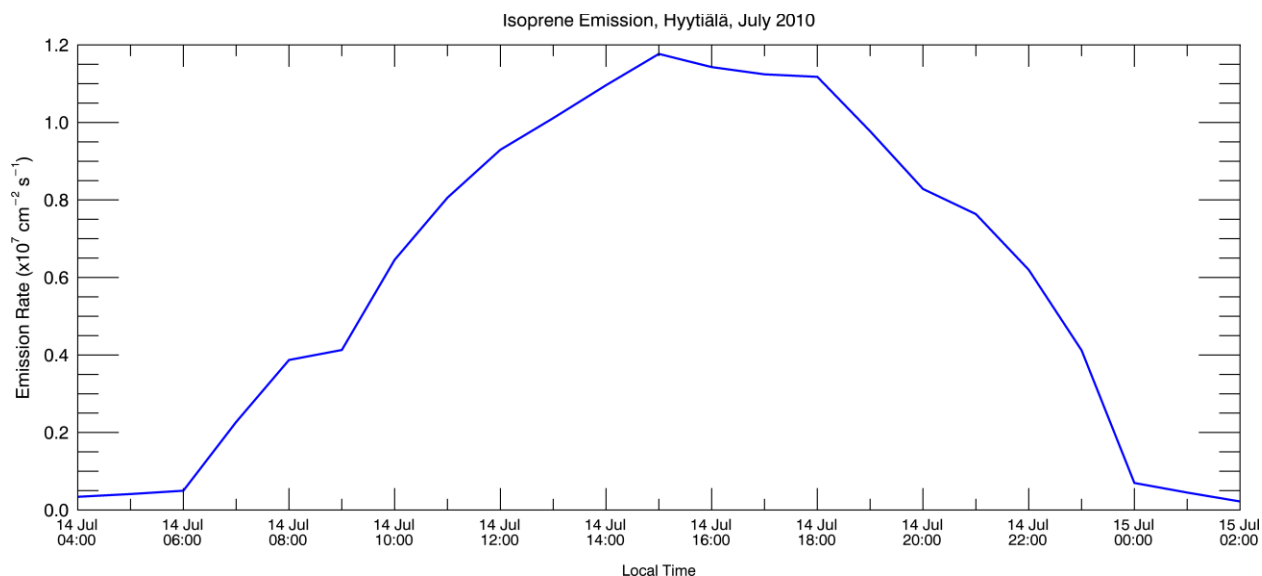
### 2.3.2 Results from Guenther Modified in SILAM

At the beginning of this project, test runs of SILAM (using meteorology from the Enviro-HIRLAM control/reference mode as the input) on 14 July 2010 were performed as a general validity check and to determine the effectiveness of the GM-algorithm in the boreal forest of the Hyytiälä station region. Preliminary results of isoprene and monoterpene emission rates are shown in Figure 4 below. The SILAM model run to generate this figure was performed using 10 minute time steps, with one hour emission rate forecasts.



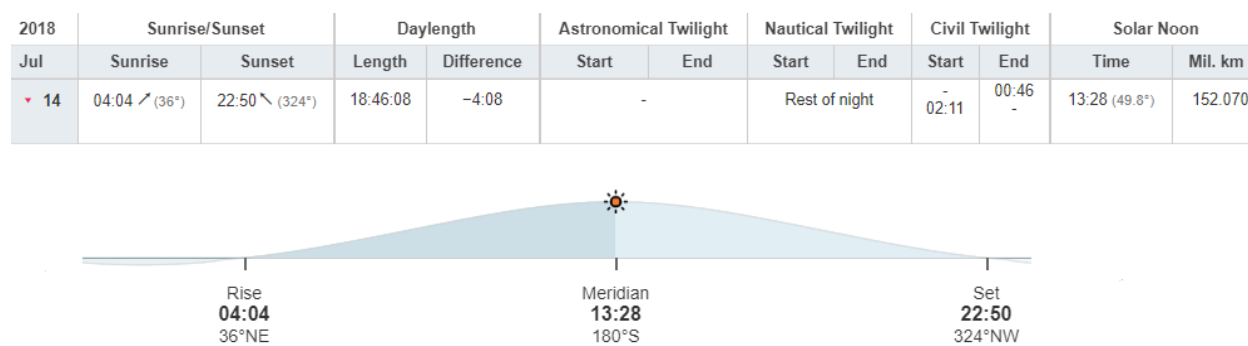
**Figure 4:** Isoprene emission rates calculated by one-hour SILAM runs (using meteorology from Enviro-HIRLAM control/reference mode) on 14 July 2010. In this run, SILAM uses the USGS Land Cover Classification to determine the emission rates.

This model run was performed for the entire day, using the same parameters described above. Emission rates of isoprene over Hyytiälä are shown in Figure 5 below.



**Figure 5:** Modelled isoprene emission rate over Hyytiälä, 14 July (shown in Finnish local time, which is UTC+3, taking into account European Summer Time), using SILAM and control/reference mode of Enviro-HIRLAM meteorology forecast.

Overall, this plot illustrates that isoprene is dependent on sunlight, and thus the emission should peak in the middle of the day. This is the expected result. However, the timing of emission rate does not appear to correlate exactly with the daylight hours, shown in Figure 6 below.



**Figure 6:** Astronomical information for 14 July 2010 at Hyytiälä (source: timeanddate.com).

Figure 6 is sourced from the website timeanddate.com with the input of Hyytiälä (61.847°N, 24.294°E), 14 July 2010. Data in this figure are also shown in local time.

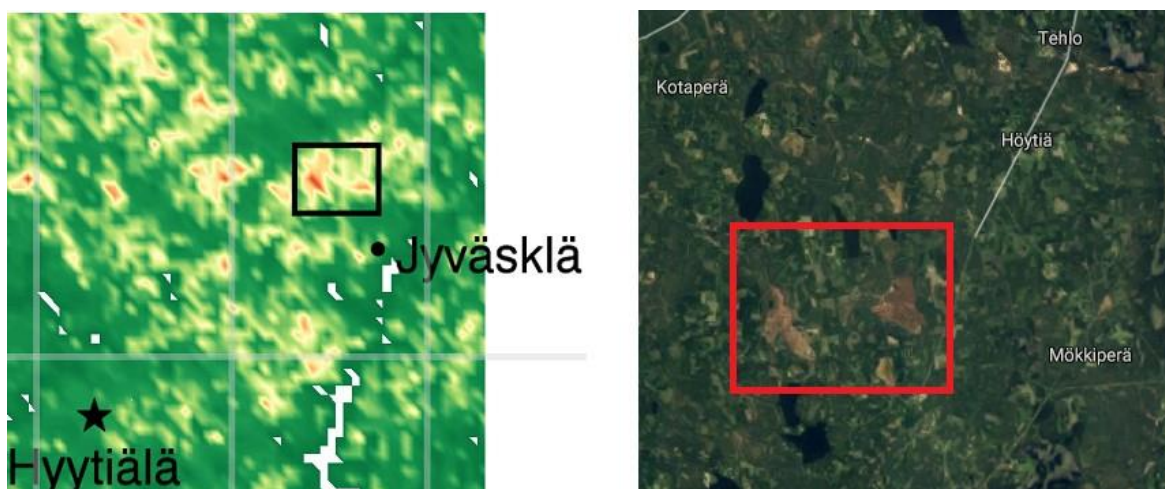
At the beginning of the day, the emission rate in Figure 5 lines up well with the solar information in Figure 6. Sunrise is at 4:04 am, and the modelled emission rate appears to increase after this time. However, solar noon is at 13:28, yet the peak emission rate in the model is observed later at 15:00. Furthermore, the model shows significant emission all the way until midnight, which extends after the sunset at 22:50. This shift between emission rate and solar time indicates that SILAM is possibly making overly generic assumptions leading to this result.

Nonetheless, even though the alignment is slightly shifted from solar noon, the overall diurnal pattern is, in general, what would be expected.

The model shows a very small amount of isoprene emission even at night. That is reasonable, however, since there is twilight all night in July, and thus it is very possible to have a small amount of nighttime emission.

### 2.3.3 Results Based on Land Classification

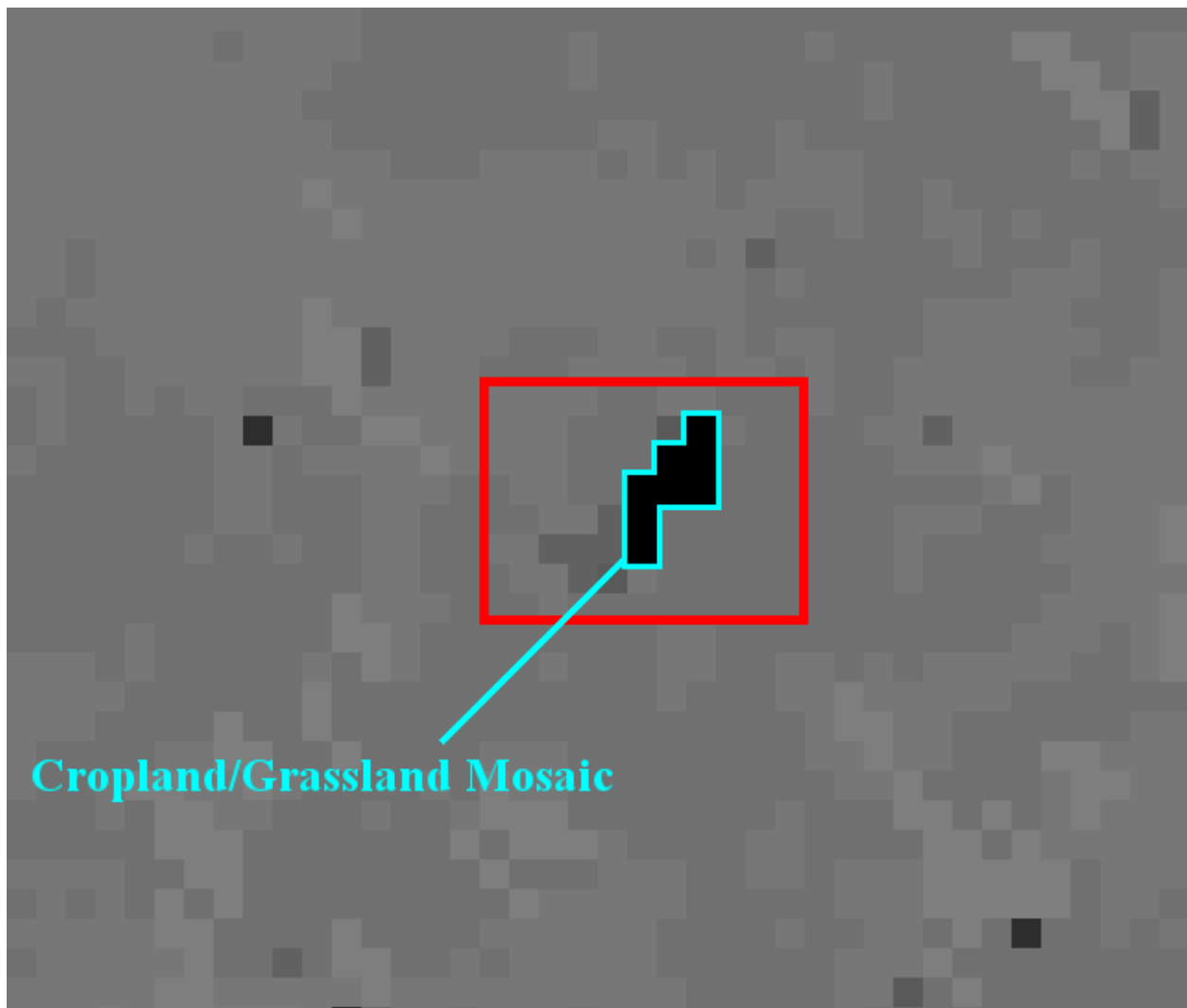
When analysing spatial distribution of the modelled isoprene emissions in Figure 4 in Section 2.3.2, it was found that emissions appear to be significantly higher in value over large agricultural areas. One notable example is an area situated approximately 20 km northwest of Jyväskylä, where there are several large agricultural fields, which are significantly larger than the 0.01 x 0.01 degree area (horizontal resolution used in the model runs). This location is outlined in Figure 7 below, with a side-by-side comparison with a satellite image (the left-hand image is from Figure 4 at 00:00UTC, zoomed in on the area of interest). The imagery is sourced from the Landsat satellite and was accessed through Google Earth (imagery date 5 August 2014).



**Figure 7:** Outlined area of interest northwest of Jyväskylä (left), juxtaposed against satellite imagery (right). This location has noticeably high isoprene emissions modelled by SILAM, and the location appears to be a large agricultural field.

Based on analysis and interpretation of the satellite image above (Figure 7-right), the area of interest appears to be composed of several large agricultural farms or cultivated fields.

The land cover file, which was downloaded from the USGS EarthExplorer interface, is shown in Figure 8 below. Note that this is a grayscale GeoTIFF (georeferenced Tagged Image File Format) file, not an RGB image. Each pixel value is a single byte (valid values are in range 1-254), corresponding to a single land use type. This is different than Figure 3, which was colorized for easier visual interpretation.



**Figure 8:** The land cover classification (file downloaded from USGS EarthExplorer) zoomed in over the area of interest northwest of Jyväskylä. The outlined area where SILAM is modelling high isoprene emissions is classified as Cropland/Grassland Mosaic.

When zooming in on the area of interest northwest of Jyväskylä, performing a pixel analysis, and referencing the USGS Land Use Classification Database, it was found that this area is classified as “Cropland/Grassland Mosaic.” Therefore, the land cover class with (agricultural) crops and grass appears to be the contributor to the area of high isoprene emission.

Studies by Steinbrecher et al. (2013) and Guenther et al. (2006) underlined that in general, agricultural areas do in fact emit more isoprene than forested areas, which is on par with modelled emission rates by SILAM.

However, “Cropland/Grassland Mosaic” is a very vague classification and does not take into account variability, such as latitude, length of day, climate, or plant species. As an example, a cropland in a tropical location and a cropland in Finland may have the same classification designated above, but one would not necessarily expect emissions from those two different croplands to be the same. Steinbrecher et al. (2013), for instance, focused mainly on comparing tropical forests to croplands, and at the current moment there is less study and understanding on northern boreal forests.

Another consideration is small-scale farms that are smaller in size than the resolution in the USGS file. It appears that in these cases, the land is classified as forest. The land containing these small farms might emit isoprene at a higher rate than it would if it was fully covered in forest, and SILAM overlooks this.

Furthermore, only lakes larger than one grid cell are classified as “Water Bodies” in the classification file, and smaller lakes do not appear to be properly accounted for. As an example, the lake adjacent to the Hyytiälä field station, which is only 250 meters wide, does not appear to have any classification of water type, rather it is encapsulated by the classification of the surrounding forest. Since there is no isoprene emission from a lake considered in SILAM, this generalization of homogeneity within each grid cell is likely to incorrectly forecast BVOC emissions areas where a small lake is included in the classification of a forest.

## **2.4 MEGAN as Proposed Replacement**

In SILAM, alongside the Guenther Modified code described above, there is a placeholder in the SILAM code where there is the possibility and presumably intention to add new options, e.g. MEGAN, for calculating BVOC emissions. This is done so using Switch/Case statements. Currently, Guenther Modified is the only option, but others can be easily added. This design that is already in SILAM makes implementing MEGAN inside of this placeholder convenient.

