aeromap

# Bulletin 1

## Mid-term report

# 1 Publishable summary

## 1.1 A summary description of project context and objectives

The International Panel on Climate Change (IPCC) has identified that the largest uncertainty in current estimates of planetary radiation forcing is due to atmospheric aerosols and has called for an urgent expansion of global studies to monitor and characterise them. Aerosols are characterised by their optical and microphysical properties and while the aerosol robotic network (AERONET) of remote sensing instruments provides accurate values for them, AERONET's coverage of the Earth surface is sparse (mostly city-based) and patchy. The AEROMAP project is designed to provide a solution to overcome this lack of information on global aerosols without the need to invest in hundreds of new AERONET sites. To achieve this, AEROMAP will harness and capitalise on the high resolution full-Earth measurements of aerosol optical depth provided daily by satellite remote sensing instruments. Furthermore, by using satellite-driven data, AEROMAP will be able to gain access to aerosol characteristics over the oceans which are of paramount importance to the estimate of the overall radiation budget. As a result, the retrieval of aerosol optical and microphysical properties by AEROMAP will significantly extend the efforts to globally monitor and characterise atmospheric aerosols.

AEROMAP has developed and validated data mining tools based on neural networks to convert satellite measurements of aerosol optical depth and columnar water vapour into aerosol optical and microphysical properties like the single scattering albedo, particle asymmetry factor, complex refractive index and the aerosol size distribution. This is a very challenging problem as satellite instruments, unlike their ground-based counterparts in AERONET, do not yet have the capability to provide polarisation information obtained from the diffuse radiation field. In essence, AEROMAP is attempting to produce a robust inversion algorithm for the retrieval of aerosol parameters analogous to that used by AERONET but without this source of information. AEROMAP's inversion algorithm will then open the door to global aerosol monitoring and characterisation. The inputs to AEROMAP are a small but specific set of satellite measurements. Specifically, daily, full-Earth measurements of the aerosol optical depth (AOD) provided by the MODerate resolution Imaging Spectrometer (MODIS) satellite instrument in 3 wavelength bands (438-448nm, 673-683nm, 862-877nm) spanning the visible spectrum and the near-infrared taken together with a near-infrared measurement of the columnar water vapour, are used as well as the best estimate of the absorption aerosol optical depth (AAOD) at 500nm provided by the Ozone Monitoring Instrument (OMI) satellite. The outputs from AEROMAP are retrievals of over 40 optical and microphysical parameters needed to characterise aerosols. This is accomplished by feeding the satellite data into neural networks that have been trained on AERONET "ground-truth" data to learn the relation between the inputs and output parameters. The result will be global maps of aerosol optical and microphysical parameters at a resolution of $1^o$ x $1^o$ (50km x 50km) which will be used to monitor and classify aerosols as they move daily across the Earth's surface.

The main objectives of AEROMAP are to code neural networks to learn the relation between existing satellite inputs and aerosol optical and microphysical parameters for different types of aerosol. AEROMAP will use co-located and synchronous ground-truth data to test that the trained networks can extrapolate the aerosol parameters retrieved from the satellite inputs at the training data location to new and distant geo-locations. AEROMAP will then generate daily-updated global maps of aerosol optical and microphysical parameters and classify aerosol in each pixel by type. AEROMAP will apply the networks to satellite measurements of important extreme aerosol events such as desert dust storms, forest fire outbreaks, urban brown cloud episodes, volcanic eruptions and radiation clouds to assess whether or not it is possible to monitor and track their spatio-temporal characteristics. By updating the global maps with daily satellite over-pass data, AEROMAP will function as a near-real-time monitor of aerosols whose maps and data files will be made publicly-available at the project website/portal. The global aerosol maps will be assessed to develop an aerosol impact scale that will be used for the issuing of alerts and early-warning information about aerosol-related hazards.

AEROMAP is unique in many ways. Firstly, it will address the need for global characterization of aerosols by producing the first global maps of aerosol optical and microphysical parameters. Secondly, AEROMAP will provide detailed information about aerosols over the oceans for the first time. Thirdly, the global classification of aerosols provided by AEROMAP will help inform decisions about where to place new ground-based remote sensing instruments for detailed local studies. Finally, AEROMAP innovates highly in the field of atmospheric physics by applying contemporary methods of machine learning, mathematical data analysis, and statistics to substantially increase the understanding of the global aerosol system.

AEROMAP capitalizes both on the skills of the fellow Dr Michael Taylor, an experienced researcher in the fields of machine learning and mathematical data analysis, and the expertise of the scientist in charge (SIC) Dr Stelios Kazadzis in the fields of satellite remote sensing, ground-based aerosol measurement and retrieval at the National Observatory of Athens - a world leading centre for aerosol monitoring. AEROMAP is a highly multidisciplinary project that will use techniques, methods and analyses from the following fields: machine learning, aerosol science, multivariate statistics, computer science (spatio-temporal mapping and real-time monitoring), and science communication (project website/portal). Great emphasis will be placed on the latter in order to engage the public on the issue of climate change and the emerging central role being played by atmospheric aerosols.

## 1.2 Description of the work performed and the main results achieved so far

AEROMAP has obtained the complete data record of daily-averaged data from 807 ground-based remote sensing instruments federated to AERONET spanning the years 1996-2011 and has pre-processed the data so that its sky radiance measurements are aligned in wavelength to those available from satellite inputs. Satellite inputs for the entire global domain spanning the years 2000-2011 have been obtained and code has been written to select the data between the two datasets that is co-located and synchronous. Neural networks have been optimised and trained to learn the functional relation between ground-based AERONET direct sun AOD measurements and AERONET aerosol optical and microphysical parameters for three key aerosol types: desert dust, biomass burning and urban sulphate aerosols. Retrievals obtained by feeding the networks with AERONET inputs from stations distant from training data sites were validated against those provided by AERONET's inversion algorithm – allowing for i) an independent check of the results of the AERONET algorithm and ii) a test of the ability of the networks to extrapolate to other regions of similar aerosol type. With only AOD inputs, the networks correctly recovered the mean aerosol size distribution as well as the mean values of optical and microphysical parameters. However, uncertainty analysis of the mean values obtained at timescales ranging from 1 day to 1 year revealed that the networks performed poorly in the retrieval of the temporal variation of these parameters. It was found that this could be solved by addition of columnar precipitable water and absorption AOD to the inputs. With 5 inputs, the networks successfully recovered the temporal variation of all parameters apart from the real part of the complex refractive index. In preparation for application of the networks to satellite data, co-located and synchronous MODIS AOD data at a selection of sites (of different aerosol type) was regressed on equivalent wavelength AERONET AOD to obtain a functional relation between the two, and to establish the nature of a reported bias. It was found that, while a bias is present with MODIS AOD values being higher than AERONET extrapolated AOD, the effect is linear and could easily be removed by rotating the data. The 5-input networks were then fed with daily satellite data co-located and synchronous with AERONET aerosol retrievals at the same location. As for case of the earlier study using AERONET inputs, the networks retrieved the temporal variability of aerosol optical and microphysical parameters for all variables apart for the real part of the complex refractive index. The ability of the networks to extrapolate retrievals to distant regions of the same aerosol type was tested by feeding them with satellite inputs at locations distant from the training data (for three different aerosol types: desert dust, biomass burning products and urban sulphate aerosol). Uncertainty analysis compared the satellite retrievals against ground-truth data at different timescale and verified that the networks are capable of recovering the expected temporal variation of aerosol parameters thereby allowing for their extrapolation to locations where no ground sites exist - something not currently possible with existing methods. As a result of these investigations, a number of new findings were made. A new coefficient of variability was constructed that ranks global sites so that locations of high or low aerosol variability (for all or selected sites of a given aerosol type) can be identified. This will be used to assess the variability of aerosols on the global maps to be produced by AEROMAP. Complementing this, the contribution of different types of aerosols to AOD extinction provided by the GOCART model was used to rank global AERONET sites by their percentage mixture of desert dust, sea salt, organic and black carbon, and sulphate ($SO_2$). This ranking was used to identify sites for training and validating networks with data corresponding to the same type of aerosol – effectively implementing a new form of aerosol classification. A new technique for analysing the aerosol size distribution, based on Gaussian mixture models, was developed that allows for the automatic identification of the number of distinct aerosol modes (having different characteristic/effective particle sizes) comprising the aerosol size distribution. This goes beyond the current limitation of research in the field to only two modes – the fine mode and the coarse mode – and has been applied to a number of extreme aerosol episodes to show how they influence the background conditions. Finally, techniques developed to present and analyse the aerosol size distribution have been applied to a comparison of aerosol microphysical retrievals from AOD measurements by AERONET and by the World Meteorological Organisation's precision filter radiometer in a study completed with colleagues at the Physics Instrumentation Centre in Moscow and the World Radiation Centre at Davos.

### 1.2.1 The expected final results

AEROMAP will also train and validate a neural network for retrieving aerosol parameters over the oceans – perhaps the greatest source of uncertainty. AEROMAP will investigate how to combine networks so as to model mixtures of aerosols. Armed with this knowledge and daily full-Earth satellite inputs, AEROMAP will then be able to use the networks to render the first accurate, daily-updated global maps of key aerosol optical and microphysical parameters at $1^{o}x1^{o}$ (≈50kmx50km) resolution with each pixel being classified by aerosol type/mixture. These maps will be used to help characterise and monitor aerosols at European historical sites (Rome and Athens) as well as island sites (Sicily and Crete) to assess the potential impact of aerosols on tourism, fishing and agriculture – research which has yet to be conducted. AEROMAP maps will be produced daily and constitute a new real-time monitor of global aerosols and its maps will be available for download as high-resolution JPEG images (the raw data will be made available on request via FTP) from the project website/portal. Scientists will be able to collect maps of large-scale pollution episodes on the timescale of days to weeks and easily create AVI movies (1 frame per day) to trace their dispersal. Videos of climatologically and/or socio-economically important strong aerosol events including dust storms, forest fire outbreaks, urban brown clouds and volcanic eruptions will be used to showcase AEROMAP and to demonstrate its ability to monitor and characterise global aerosols and their temporal variation. In addition, AEROMAP will develop a scale and a set of indices to measure aerosol levels at different locations. Global maps of these indices will then be used to issue early-warnings worldwide and on a daily basis so that local authorities can take evasive action. Finally, following the development of the project website/portal and implementation of the real-time monitor, AEROMAP will engage the public with a talk entitled "Marie-Curie and atmospheric chemistry" aimed at building a bridge between the birth of the field and the state of the art maps produced by the project. The project will conclude with a public debate on climate change to discuss the role played by atmospheric aerosols.

### 1.2.2 Socio-economic impact of AEROMAP's results and the wider societal implications

#### *1.2.2.1 Contribution to European excellence and European competitiveness*

By contributing to one of the most important climate change parameters - the global characterization of aerosols, AEROMAP will help significantly raise the profile of the ERA for researchers. The deliverables and end-products of this project will contribute to European excellence and competitiveness by demonstrating that European centres like the host institute NOA are pioneering new mathematical/computational methods that will allow for exploitation of satellite remote sensing datasets. The fellow, the host institute and hence Europe as originator of such methodologies will act as a pole of attraction to scientists from outside the ERA who are interested in learning such techniques with a view to applying them to their own fields of research. Furthermore, the real-time monitor and alerting service, together with other data products such as global maps of aerosol optical and microphysical properties are unique. Their showcasing at the project website/portal will be a strong advertisement for the project, the fellow, the host institute NOA and Marie Curie actions in general. Their potential impact on environmental policy-making decisions may have worldwide repercussions that will further turn eyes towards the ERA.

#### *1.2.2.2 How AEROMAP will produce long-term synergies and/or structuring effects*

Considering the cost of satellite data acquisition, and especially the large data quantities involved with global modelling, AEROMAP will aim towards efficient data exploitation. Until now, it has not been possible to exploit the full-Earth coverage provided by satellite remote sensing in order to globally-characterise aerosols via their microphysical properties. As a result Europe has followed the example of other continents in building instruments and investing in ground stations federated to AERONET. While the methodology proposed here to extrapolate from local AERONET sites to non-existent site locations will, it is hoped, yield accurate deductions of aerosol properties and aerosol type, there is no doubt that the best way to validate neural network-derived retrievals, is with co-located AERONET data. One direct long-term structuring effect of AEROMAP is that its products will allow for determination of the best sites to locate such stations. Small scale but high impact projects like AEROMAP will help to ensure that Europe can meet the challenges of the 21st century, where the quality of the environment affects everyday life and the impact of aerosols on climate change, air quality and business (e.g. aviation) is high in the public conscience. AEROMAP represents a concerted effort to bring satellite remote sensing and ground-based data-providers and scientists together to maximise their utility and impact and, as a result, AEROMAP and its project website/portal will be a showpiece for European science and fellowships and actions like the Marie Curie IEF.

#### *1.2.2.3 Impact of outreach activities*

Central to increasing the national, European and international impact of AEROMAP are outreach activities led by the fellow. The fellow participated in the European Open Science Forum 2012 in July in Dublin, Ireland and presented a digital poster explaining to the public what the project is about. The first week in September, he presented a poster at the European Aerosol Conference 2012 in Granada, Spain displaying the first results of the project. At the end of September his work was displayed as part of the activities of the ACP Group at NOA at the FP7-funded Researcher's Night 2012 at the National Hellenic Research Foundation designed to attract graduate students to apply for postgraduate studies and research posts at universities and institutes. Within the ACP Group, the fellow has given two seminars reporting back on the progress of the project and the results of Phases A and B. On the social media sphere, the fellow has an active profile page on the ACP Group website where copies of his talks and conference posters are exhibited. The fellow has created a Twitter account for tweeting updates and news about the project and he has designed and developed a dedicated website/portal for the project which has a News column and houses resources and publications produced by AEROMAP. He is currently busy writing up the initial results in three articles for publication in peer-reviewed journals. He is planning a big social media splash once his articles are accepted for publication. The science communication part of AEROMAP will reach a crescendo in the next couple of months and it is hoped that the carefully-planned public outreach action plan to be followed will increase awareness of the important role played by Marie-Curie actions in making it facilitating European research and mobility, the role of scientific research on society, the project AEROMAP, the fellow, the host institute, and work currently being funded and undertaken in the ERA on the global aerosol system.

| |
|---|
| The project website and data portal can be accessed via: ***http://apcg.space.noa.gr/aeromap*** and you can follow updates on the progress of the project also @Twitter: ***https://twitter.com/_AEROMAP***. |

### 1.1.3 Diagrams or photographs illustrating and promoting the work of the project

Since AEROMAP's real-time monitor and database will potentially have many users, it is important that its products will be immediately recognisable. The fellow has designed a logo and a colour-coded scheme for the cover pages of documents disseminating the project's deliverables such as peer-reviewed research publications, conference and educational posters, project reports and bulletins as well as the User's Guide for the AEROMAP software which will be made open source on completion of the project:

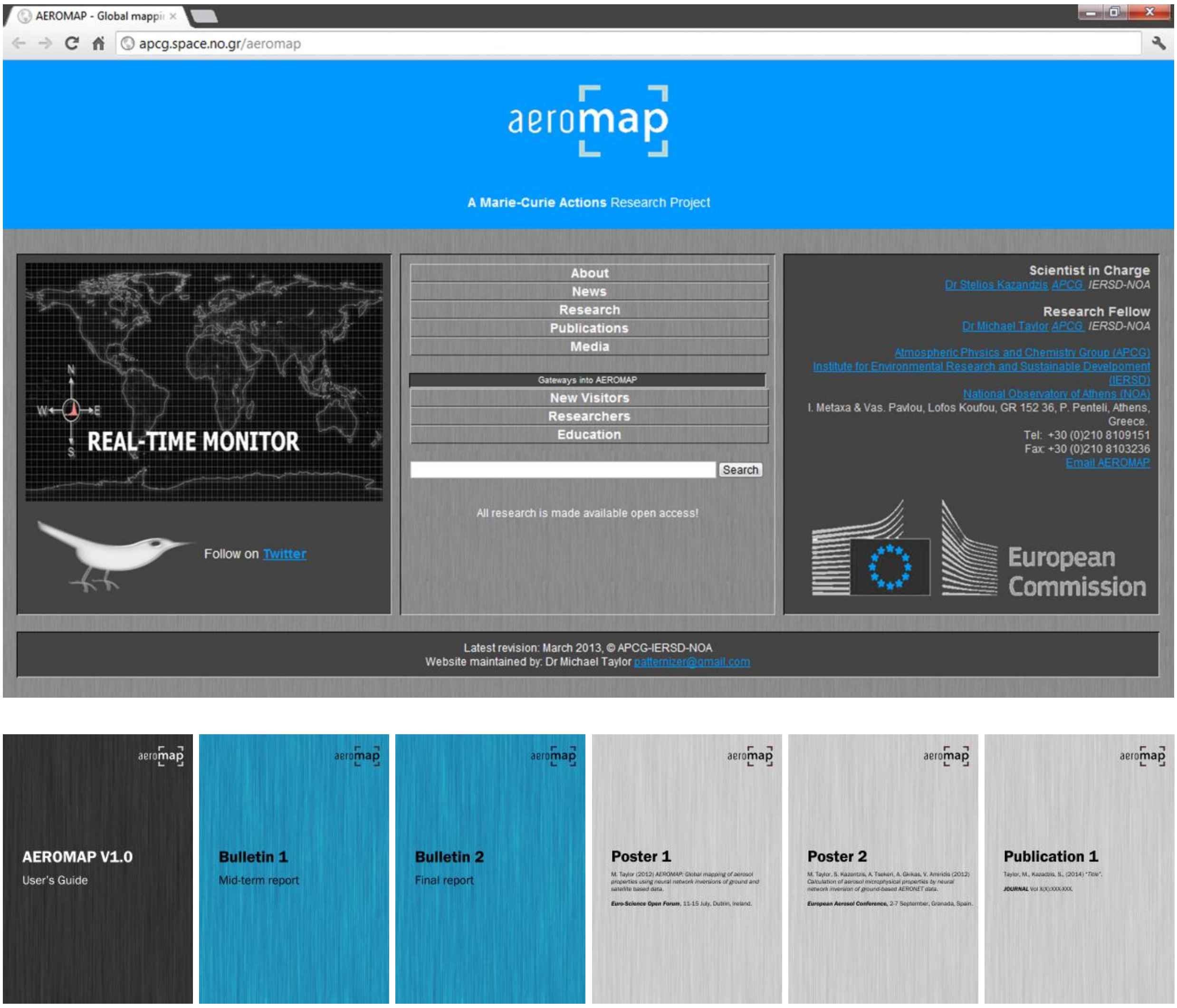
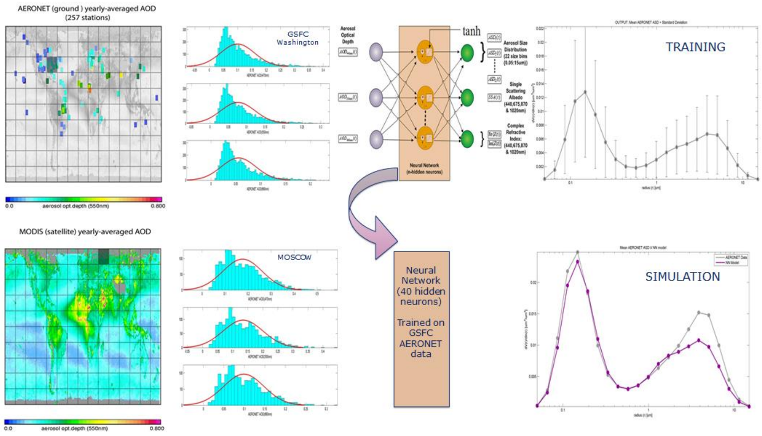


***Figure 1a***: *The website/portal home page bearing the project logo, and dissemination resources' cover pages*

In addition, a ***graphical abstract*** has been produced for online journal publications and also to help researchers have a quick visual insight into the methodology adopted by the project:




***<u>Figure 1b</u>**: Graphical abstract illustrating the AEROMAP methodology*

### 1.1.4 Contact details

Dr Stelios Kazadzis (***SIC***) APCG*, IERSD-NOA*
Dr Michael Taylor (***fellow***) APCG, *IERSD-NOA (*email: patternizer@gmail.com)

Atmospheric Physics and Chemistry Group (APCG),
Institute for Environmental Research and Sustainable Development (IERSD),
National Observatory of Athens (NOA),
I. Metaxa & Vas. Pavlou, Lofos Koufou, GR 152 36, P. Penteli, Athens, Greece.
Tel: +30 (0)210 8109151
Fax: +30 (0)210 8103236

# 2 Project objectives, work progress and achievements & project management

## 2.1 Project objectives for the period

The reporting period spans one year from 01/03/2012 to 28/02/2013 and includes all or part of several of the following stated phases of the project as shown in the project timeline below:

Phase A (months 1-6): Machine learning, function approximation and generalisation
Phase B (months 6-12): Independent assessment of results using inversion algorithms
Phase C (months 9-15): Cluster analysis, aerosol typing and case studies
Phase D (months 15-24): 3D spatio-temporal mapping and real-time monitoring/alerting
Phase E (Months 1-24): Website design, maintenance and public engagement

| Timeline(months) /Phases | 0-3 | 3-6 | 6-9 | 9-12 | 12-15 | 15-18 | 18-21 | 21-24 |
|---|---|---|---|---|---|---|---|---|
| Phase A: | A1,A2 | A2,A3 | | | | | | |
| Phase B: | | | B1,B2 | | | | | |
| Phase C: | | | | C1,C2 | C2, C3 | | | |
| Phase D: | | | | | | D1 | D2 | D2 |
| Phase E: | E1 | E1 | E1 | E2 | E3 | E3 | E3 | E4 |
| Milestones | | | M1 | M2 | | | M3 | M4 |

The project objectives, relevant to the first 12 months of the project, span Phases A1-4, B1-2, C1-2 and E1-2 and are as follows:

***Objective 1***: to train neural networks to identify the functional relationship between AERONET direct sun AOD and AERONET retrieval data

***Objective 2***: to investigate the ability of neural networks trained and validated for one site to extrapolate to qualitatively-similar sites worldwide for key aerosol types (e.g. desert dust, biomass burning, and urban industrial pollution)

***Objective 3***: to train neural networks to identify the functional relationship between MODIS satellite AOD data and AERONET direct sun AOD data

***Objective 4***: to investigate the ability of neural networks trained and validated on co-located sites to extrapolate to regions where no sites exist using MODIS AOD and neural network derived functions

***Objective 9***: to engage in science communication to inform the public of the project, its results and the potential impact of global aerosol characterisation on European Research Area (ERA) environmental policy.

## 2.2 Work progress and achievements during the period

***TASK: A1 (months 0-3): data acquisition and pre-processing***

***Satellite data***

The complete record of daily-averaged MODIS data for the period 01/03/2000-24/07/2011 was downloaded in hierarchical data format (HDF) from the Level 3, Collection 5.1 Product that contains $1^{o}$ x $1^{o}$ grid-average values of atmospheric parameters related to aerosols. The level 3 data are averaged to a 1° latitude/longitude grid. In addition, level 2 (50x50 Km) averaged data were used in areas with complex terrain where 1°x1° pixels are not representative. The averaged data was kindly supplied by Dr Antonios Gkikas at the University of Ioannina. From the dataset, two aerosol optical property measures were extracted: the AOD at 550nm and the Ångström Exponent å(470nm/660nm), and used to interpolate values for the AOD(470nm) and AOD(660nm). Where the land å values from Terra were not available, corresponding sea values at å (550nm/865nm) were used for interpolation. As a result, raw satellite data for the 3 AOD inputs: AOD (470nm), AOD (550nm) and AOD(660nm) over the entire global domain was collected. In addition, for specific pairs of training and validation sites (see Task A3), the fellow downloaded daily, Level 2 near-infrared total column precipitable water mean values from the Terra satellite over land (dataset MOD05_L2) and the Aqua satellite over ocean (dataset MYD05_L2) from the Giovanni Multi-sensor Aerosol Products Sampling System (MAPSS). The Level 2 water data are generated at the 1km spatial resolution of the MODIS instrument using the near-infrared algorithm during the day, and at 5km x 5km pixel resolution both day and night using the infra-red algorithm when at least nine fields of view are cloud free. Finally with respect to satellite data, the best daily estimate of the absorption AOD at 500nm from the OMI near-UV aerosol absorption and extinction optical depth and single scattering albedo OMAERUV product was also downloaded from the Giovanni Multi-sensor Aerosol Products Sampling System (MAPSS) for days collocated and synchronous with the MODIS water vapour data. The fellow then coded a script in the MATLAB object-oriented programming language to read in the satellite data files and to store them locally in matrices.

**BI-PRODUCT**: the script: ***call_load_MODIS.m*** is open source and available for the community from the ***Researchers Gateway*** of the project website.

***AERONET inversion data***

The AERONET inversion retrievals represent a wide number of parameters and characteristics that are important for comprehensive interpretation of the optical aerosol regime. Sun photometer direct sky radiance measurements from AERONET are inverted in the Level 2 Inversion Product Version 2.0 and provide raw data from all collaborating sites. The data has a standard format and provides retrievals for 150 different aerosol property parameters including the aerosol optical depth, single scattering albedo, asymmetry factor and complex refractive index centred at 4 wavelengths: 440nm, 675nm, 870nm and 1020nm. The L2 Inversion Product also provides the volume particle size distribution dV(r)/dlnr (µm3/µm2) in 22 logarithmically equidistant bins spanning the particle size range: $0.05\mu m \leq r \leq 15\mu m$. The inversions are obtained from an inversion algorithm which is subject to certain constraints: moderate aerosol loads (AOD>0.4), high solar zenith angles ($>50^{o}$), a simultaneous measurement of the AOD at all 4 wavelengths and within ±16 minutes of the almucantar measurement. When these conditions are not satisfied the inversion results are less reliable or not available. In this case, Level 1.5 AOD data and inversions can be used which are cloud-screened and also available daily. For very low AOD loads and during cloudy conditions AERONET's spectral de-convolution algorithm (SDA) provides AOD measurements and the size distribution but not the aerosol optical parameters. Another difficulty is working with this data is that the raw data files are text files with headers and are not easily machine-readable. In order to have access to the whole spectrum of data so as to be able to train networks for diverse daily conditions, the fellow downloaded the complete data record from 01/03/1996 to 10/11/2012 for all three datasets: i) the Level 2.0 Version 2.0 inversions, ii) the level 1.5 Version 2.0 inversions and iii) the SDA algorithm. The fellow then wrote a complex script in MATLAB capable of parsing all three datasets and to store the daily AERONET parameter data locally in matrices.

**BI-PRODUCT**: the script: ***call_load_AERONET.m*** is open source and available for the community from the ***Researchers Gateway*** of the project website.

Since there is a small difference between the central wavelengths of the MODIS AODs and the AERONET AODs, the Ångström Exponents: å(675nm/440nm), å(870nm/440nm), å(1020nm/440nm), å(1020nm/675nm) and å(550nm/870nm) were calculated and used to extrapolate AERONET AODs to new values centred at the MODIS wavelengths: AOD(470nm), AOD(550nm) and AOD(660nm). Finally, although AERONET's Level 2 Version 2.0 Inversion Product provides the volume concentrations of the fine(f) and coarse(c) modes, the fine mode fraction η is not explicitly listed. These parameters are derived from the raw AERONET values and the fellow wrote a script to calculate and append them to the AERONET dataset.

**BI-PRODUCT**: the script: ***call_add_parameters.m*** is open source and available for the community from the ***Researchers Gateway*** of the project website.

Central to the success of tasks A3 and B1-B2 (described below) it was necessary to obtain a homogeneous sample of AERONET data containing complete data for the 5 inputs (AOD, water vapour and absorption AOD) as well as the 38 output parameters (22 bins of the size distribution and the single scattering albedo, asymmetry factor, real part of the refractive index and the imaginary part of the refractive index at four wavelengths).