MEGAN was considered in this study because it is a comprehensive BVOC emission model, which includes state-of-the-art research, and it is one of the most detailed models of its kind in

the world. Additionally, the University of Helsinki has close ties with this model and has integrated MEGAN into other models.

### **2.4.1 Overview of MEGAN**

The Model of Emissions of Gases and Aerosols in Nature (MEGAN) is a high-resolution modelling system for estimating emissions from terrestrial ecosystems. This includes BVOC emission, which is of interest for this project. MEGAN takes in several parameters for its calculations, and it can be run at 1 km resolution (Guenther et al., 2006). MEGAN3 calculates over 200 chemical compounds; however, this project focused only on isoprene and monoterpene.

#### **2.4.1.1 History and Current Version of MEGAN**

This project uses MEGAN version 3.0 (MEGAN3), the newest release. Code and documentation for MEGAN3 are located at <https://bai.ess.uci.edu/megan/versions/megan3>. MEGAN3 was released in July 2017 as a beta version and officially released in March 2018. The first study using MEGAN3 was published in June 2018, which used MEGAN3 to model the effects of drought on isoprene emission in the Ozark Mountains of Missouri, USA (Jiang et al 2018).

The previous version of MEGAN was version 2.1, which was introduced in 2012 (Guenther et al 2012). The previous version to that was MEGAN 2.04, introduced in 2007. These versions can be found at the following website from Washington State University:  
<http://lar.wsu.edu/megan/guides.html>

MEGAN is now developed at the University of California at Irvine, and the latest code and user guides are available at the following website: <https://bai.ess.uci.edu/megan/versions/megan3>

In addition to various updates to the emissions algorithms, new input tables and new chemical species added to the model, one of the main differences between MEGAN3 and the previous version, MEGAN 2.1, is that part of the code, the Emission Factor Processor (EFP), has been rewritten in the Python language. Therefore, it must be run separately between the MEGAN3 pre-processor and the main MEGAN3 code. In MEGAN2.1, the preprocessing, emission factor calculation, and emission algorithms were all performed in a single program.

MEGAN3, which is the latest version, was used in this project and is proposed as a possible improvement to SILAM.

### 2.4.1.2 MEGAN Model Components

MEGAN3 is divided into three separate components (listed here in order by calling sequence):

- Pre-processor - Regrids input data for growth form, crops, grass, shrub, total tree, needleleaf tree fraction, tropical tree fraction, and LAIv into the selected spatial and temporal domain. In this case, this is the domain for running the SILAM model. LAIv was downloaded and processed from the MODIS satellite (see Appendix B), and the others are available from the MEGAN3 website.
- Emission Factor Processor (EFP) - Calculates the emission factors based on the regridded output of the pre-processor and creates tables that are then passed into the MEGAN3 Emission Calculator. Unlike the Pre-processor and Emission Calculator, the EFP is written in the Python language. Although this means the EFP does not require compilation and is easier to run than Fortran code, it requires the system to have Python already installed. Additionally, the Python code requires the geo-env module to be loaded on the system.
- Emission Calculator - Estimates emissions for over 200 chemical species, using information from the two components above.

MEGAN3 expects meteorology forecast input from either the Weather Research and Forecasting Model (WRF; NCAR, USA) or the Fifth Generation Mesoscale Model (MM5; Penn State University and NCAR, USA).

MEGAN3 is not designed to accept meteorological forecast as input from other NWP models. A challenge here is that in order to use MEGAN3 with the Enviro-HIRLAM input, the meteorology input must first be rewritten in a NetCDF file which contains the same variables names, attributes, dimensions, and file structure as a WRF file. In other words, the forecast data had to be copied into a new file that appears to be a WRF file from the viewpoint of MEGAN3. This “mimicked” WRF file is then passed into MEGAN3. Additionally, some modifications to the MEGAN3 code were made for this project so that MEGAN3 recognized the file, reformed arrays as necessary, etc. See Appendix C for more technical details on this file conversion.

Other necessary inputs for MEGAN3 include global coverage files containing data tables of crops, grass, needleleaf tree fraction, tropical tree fraction, and shrubbery. The ozone database was also downloaded from the MEGAN3 website without modification. These global files were downloaded from the MEGAN website were not modified for this project. Other input parameters and settings for MEGAN3 were left as their default.

Additionally, MEGAN expects the WRF or MM5 data to be processed through the Meteorology-Chemistry Interface Processor (MCIP) before being passed into the MEGAN3 main code. See Otte et. al. (2010) and Byun et. al. (1999) for details of MCIP.

## **2.4.2 MODIS Input to MEGAN**

Moderate Resolution Imaging Spectroradiometer (MODIS) datasets were used for the Leaf Area Index (LAI) and Vegetation Cover Fraction (VCF, not to be confused with Vegetation Continuous Fields, an unrelated MODIS product) input into the MEGAN pre-processor. Before providing this data directly into the pre-processor, LAI from MODIS must be both regridded into latitude/longitude coordinates and also converted into LAIv. LAIv is defined as the LAI divided by the VCF. The LAIv represents actual total leaf area at the location, whereas raw LAI is only relative. The calculated LAIv data was then saved into a new NetCDF file that is then provided to the MEGAN pre-processor.

Regridding was done based on the equations in Subirana et. al. 2011.

The reason this calculation was performed from the data downloaded from MODIS, rather than the University of Maryland's Global Land Surface Satellites (GLASS) website (<http://glcf.umd.edu/data/lai/>), per suggestion on the MEGAN website, is because the GLASS data is only available for North American domains.

The MODIS data used for this project is an 8 day average over the period of interest, in this case 12-19 of July, 2010.

For a detailed description of the MODIS input data and its source, see Appendix B (Description of MODIS Data & Calculation of LAIv from MODIS).

For more detailed description of the calculation that was performed to determine the LAIv and the MODIS latitude/longitude points from this data, including a description of the NetCDF files that were generated for the MEGAN input, also see Appendix B.

## **2.4.3 J-Rating**

One of the options in the MEGAN EFP is the J-rating. This is a form of confidence checking, which is based on the input MEGAN databases. A J-rating of 4 is the highest confidence, and the EFP will only use data with high confidence. A J-rating of 0 is the lowest confidence, and the EFP will use all data.

For this project, 0 was used as the J-rating so that all input data is used in the desired domain.

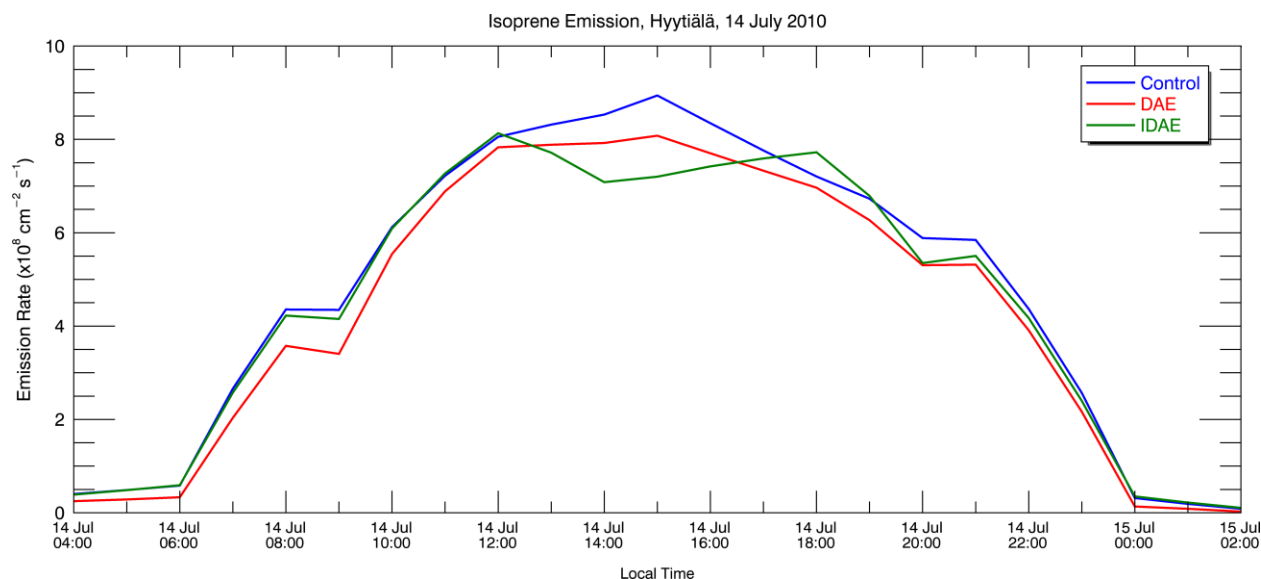
## 3. Results and Discussion

### 3.1 Findings in the Current SILAM Version

Using the existing SILAM distribution, which uses the GM algorithm for calculating BVOC emissions, models were run first comparing the SILAM emissions with different meteorological input from Enviro-HIRLAM, and then a comparison to measured BVOC flux data was performed.

#### 3.1.1 Comparison of Enviro-HIRLAM modes in SILAM

The first result is a comparison between the aerosol modes of Enviro-HIRLAM used for the meteorological forecast input to SILAM. This is based on the same parameters as in Figure 5 in Section 2.3.2, with the only thing different being the Enviro-HIRLAM mode. This comparison is shown in Figure 9 below.



**Figure 9:** Isoprene emission rate over Hyytiälä on 14 July 2010 modelled by SILAM using the following Enviro-HIRLAM modes for the meteorological input: Control/Reference (blue), Direct Aerosol Effects (red), and Indirect Aerosol Effects (green). All other parameters are the same. Times are in Finnish local time.

Overall, the results from each of these three SILAM model runs are very similar to each other and follow a very similar overall pattern on a diurnal cycle, except a few differences. The main difference noticed visually in Figure 9 is that at the time of peak emissions in the

control/reference run, the runs using the meteorological forecast with aerosol effects included have noticeably lower emission rates than in the control/reference case.

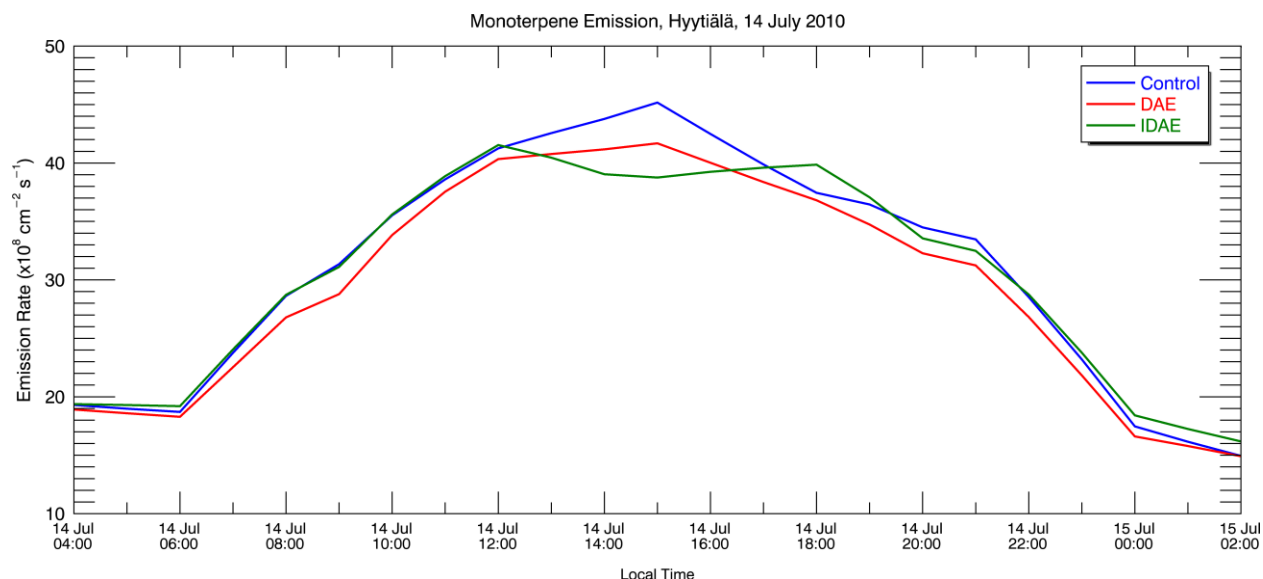
On this particular day at 15:00, the isoprene emission rate in the DAE case was 9.6% lower than the control/reference case, and the emission rate was 19.5% less in the IDAE case.

This is to be expected because the presence of aerosols (in the DAE case) and possible clouds (in the IDAE case) would block/reflect some of the sunlight, thus resulting in lowering emission rates than if the aerosols didn't exist. Additionally, cloud formation typically begins approximately midday, and thus this is when there is expected to be a difference between models that do and don't take indirect aerosol effects into account.

Furthermore, the isoprene emission rate in the DAE case is less throughout the day, not just during peak midday hours. For instance, it is 21.7% less at 09:00 local time and 15.7% less at 23:00 local time. This indicates the continuous presence of aerosols, which reflect/block sunlight, thus reducing the emission rate of isoprene from the forest throughout the day.

However, this is not nearly as much the case in the IDAE case, where there is likely not the presence of clouds during the morning and late evening. Isoprene emission in the IDAE case is only 4.4% less at 09:00 and 6.6% less at 23:00

Figure 10 below shows the same plot for monoterpene emission rates from the same SILAM model runs.



**Figure 10:** Monoterpene emission rate over Hyytiälä on 14 July 2010 modelled by SILAM using the following Enviro-HIRLAM aerosol modes for the meteorological input: Control/Reference (blue), Direct Aerosol Effects (red), and Indirect Aerosol Effects (green). All other parameters are the same. Times are in Finnish local time.

This figure shows that monoterpene emission follows a nearly identical pattern in emission rates as isoprene. The aerosol effects result in somewhat lower emission rates during the peak hours of the day. In this case, the difference in emission rate was as follows:

- DAE:
  - 8.2% less at 09:00
  - 7.7% less at 15:00
  - 5.9% less at 23:00
- IDAE:
  - 0.8% less at 09:00
  - 14.2% less at 15:00
  - 2.4% **more** at 23:00

Therefore, monoterpene emission rate is also decreased throughout the day due to aerosol presence in the DAE case. In the IDAE case, however, as with isoprene, the effects caused by clouds are only significant at midday, indicating the presence of midday clouds but not in the morning or evening.

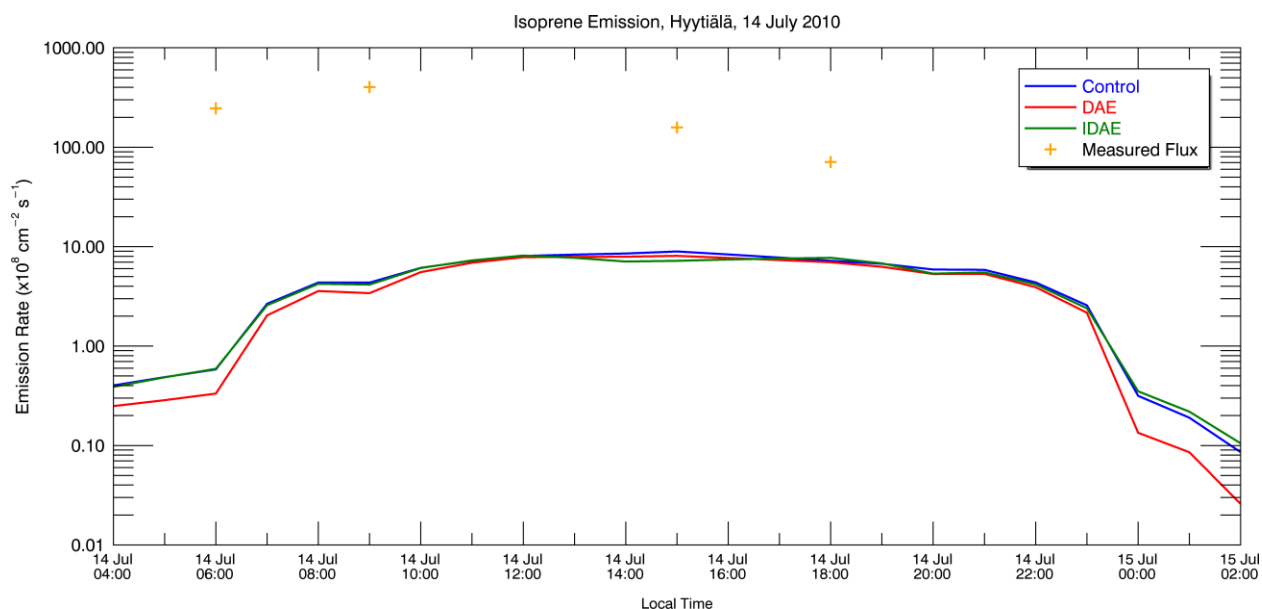
### 3.1.2 Comparison of current SILAM implementation with GM algorithm to measured BVOC flux data

Flux data of isoprene and monoterpenes were collected once every three hours at SMEAR-II. In order to determine the accuracy of the SILAM model with the GM algorithm, the measured data have been compared to the modelled data shown above.

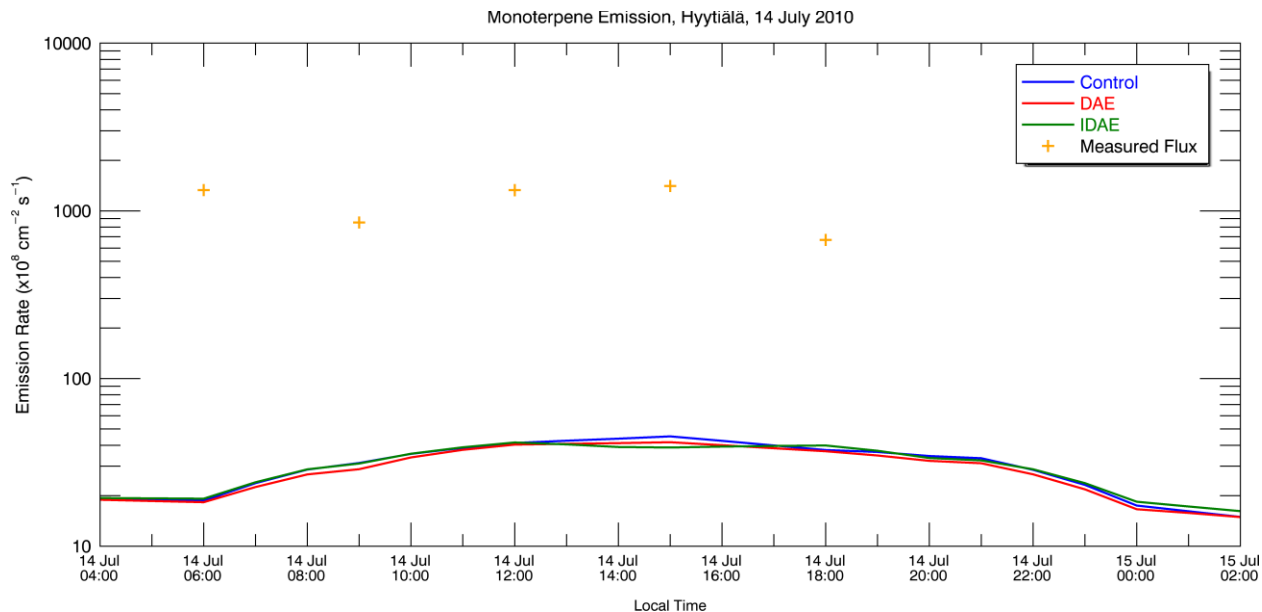
Although flux is somewhat different from actual emissions, flux is a reasonable approximation for emission. Flux is measured at the top of the canopy rather than on individual trees. There are possibly some small losses within the canopy due to deposition and chemical reactions. However, deposition is usually orders of magnitude smaller than emission, and the chemical lifetime of both isoprene is typically on the order of hours. Thus, it is assumed in this study that the losses within the canopy are significantly small compared to the emission. Hence, flux measurement is a reasonable estimate for actual emission rates at the location.

It's worth noting that this study, along with many other studies, assumes that Hyytiälä is a perfectly homogeneous forest, and thus flux measured at the SMEAR-II station represents the entire 0.01 x 0.01 degree grid cell surrounding the site.

Figure 11 shows the modelled isoprene emission rates (same as in Figure 9), with measured isoprene flux data plotted on top of it, plotted on a y-logarithmic scale. Likewise, Figure 12 shows modelled monoterpene emission rates (same as in Figure 10), with measured monoterpene flux plotted on top of it, also plotted on a y-log scale.



**Figure 11:** Isoprene emission rate (same as Figure 9 above), with isoprene flux data measured once every three hours plotted in orange on top of it.

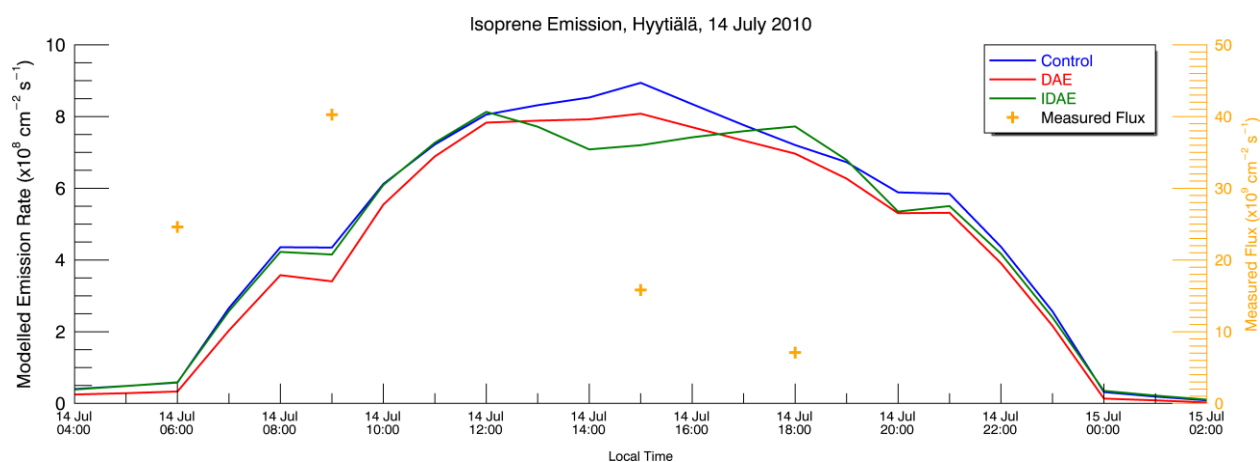


**Figure 12:** Monoterpene emission rate (same as Figure 10 above), with isoprene flux data measured once every three hours plotted in orange on top of it.

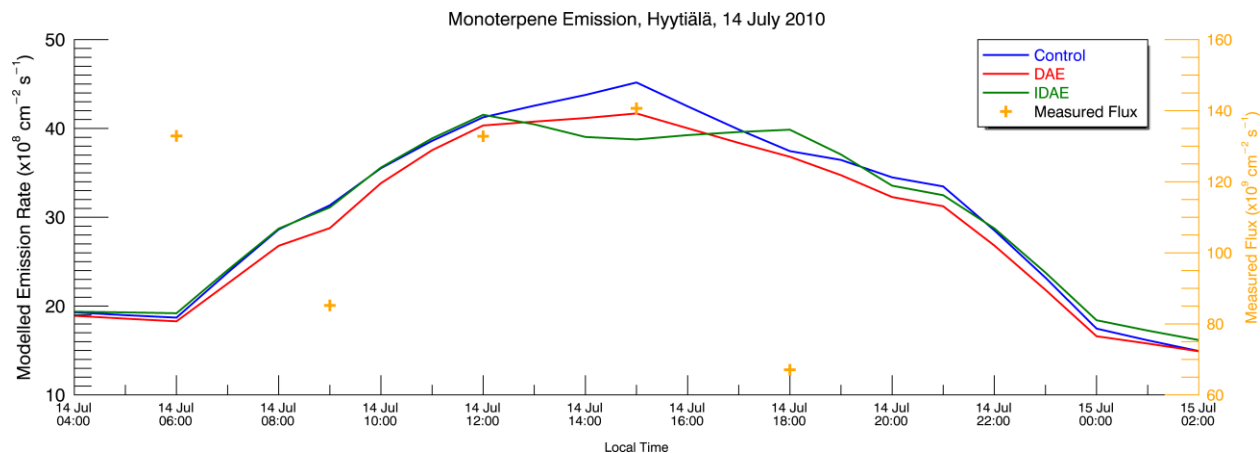
The flux measurements for both isoprene and monoterpene are one to two orders of magnitude larger than the modelled results, so the measurements unfortunately dwarf the modelled results. Therefore, it is evident that the SILAM modelled results using the GM algorithm are significantly underestimating the emission rates at this geographical location for the specified summer period of July 2010.

Note that these flux measurements from SMEAR-II are the same that are used in Zhou et al 2017.

Figures 13 and 14 below show the same data as Figures 11 and 12, respectively, on a dual-axis scale. This outlines the relationship between the diurnal cycle of the modelled versus measured data.



**Figure 13:** Isoprene emission rate and measured isoprene flux (same as Figure 11 above), plotted on dual-axes, with modelled results on the left axis and measurements in orange on the right axis.



**Figure 14:** Monoterpene emission rate and measured monoterpene flux (same as Figure 12 above), plotted on dual-axes, with modelled results on the left axis and measurements in orange on the right axis.

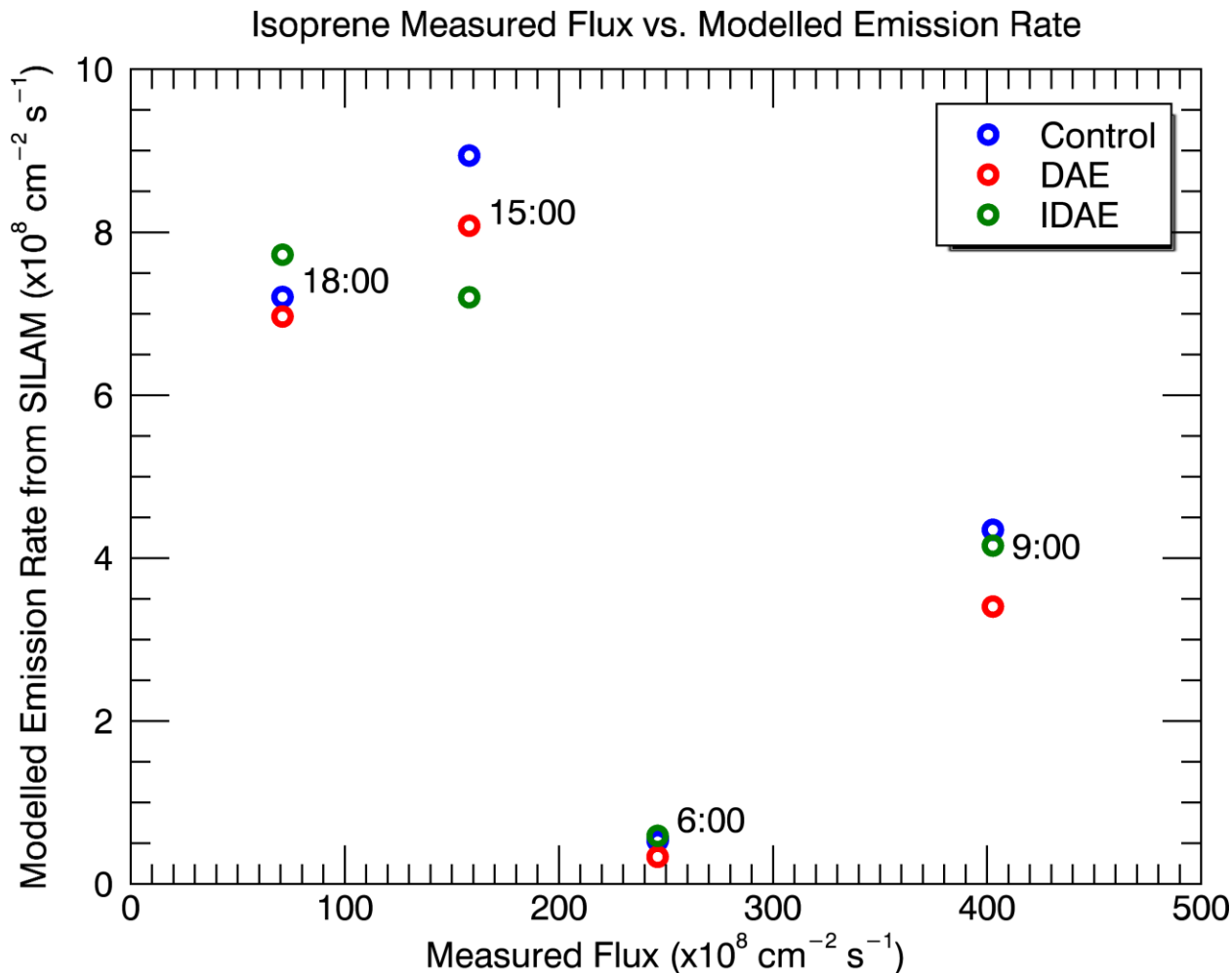
In Figure 13, there appears to be increasing isoprene flux measurements as the morning progresses, followed by a decreasing emission rate through the afternoon into evening. Unfortunately, the noontime measurement of isoprene was not available on this day, but nonetheless, the peak emission rate appears to likely be earlier than 15:00 as predicted by the models.

On the other hand, in Figure 14, measurements of monoterpene flux show the peak is at 15:00, which agrees with the models in terms of pattern at midday.