***Analysis of the global AERONET inversion data***

The complete dataset of daily-averaged AERONET Level 2 Version 2.0 data spans 807 sites and the years 1996-2012. However, it is very inhomogeneous since, as mentioned above, many sites have days where the inversion algorithm could not be run due to cloudy conditions, low solar zenith angles but mainly due to low aerosol loads (AOD<0.4). In the process of selecting sites for training networks, it was necessary to rank the list of sites not by number of entries but by the numbers of days where complete records of training data (i.e. containing simultaneous data for the 3 AODs extrapolated to MODIS wavelengths, water vapour, the absorption AOD at 440nm plus the microphysical parameters: the aerosol size distribution and the complex refractive index, as well as the optical parameters: the single scattering albedo and asymmetry factor). As a result of this analysis, the number of sites containing at least one full record dropped from 807 to 623. The distribution of the number of records was found to follow a power law. The top 73 sites accounted for 50% of the entire record. Moreover, only 162 sites had at least 1 year of data (365 days of data – not necessarily falling entirely within a single year). For sites having at least 3 years of data the number fell to just 37 sites. This meant that site selection, as well as taking into account aerosol type (e.g. location in a desert source region – see below) also required consideration of the number of days of data available to train the networks. Since this analysis can potentially assist other researchers involved in site selection for studies, an EXCEL file of the ranked data has been produced and made available at the project website. As an illustration, the first few entries are as follows:

| Kb | Site Name | Cumulative Kb | Cumulative % | N lines | N Sum |
|---|---|---|---|---|---|
| 3987305 | Solar_Village.dubovikday | 3987305 | 1.73 | 3065 | |
| 3425856 | SEDE_BOKER.dubovikday | 7413161 | 3.22 | 2663 | |
| 3231976 | GSFC.dubovikday | 10645137 | 4.62 | 2506 | |
| 2967987 | Banizoumbou.dubovikday | 13613124 | 5.91 | 2283 | |
| 2,609,075 | Sevilleta.dubovikday | 16,222,199 | 7.04 | 2,039 | 12,556 |
| 2252398 | Mauna_Loa.dubovikday | 18474597 | 8.02 | 1769 | |
| 2238457 | Carpentras.dubovikday | 20713054 | 8.99 | 1742 | |
| 2,187,753 | Venise.dubovikday | 22,900,807 | 9.94 | 1,700 | 17,767 |

**BI-PRODUCT**: the EXCEL file: ***AERONET_global_ranked_data.xlsx*** is open source and available for the community from the ***Researchers Gateway*** of the project website.

TASK A1 was completed by the end of month 3 and generated four bi-products – the 3 scripts above for loading, parsing and post-processing AERONET and satellite data plus the EXCEL file ranking AERONET sites by available inversion data.

***TASK A2 (months 0-3): Function approximation of the relation between MODIS AOD and AERONET AOD***

A simple approach was implemented to identify the mathematical correlation mapping between MODIS satellite AOD data and AERONET direct sun AOD data. As mentioned above, thanks to recently completed work on the sampling of MODIS data pixels at the University of Ioannina by Dr Antonios Gkikas, accurate

MODIS AOD data centred at the wavelengths 470nm, 550nm and 660nm were extracted for the complete record of MODIS data spanning the period 01/03/2000-24/07/2011 and covering the entire global domain.

For our study, MODIS AOD at 470nm, 550nm and 660nm are used as inputs. However, not all AERONET stations use the same CIMEL sun photometer or filters and as a result, the vast majority of AERONET stations provide data at 440nm, 670nm, 870nm, 936nm and 1020 nm, but not at the wavelengths of the MODIS AOD (470nm, 550nm and 660nm). As mentioned earlier, in order to interpolate AERONET AODs to match those available at MODIS wavelengths, we calculated the values of Angstrom Exponents so that the following extrapolated values of the AOD could be deduced:

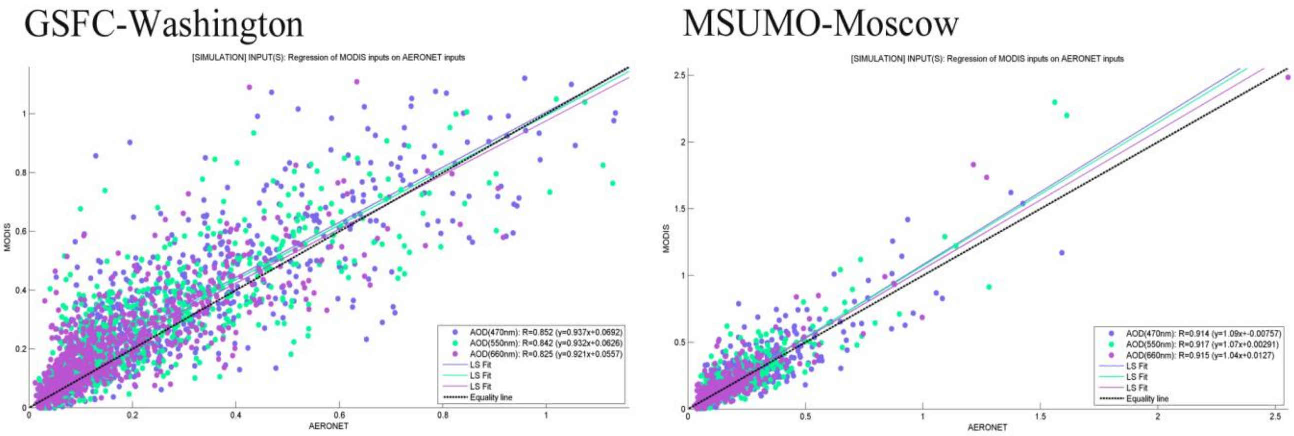
$$AOD(470nm) = AOD(440nm)\left(\frac{470nm}{440nm}\right)^{-å(675nm,440nm)}$$
$$AOD(550nm) = AOD(440nm)\left(\frac{550nm}{440nm}\right)^{-å(675nm,440nm)}$$
$$AOD(660nm) = AOD(440nm)\left(\frac{660nm}{440nm}\right)^{-å(675nm,440nm)}$$

for all data at the two sites. However, there is fundamental physical difference in the way AERONET's ground-based CIMEL sun photometers and the MODIS instrument's spectro-radiometers measure the AOD. Sun photometers perform almucantar scans of light radiation based around the pointing direction to the sun (zenith angle) whereas spectro-radiometers measure the intensity of solar radiation reflected vertically by the Earth system (the planetary surface and the atmosphere). As a result, the light paths are very different and sample slightly different angular variations of aerosol. This effect is minimized when the sun is overhead but tends to increase close to sunrise and sunset and creates a bias in the comparison of MODIS AOD and AERONET AOD. While it was originally envisaged that it would be necessary to train a neural network specifically for the purpose of transforming MODIS AOD onto AERONET AOD, it was decided first to make a comparison for two well-established sites have long time series of both ground-based and satellite-based AOD measurements. The sites selected were: Washington GSFC and Moscow MSU-MO from which co-located and synchronous MODIS and AERONET AOD data was extracted. The results of the regression analysis are as follows:

GSFC-Washington
AOD (470nm): R=0.852, MODIS (y) = 0.937 X AERONET (x) + 0.0692
AOD (550nm): R=0.842, MODIS (y) = 0.932 X AERONET (x) + 0.0626
AOD (660nm): R=0.825, MODIS (y) = 0.921 X AERONET (x) + 0.0557

MSUMO-Moscow
AOD (470nm): R=0.914, MODIS (y) = 1.09 X AERONET (x) + 0.00757
AOD (550nm): R=0.917, MODIS (y) = 1.07 X AERONET (x) + 0.00291
AOD (660nm): R=0.915, MODIS (y) = 1.04 X AERONET (x) + 0.0127

as shown in Figure 1c below:




***<u>Figure 1c</u>****: Plots of the regression of MODIS AOD on AERONET AOD values at 470nm, 550nm and 660nm for two well-establish AERONET sites having long time series. The linear regression best fit lines and correlation coefficients R are also indicated.*

Assuming the validity of such linear regressions, the equations can be inverted by solving for x=f(y) so as to transform MODIS AOD measurements to "expected" AERONET AOD values ("Transformed MODIS"):

GSFC-Washington
AOD (470nm): Transformed MODIS = (MODIS - 0.0692) / 0.937
AOD (550nm): Transformed MODIS = (MODIS - 0.0626) / 0.932
AOD (660nm): Transformed MODIS = (MODIS - 0.0557) / 0.921

MSUMO-Moscow
AOD (470nm): Transformed MODIS = (MODIS - 0.00757) / 1.09
AOD (550nm): Transformed MODIS = (MODIS - 0.00291) / 1.07
AOD (660nm): Transformed MODIS = (MODIS - 0.0127) / 1.04

This confirmed the linear nature of the bias between MODIS AOD and AERONET AOD and allows for MODIS AOD to be easily transformed onto AERONET AOD values. MODIS AOD data, transformed in this way, are used as simulation inputs in the networks of tasks A3 and B1-B2. This method circumvented the need to train a neural network specifically to learn the relationship between MODIS AOD and AERONET AOD and which would incorporate a new source of uncertainty into the input data set. It is important to note that, while the neural networks are trained on AERONET data only, the extrapolation of MODIS AODs onto AERONET AOD wavelengths is what allows satellite data to be used as inputs for deducing aerosol parameters at locations where only the satellite inputs are available. The validity of this approach was verified in Task A3 by training a neural network at Washington GSFC and testing the predictions using transformed MODIS AOD inputs against AERONET ground-data at Moscow MSU-MO.

TASK A2 was completed by the end of month 3 and satisfied Objective 3. The outcome of this task was Deliverable 1 and the mapping has been incorporated into the AEROMAP code.

***Deliverable 1***: Identification of the mathematical correlation mapping between MODIS satellite AOD data and AERONET direct sun AOD data

***TASK A3 (months 3-6): Function approximation of the relation between AERONET AOD and AERONET inversion products at selected sites***

Task A3 required a careful choice of AERONET sites for the training and validation of neural networks. In particular, in order to test the performance of the networks for different aerosol types it was vital to identify data that was clearly dominated by desert dust, biomass burning aerosol and urban sulphates.

***Aerosol typing and analysis of global GOCART aerosol type data***

An initial literature survey allowed for identification of sites that are known to be dominated by desert dust and biomass burning aerosol during various seasons of the year. In particular, sites in the Sahara desert and on the Arabian Peninsula are consistently dominated by desert dust all year round. However, biomass burning is very seasonal and occurs during very different periods in the African Savannah, the Amazon Cerado, in Australia and in the Boreal forests of North America and Southern Canada. The situation is complicated further by the lack of a standard naming convention for aerosol types. For example, it was commonplace to come across articles describing rural pollution or continental background pollution and polluted dust or dirty pollution for the same location. Furthermore, no consensus has been reached yet on which optical or microphysical parameters to use to universally classify aerosol types – and to deal with the case of aerosol mixtures. Since it is necessary to train and assess the performance of networks on distinct aerosol types, the issue of aerosol classification was one of the first that needed to be addressed before coding could begin. Initially, the classification scheme adopted by the CALIOP mission was applied to the homogeneous dataset containing full daily data records. This scheme associates desert dust, smoke (from biomass burning) and urban pollution with specific value ranges of the colour ratio and LIDAR ratio. While it is capable of selecting days associated with these types independently of the site or location, the CALIOP classification has a severe drawback in the context of AERONET data as the LIDAR ratio (although calculable from a combination of the single scattering albedo and the phase function at $180^{o}$ provided by AERONET) essentially measures backscatter along the line of sight but *away* from the ground and therefore not likely to be accurate. As a result, it was decided that a different approach should be adopted for AEROMAP. The GOCART model used by the NASA GEOS-5 mission, provides 3 hourly measurements of the extinction AOD as well as the contribution to extinction AOD due to desert dust grains, sea salt, organic and black carbon and sulphate ($SO_2$). GOCART data spanning the years 2001-2005 (inclusive) are available from the AERONET data synergy portal and was downloaded for the first 155 AERONET sites (75% of all inversion records) ranked by the number of days of available data. Since GOCART provides 8x3-hourly measurements per day, these were averaged to produce daily averaged percentages. Then the ratio of the contribution of individual aerosol types (desert dust grains) to the total extinction AOD was calculated and an EXCEL file was produced that ranked sites by the percentage makeup of each aerosol type. Sorting by type then allowed for suitable site selection for the training of networks by AEROMAP. The top 10 ranked “desert dust” AERONET sites using this method are:

| Site Name | N (L2 Inv.) | <AOD> | % SO2 | % OC | % BC | % Dust | % Sea Salt | % OC+BC | % DUST + SO2 |
|---|---|---|---|---|---|---|---|---|---|
| Tamanrasset_INM | 407 | 0.793 | 4.54% | 1.39% | 0.63% | 93.44% | 0.13% | 2.02% | 97.98% |
| Agoufou | 1028 | 0.973 | 3.70% | 2.47% | 0.82% | 92.91% | 0.10% | 3.29% | 96.61% |
| Banizoumbou | 2283 | 0.92 | 4.57% | 3.48% | 1.09% | 90.76% | 0.11% | 4.57% | 95.33% |
| DMN_Maine_Soroa | 680 | 0.967 | 5.27% | 3.52% | 1.14% | 90.07% | 0.10% | 4.65% | 95.35% |
| IER_Cinzana | 1469 | 0.823 | 4.86% | 4.62% | 1.22% | 89.19% | 0.12% | 5.83% | 94.05% |
| Dakar | 1583 | 0.705 | 7.38% | 5.53% | 1.42% | 84.82% | 0.71% | 6.95% | 92.20% |
| Ouagadougou | 966 | 0.776 | 6.06% | 7.47% | 1.93% | 84.41% | 0.13% | 9.41% | 90.46% |
| Solar_Village | 3065 | 0.498 | 13.25% | 3.41% | 1.61% | 81.53% | 0.20% | 5.02% | 94.78% |
| Capo_Verde | 1490 | 0.596 | 9.73% | 4.36% | 1.17% | 81.21% | 3.36% | 5.54% | 90.94% |
| Mussafa | 763 | 0.41 | 13.90% | 3.41% | 2.20% | 79.76% | 0.24% | 5.61% | 93.66% |

While no site of course has pure dust, the list agrees well with reports in the literature and observations by the TOMS satellite of dust episodes at these locations – most of which are in North Africa apart from Solar Village and Mussafa which are on the Arabian Peninsula. From this list, Banizoumbou and Solar Village were selected since they both had mean daily percentages of dust>80%, and over 5 years of AERONET data records combined with the fact that they also having the longest record of co-located and synchronous satellite data measurements.

The top 10 ranked “biomass burning” AERONET sites using this method are:

| Site Name | N (L2 Inv.) | <AOD> | % SO2 | % OC | % BC | % Dust | % Sea Salt | % OC+BC | % DUST + SO2 |
|---|---|---|---|---|---|---|---|---|---|
| Alta_Floresta | 901 | 0.243 | 18.93% | 66.67% | 11.11% | 2.47% | 0.82% | 77.78% | 21.40% |
| Mongu | 1573 | 0.216 | 26.85% | 59.72% | 11.57% | 0.93% | 0.93% | 71.30% | 27.78% |
| Campo_Grande_SONDA | 620 | 0.193 | 27.98% | 60.62% | 9.33% | 1.55% | 0.52% | 69.95% | 29.53% |
| Rio_Branco | 392 | 0.210 | 23.81% | 61.43% | 8.10% | 5.24% | 0.95% | 69.52% | 29.05% |
| CUIABA-MIRANDA | 506 | 0.167 | 28.74% | 58.68% | 8.98% | 2.40% | 1.20% | 67.66% | 31.14% |
| Yakutsk | 370 | 0.384 | 24.48% | 46.61% | 9.38% | 18.49% | 1.04% | 55.99% | 42.97% |
| Lake_Argyle | 1242 | 0.129 | 25.58% | 44.96% | 8.53% | 18.60% | 3.10% | 53.49% | 44.19% |
| Darwin | 430 | 0.125 | 31.20% | 44.80% | 8.00% | 8.80% | 5.60% | 52.80% | 40.00% |
| Jabiru | 737 | 0.117 | 32.48% | 43.59% | 8.55% | 8.55% | 6.84% | 52.14% | 41.03% |
| Sao_Paulo | 590 | 0.163 | 47.85% | 38.04% | 8.59% | 1.84% | 4.29% | 46.63% | 49.69% |

Note that columns have been added for two combinations of aerosol: the combined percentage of organic and black carbon %OC+BC and also %dust + SO2. Although SO2 accompanies the burning of biomass products, the combination %OC+BC distinguishes biomass burning sites from urban sites which are dominated by SO2. The combination %dust + SO2 is commonly referred to in the literature as polluted dust or dirty pollution as has been added for completeness. The same analysis was performed for ranking sites by “urban SO2” and “sea salt”. In the case of urban SO2, Washington GSFC and Moscow MSU-MO were selected since they both had mean daily percentages of SO2>50%, at least 2 years of AERONET data records combined with the fact that they also have the longest record of co-located and synchronous satellite data measurements. The full analysis is contained in an EXCEL file and made available for researchers in the community at the project website.

**BI-PRODUCT**: the EXCEL file: ***GOCART_global_aerosol_mixtures.xlsx*** is open source and available for the community from the ***Researchers Gateway*** of the project website.

This additional task proved to be invaluable in providing an unambiguous choice of sites corresponding to distinct and dominant aerosol types used in tasks A3 and B1-B2. To this end, a detailed investigation into the known global distribution of aerosol types was initiated to identify representative desert dust, biomass burning and urban SO2 sites to be studied using the outputs from the GOCART aerosol type model incorporated into the NASA GEOS-5 mission. The result of this analysis, taking into account also the availability of AERONET inversion and co-located synchronous satellite input data, led to the selection of the following pairs of sites:

| | | |
|---|---|---|
| DESERT DUST aerosol: | Training site = Banizoumbou | Simulation site=Solar Village |
| BIOMASS BURNING aerosol: | Training site = Alta Floresta | Simulation site=Mongu |
| URBAN SO2 aerosol: | Training site = Washington GSFC | Simulation site=Moscow MSU-MO |

***Deduction of the optimal neural network architecture***

Feed-forward neural networks (NN) with a single layer of hidden neurons were coded in MATLAB’s object-oriented scripting language. The NNs have 3 inputs: one for each AOD measurement at 470nm, 550nm and 660nm, and 38 outputs: one for the value of the volume particle size distribution dV(r)/dlnr (µm3/µm2) in each of 22 logarithmically-equidistant bins of the ASD in the range of sizes 0.05µm ≤ r ≤ 15 µm, and a further 16 outputs for the single scattering albedo, the real and imaginary parts of the complex refractive index and the asymmetry factor at wavelengths 440nm, 675nm, 870nm and 1020nm. The inputs and outputs were connected through two network layers, the first containing hidden neurons with hyperbolic tangent (Tanh) activation functions and the second containing linear neurons as shown in Figure 1d below:

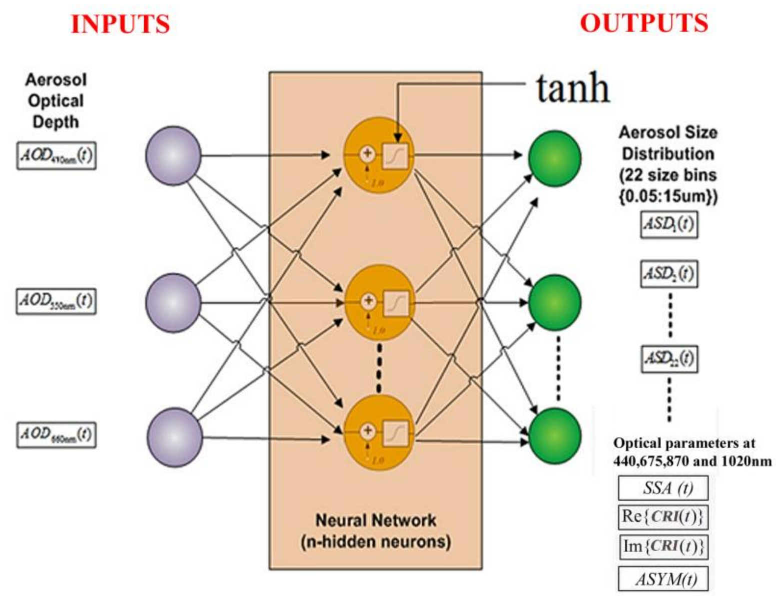



**Figure 1d**: The generic architecture of the neural networks created by AEROMAP

Networks containing differing numbers of hidden neurons effectively model the functional relation between input and output variables differently. There is no consensus on exactly how many neurons should be used for a given problem. AEROMAP has developed a new technique to solve this problem. A grid of neural networks whose weight and bias values were updated (i.e. trained) using the Levenberg-Marquardt back-propagation optimisation algorithm, was created by varying the number of neurons from 3 (the number of inputs) to 38 (the number of outputs) and by varying the % of training data relative to validation data. Once run, the combination of neurons and % training data that has the lowest total (training + validation) mean squared error between the NN outputs and the training outputs (i.e. AERONET data), is selected. To test this approach, a batch run was performed that generated a grid of 180 NNs permuting through a range of architectures such that the number of hidden neurons ranged from 4-38 (in steps of 2) and training fractions ranged from 40% to 90% (in steps of 5%). Each NN in the grid was presented with normalised (min-max) AOD inputs and optical and microphysical parameter outputs from the AERONET station at Washington GSFC is shown in Figure 1e below:

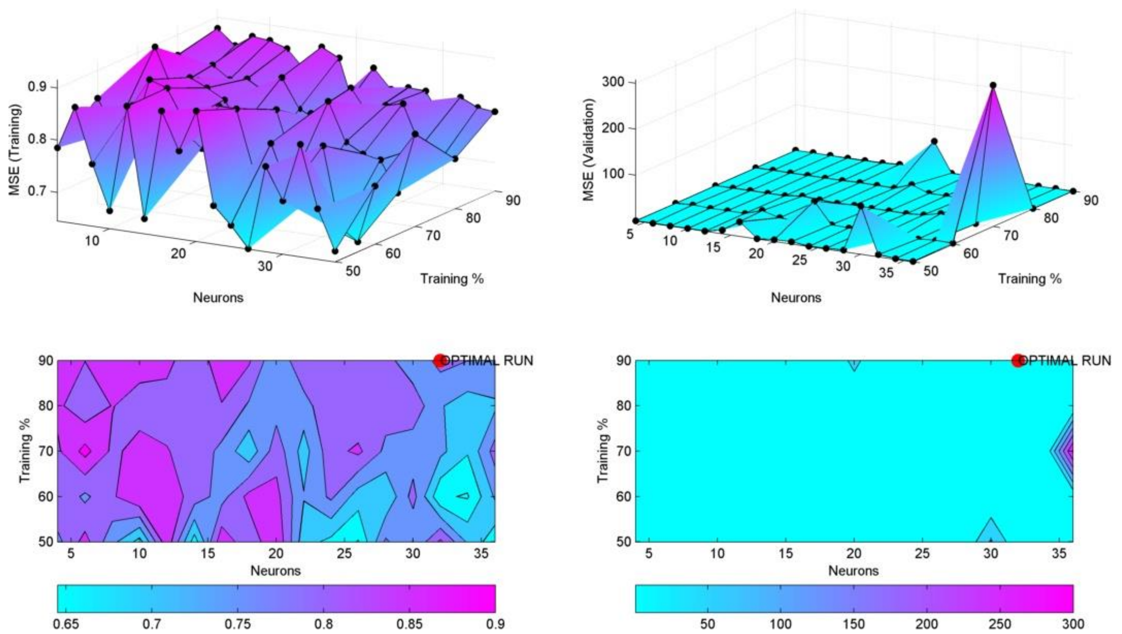


***Figure 6**: The NN mean squared error (MSE) as a function of the number of nonlinear (Tanh) hidden neurons and the fraction of data used (training %). The figures at left show the MSE for the training subset while the figures at right show the MSE for the validation subset ( = 100% - training%). The optimal NN architecture is (32,90%).*

The optimal 32 neuron NN was trained with a goal set at 1/100 of the mean variance of the AERONET data. For this dataset, MSE (goal) = 0.00000311 and the best training performance was achieved after 20 training iterations with a value of MSE = 0.00014419.

***Assessing the temporal and spatial generalisation power of the neural networks***

The networks, trained in this way were then tested by feeding it with ground-based AERONET AOD data at other sites. For the case of the Washington GSFC network above, testing was performed at Moscow MSU. The network successfully reproduced the mean values of all aerosol microphysical parameters as shown in Figures 2a and 2b below.

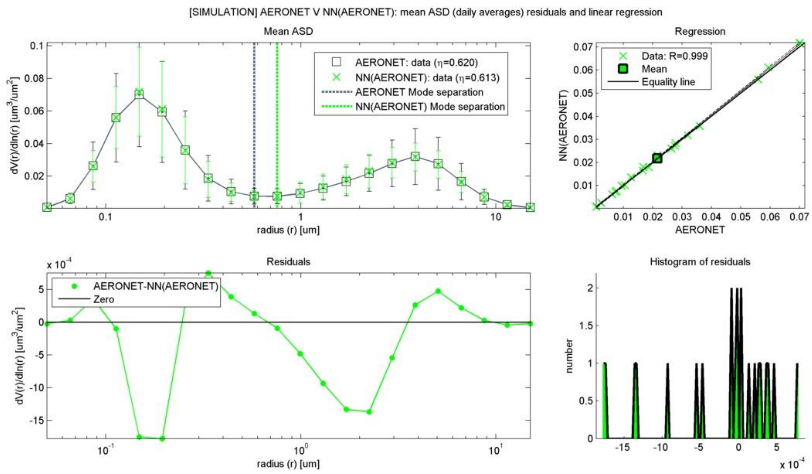



***Figure 2a**: Mean AERONET size distribution at Moscow MSU-MO overlaid with the prediction resulting from feeding the NN trained with AERONET data at Washington-GSFC with ground-based AERONET inputs at Moscow MSU-MO. The vertical dotted lines show the point separating fine and coarse particles.*

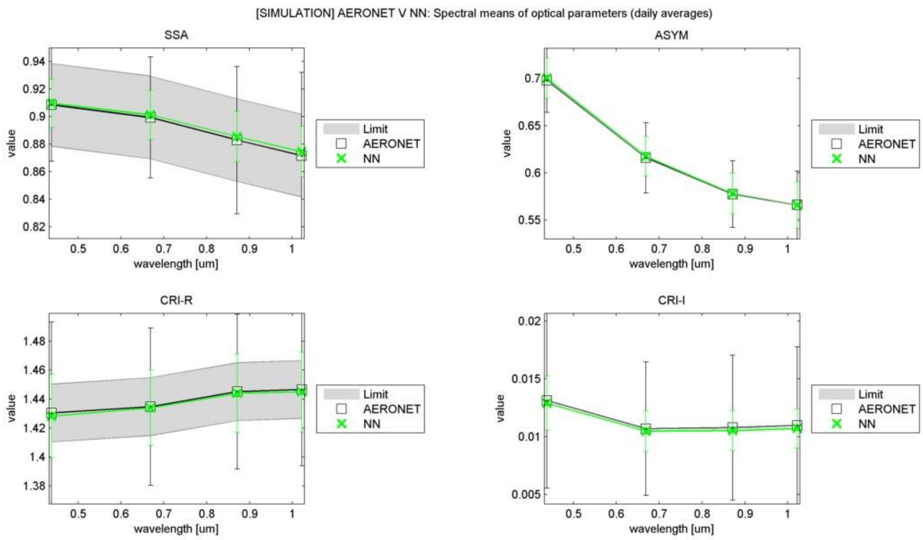



***Figure 2b**: Spectral mean AERONET optical parameters and NN predictions at Moscow MSU. The grey shaded bands represent operational levels of accepted uncertainty*

In order to test the predictions of the network using *satellite* inputs, co-located and synchronous AOD data over Moscow MSU-MO from the MODIS satellite were fed to network. The network successfully predicted the mean values of the aerosol microphysical and optical parameters as shown in Figures 2c and 2d below. It was noted though that the network was slightly under-predicting the mean values of the real part of the complex refractive index as shown in Figure 2d.