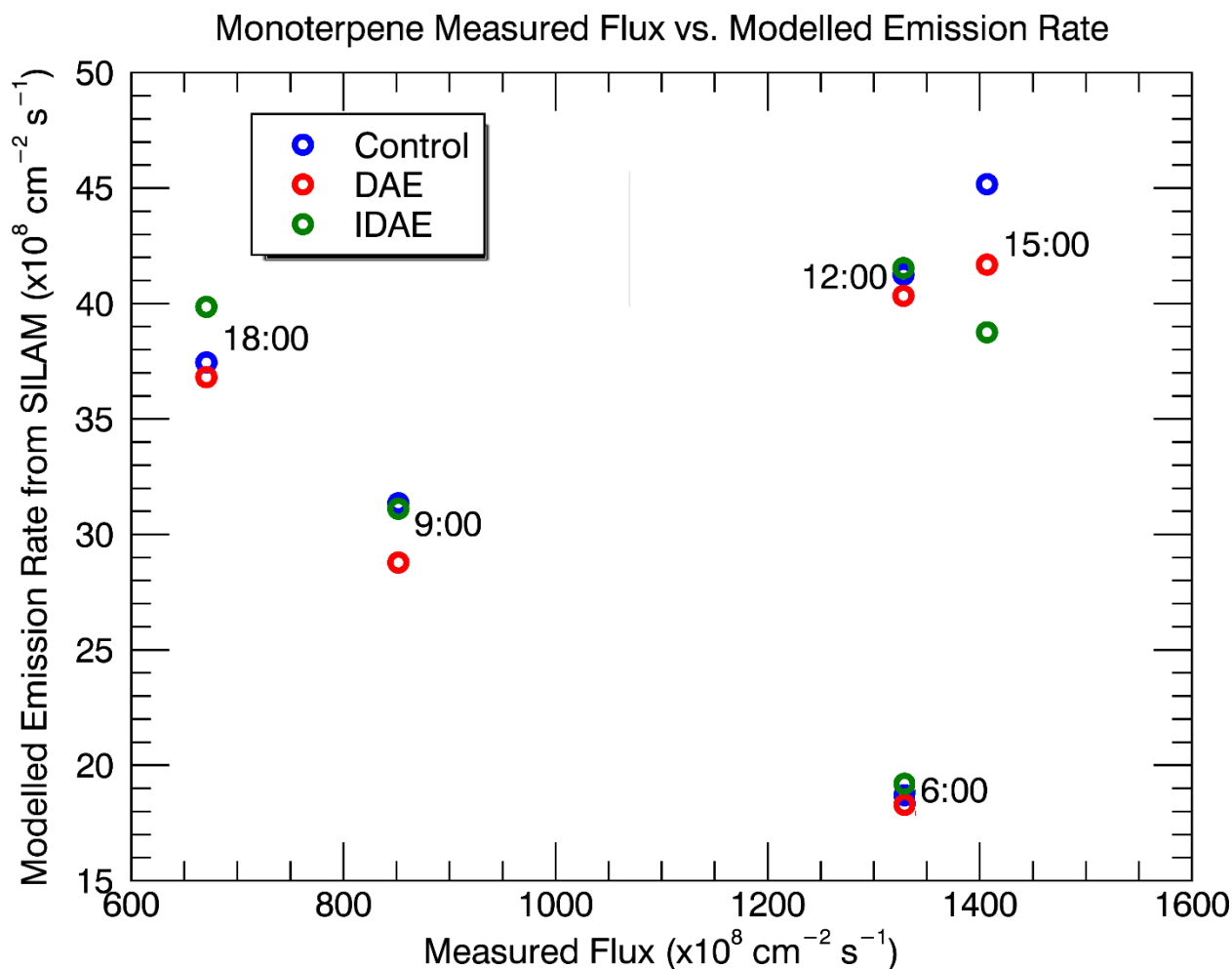
In both the isoprene and monoterpene flux measurements, the sudden drops could be due to the presence of clouds, in particular in the mid and late afternoon.

Nonetheless, these observations are based on only four data points for isoprene and five data points for monoterpene over the course of a single day.

Figures 15 and 16 below show scatter plots for measured flux vs. modelled emission rates.



**Figure 15:** Scatter plot of measured isoprene flux vs. isoprene emission rate modelled by SILAM using the Control/Reference, DAE, and IDAE modes of Enviro-HIRLAM meteorological forecast data Hyytiälä on 14 July 2010.



**Figure 16:** Scatter plot of measured monoterpene flux vs. monoterpene emission rate modelled by SILAM using the Control/Reference, DAE, and IDAE modes of Enviro-HIRLAM meteorological forecast data Hyytiälä on 14 July 2010.

Figures 15 and 16 above show that modelled emission rate of both monoterpene and isoprene has the lowest differences between the results from the different Enviro-HIRLAM modes in the early morning and the highest difference at 15:00. Additionally, at 15:00, the modelled emission rates for both compounds are noticeably lower in the DAE and IDAE cases, with the IDAE case being the lowest. As with Figures 9 and 10, this indicates that the DAE in the models estimates a lower emission rate due to aerosols blocking radiation during the day. Likewise, the lower values in the IDAE case indicate the likely presence of clouds, which also blocks solar radiation, resulting in lower emission rates.

There is no apparent correlation in either of these scatter plots, indicating that there is not a specific bias in the model. However, there are numerous variables in the real-life setting that could influence the measured results.

It was nonetheless noted as a general observation from the plots that the largest differences between the modelled emission rates between the different Enviro-HIRLAM modes was during the midday, and during the daytime the modelled results were closer to the measured fluxes than in the early morning hours, especially for isoprene.

## 3.2 Initial Steps in Integration of MEGAN3 with SILAM

In this project, the following steps have been performed in working towards designing practical approach for an implementation of MEGAN3 in SILAM:

1. **Meteorology:** Weather forecast data from Enviro-HIRLAM's control/reference, DAE, and IDAE modes were collected with a SILAM "dry run" and were provided to the MEGAN3 pre-processor.
2. **Satellite Data:** Data from the MODIS satellite for LAI downloaded, regridded, and provided to MEGAN3.
3. **Pre-Processor:** The MEGAN3 pre-processor was run for the spatial and temporal domain of this project. The output from Steps 1 and 2 above were provided. Output from this step was provided to the MEGAN EFP.
4. **MEGAN3 EFP:** This step was setup and run using the output of the MEGAN3 pre-processor. Other than setting the domain and inputs from the pre-processor, all parameters were left at their defaults.
5. **MCIP:** Meteorology data from Step 1 was provided to MEGAN3 as an alternative to MCIP or MM5. This involved several alterations to the MEGAN3 and associated external code.

Technical details of the steps above are outlined in Appendix E.

This practical approach addresses some of the complex technicalities of integrating the MEGAN3 emission model with the SILAM forecast model. One complication we were able to overcome is the challenge of combining a research model conglomerate with an operational forecast model, both of which are setup and programmatically initiated in much different fashions.

The key difference between these two models is that SILAM, which is an operational model intended to be setup and run on short notice, is driven by a control file, based on a selection of input scenarios and output parameters. In SILAM, the external dependence (e.g. files downloaded for specific spatiotemporal domains, data tables, intermediate files, etc; the exception being meteorology, which is specific to each model run) is intended to be minimal so

that every model run is quick and easy to setup. This on-the-spot approach is necessary for daily forecasts and is especially critical in emergency response situations.

On the other hand, MEGAN3 is a research model, which is not designed with the intention of on-the-spot runs. MEGAN3 not only has heavy dependence on external code (e.g. open source libraries and external models) but also files downloaded and/or created for specific spatiotemporal domains. The steps described above and in Appendix E outline the method needed to compile and arrange SILAM with MEGAN3 included in it. These steps also outline the details needed to set up each MEGAN3 run before initiating the SILAM forecast run.

## 4. Conclusions

### 4.1 Primary Findings

During this study, the following findings were made

- The System for Integrated modelLling of Atmospheric coMposition (SILAM) uses the Guenther Modified (GM) algorithm to calculate isoprene and monoterpene emissions. This is a very simplified algorithm, which is based on MEGAN but is not a full implementation of MEGAN. This algorithm uses the USGS ELCC map combined with three meteorological parameters (temperature, radiation, and relative humidity).
- A comparison of multiple SILAM runs using Control/Reference, DAE, and IDAE in Enviro-HIRLAM as the meteorological input to SILAM showed the following:
  - There was somewhat less emission of both isoprene and monoterpene in the case with DAE input because aerosols will decrease sunlight, and to an extent temperature, resulting in somewhat lower emissions compared to the control case.
  - There was a dip in the midday of the IDAE case. This is likely due to cloud formation of the aerosols, resulting in both less solar radiation and possibly slightly lower temperature, which would result in a drop in emissions.

Moreover, from a practical point of view of the MEGAN3 integration, several key steps were successfully performed:

- The MEGAN3 pre-processor was successfully run using the SILAM dry run by saving the data in the same format as WRF (note that the MEGAN3 pre-processor expects meteorology forecast data from either WRF or MM5, so this had to be mimicked).
- In order to run the MEGAN3 model for a domain outside North America (the GLASS data is only available for North America), LAIv must be manually downloaded, calculated, and regridded from the desired satellite (in this case MODIS). This step was successfully performed, and the MEGAN pre-processor was able to run with the MODIS LAIv input.
- Once the MEGAN3 pre-processor successfully runs with the input above, the MEGAN3 EFP can also run successfully using the global data files downloaded from the MEGAN3 website.
- MCIP, which is one of the required inputs, requires a full run of either WRF or MM5. In order to use the SILAM dry run instead, an alternative will need to be developed, which includes modification to the MEGAN3 code, so that it will not rely on MCIP.

## 4.2 Current Status

At the time of publishing this thesis, the current work has been to develop an alternative in MEGAN3 to MCIP, so that meteorology data can be taken from the SILAM dry run instead from MCIP. This involves adaptation of the MEGAN3 code, along with additional modifications to the I/O API library, which is an open-source external library separate from MEGAN.

Alternatively, the dependence on the I/O API library, which is intended for use in combination with MCIP, could be removed, and MEGAN3 could use direct NetCDF calls instead. This would be a cleaner approach as it wouldn't require modification of the I/O API library, but it would require significant modification to MEGAN3's file-reading subroutines.

## 4.3 Future Steps

MEGAN is a very complex and state-of-the-art model, and it has heavy dependence on external libraries, modules and models, including I/O API, MCIP, and either WRF or MM5. Another complication is that MEGAN3 first requires the meteorological forecast data in NetCDF format. This requires the SILAM dry run.

To run the MEGAN code in a stand-alone mode (i.e. without integration into the SILAM code), a sequence of separate runs will be needed. The first step is to produce an output NetCDF file from the SILAM model containing the interpolated meteorology forecast data. This is required for MEGAN3 as its necessary meteorological input. Second is to run the three steps of MEGAN3 (pre-processor, EFP, and MEGAN3 main code, respectively). A linkage between these two, with writing and reading files connecting each step, can be achieved through integration into common script, where written parts of the codes in different programming languages can be linked through calls to the respective executables.

To run the MEGAN3 code as part of the SILAM modelling, a step-by-step implementation and integration of the code (or parts of the code) into SILAM is required. This process is not a straightforward task, but it is series of subsequent tasks. It includes not only changes and modifications to the SILAM model code, along with outside scripts to interpolate MODIS and meteorology, but also linking and integration of additional datasets required by the MEGAN model. Doing so would add significant complexity to the SILAM model. Future steps to do so would include the steps recommended in the following section.

### 4.3.1 Integration of MEGAN3 into SILAM

First, in order to integrate MEGAN3 into SILAM, the MEGAN3 code will need to be described in Section 4.2 above. Then SILAM will need to be compiled with the MEGAN pre-processor and MEGAN3 main code. The I/O API library would also need to be linked in SILAM because MEGAN3 depends on it. Furthermore, in addition to the Fortran compiler, the Python software must be available on the operational computer system.

Then a script will need to be developed that includes the following steps:

1. Run the SILAM “dry run” to gather only the meteorology over the defined gridded domain and numerical parameters.
2. Call to external code written for this project that interpolates into the SILAM gridded domain and converts the meteorology from GRIB to netCDF-format in the previous step to mimic a formatted file (i.e. file which similar to those files produced by WRF or MM5 models).
3. Call to external code written for this project that calculates and regrids the MODIS LAIv. If the model is run over North America, then LAIv files may be downloaded directly from the GLASS website, in which case this step can be skipped. In principle, this task can be performed during initialisation stage of the SILAM model.
4. Call to MEGAN3 pre-processor to obtain the necessary input data (e.g. LAIv, spatial and temporal domain) and landcover and environmental data for the specific model run. This regrids the base data into “table” files for each grid point in the model. These tables will be used by the MEGAN3 EFP and MEGAN3 main code.
5. Call to the MEGAN3 EFP Python Code to calculate light-dependent factors, based on the specific vegetation defined in the previous step
6. Call to MEGAN3 to calculate the emission rates of up to 200 chemical species based on the previous two steps.
7. Run SILAM again, with the same inputs and parameters as Step 1, now including MEGAN3 specified as the BVOC emission scheme in the control file, with the result from the previous step used as the input to this step.

### 4.3.2 Recommendations

The first recommendation would be that this project should be continued in order to complete the necessary steps to fully implement and integrate MEGAN3 into SILAM. The work needed to be done to perform this task is beyond the scope of this MSc thesis, and it will consist of both advanced technical computer program development and continued scientific work. With additional time and man-hours, this work could be realized in the future.

From there, we could analyse the results compared to the current GM algorithm and recommend whether or not the additional complexity of MEGAN is worth the improved outcome of BVOC emissions from the basis of model and computer cost-effectiveness.

One recommendation based on the complexity of MEGAN is that perhaps GM is a reasonable “Pareto Principle” solution (also known as the 80/20 rule) for modelling BVOCs. In other words, GM is an acceptable option and is better than nothing, but a user would need to consider the cost-effectiveness (both in complexity and computing resources) of a full MEGAN3 implementation inside of SILAM. Some users may consider GM to be a reasonable compromise, and for that reason, it should still be an available option in SILAM.

A consideration would be to use the previous version of MEGAN, which is 2.1. This version does not have the EFP in Python, which would reduce the number of steps in Section 4.3.1. Doing so could make the integration of MEGAN into SILAM somewhat simpler. MEGAN3 is a newer version, which is presumably more accurate; however, it is not only more advanced but also more complicated. Using the previous stable version could be a reasonable alternative solution. Moreover, older versions of MEGAN (or parts of its code) have already been implemented in several models, such as WRF-Chem, CESM, MOZART, GEOS-Chem, MALTE, SURFEX, BRAMS, and others.

Furthermore, once BVOC emissions are updated with adding MEGAN to SILAM, the modelled chemical reactions in SILAM should be reviewed and updated/modified if necessary to avoid double or missing effects. One example of what could be done would be integrating the University of Helsinki Multicomponent Aerosol (UHMA) model. However, this study has been focused only on emissions and has not yet performed any specific numerical analysis of SOA formation. A comparison of chemistry between MEGAN and GM in SILAM runs will provide additional insight into the cost-effectiveness of adding MEGAN to SILAM.