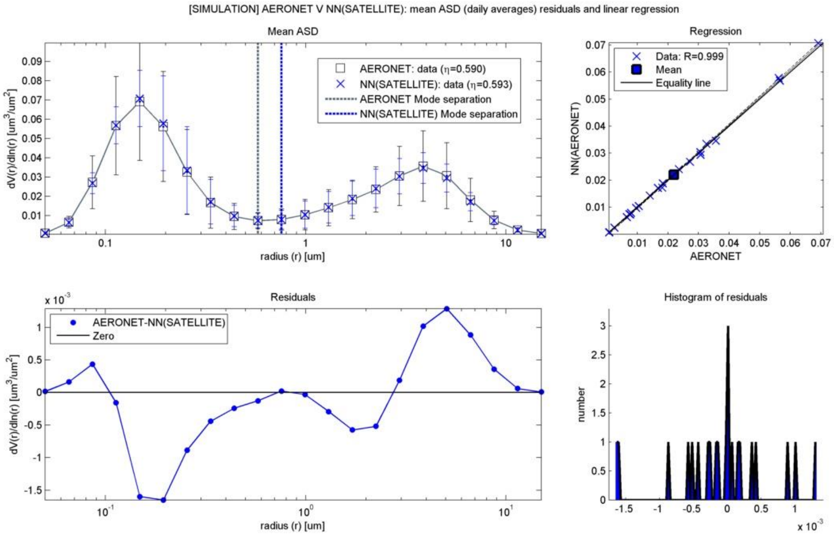



***Figure 2c***: *Mean AERONET size distribution and NN prediction at Moscow MSU-MO resulting from feeling the NN with MODIS satellite AOD inputs*

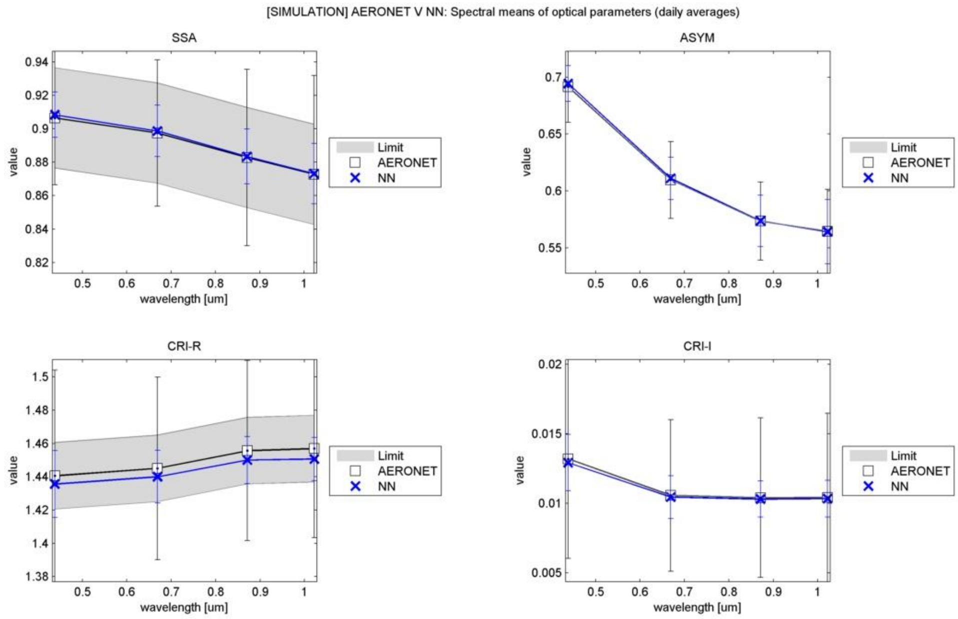



***Figure 2d***: *Spectral mean AERONET optical parameters and NN predictions at Moscow MSU. Notice that the NN is slightly under-predicting the mean values of the real part of the refractive index (CRI-R).*

| |
|---|
| The results of this work were presented in a poster session at the European Aerosol Conference 2012 in Granada, Spain. |

The neural network trained at the urban Washington GSFC site was tested at Moscow MSU-MO, another urban city. This verified the ability of the neural network to extrapolate over a large spatial distance and to retrieve the mean values of aerosol optical and microphysical parameters. With regard to uncertainty analysis it was found that, while the neural network was able to retrieve the average values of the microphysical parameters, it failed to retrieve the temporal variation of the parameters as shown for example by the single scattering albedo in the figure below.

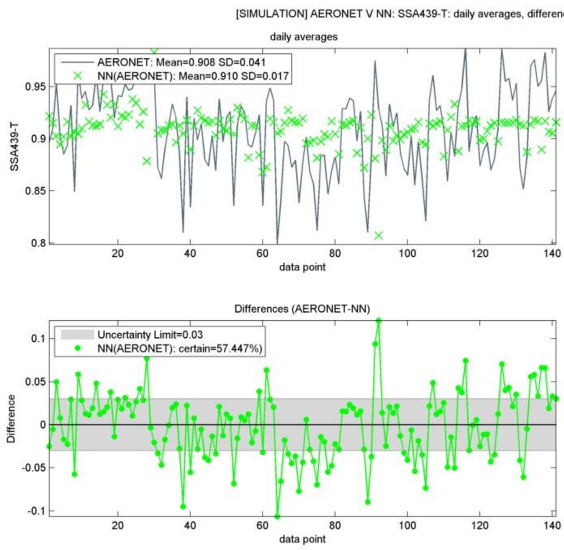



***Figure 2e***: *Daily variation of the AERONET single scattering albedo (439nm) and NN prediction at Moscow MSU. While 57% of the data is within the accepted level of uncertainty, the regression coefficient of AERONET values on NN values is low (R=0.302).*

This issue was addressed following discussions with members of the APC Group in the first Power-point presentation of the results and feedback from the EAC 2012 conference. It was decided that the average values of the microphysical parameters should be calculated over a range of timescales: daily, weekly, monthly, tri-monthly (seasonally) and yearly – so as to assess the operational limit of the neural network.

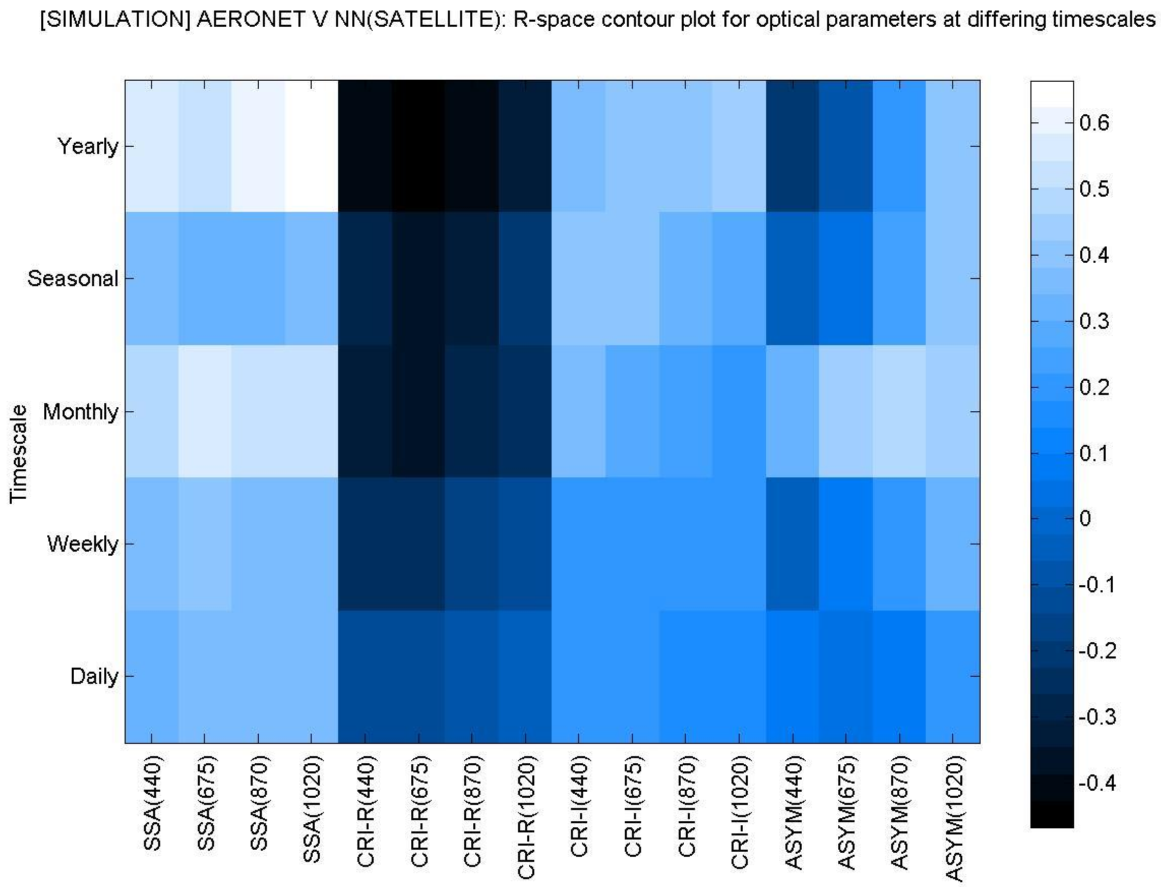


***Figure 2f***: *Values of the regression coefficient (R) between AERONET and NN measurements at Moscow MSU-MO of the aerosol optical parameters resulting from averaging the daily date at 5 different timescales. The NN predictions are calculated using co-located and synchronous AOD inputs from MODIS over Moscow. Note that the best performance appears to be at the monthly scale.*

This timescale analysis was performed and it was found that the neural network trained at GSFC was able to recover the *monthly* variation of the aerosol parameters at Moscow MSU-MO.

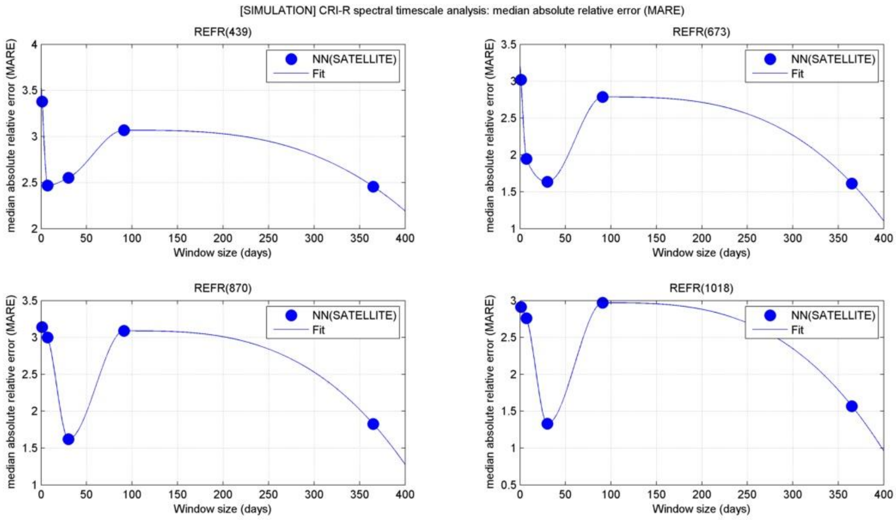
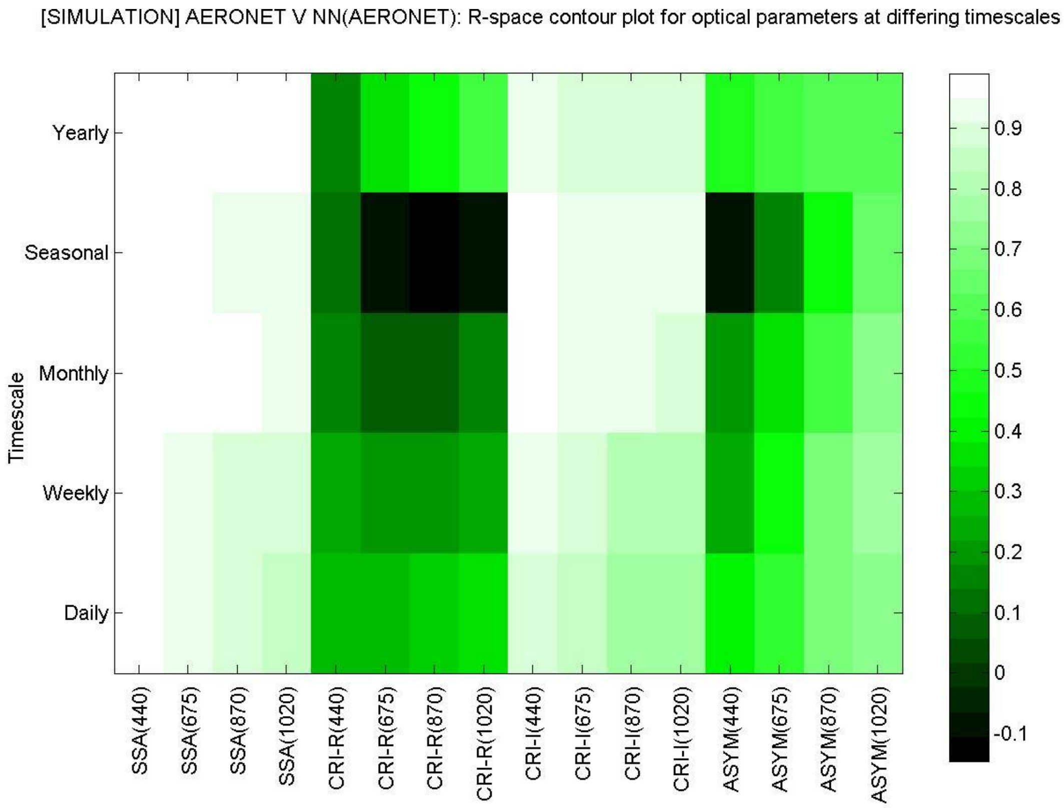



***Figure 2g***: *Values of the median absolute relative error % calculated from the difference between AERONET and NN measurements at Moscow MSU-MO for the imaginary part of the complex refractive index at 5 different timescales. While the error never exceeds 4% the best performance appears to be at the monthly scale (≈30 days).*

In order to improve the performance of the neural network at the shorter timescales required by a daily real-time monitor, it was discovered that the networks required further information in terms of additional inputs (water vapour and absorption AOD) to complement the 3 AOD inputs used.




***<u>Figure 2h</u>**: Values of the regression coefficient (R) between AERONET and NN measurements at Moscow MSU-MO of the aerosol optical parameters resulting from averaging the daily date at 5 different timescales but with 5 inputs. The NN predictions are calculated from AERONET values of these inputs at Moscow MSU-MO. The performance is best (R>0.8) for parameters related to absorption: the single scattering albedo (SSA) and imaginary part of the complex refractive index (CRI-I), but slightly poorer for parameters relating to the nature of the aerosol: the real part of the complex refractive index (CRI-R) and the asymmetry factor (ASYM).*

The networks were found to show the ability to perform well at timescales ranging from 1 day to 1 month. They also strongly show the ability to recover AERONET parameter values at distant sites (of the same aerosol type). These findings confirm that the trained neural networks have i) temporal generalisation power and ii) spatial generalisation power (by aerosol type). The outcome of Tasks A1-A3 was very satisfactory and demonstrated the potential of networks to predict the values of aerosol optical and microphysical parameters with confidence even at distant locations.