## 5. References

- Baklanov, A., Korsholm, U. S., Nuterman, R., Mahura, A., Nielsen, K. P., Sass, B. H., Rasmussen, A., Zakey, A., Kaas, E., Kurganskiy, A., Sørensen, B., and Gonzalez-Aparicio, I. (2017). “Enviro-HIRLAM online integrated meteorology–chemistry modelling system: strategy, methodology, developments and applications (v7.2)”. *Geosci. Model Dev.*, 16:2971–2999.
- Baklanov, A., Schlunzen, K., Suppan, P., Baldasano, J., Brunner, D., Aksoyoglu, S., Carmichael, G., Douros, J., Flemming, J., Forkel, R., Galmarini, S., Gauss, M., Grell, G., Hirtl, M., Joffre, S., Jorba, O., Kaas, E., Kaasik, M., Kallos, G., Kong, X., Korsholm, U., Kurganskiy, A., Kushta, J., Lohmann, U., Mahura, A., Manders-Groot, A., Maurizi, A., Moussiopoulos, N., Rao, S. T., Savage, N., Seigneur, C., Sokhi, R. S., Solazzo, E., Solomos, S., Sørensen, B., Tsegas, G., Vignati, E., Vogel, B., and Zhang, Y. (2014). “Online coupled regional meteorology chemistry models in Europe: current status and prospects”. *Atmospheric Chemistry and Physics*, 14(1):317–398.
- Bannai, T. (2011). “INES (the International Nuclear and Radiological Event Scale) Rating on the Events in Fukushima Dai-ichi Nuclear Power Station by the Tohoku District - off the Pacific Ocean Earthquake”. *Japanese Ministry of Economy, Trade and Industry, News Release*.
- Bengtsson, G. (1986). “Chernobyl - its impact on Sweden”. *Statens stralskyddsinstitut, SSI-rapport*, 86-12.
- Byun, D.W., Pleim, J.E., Tang, R. T., and Bourgeois, A. (1999). “Meteorology-Chemistry Interface Processor (MCIP) for Models-3 Community Multiscale Air Quality (CMAQ) Modeling System. In Science algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) Modeling System”. US Environmental Protection Agency Report, EPA-600/R-99/030, 12-1–12-91.
- Diem, J.E. and Comrie, A.C. (2001). “Air Quality, Climate, and Policy: A Case Study of Ozone Pollution in Tucson, Arizona.” *The Professional Geographer*. 0033-0124, pg. 1467-9272.
- Dimiceli, C., Carroll, M., Sohlberg, R., Kim, D.H., Kelly, M., Townshend, J.R.G. (2015). “MOD44B MODIS/Terra Vegetation Continuous Fields Yearly L3 Global 250m SIN Grid V006 [Data set]”. NASA EOSDIS Land Processes DAAC.  
doi:10.5067/MODIS/MOD44B.006
- FMI (2011a). “The Finnish Meteorological Institute is monitoring the dispersion of ash from the volcanic eruption in Iceland”. *FMI Press Release*.
- FMI (2011b). “Press release archive 2011: The Finnish Meteorological Institute monitors the situation in Japan and conveys information to the other authorities”. *FMI Press Release*.
- Guenther, A., Hewitt, N., Erickson, D., Fall, R., Geron, C., Graedel, T., Harley, P., Klinger, L., Lerdau, M., McKay, W., Pierce, T., Scholes, B., Steinbrecher, R., Tallamraju, R., Taylor, J., Zimmerman, P. (1995). “A global model of natural volatile organic compound emissions”. *J. Geophys. Res.* 100, 8873-8892.

- Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., and Wang, X. (2012). “The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions”. *Geosci. Model Dev.*, 5, 1471-1492, <https://doi.org/10.5194/gmd-5-1471-2012>
- Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., and Geron, C. (2006). “Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature)”. *Atmospheric Chemistry and Physics*, 3:3181–3210.
- Hermansson, E., Roldin, P., Rusanen, A., Mogensen, D., Kivekas, N., Väänänen, R., Boy, M., and Swietlicki, E. (2014). “Biogenic SOA formation through gas-phase oxidation and gas-to-particle partitioning – a comparison between process models of varying complexity”. *Atmospheric Chemistry and Physics*, 14:11853–2014.
- Jiang, X., Guenther, A., Potosnak, M., Geron, C., Seco, R., Karl, T., Kim, S., Gu, L. and Pallardy, S. (2018). “Isoprene emission response to drought and the impact on global atmospheric chemistry”. *Atmospheric Environment*, 183:69-83
- Korhonen, H., Lehtinen, K. E. J., and Kulmala, M. (2004). “Multicomponent aerosol dynamics model UHMA: model development and validation”. *Atmos. Chem. and Physics*, 4:757–771.
- Kouznetsov, R., & Sofiev, M. (2012). “A methodology for evaluation of vertical dispersion and dry deposition of atmospheric aerosols”. *Journal of Geophysical Research*, 117(D01202).
- Kulmala, M., Suni, T., Lehtinen, K. E. J., Maso, M. D., Boy, M., Reissell, A., Rannik, U., Aalto, P., Keronen, P., Hakola, H., Back, J., Hoffmann, T., Vesala, T., and Hari, P. (2004). “A new feedback mechanism linking forests, aerosols, and climate”. *Atmos. Chem. and Physics*, 4:557–562.
- Liang, S. and Xiao, Z. (2012). “Global Land Surface Products: Leaf Area Index Product Data Collection (1985-2010)”, Beijing Normal University. doi:10.6050/glass863.3004.db
- MACC (2010). “Eruption of the Eyjafjallajökull volcano”. *MACC*.
- McPherson, E.G., Scott K.I. and Simpson, J.R. (1997). “Estimating Cost Effectiveness of Residential Yard Trees For Improving Air Quality In Sacramento, California, Using Existing Models.” *Atmospheric Environment*. 32, No. 1, pp. 75 84.
- Min. H., Carmichael, G. R. Crawford, J. H., Wisthaler, A., and Zhan, X. (2017). “Biogenic isoprene emissions driven by regional weather predictions using different initialization methods: case studies during the SEAC 4 RS and DISCOVER-AQ airborne campaigns”. *Geoscientific Model Development*, 10: 3085-3104. DOI:10.5194/gmd-10-3085-2017
- Myneni, R., Knyazikhin, Y., Park, T. (2015). “MCD15A2H MODIS/Terra+Aqua Leaf Area Index/FPAR 8-day L4 Global 500m SIN Grid V006 [Data set]”. NASA EOSDIS Land Processes DAAC. doi:10.5067/MODIS/MCD15A2H.006
- NERIS (2015). “European Nuclear and radiological Emergency management and Rehabilitation strategies Information (NERIS) Workshop 2015”. *NERIS*.
- Nerobelov G., Sedeeva M., Mahura A., Nuterman R., Mostamandi S., Smyshlyaev S. (2018). “Online integrated modeling on regional scale in North-West Russia: Evaluation of aerosols

- influence on meteorological parameters”. *Geography, Environment, Sustainability*, 11(2):73-83
- Otte, P.M. and Pleim, J.E. (2010). “The Meteorology-Chemistry Interface Processor (MCIP) for the CMAQ modeling system: updates through MCIPv3.4.1”. *Geosci. Model Dev.*, 3, 243–256.
- Parachini, J. (2003). “Putting WMD terrorism into perspective”. *Washington Quarterly*, 26:4:37–50.
- Seinfeld, J. H. and Pandis, S. N. (2006). *Atmospheric Chemistry and Physics, from Air Pollution to Climate Change, Second Edition*. Wiley.
- Sofiev, M. (2010). “Specification for the flexible PBAP-subroutine”. *MACC Deliverables*, R-ENS 1.7.1.
- Sofiev, M. (2011). “Aerosol Species in the Air Quality Forecasting System of FMI: Possibilities for Coupling with NWP Models”. *Integrated Systems of Meso-Meteorological and Chemical Transport Models*, pages 159–166.
- Sofiev, M., Kouznetsov, R., Prank, M., Soares, J., Vira, J., Tarvainen, V., and Sofieva, V. (2017). “A long-term re-analysis of atmospheric composition and air quality”. *Air Pollution Modeling and its Application XXV*, pages 55–61.
- Sofiev, M. and Siljamo, P. (2014). “Forward and Inverse Simulations with Finnish Emergency Model Silam”. *Air Pollution Modeling and Its Application XVI*, pages 417–425.
- Sofiev, M., Vira, J., Kouznetsov, R., Prank, M., Soares, J., and Genikhovich, E. (2015). “Construction of the SILAM Eulerian atmospheric dispersion model based on the advection algorithm of Michael Galperin”. *Geosci. Model Dev.*, 8:3497–3522.
- Steinbrecher, R., Geunther, A., and Seufert, G. (2013). “Isoprene and other isoprenoids”. *Trace Gas Exchange in Forest Ecosystems*, pages 175–190.
- Subirana, J.S., Zornoza, J.M.J. and Hernández-Pajares, M. (2011) “Ellipsoidal and Cartesian Coordinates Conversion.” *European Space Agency (ESA) Navipedia*.  
[https://gssc.esa.int/navipedia/index.php/Ellipsoidal and Cartesian Coordinates Conversion](https://gssc.esa.int/navipedia/index.php/Ellipsoidal_and_Cartesian_Coordinates_Conversion)
- Toll, I. and J.M. Baldasano (2000). “Modeling of photochemical air pollution in the Barcelona area with highly disaggregated anthropogenic and biogenic emissions.” *Atmospheric Environment*. Volume 34, Issue 19, pg. 3069-3084.
- Tunved, P., Hansson, H.-C., Kerminen, V.-M., Strom, J., Maso, M. D., Lihavainen, H., Viisanen, Y., Aalto, P. P., Komppula, M., and Kulmala, M. (2006). “High Natural Aerosol Loading over Boreal Forests”. *Science*, 312:261–263.
- Undén, P., Rontu, L., Järvinen, H., Lynch, P., Calvo, J., Cats, G., Cuxart, J., Eerola, K., Fortelius, C., Garcia-Moya, J.A., Jones, C., Lenderlink, G., McDonald, A., McGrath, R., Navascues, B., Nielsen, N.W., Ødegaard, V., Rodriguez, E., Rummukainen, M., Rõõm, R., Sattler, K., Sass, B.H., Savijärvi, H., Schreur, B.W., Sigg, R., The, H., Tjijm, A. (2002). “The HIRLAM version 5.0 model. HIRLAM documentation manual (HIRLAM Scientific Documentation).  
[http://hirlam.org/index.php/hirlam-documentation/doc\\_download/308-unden-et-al-2002](http://hirlam.org/index.php/hirlam-documentation/doc_download/308-unden-et-al-2002)
- USGS ELCC (2018). “Eurasia Land Cover Characteristics Data Base Version 2.0.”. *USGS*.  
[https://lta.cr.usgs.gov/glcc/eadoc2\\_0](https://lta.cr.usgs.gov/glcc/eadoc2_0)
- Williams, J., Crowley, J., Fischer, H., Harder, H., Martinez, M., Petajä, T., Rinne, J., Bäck, J., Boy, M., Dal Maso, M., Hakala, J., Kajos, M., Keronen, P., Rantala, P., Aalto, J., Aaltonen, H.,

Paatero, J., Vesala, T., Hakola, H., and Lelieveld, J. (2011). “The summertime Boreal forest field measurement intensive (HUMPPA-COPEC-2010): An overview of meteorological and chemical influences”. *Atmos. Chem. Phys.*, 11:10599–10618.

WorldBank (2018). “DataBank”. *World Bank Open Data*.

Zhou, P., Ganzeveld, L., Taipale, D., Rannik, U., Rantala, P., Rissanen, M.P.,

Chen, D., and Boy, M. (2017). “Boreal forest BVOC exchange: emissions versus in-canopy sinks.” *Atmos. Chem. Phys.*, 17:14309–14332.

## 6. Acknowledgements

I first greatly acknowledge Academician Prof. Markku Kulmala as the head of the Institute for Atmospheric and Earth System Research (INAR) for providing me the opportunity to conduct my master's thesis research at the University of Helsinki. It has been an honour to be invited to this institute to further my academic career.

I am extremely grateful to Dr. Michael Boy for his continued support and supervision of this research project. Michael has been a wonderful support for the project during my time at the University of Helsinki.

I would also like to thank Dr. Alexander Mahura and his colleagues at the HIRLAM Consortium, who provided Enviro-HIRLAM model meteorological output for studied period of summer 2010 for this study for the SILAM model simulations, along with support and advice on the Enviro-HIRLAM and SILAM models and the interface between the two. Alexander Mahura has also been a wonderful support and co-supervisor for this project.

I would additionally like to thank many of the INAR and Faculty of Science at the University of Helsinki colleagues (with whom I have communicated in person and when it was necessary) for their useful discussions and comments and constructive criticism.

I would further like to acknowledge the Finnish Meteorological Institute (FMI), especially the FMI's SILAM development group, including Dr. Mikhail Sofiev, Dr. Rostislav Kouznetsov, and the rest of their colleagues. This also includes Dr. Joana Soares, who is now continuing her research in Canada. They have provided all of the code for SILAM and have also provided support in debugging, setting up model runs, turning on modules, and much more.

I would like to acknowledge the Centre for Scientific Computing (CSC, <https://www.csc.fi/>) which provided computing services, including access to the Sisu and Taito supercomputing facilities. Model tests and runs, visualisation of modelling results, and data analysis for this study have been performed using CSC resources.

Furthermore, I would like to thank Prof. Alex Guenther, primary developer of MEGAN, from the University of California – Irvine for providing the MEGAN3 model as free and open code. In addition to providing the model code, I would like to thank both him, his student Xiaoyan Jiang for providing assistance with compiling and running the model.

MODIS data products from the Aqua and Terra satellites were retrieved from the NASA EarthData website (<https://search.earthdata.nasa.gov/>). Example images shown from this dataset

are courtesy of NASA and are taken directly from the EarthData descriptions. I would like to acknowledge NASA for the use of this free, openly available data.

Sunrise and sunset information was found at timeanddate.com: “Sunrise and Sunset Calculator - City Lookup” (<https://www.timeanddate.com/sun/>).

The colorized USGS ELCC file was downloaded from the USGS ELCC website. The equivalent data file (geoTIFF byte value, not colorized) was downloaded from the USGS EarthExplorer website: <https://earthexplorer.usgs.gov/>. I would like to acknowledge the USGS for free and open access to this dataset.

Landsat/Copernicus imagery was used for Figure 7, which was accessed using the Google Earth software. Image © 2018 Digital Globe and is not intended for commercial use.

Plots have been generated using the Interactive Data Language (IDL) software/tool, run on the CSC’s Taito supercomputer.

# 7. Appendix

## Appendix A: Glossary of Terms and Acronyms

BVOC: Biogenic Volatile Organic Compound(s).

CSC: Centre for Scientific Computing – The organization that provided the high-performance computing services for this project. See Acknowledgements section for more information about the services they provided.

ECMWF: European Centre for Medium-range Weather Forecasts – A weather forecast, not used for this project, but could be used as the meteorological input for SILAM. ECMWF in addition to HARMONIE is what is used for SILAM's daily public operational runs.

Enviro-HIRLAM: Environment - High Resolution Limited Area Model.

FMI: Finnish Meteorological Institute.

geoTIFF: georeferenced Tagged Image File Format – the format in which the USGS ELCC is provided. geoTiff is a standard TIFF file with geographical information included in it. TIFF is a public domain image file format developed and managed by Adobe. Additional information about TIFF is at: <https://www.adobe.io/open/standards/TIFF.html>

GM: Guenther Modified (GM-Algorithm): The implementation for BVOC emissions in the current public release of SILAM prior to this project. This method is based on the USGS Eurasia Land Cover Classification Map, combined with a lookup table and simple equations.