TASK A3 was completed by the end of month 7 (taking one month longer than expected) and satisfied Objective 1. The outcome of this task was Deliverable 1 which has been incorporated in the AEROMAP code. Task A3 also part contributed to Deliverables 3 and 4 (see Tasks B1-B2 for details).

| |
|---|
| ***Deliverable 1***: Identification of the functional relationship between AERONET direct sun AOD and AERONET retrieval data |

***TASK B1 (months 7-9): validation of extrapolation power by training with capped data &***
***TASK B2 (months 7-9): validation of extrapolation power by training with one aerosol type region and validating in another global location***

The neural networks, trained and validated on synchronous and co-located data at individual sites in Phase A (tasks A1-A3), were re-trained for datasets comprising multiple sites (of the same aerosol type) and therefore larger and more diverse training data. Since the aim of AEROMAP is to extrapolate retrievals across the globe to regions where no ground stations exist, it is essential to test that networks for different aerosol types are capable of performing well in the presence of data whose aerosol type is less pure and may be of higher variability. Since neural networks are powerful *interpolators* i.e. they are best suited to approximating a function between values that fall in the [minimum, maximum] range of the training data. The potential of neural networks to generalize outside this range – i.e. to *extrapolate* numerically depends on the curvature of the fitting function itself. For monotonic and smoothly-varying fitting functions, extrapolation can sometimes be surprisingly accurate. To test this, 2 networks were trained on more diverse datasets for two aerosol types: desert dust and biomass burning aerosol. The desert dust network was trained on data from 7 sites (1=Agoufou, 2=Banizoumbou, 3=DMN Maine Soroa, 4=Dakar, 5=IER Cinzana, 6=Ouagadougou and 7=Tamanrasset INM) in Northern Africa in the region of maximum global dust extinction (see the map below) and tested at Solar Village in the Arabian Peninsula and at Dalanzadgad in the Gobi desert as illustrated in Figure 3a below:

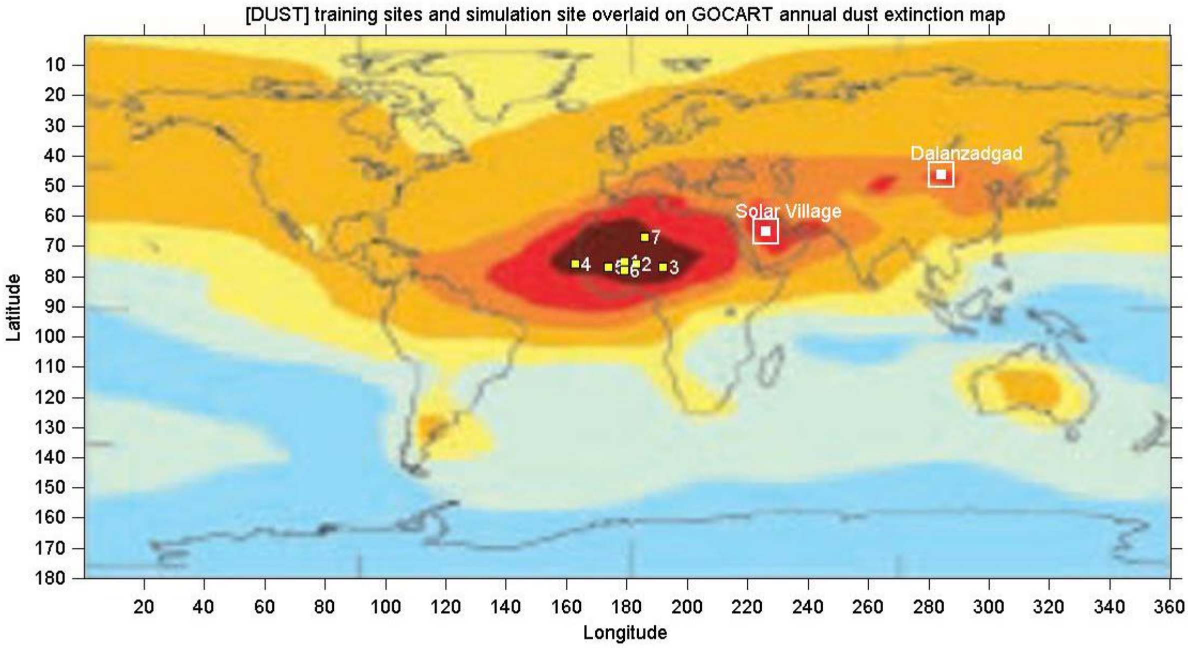
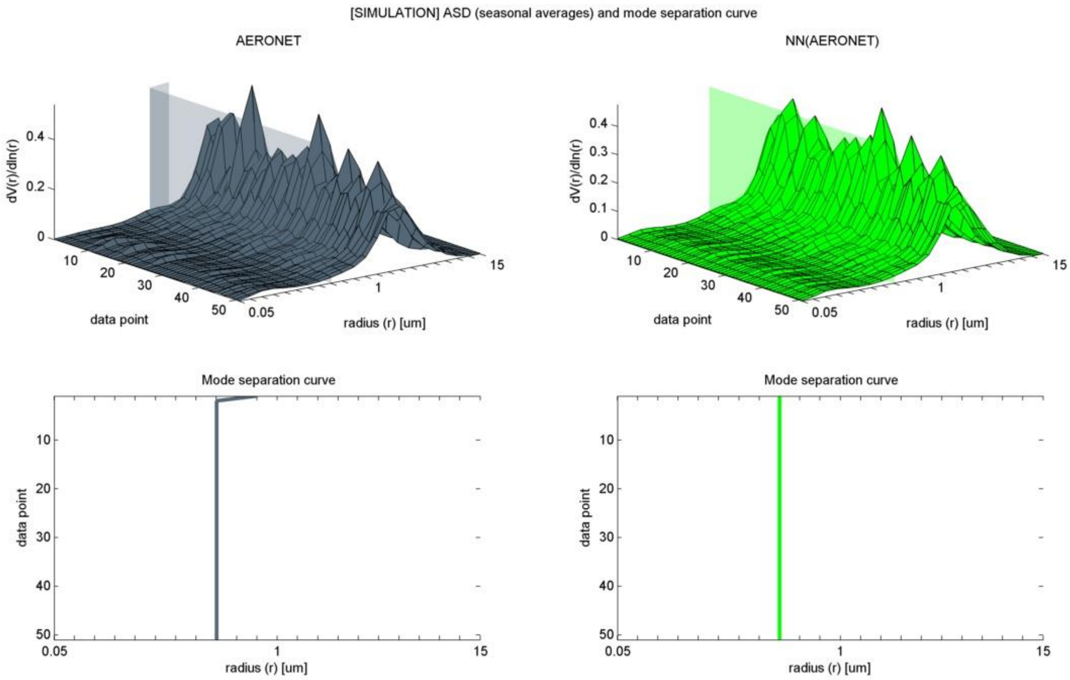
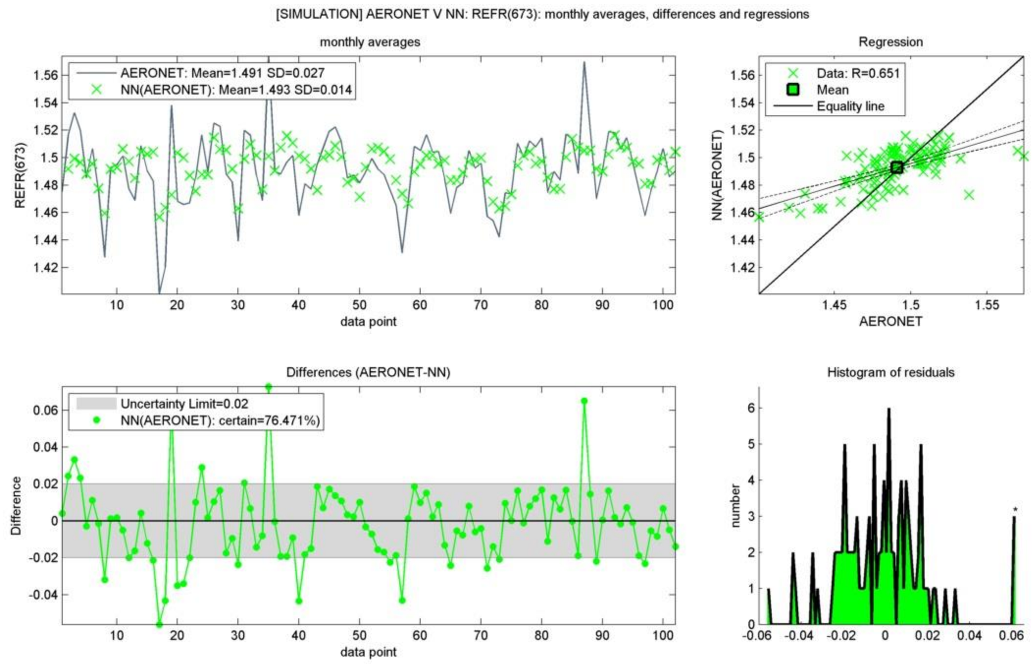



***Figure 3a**: The desert dust training sites (1-7) in the Sahara and the two test sites Solar Village and Dalanzadgad located in the Arabian and Gobi deserts overlaid on the global GOCART map of extinction AOD.*

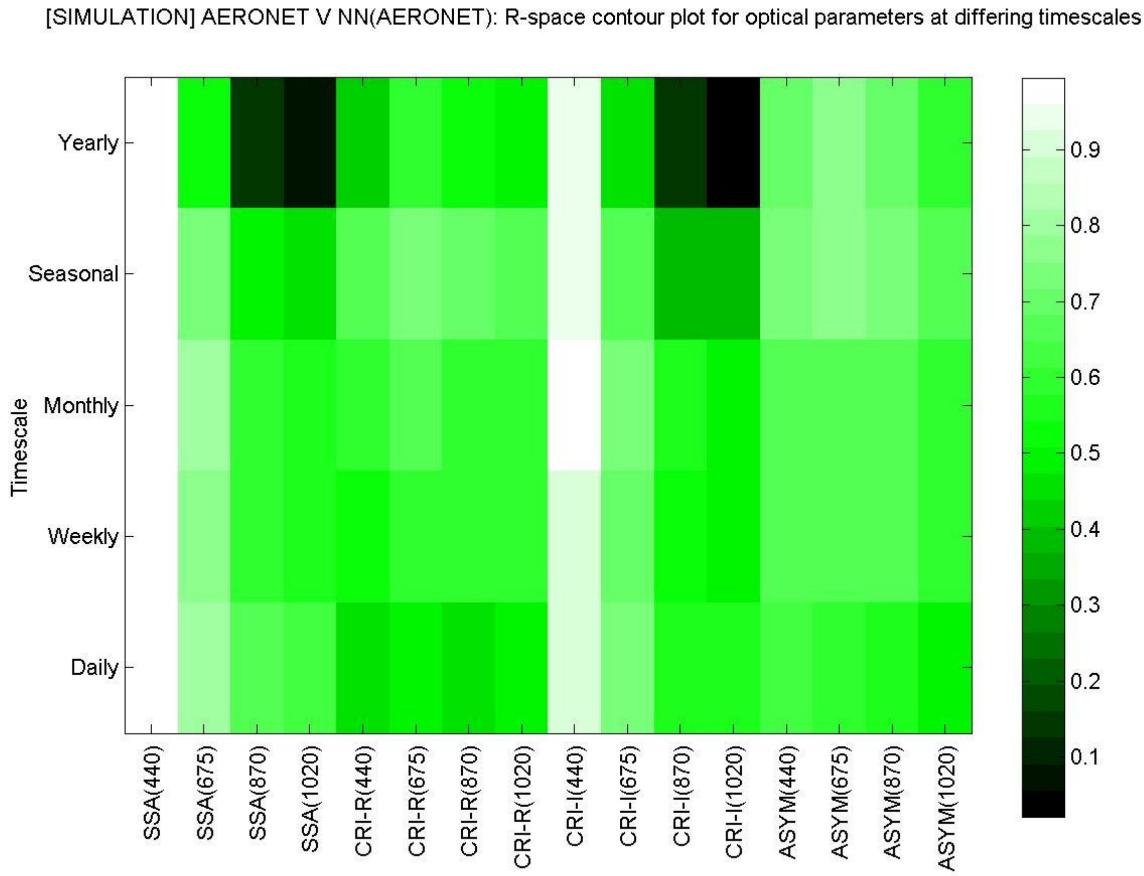
Half of the training dataset was selected at random and used as capped data for training. The other half was used to test the *interpolation* power of the network with AERONET inputs. The trained network was then tested on satellite inputs at the two test sites: Solar Village and Dalanzadgad to test its *generalisation* power. Figures 3b-3f show some of the key findings:




***<u>Figure 3b</u>****: The daily size distribution retrieved by the NN having AERONET input from the remaining 50% of capped data used to train the North African desert dust sites (1-7)*. There is excellent agreement also in the location of the mode separation curve.




***<u>Figure 3c</u>****: The daily real part of the refractive index retrieved by the NN having AERONET input from the remaining 50% of capped data used to train the North African desert dust sites (1-7)*. There is reasonable agreement for this notoriously difficult to retrieve parameter as reflected in the value of R=0.651.




***<u>Figure 3d</u>**: The timescale analysis and regression coefficients of the optical parameters for the same run as Figures 3b and 3c*. The fact that regression is seen to improve when moving to smaller timescales (more average values) suggests that the NN is interpolating very well on this data.