HYSPLIT: Hybrid Single Particle Lagrangian Integrated Trajectory – A trajectory and dispersion model developed by NOAA, which can be run in either forward or reverse mode. More information is at the website <https://www.ready.noaa.gov/HYSPLIT.php>

LAI: Leaf Area Index – Amount of leaf area that exists with respect to the vegetation.

LAIv: Leaf Area Index divided by the Vegetation Cover Fraction (VCF). This determines the amount of actual leaf coverage exists in each grid box (i.e. total surface area of leaves).

MACC: Monitoring Atmospheric Composition and Climate – Organization that provides both real-time and archives of various atmospheric composition forecasts and climate data. MACC is primarily a data provider. More information is at the website <http://www.gmes-atmosphere.eu>

MEGAN: Model for Emissions of Gases and Aerosols in Nature – A BVOC emission model used as the replacement to the Guenther Modified code in SILAM.

MEGAN EFP: The MEGAN Emission Factor Processor – An additional processing tool, written in Python, that takes some of the input files and processes them into tables for reading from the MEGAN3 emission calculator.

MODIS: MODerate resolution Imaging Spectrometer – A sensor aboard NASA’s Terra and Aqua satellites. In this project, MODIS is used as the source for LAI and VCF data, which is used for the input to MEGAN.

NCAR: National Center for Atmospheric Research, Boulder (Colorado, USA) – Research institute and developer of the NetCDF file format and the WRF and MM5 models.

NetCDF: Network Common Data Format – File format in which output from the SILAM model is stored. Input for MEGAN3 is also in NetCDF format. NetCDF is developed by NCAR. More information is found at: <https://www.unidata.ucar.edu/software/netcdf/>

SILAM: System for Integrated modelLling of Atmospheric coMposition.

SMEAR: Station for Measuring Ecosystem-Atmosphere Relations

SMEAR-II: The second SMEAR station constructed, located in Hyytiälä, southern Finland.

USGS ELCC: Eurasia Land Cover Classification Map – A file provided by the United States Geological Survey (USGS) that contains a 1km x 1km resolution map over Europe and Asia, defining land cover classification/land use. This map is based on the USGS’s Eurasia Land Cover Characteristics Data Base Version 2.0. Further information is at [https://lta.cr.usgs.gov/glcc/eadoc2\\_0](https://lta.cr.usgs.gov/glcc/eadoc2_0)

VCF: Vegetation Cover Fraction **or** Vegetation Continuous Fields – Amount of land in a given grid box is covered by vegetation. This is subdivided into three categories: Percent tree cover, percent non-tree vegetation cover, and percent non-vegetated. For every grid box, these three add up to 100%. The term Vegetation Cover Fraction is used in the MEGAN documentation, and the Vegetation Continuous Fields is used in the MODIS. The difference is that the MODIS Vegetation Continuous Fields is split up into three categories: percent tree cover, percent non-tree cover, and percent non-vegetated (bare) (Dimiceli et al, 2015). Vegetation Cover Fraction, however, is the sum of percent tree cover and percent non-tree vegetation. **In this thesis, VCF refers to Vegetation Cover Fraction.**

## Appendix B: Description of MODIS Data & Calculation of LAI<sub>v</sub> from MODIS

Data from UMD's GLASS website (Liang and Xiao, 2010) is unfortunately not available outside of North America, so the equivalent product must be created for the area of interest in Finland. To do so, this project used data from the MODIS satellite.

Data for MODIS was downloaded from NASA's EarthData website (<https://search.earthdata.nasa.gov/>). This data is free to download, though it requires registration with an EarthData account (registration is also free).

The Leaf Area Index (LAI) data downloaded for this project is of the type:

**MCD15A2H: MODIS/Terra+Aqua Leaf Area Index/FPAR 8-Day L4 Global 500 m SIN Grid V006**

A description of this data is at the following website:

[https://lpdaac.usgs.gov/dataset\\_discovery/modis/modis\\_products\\_table/mcd15a2h\\_v006](https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mcd15a2h_v006)

(Myneni et al, 2015).

The Vegetation Continuous Fields, not to be confused with Vegetation Cover Fraction (see Appendix A: Glossary of Terms and Acronyms) data from MODIS is the product of the type:

**MOD44B: MODIS/Terra Vegetation Continuous Fields Yearly L3 Global 250 m SIN Grid V006**

A description of this data is at the following website:

[https://lpdaac.usgs.gov/dataset\\_discovery/modis/modis\\_products\\_table/mod44b\\_v006](https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod44b_v006) (Dimiceli et al, 2015).

The datasets were downloaded on 3 July 2018.

## Appendix C: SILAM Meteorology for MEGAN

MEGAN requires meteorology input for its BVOC calculations. By default, MEGAN uses WRF, but other inputs are acceptable if they are provided in NetCDF format with the same structure and variable names as SILAM.

In this project, a “dry run” run of SILAM was performed (without any BVOC or chemistry running) for the specific days, and this was used for the meteorological input as an alternative to WRF. The main reason for this is so that the data is already gridded, and so that the same simulated meteorological conditions are used in this analysis as in other parts of SILAM. This dry run SILAM output was used as the input to the MEGAN pre-processor. This was only done once, as the preprocessed files are then stored and used by MEGAN inside SILAM, regardless of the type of SILAM run performed.

Because the structure and variable names in SILAM’s output are slightly different from WRF, a short IDL script was written to convert the SILAM output into the file structure that MEGAN is expecting. Below are tables showing the dimension and variable names (note that dimension and variable names in netCDF files are case-sensitive).

Below is a table of variable, dimension, and attribute names that were converted.

Dimension	Length	Name in SILAM	Name in WRF
X (longitude)		lon	west_east
Y (latitude)		lat	south_north
height		height	bottom_top
time	Unimited in both WRF and SILAM.	time	Time (Note capitalization)

Variable	Units	Dimensions	Name in SILAM	Name in WRF
Latitude	Degrees North	SILAM: (lat) WRF <sup>2</sup> : (time, lat, lon)	lat	XLAT
Longitude	Degrees East	SILAM: (lat) WRF <sup>2</sup> : (time, lat, lon)	lon	XLONG
Time	SILAM: Seconds WRF <sup>1</sup> : Minutes	(time)	time	XTIME
Height	Meters	<b>(height)</b>	height	HGT
Temperature at 2 meters	Kelvin	(time, lat, lon)	temp_2m	T2
Ground Surface Temperature	Kelvin	(time, lat, lon)	ground_surface_temp	SOIT1
Soil Moisture Content <sup>3</sup>	m <sup>3</sup> /m <sup>3</sup>	(time, lat, lon)	soil_moisture_content	SOIM1
Soil Type	Classification (unitless)	(time, lat, lon)	soiltype	SLTYP
Relative Humidity	Percent	(time, height, lat, lon)	rel_humid	RH
Water Mixing Ratio	kg/kg	(time, height, lat, lon)	humidity_mixing_ratio	QV
Pressure	Pa	(time, height, lat, lon)	surface_pressure	PRES
SWDOWN <sup>4</sup>	W/m <sup>2</sup>	<b>(time, lat, lon)</b>	SW_int_net_surf	PAR

<sup>1</sup>Time will need to be converted from seconds to minutes when creating the new file.

<sup>2</sup>Latitude and longitude will need to be converted (replicated) from an array to a grid for the new file. There is no need to regrid because MEGAN will handle this.

<sup>3</sup> Soil moisture content in SILAM only uses the top layer.

<sup>4</sup>PAR is not yet implemented in SILAM. Therefore, solar irradiation is used instead (surf\_sw\_net\_radiation), however there is a conversion of solar shortwave downward radiation to PAR in MEGAN.

Items in bold above need to be reformed to match the WRF dimensions.

Attributes also need to be added to the file. The following table contains a list of attributes on each variable that need to be added to the new file.

<b>Variable</b>	<b>Attribute</b>	<b>Name in File</b>	<b>Data Type</b>	<b>Value</b>
XLAT	Field Type	FieldType	int	104
	Memory Order	MemoryOrder	string	“XY”
	Description	description	string	“LATITUDE, SOUTH IS NEGATIVE”
	Units	units	string	“degree north”
	Stagger	stagger	string	(empty string)
XLONG	Field Type	FieldType	int	104
	Memory Order	MemoryOrder	string	“XY”
	Description	description	string	“LONGITUDE, WEST IS NEGATIVE”
	Units	units	string	“degree east”
	Stagger	stagger	string	(empty string)
XTIME	Field Type	FieldType	int	104
	Memory Order	MemoryOrder	string	“0”
	Description	description	string	“minutes since simulation start”
	Units	units	string	“MINUTES”
	Stagger	stagger	string	(empty string)
HGT	Field Type	FieldType	int	104
	Memory Order	MemoryOrder	string	“XY”
	Description	description	string	“height”
	Units	units	string	“m”
	Stagger	stagger	string	(empty string)
T2	Coordinates	coordinates	string	“XLONG XLAT”
	Field Type	FieldType	int	104
	Memory Order	MemoryOrder	string	“XY”
	Description	description	string	“TEMP at 2 M”
	Units	units	string	“K”
RH	Stagger	stagger	string	(empty string)
	Coordinates	coordinates	string	“XLONG XLAT”
	Field Type	FieldType	int	104
	Memory Order	MemoryOrder	string	“XY”
	Description	description	string	“relative humidity”
	Units	units	string	“PERCENT”
	Stagger	stagger	string	(empty string)
	Coordinates	coordinates	string	“XLONG XLAT”

Additionally, the following global variables need to be stored in the new file:

Attribute Name in File	Description	Data Type	Value
MAP_PROJ <sup>1</sup>	Map projection used for the data.	int	See note below
MAP_PROJ_STR <sup>2</sup>	String name of map projection.	string	See note below
CEN_LON	Center longitude	float	Determined by data
CEN_LAT	Center latitude	float	Determined by data
STAND_LON			
POLE_LAT	Latitude of North Pole	float	90.0
TRUELAT1		float	
TRUELAT2		float	
DX		float	

Notes:

<sup>1</sup>. See note 2 below for MAP\_PROJ, which corresponds to MAP\_PROJ\_STR.

<sup>2</sup>. The MAP\_PROJ\_STR attribute is the same as the MAP\_PROJ attribute, just in string form for human understanding. The translation for valid values is as follows:

MAP\_PROJ = 0 --> "CylindricalEquidistant"

MAP\_PROJ = 1 --> "LambertConformal"

MAP\_PROJ = 2 --> "Stereographic"

MAP\_PROJ = 3 --> "Mercator"

MAP\_PROJ = 6 --> "Lat/Lon"

This attribute is not used by MEGAN because MEGAN uses MAP\_PROJ. If one is changed, the other needs to be changed as well. For using SILAM meteorology, 6 should be used.

Complete description of the above attributes is here:

[https://www.ncl.ucar.edu/Document/Functions/Built-in/wrf\\_ij\\_to\\_ll.shtml](https://www.ncl.ucar.edu/Document/Functions/Built-in/wrf_ij_to_ll.shtml)

## Appendix D: Control File and BVOC Emission File for SILAM Run

### CONTROL\_V5\_3

```
LIST = general_parameters
  case_name = hyytiala_2010
  direction_in_time = FORWARD
  simulation_type = EULERIAN
  start_time = 2010 07 14 23 00 0.0
  computed_period = 1 sec
  time_step = 1 sec
  cut_area_source_if_outside_meteo_grid = YES
  computation_accuracy = 10
END_LIST = general_parameters

LIST = mpi_parallel_parameters
  x_divisions = 1
  y_divisions = 1
  max_wind_speed = 100 # [m/s]
  use_mpiio = NO
  use_mpiio_netcdf = NO
END_LIST = mpi_parallel_parameters

LIST = dispersion_parameters
  grid_method = OUTPUT_GRID
  vertical_method = OUTPUT_LEVELS
END_LIST = dispersion_parameters

LIST = meteo_parameters
  # Uncomment which mode of Enviro-HIRLAM is desired.
  # Control/Reference
  #dynamic_meteo_file = GRIB
/wrk/foreback/DONOTREMOVE/ENVHMETEO201007/%ay4/%am2/%ad2/%ah2/fc%ay4%am2%ad2_%ah2+%f3
  #dynamic_meteo_file = GRIB
/wrk/foreback/DONOTREMOVE/ENVHMETEO201007/%ay4/%am2/%ad2/%ah2/fc%ay4%am2%ad2_%ah2+%f3ve
  #dynamic_meteo_file = GRIB
/wrk/foreback/DONOTREMOVE/ENVHMETEO201007/%ay4/%am2/%ad2/%ah2/fc%ay4%am2%ad2_%ah2+%f3md
  #static_meteo_file = GRIB
/homeappl/home/foreback/appl_sisu/silam_runs/static_files/EC_orography_2012.fixed.cut

  # Direct Aerosol Effects (DAE)
  #dynamic_meteo_file = GRIB
/wrk/foreback/DONOTREMOVE/ENVHMETEO201007/DAE%ay4/%ay4/%am2/%ad2/%ah2/fc%ay4%am2%ad2_%ah2+%f3
  #dynamic_meteo_file = GRIB
/wrk/foreback/DONOTREMOVE/ENVHMETEO201007/DAE%ay4/%ay4/%am2/%ad2/%ah2/fc%ay4%am2%ad2_%ah2+%f3ve
  #dynamic_meteo_file = GRIB
/wrk/foreback/DONOTREMOVE/ENVHMETEO201007/DAE%ay4/%ay4/%am2/%ad2/%ah2/fc%ay4%am2%ad2_%ah2+%f3md
  #static_meteo_file = GRIB
```

```
/homeappl/home/foreback/appl_sisu/silam_runs/static_files/EC_orography_2012.f  
ixed.cut
```

```
# Indirect Aerosol Effects (IDAE)  
dynamic_meteo_file = GRIB  
/wrk/foreback/DONOTREMOVE/ENVHMETEO201007/IDAE%ay4/%ay4/%am2/%ad2/%ah2/fc%ay4  
%am2%ad2_%ah2+%f3  
dynamic_meteo_file = GRIB  
/wrk/foreback/DONOTREMOVE/ENVHMETEO201007/IDAE%ay4/%ay4/%am2/%ad2/%ah2/fc%ay4  
%am2%ad2_%ah2+%f3ve  
dynamic_meteo_file = GRIB  
/wrk/foreback/DONOTREMOVE/ENVHMETEO201007/IDAE%ay4/%ay4/%am2/%ad2/%ah2/fc%ay4  
%am2%ad2_%ah2+%f3md  
static_meteo_file = GRIB  
/homeappl/home/foreback/appl_sisu/silam_runs/static_files/EC_orography_2012.f  
ixed.cut
```

```
meteo_time_step = 3 hr  
if_wait_for_data = YES  
abl_parameterization_method = FULL_PARAM      # DRY_ABL, FULL_PARAM  
number_of_precipitation_fields = 2  
END_LIST = meteo_parameters
```

```
LIST = emission_parameters
```

```
emission_source = EULERIAN bvoc/src_bio_voc.ini  
  
cut_area_source_if_outside_meteo_grid = YES  
  
if_technical_source_dump = NONE      # NONE / ORIGINAL_GRID / DISPERSION_GRID
```

```
END_LIST = emission_parameters
```

```
LIST = initial_and_boundary_conditions  
#initialize_quantity = concentration      ! if no such line, initial conditions  
are void  
#initialization_file = NETCDF init.super_ctl  
#boundary_type = DIRICHLET                ! ZERO / DIRICHLET  
#if_lateral_boundary = YES                ! YES/NO  
#if_top_boundary = YES                    ! YES/NO  
#if_bottom_boundary = NO                  ! YES/NO  
#boundary_time_step = 3 hr                ! timestep unit  
#boundary_header_filename = ../common/boundary_cb4.ini  
END_LIST = initial_and_boundary_conditions
```

```
LIST = transformation_parameters  
# transformation = PASSIVE EULERIAN  
# transformation = PM_GENERAL EULERIAN  
# transformation = DMAT_SULPHUR EULERIAN  
transformation = CB4_SOA EULERIAN  
# transformation = POP_GENERAL EULERIAN  
# transformation = ACID_BASIC EULERIAN  
# transformation = RADIOACTIVE EULERIAN
```

```
aerosol_dynamics = VBS EULERIAN
```

```

biogenic_SOA_aging_rate = 4.0E-11    # cm3 molecule-1 s-1
anthropogenic_SOA_aging_rate = 4.0E-11    # cm3 molecule-1 s-1
intermediate_volatility_OC_aging_rate = 4.0E-11    # cm3 molecule-1 s-1
if_monoterpene_products = 1.0    # if monoterpene emission influences gas
phase chemistry

dry_deposition_scheme = KS2011_TF
wet_deposition_scheme = STANDARD_3D_SCAVENGING

if_actual_humidity_for_particle_size = YES
default_relative_humidity = 0.8
compute_thermodiffusion = NO
mass_low_threshold = STANDARD_ACCURACY    # CRUDE_ACCURACY,
STANDARD_ACCURACY, HIGH_ACCURACY

passive_subst_ref_lifetime = 500 day
passive_subst_ref_tempr = 288
passive_subst_dLifeTime_dT = -1 min/K

ADB_if_compute_nucleation = YES
ADB_nucleation_scheme = KINETIC    # BINARY, TERNARY, KINETIC, ACTIVATION
ADB_if_compute_coagulation = YES
ADB_if_compute_condensation = YES
ADB_if_compute_cloud_activation = NO
ADB_if_recalc_wet_d = YES

if_full_acid_chemistry = YES

make_coarse_no3 = ss1t    0.03    ! material of aerosol to make it on and
stickiness coef

END_LIST = transformation_parameters

LIST = optical_density_parameters
optical_coefficients_depend_on_relative_humidity = YES
optical_coefficients_depend_on_temperature = YES
if_split_aerosol_modes = YES    ! doesn't work yet
if_narrow_wave_bands = YES    ! doesn't work yet
END_LIST = optical_density_parameters

LIST = output_parameters
source_id = NO_SOURCE_SPLIT    #
SOURCE_NAME    SOURCE_SECTOR    SOURCE_NAME_AND_SECTOR
vertical_method = CUSTOM_LAYERS
level_type = HEIGHT_FROM_SURFACE
layer_thickness = 25. 50. 100. 200. 400. 750. 1200. 2000. 2000    # output
levels [m]/[pa]/[hybrid_nbr], reals
output_time_step = 1 sec
output_times = REGULAR
output_format = NETCDF3
time_split = ALL_IN_ONE
template = output/${CASE}/%case
variable_list = output_config.ini
grid_method = CUSTOM_GRID
grid_type = lon_lat
grid_title = GEMS output grid
resol_flag = 128

```

```

    ifReduced = 0
    earth_flag = 0
    wind_component = 0
    reduced_nbr_str = 0

# 61.8N 24.3E
    lon_start = 23.3
    lat_start = 60.8
    nx = 100
    ny = 100
    lon_end = 25.3
    lat_end = 62.8

    lat_s_pole = -90.
    lon_s_pole = 0.
    lat_pole_stretch = 0.
    lon_pole_stretch = 0.
END_LIST = output_parameters

LIST = standard_setup
    #horizontal_advection_method_eulerian = EULERIAN_HORIZ_V5
    #vertical_advection_method_eulerian = EULERIAN_VERT_V5
    advection_method_eulerian = EULERIAN_V5
    kz_profile_method = SILAM_ABL_EC_FT_KZ
    advection_method_lagrangian = LAGRANGIAN_WIND_ENDPOINT_3D
    random_walk_method = FULLY_MIXED
    advection_method_default = EULERIAN
    abl_height_method = COMBINATION
    continuity_equation = anelastic_v2
    horizontal_interpolation = LINEAR
    vertical_interpolation = LINEAR
    time_interpolation = LINEAR
    nuclide_database_fnm = ini/silam_nuclides.dat
    chemical_database_fnm = ini/silam_chemicals.dat
    standard_cocktail_fnm = ini/standard_chemistry_cocktails.ini
    standard_cocktail_fnm = ini/standard_auxillary_cocktails.ini
    standard_cocktail_fnm = ini/standard_aerosols_cocktails.ini
    standard_cocktail_fnm = ini/standard_radioactive_cocktails.ini
    grib_code_table_fnm = ini/grib_code_table_v5.silam
    netcdf_name_table_fnm = ini/netcdf_name_table.silam
    land_use_data_meta_file = data/physio/land_use_features_USGS_Eurasia.dat
    optical_properties_meta_data_file = ini/optical_properties.dat
    photolysis_data_file = ini/photolysis_finrose.dat
    timezone_list_fnm = ini/tzindex.dat
    allow_zero_forecast_length = NO
    precipitation_low_limit = 0.1 mm/hr
    print_debug_info = DEBUG_INFO_YES
    cloud_report_interval = 1
    disregard_meteo_data_sources = YES
END_LIST = standard_setup

END_CONTROL_V5_3

```

# Appendix E: Detailed Steps of Installing, Compiling and Running MEGAN on CSC

## E1. Overview of Steps Performed

Thus far, the following steps have been performed in working towards designing practical approach for an implementation of MEGAN3 in SILAM:

- Perform “dry runs” of SILAM using Enviro-HIRLAM meteorological output from the control/reference, DAE, and IDAE modes, that is no BVOCs were calculated, rather the dry run was performed only for the purpose of gathering and interpolating necessary meteorology data. The resulting meteorology data was stored in a NetCDF format for use as input to MEGAN3, as described in Section 2.4.1.2 and further described in Appendix C.
- MODIS data was downloaded and regridded for use in MEGAN3, as described in Section 2.4.2 and further described in Appendix B.
- The MEGAN3 pre-processor was run, and the table files were generated for the domain and time scale of the model. Additional required inputs for the pre-processor include global coverage data files for crops, grass, needleleaf tree fraction, tropical tree fraction, and shrubbery. These input files were downloaded from the MEGAN3 website and have not been modified.
- The MEGAN3 EFP (Python code) was setup and run after the MEGAN3 pre-processor. Other than setting the domain and inputs from the pre-processor, all parameters were left at their defaults.
- The SILAM dry run meteorology was also placed in a file that mimics files produced by the WRF model to be used by the MCIP processor. This is a prerequisite for MEGAN3.
  - Following another approach, the MEGAN3 main code was additionally modified to read the file above instead of an MCIP output file. The reason for this is because MCIP requires a full WRF or MM5 run to be performed, thus the simulated file was not sufficient. Note that running MCIP was not desired for this project due to its heavy dependence on WRF or MM5 NWP models.
  - Additionally, the I/O API library (used for reading the NetCDF files in MEGAN) was also modified to be able to accept the meteorology interpolated by SILAM rather than WRF or MM5. For this particular case study, some of the parameters were excluded or assumed to be zero because they are not relevant or needed by the MEGAN3 main code.

## E2. Modules, Packages and Installations

Modules Needed (this is on Taito, the names might be slightly different on Sisu; they need to be loaded in this order):

- hdf5-par
- netcdf4
- geo-env

Additionally, we need to module swap *intel* with *gcc* compilers (need to be consistent with SILAM).

Packages to install:

- IO/API (described below)
- MCIP (described below)

Additionally, Python needs to be installed and available on the system, which is already the case on CSC.

## E3. Compiling and running the MEGAN3 pre-processor

Before compiling with *gcc*, a few minor modifications to the Fortran syntax needed to be made to fix syntax errors occurring due to syntax that is allowed with the PGI compiler. MEGAN is built for but is proprietary and not available on CSC, so some of the PGI syntax is not allowed with *gcc*. There were about a couple dozen of these errors, which were relatively easy to figure out when attempting to compile with *gcc*. I added a few comments in the code mentioning my changes.

These syntax changes do not affect the behaviour of the code.

MODIS was used for the LAI<sub>v</sub> input. This includes the LAI and the vegetation fraction from 8-day MODIS average over the time range of interest. LAI<sub>v</sub> is calculated as the LAI divided by the vegetation fraction. was regridded, centered on the desired domain. See Appendix B for a description of how this was done.

SILAM meteorology (dry run of SILAM, without any emissions simulations) was used as the meteorology input for the MEGAN pre-processor. The same domain as the future SILAM run was used. I wrote a short script that converted the variable names and structure in the SILAM dry-run output to the layout that MEGAN needs.

The global files for crops, grass, needleleaf tree fraction, tropical tree fraction, and shrubbery downloaded from the MEGAN website were used for these inputs and have not been modified.

## **E4. MEGAN3 EFP**

Python code, does not require compiling. No modifications in the Python code for this project, except the directory and file names of inputs.

The global W126 (ozone data) file provided on the MEGAN website was used, without any modification. This is optional and probably negligible, but I used it anyway because it was available.

Used the growth form output from the Pre-processor as the growth form input to the EFP.

All other input parameters and user options for the EFP were left at their default.

## **E5. MCIP**

MCIP must be built, and first geogrid must be created.

MCIP must be built and run using the intel compiler on Taito, not *gcc*.

## **E6. MEGAN3 Main Code**

### **E6.1 Overview**

For compiling the main MEGAN3 code, I followed the directions on pages 6-7 of the MEGAN3 user guide, using 64-bit Linux mode. A major step changing *pgi* to *gfortran* (the *gcc* compiler) because there were numerous places where the syntax had to be changed to be accepted by *gfortran*. Doing a mass search-and-replace for certain pieces of text made this easier, but it had to be done in every single one of the MEGAN3 source files. See below for the compilation steps.

Like with the pre-processor, a SILAM dry-run was used for the meteorology input.

MEGAN can accept downward shortwave radiation rather than PAR, and it performs the calculation internally.

The option for 64-bit needs to be set when calling the build script. The Taito and Sisuh supercomputers both use 64-bit architecture.

Soil type has been omitted for the time being.

The outputs from the pre-processor were used as input to MEGAN3.

Additionally, WRF-formatted meteorology output from the SILAM dry-run was processed through the MCIP processor, and the result was set the met2mgn script.

## E6.2 Compilation

MEGAN is written for the PGI compiler, whereas for SILAM on the CSC supercomputers, the GCC (gfortran) compiler is used. A number of syntax changes to the MEGAN pre-processor had to be made in order to compile. This applies to the MEGAN3 main code as well. Notable practical changes include:

- Lines beginning with the letter “c” are comments, which is not recognized by the GCC Fortran90 compiler. These had to be changed to the modern standard “!” comment character.
- The line continuation & had to be moved to the end of the previous line, not at the beginning of the new line.
- The INCLUDE and LIB paths for NetCDF were set to the correct location on the CSC network. NetCDF4 was used instead of NetCDF3 to keep it modern and efficient. NetCDF4 is back-compatible with NetCDF3, so this does not affect the program’s behaviour. The Models-3 I/O API was downloaded from the website above and was compiled using the gfortran option and Linux 64-bit settings. The result was used for the INCLUDE and LIB path in MEGAN. The BIN variable was modified to use gfortran on CSC.
- The following compiler flags were changed from the PGI flags to gfortran:

PGI Flag	Equivalent gfortran Flag
-Mvect=sse	mfpmath=sse
-Mextend	-ffree-form
-Msecond_underscore	-fsecond-underscore
-Bstatic_pg	<i>(Not Applicable for gfortran)</i>
-Mlfs	<i>(Not Applicable and ignored on 64-bit Linux)</i>
-Mbyteswapio	-fconvert=swap
-lnetcdf	-lnetcdf

Additionally, the **-fopenmp** flag was added in order to include the OpenMP library, which is the default for the PGI compiler but not gfortran. I also added the **-w** flag to hide warning messages.

Finally, the MCIP processor was downloaded from <http://www.cmascenter.org/>. Alternatively, it is available on GitHub. This had to be installed using gfortran and 64-bit Linux settings.

### E6.3 Setup of MEGAN

First, make sure that the netcdf module is loaded (on CSC call module load netcdf4. hdf5-par module must be already loaded).

The following two lines were added to the GRIDDESC file in the /MEGAN3/work directory:

```
'hyytiala_201007_1km'  
'GEO_61.37N24.35E' 24.35 61.37 0.01 0.01 200 200 1
```

These settings describe the domain used for this project, with the variables being as follows: Name, xstart (bottom left longitude in degrees) ystart (bottom left latitude in degrees) xGridSize (in degrees), yGridSize (in degrees), nx, ny (number of grid points along longitude and latitude), externalBoundarySize.

### E6.4 Steps to initiate a MEGAN3 run

Directions on page 8 of the MEGAN user guide were used, with the following parameters set:

Step 1, txt2api:

EXEDIR set to the /bin directory

GDNAM3D set to the name “hyytiala\_201007\_”

dom set to “1km” (only one value)

scen set to “J0” (the only scenario used for this project, but J1, J2, etc. could be used in the future).

Note: When using “LATLON” as the coordinate system name, the function throws a warning saying that the coordinate system is missing. However, the program still ran to completion

Step 2, met2mgn:

Followed directions in this step and also modified the EXE variable. I use the *run.met2mgn.v3\_rad45.csh* script instead of *run.met2mgn.v3*, per Xiaoying’s suggestion. Also, the function *yj2ymd* wasn’t provided anywhere in MEGAN, so the julian package had to be download, and the missing function was substituted

with some math and the j2ymd function. Finally, the MCIP file had to be created from the SILAM dry run's WRF-type formatted output using the MCIP processor.

Step 3, daymet:

I used a separate run of MCIP rather than trying to use a SILAM dry run and then convert it to the format of MCIP. I downloaded and ran MCIP version 3.6 rather than the latest one because this is what the instructions in the MEGAN use guide suggest. Other than this, the daymet step is straightforward.

Step 4, megsea (soil info):

Use defaults.

Step 5, megcan (canopy info):

Use defaults.

Step 6, megvea (stress factors):

Use defaults.

Step 7, mgn2mech (calculator for species emission):

Use defaults. This is the final output of MEGAN3.

Step 8, ioapi2uam (file format conversion; optional). Changes file format of the output if desired in a format other than NetCDF.

## **E6.5 Challenges with not running MEGAN on MPI**

MEGAN3 is not currently set up to run with MPI (although it nonetheless requires the MPI library to be linked). For a relatively small domain, as is the case in this project, using serial vs. MPI is not a significant concern. When running MEGAN3 separately from SILAM, with SILAM only reading the output files from MEGAN from disk (rather than making MEGAN calls from within the SILAM code) this is not a major problem because SILAM could be run with MPI separately from MEGAN3. However, for a larger domain, being able to use MPI is a consideration to take into account when evaluating the effectiveness of MEGAN.