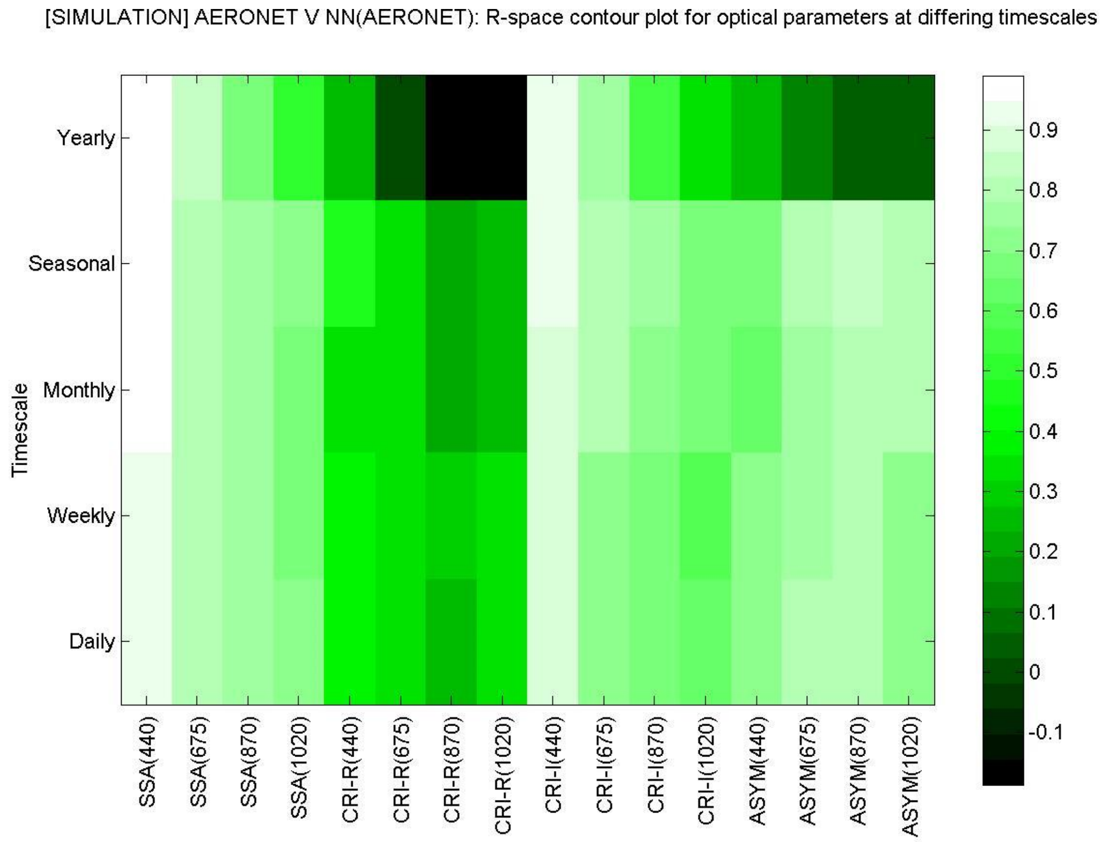



***Figure 3e***: The timescale analysis and regression coefficients of the optical parameters using the same NN as in Figures 3b-3d but with AERONET inputs at **Solar Village**. Regression peaks around the weekly timescale with R>0.5 for all parameters. There is a clear under-performance with respect to the real part of the refractive index (CRI-R) at all wavelengths compared to other parameters.

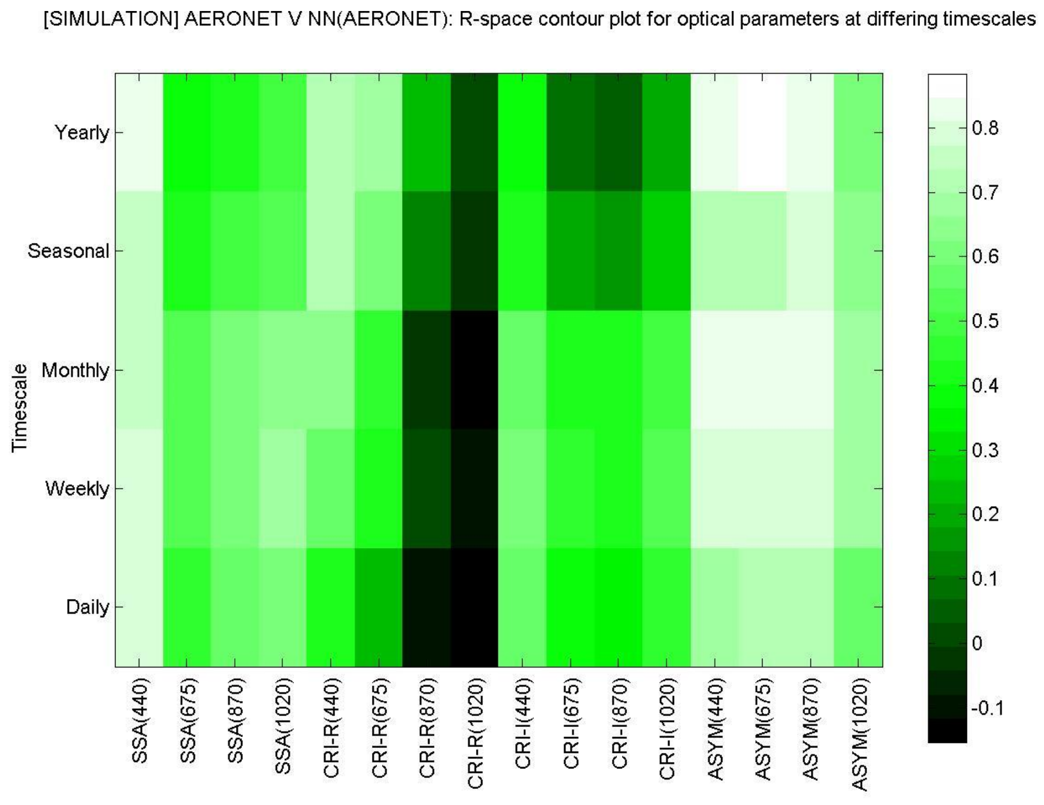


***Figure 3f***: The timescale analysis and regression coefficients of the optical parameters using the same NN as in Figures 3b-3d but with AERONET inputs at **Dalanzadgad**. Regression peaks around the monthly timescale with R>0.4 for all parameters except the real part of the refractive index (CRI-R) at 870nm and 1020nm. This suggests that the aerosol content at Dalanzadgad is different to that in the North African source region used to train the NN.

The biomass burning network was trained on data from 5 sites (1=Alta Floresta, 2=CUIABA-MIRANDA, 3=Campo Grande SONDA, 4=Ji Parana SE, 5= Rio Branco) in central Amazon in South America in the region of maximum global biomass burning black carbon extinction (see the map below) and tested at Mongu in the African Savanna as illustrated in Figure 4a below:

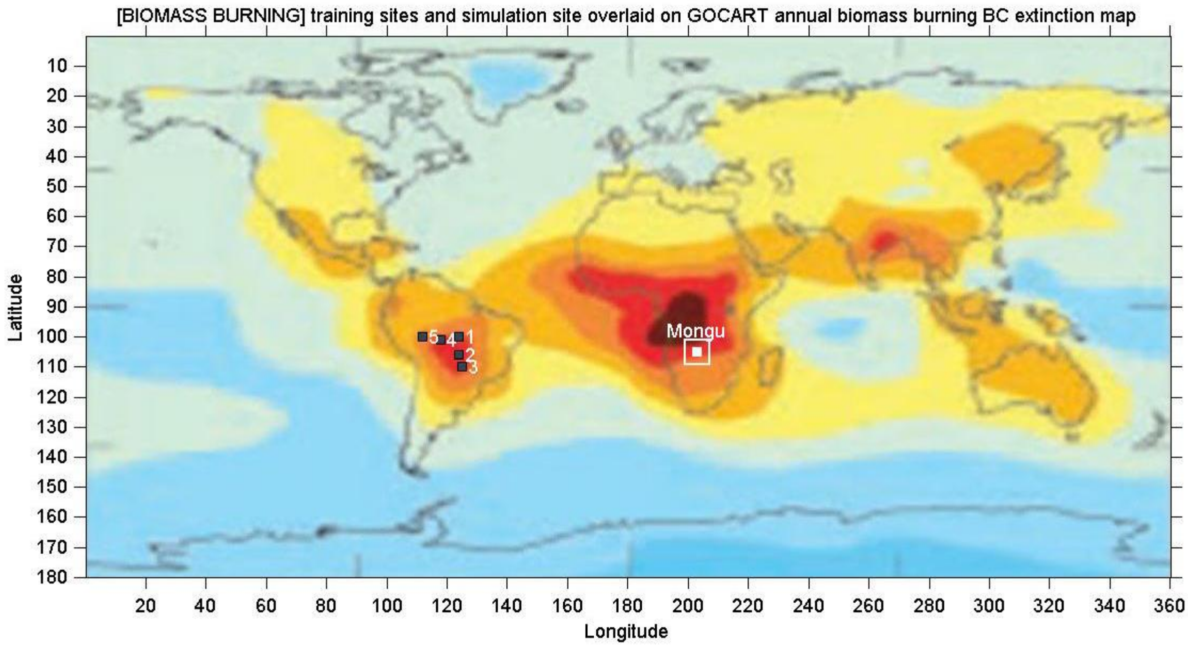



***Figure 4a****: The biomass burning sites (1-5) in the Amazon used to train a neural network for this type of aerosol and the test site Mongu located in the African Savannah overlaid on the global GOCART map of extinction AOD for black carbon (BC).*

As for the study of desert dust, half of the training dataset was selected at random and used as capped data for training. The other half was used to test the *interpolation* power of the network with AERONET inputs. The trained network was then tested on satellite inputs at Mongu to test its *generalisation* power. Figures 4b-4e show some of the key findings:

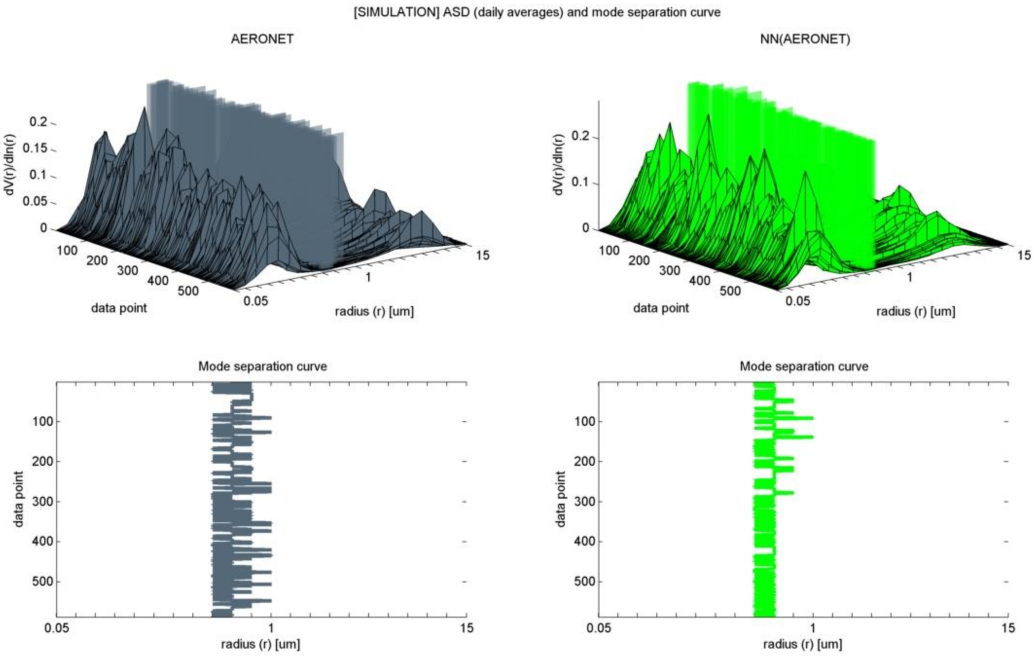



***Figure 4b****: The daily size distribution retrieved by the NN having AERONET input from the remaining 50% of capped data used to train the Amazonian biomass burning sites (1-5)*. While there is good agreement in the general location of the mode separation curve, the curve for biomass burning shows a great deal more variation than that for desert dust in Figure 4b.