## **E7. Summary of Required Inputs**

- Meteorology for Pre-Processor, from SILAM dry run and translated into WRF format.
- LAIv file for pre-processor. For North America, this is easily downloaded. However, for other parts of the world, this needs to be created by hand, using data from somewhere. I used MODIS, and doing so requires both MODIS LAI + vegetation cover. This requires an additional script to do the division and create a new file. To make it even more complicated, the MODIS data needs to be converted to lat/lon coordinates using further hand-written scripts.
- MEGAN global files (downloaded from the MEGAN3 website, over 1GB each).
- Optional ozone file.

- MCIP input. This is not generalized, so I had to write an additional script to convert SILAM's chemical output to mimic this file.

## E8. Additional Notes

- MEGAN3 appears to be designed for North America (specifically with the required meteorological models and the GLASS database for LAIv) without as much consideration for Europe or other parts of the world.
- MEGAN3 is designed specifically for WRF and similar meteorology inputs, and it does not really consider other weather forecast inputs such as the SILAM dry-run or Enviro-HIRLAM. It does not seem to have that intended generality or flexibility, and hence, it requires the user to write additional scripts for pre-pre-processing.
- If MEGAN3 was fully integrated into SILAM rather than run as a stand-alone program, it would require not only the user to have the Python and geo-env modules installed and loaded, but it would require full compilation of the IO/API library and the MCIP processor, which would add significant effort and time to build required executables for SILAM.
- The *jd2ymd* function was not available in the MEGAN3 repository. We had to re-create its functionality.
- The MEGAN global coverage files are over 1 GB each, which would be undesirable for integrating directly into SILAM simulations, but it can be used at the SILAM model initialisation stage with subsequent pre-processing.
- The current version of the MEGAN3 user guide is very limited and does not provide detailed instructions for fully installing and running MEGAN, which could confuse some of SILAM beginners-users who are not familiar with the original concept of MEGAN.
- The makefiles in each of the individual directories of the MEGAN3/source repository require extensive work to accurately configure and validate all parameters to the current operational system, and they each must be modified separately rather than having one single place and file to modify.
- There are multiple places where there are paths in scripts or hard-coded paths in the MEGAN3 Fortran & Python subroutines that point to somewhere on the developers' computer/network, which have to all be changed manually to wherever the location on another or new computer/ network is. If modified code will be passed to someone else, they would all have to make these same changes.
- The step was to run MCIP using the SILAM dry run file. However, MCIP requires additional datasets beyond what was available in the SILAM dry run. Therefore, for this project, in the interest of time and not having to write an additional script to mimic an NCIP file, a simple run of NCIP using the WRF-formatted input from the pre-processor was used, and this output was used.
- MEGAN3 is not recommended to be run on multiple processors using the MPI. As mentioned above, this leads to inefficiency. In this project, however, it was sufficient to run MEGAN in non MPI mode, along with SILAM.

MCIP – needs *gcc*, *hdf5*, and *netcdf4* modules.

eta coordinates for NWP models (such as for example – WRF) are needed for MCIP, and these are calculated from the following equation:

[https://www.shodor.org/os411/courses/\\_master/tools/calculators/etacoordinates/index.html](https://www.shodor.org/os411/courses/_master/tools/calculators/etacoordinates/index.html)

## Appendix F: List of Updates to SILAM between Latest Public Release and Version Used for this Project

The following is a log-file of all of the changes submitted to FMI's codebase between the public release of SILAM v.5.1.1 and the version currently being used in this project.

-----  
r577649 | kouzne | 2018-04-25 14:18:38 +0300 (Wed, 25 Apr 2018) | 2 lines

Definitions for MEPS input

-----  
r577582 | sofievm | 2018-04-20 19:15:05 +0300 (Fri, 20 Apr 2018) | 1 line

Minor cleaning of DA modules

-----  
r577579 | kouzne | 2018-04-20 16:02:47 +0300 (Fri, 20 Apr 2018) | 7 lines

Backport from v5\_6

Fixed a bug with Henry-law-constant in mesophill conductance (very minor improvement of scores)  
DFixed S-N flip for photorates climatology. Winter climatology left for the whole year.  
Couple of typos

-----  
r577501 | sofievm | 2018-04-16 10:44:19 +0300 (Mon, 16 Apr 2018) | 1 line

Debugging of 4dvar and related issues

-----  
r577500 | sofievm | 2018-04-16 10:42:51 +0300 (Mon, 16 Apr 2018) | 1 line

CHeck revision for Windows

-----  
r577050 | kouzne | 2018-03-22 19:06:03 +0200 (Thu, 22 Mar 2018) | 2 lines

Fix\_strang\_quantities for updated MEPS.

-----  
r576875 | kouzne | 2018-03-15 18:25:07 +0200 (Thu, 15 Mar 2018) | 6 lines

Addad LAI and z0 definitions for cosmo  
addd soil moisture for WRF  
definition for shortwave\_down in bio\_voc source

-----  
r576813 | kouzne | 2018-03-14 09:33:48 +0200 (Wed, 14 Mar 2018) | 6 lines

Workaround for nre cray (XC40) issue with launching external programs:  
check directory for existance before attempting to create it...

-----  
r576480 | kouzne | 2018-02-28 10:00:21 +0200 (Wed, 28 Feb 2018) | 2 lines

Added handler for SIMPLE\_ABL\_EC\_FT\_KZ (forgotten in some earlier commit)

-----  
r575676 | kouzne | 2018-01-25 16:07:45 +0200 (Thu, 25 Jan 2018) | 3 lines

Added a cocktail file that has noth VBS and OC.

-----  
r575527 | kouzne | 2018-01-18 13:14:45 +0200 (Thu, 18 Jan 2018) | 6 lines

Added simple\_abl\_ec\_ft\_kz ABL method (needed for global run)  
Fixed zero-byte in log coming from NetCDF attributes  
Yet another attempt to fix multiple reporting of shopping lists to stdout

-----  
r575458 | kouzne | 2018-01-15 17:18:30 +0200 (Mon, 15 Jan 2018) | 2 lines

standard\_noSOA\_cocktails.ini to run non-SOA run with SOA-enabled sources

-----  
r575406 | kouzne | 2018-01-11 18:33:10 +0200 (Thu, 11 Jan 2018) | 4 lines

more debug info on troubles with Vd (Intel compiler with arithmetic shortcuts)  
fixed shopping list reporting in the MPI mode  
~

-----  
r575404 | kouzne | 2018-01-11 18:00:54 +0200 (Thu, 11 Jan 2018) | 4 lines

Reporting of nthreads for lagrangian advection  
Made more portable trajectory output, so it works also with Intel

-----  
r575374 | kouzne | 2018-01-10 19:18:03 +0200 (Wed, 10 Jan 2018) | 2 lines

Got rid of erroneous #define directive

-----  
r575336 | sofievm | 2018-01-09 18:18:25 +0200 (Tue, 09 Jan 2018) | 3 lines

COrrrected bugs: random\_seed; Asimof field; time\_params  
New observation\_dose\_rate,  
Updated pollen source

-----  
r575335 | kouzne | 2018-01-09 18:03:36 +0200 (Tue, 09 Jan 2018) | 3 lines

Added optoipns for Intel compiler  
made VOIMA\_GNU\_BUG define for handling buggy voima compiler

-----  
r575315 | sofievm | 2018-01-09 12:24:06 +0200 (Tue, 09 Jan 2018) | 1 line

...and another bug nearby

-----  
r575295 | sofievm | 2018-01-09 10:50:42 +0200 (Tue, 09 Jan 2018) | 1 line

Corrected bug in expansion of the source term time parameters

-----  
r574673 | kouzne | 2017-11-30 14:58:21 +0200 (Thu, 30 Nov 2017) | 9 lines

Added new grads templates %fm2 and '%fd2 for forecast day and hour (as in Cosmo)

Fixed a bug with preiod-vaild fields, in grib files where accumulation statr  
time is not analysis (ERA5)

Disabled crash on failed SMS progress

-----  
r573798 | kouzne | 2017-11-07 17:10:34 +0200 (Tue, 07 Nov 2017) | 4 lines

Slightly improved reporting of success/failure, report of observations  
backtrace for linux\_intel

-----  
r573729 | kouzne | 2017-11-03 15:32:02 +0200 (Fri, 03 Nov 2017) | 6 lines

Progress reporting command.

"progress\_file = SMS" depricated  
should be replaced with  
"progress\_file = CMD progress-command"

-----  
r573541 | kouzne | 2017-10-26 18:40:21 +0300 (Thu, 26 Oct 2017) | 2 lines

Corrected grib codes for Cosmo7 GRIBs

-----  
r573540 | kouzne | 2017-10-26 18:39:42 +0300 (Thu, 26 Oct 2017) | 5 lines

Added (hackish) handling of Cosmo7 hyprid-pressure levels  
Restored fallback for no land raoughness.

-----  
r573409 | kouzne | 2017-10-21 12:56:01 +0300 (Sat, 21 Oct 2017) | 2 lines

Removed a debug "print"

-----  
r573408 | kouzne | 2017-10-21 12:26:52 +0300 (Sat, 21 Oct 2017) | 2 lines

Correctd reporting of aerosol modes..

-----  
r572960 | kouzne | 2017-10-09 15:25:45 +0300 (Mon, 09 Oct 2017) | 2 lines

Debug EnKF btroken after MPI 3DVAR

-----  
r572889 | kouzne | 2017-10-06 11:29:50 +0300 (Fri, 06 Oct 2017) | 2 lines

GRIB definitions for MEPS-flavour of Harmonie

-----  
r572888 | kouzne | 2017-10-06 11:28:49 +0300 (Fri, 06 Oct 2017) | 6 lines

Added handling of MEPS harmonie "GRIB" files centre\_MEPS  
More clear error reporting in "store\_input\_to\_supermarket"  
Optimized scavenging of empty columns.  
OMP parallel optical column depth (Not optical density yet..)

-----  
r572727 | kouzne | 2017-10-03 08:40:22 +0300 (Tue, 03 Oct 2017) | 2 lines

Fixed segfault on optical density map

-----  
r572101 | sofievm | 2017-09-15 12:47:52 +0300 (Fri, 15 Sep 2017) | 1 line

Allowing zero-source case for source\_2\_map sub

-----  
r572071 | sofievm | 2017-09-14 16:36:02 +0300 (Thu, 14 Sep 2017) | 1 line

Making source\_2\_map robust to non-treated source types

-----  
r572070 | sofievm | 2017-09-14 16:33:37 +0300 (Thu, 14 Sep 2017) | 1 line

```

Making source_2_map robust to non-treated source types
-----
r572059 | sofievm | 2017-09-14 12:09:31 +0300 (Thu, 14 Sep 2017) | 1 line
corrected memory leak in emis whole-period reporting sub
-----
r571669 | kouzne | 2017-09-11 18:00:39 +0300 (Mon, 11 Sep 2017) | 5 lines

Further debugging of 0:360 grids
killed fu_set_longitude (does wrong thing)
minor fix for reporting dispersion vertical sticking out of meteo

-----

r571598 | kouzne | 2017-09-07 15:59:42 +0300 (Thu, 07 Sep 2017) | 4 lines

Backport from v5_6:
ps%if_inside_domain for a point source

-----

r571585 | kouzne | 2017-09-07 10:21:12 +0300 (Thu, 07 Sep 2017) | 8 lines

Merge changes from eslogin.
Minor change of crash report on a_src failure
Increases in number of allowed sources from 10000 to 2000 (in stupidity check)
silencing of pollen source
NetCDF buffer_timesteps set to 1 (caused some problems on large-file
output)

-----

r571584 | kouzne | 2017-09-07 10:11:06 +0300 (Thu, 07 Sep 2017) | 2 lines

Fixed a bug with Hirlam azimoff messages.

-----

r571582 | kouzne | 2017-09-07 07:40:57 +0300 (Thu, 07 Sep 2017) | 6 lines

Fixed non-mpi compilation bug

Ported fix for 0..360 grids from v5_4.
Now both -180..180 and 0..360 grids are fine (not -180..360 though)

-----

r571535 | kouzne | 2017-09-05 16:05:46 +0300 (Tue, 05 Sep 2017) | 10 lines

Further optimization of MPI-3DVAR

Speed-up of reading observations >x100
OMP observations and injections
Optimized scatter of control vector
Optimized domain decomposition
bug in MPI reporting of shopping lists
Timing report for quasi-newton optimisation

-----

r571430 | kouzne | 2017-08-31 16:56:14 +0300 (Thu, 31 Aug 2017) | 9 lines

Cosmetic/reporting changes:

unhide restart function from interface
got rid of work arrays in md module
added Assimilate caounter for 3Dvar
disabled creation of iterations directory when not needed
More debug and retries in create_directory_tree for linux_gnu

-----

```

r571305 | kouzne | 2017-08-24 10:23:49 +0300 (Thu, 24 Aug 2017) | 13 lines

Made MPI and no-MPI logs more diff-able:

"TOTAL MASS REPORT", emission report  
DA messages

Some debugging of MPI 3DVAR: yet another problem in  
synchronizing obs-from-model fixed. Now 3DVAR seems to be MPI-clean

Fixed a bug in bioVOC emission (total\_cloud\_cover\_flag)

Fixed uninitialized pointer in wind\_blowndust

-----  
r571267 | kouzne | 2017-08-22 12:02:24 +0300 (Tue, 22 Aug 2017) | 3 lines

Work array leakage fixed in da\_interface::vector\_from\_model

-----  
r571266 | kouzne | 2017-08-22 11:11:46 +0300 (Tue, 22 Aug 2017) | 2 lines

Yet another attempt to make mkdir work reasonably.

-----  
r571259 | kouzne | 2017-08-21 20:18:16 +0300 (Mon, 21 Aug 2017) | 11 lines

Next step for MPI 3Dvar

Support for several auxiliary\_cocktail lines  
Recovered reporting of iteration costs report\_cost  
Fixed numerics in get\_Rs\_2013  
Removed call to src\_contain\_grd (caused annoying errors for some global sources)  
Enabled fMaxForce = 1.1e5 for abl\_top\_pressure\_flag (less complains)  
Disabled moments handling in inject\_in\_situ: No way doing it right  
fixed wrong call to MPI\_Get\_count in smpi\_recv

-----  
r571174 | sofievm | 2017-08-15 18:13:19 +0300 (Tue, 15 Aug 2017) | 1 line

Final touches of cosmetics for v.5.5

-----  
r571114 | sofievm | 2017-08-10 14:32:54 +0300 (Thu, 10 Aug 2017) | 1 line

...and another refinement of aerosol cocktails

-----  
r571109 | sofievm | 2017-08-10 13:17:13 +0300 (Thu, 10 Aug 2017) | 1 line

Aerosol cocktails are harmonised among CAMS, STEAM and SYKE highres emission data

-----  
r571066 | sofievm | 2017-08-09 18:10:10 +0300 (Wed, 09 Aug 2017) | 1 line

Further refinement of STEAM netcdf nametable

-----  
r571065 | sofievm | 2017-08-09 16:50:59 +0300 (Wed, 09 Aug 2017) | 1 line

STEAM nametable updated  
-----