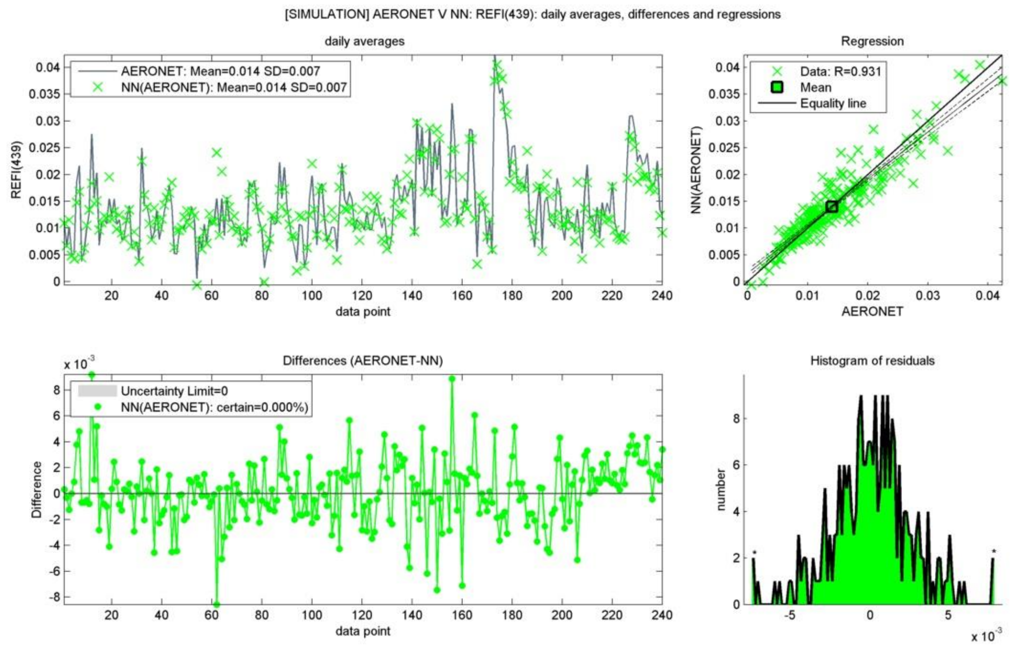
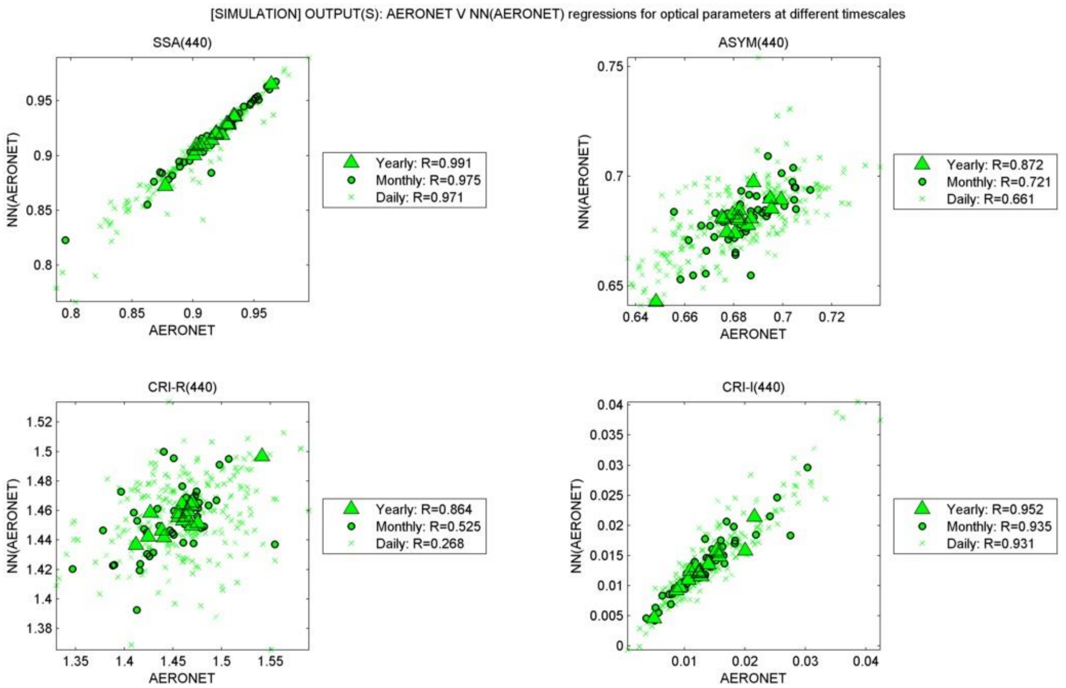
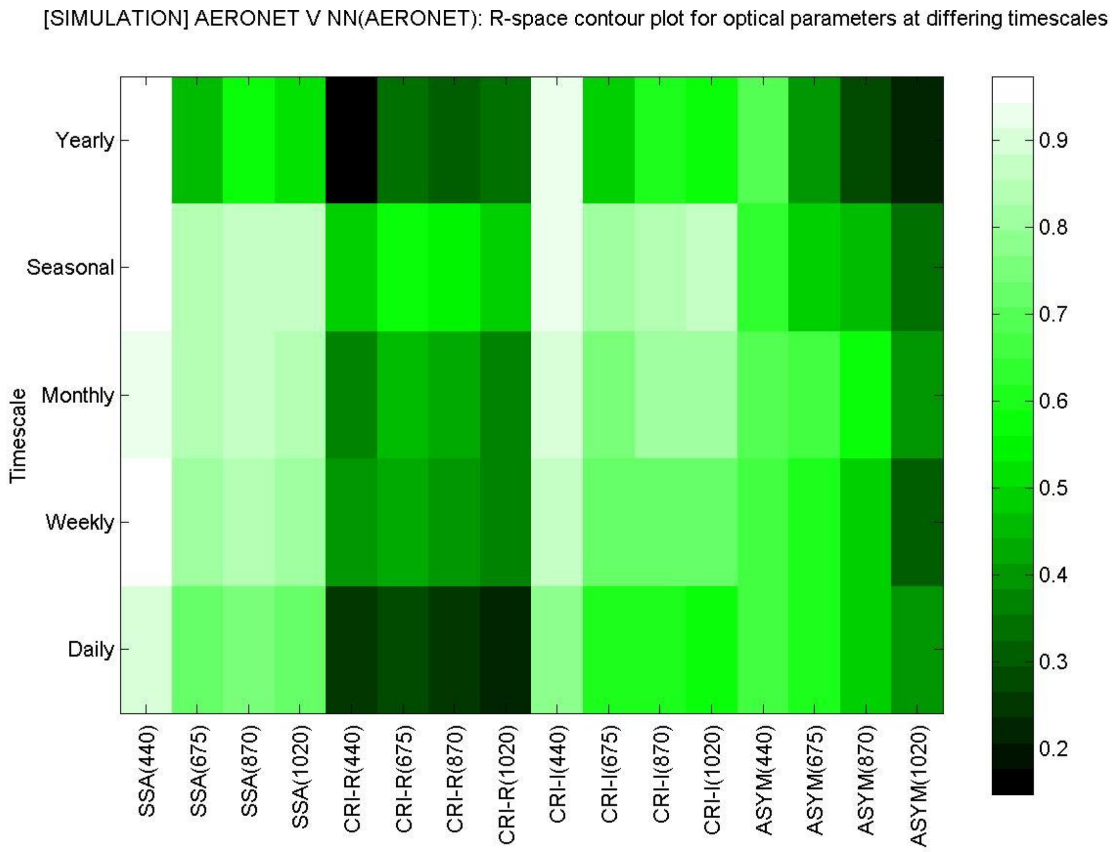



***Figure 4c****: The daily imaginary part of the refractive index at 400nm retrieved by the NN having AERONET input from the remaining 50% of capped data used to train the Amazonian biomass burning sites (1-5)*. There is excellent agreement for this absorption parameter.




***<u>Figure 4d</u>:*** The timescale analysis and regressions at the yearly, monthly and daily timescales for optical parameters at 440nm. Strong correlation is observable at all timescales for the single scattering albedo (SSA) and the imaginary part of the refractive index associated with the absorbing nature of fine biomass burning aerosol.




***<u>Figure 4e</u>***: The timescale analysis and regression coefficients of the optical parameters using the same NN as in Figures 4b-4d but with AERONET inputs at **Mongu**. Regression at the daily timescale is particularly high (R>0.8 for the

SSA(440nm) and the CRI-I(440nm). This is likely due to the fact that biomass burning is dominated by the fine mode that is more absorbing at short wavelengths thereby correlating strongly with the absorption AOD(500nm) which is used as a satellite input. A significantly lower is again observed for the real part of the refractive index (and also the asymmetry factor at longer wavelengths).

Regarding the variability in the mode separation curve seen in Figure 4b, the same algorithm used by AERONET to deduce secondary microphysical parameters from the size distribution was coded into AEROMAP. These parameters measure the effective radius and variance of fine and coarse particle modes which are separated by the mode separation point. In addition, the fraction of the total volume occupied by the fine mode (the fine mode fraction FMF, also denoted by the Greek symbol η) is a further parameter provided by AERONET and which was also calculated by our code. In Figures 5a-c below we show comparisons of the values of these parameters from AERONET against those deduced from the size distribution predicted by the NN:

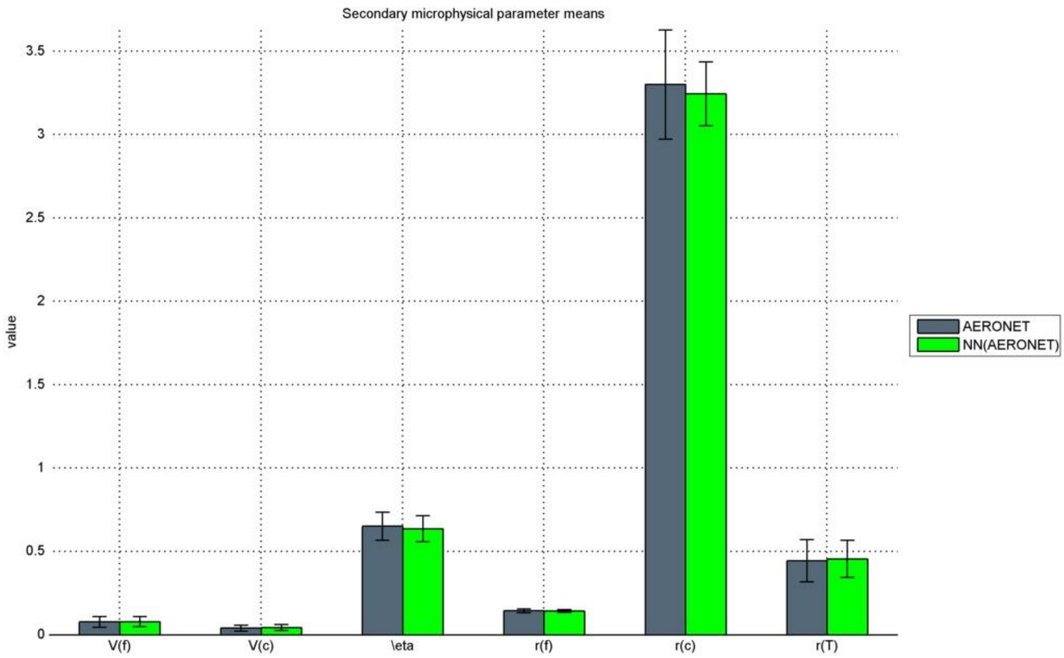



***Figure 5a***: Comparison of the NN deduced secondary parameters describing the fine and coarse particle modes against AERONET values for the NN simulated with AERONET inputs at **Mongu**. Most notably the NN is underpredicting the effect radius r(c) of coarse mode particles for this site.

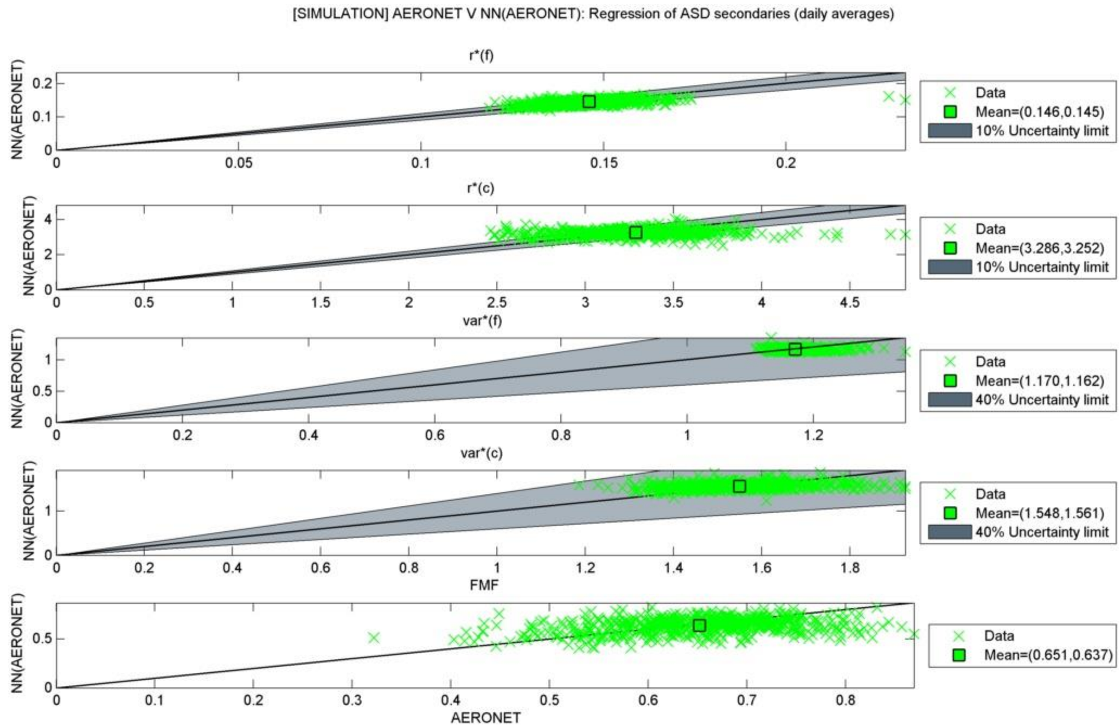



***Figure 5b***: Regression of the NN deduced secondary parameters describing the fine and coarse particle modes against AERONET values for the NN simulated with AERONET inputs at **Mongu**. The grey bands represent acceptable levels of uncertainty. There is generally good agreement and it should be noted that the horizontal nature of the coarse mode effective radius retrieval appears to be due to the NN over-predicting daily values below the mean and under-predicting daily values above the mean.

The size distribution is known to be best fit by a combination of two log-normal distributions – or equivalently a pair of Gaussians in the domain of log(radius). Figure 5c shows the bi-modal fit to the mean size distribution for the NN simulated at Mongu:

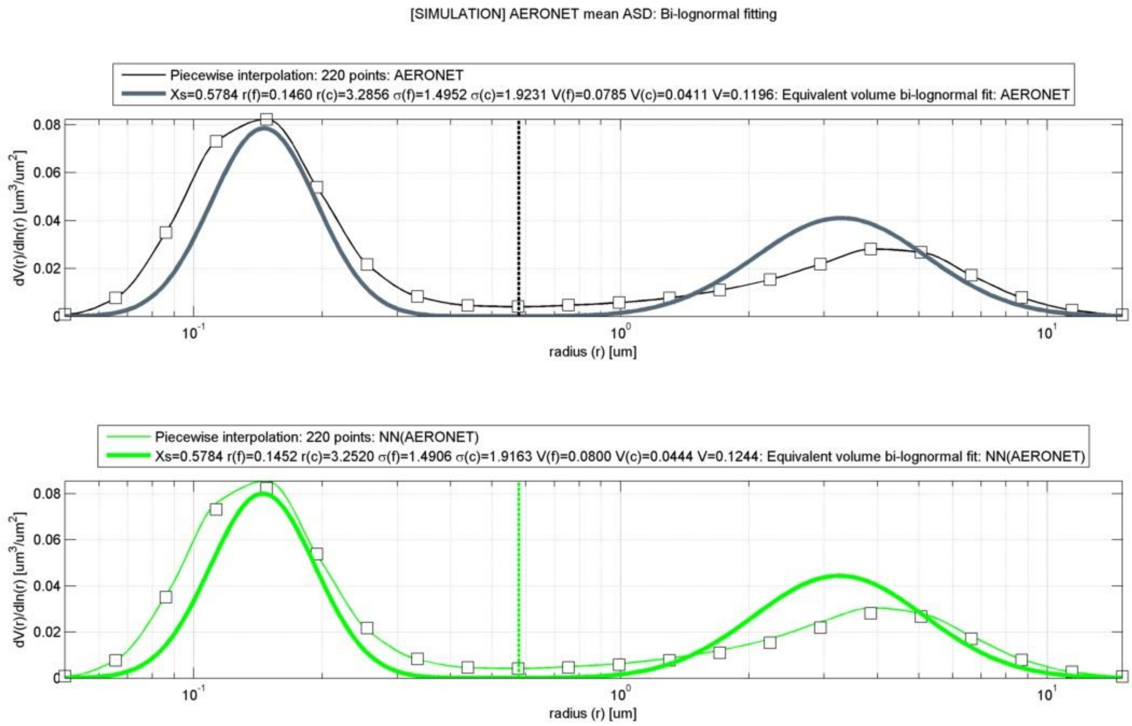



***Figure 5c***: Comparison of the interpolated mean size distribution from AERONET and predicted by the NN at Mongu, and the bi-log-normal fit obtained using the mean values of the effective radii and variances of the two modes. It is clear that both AERONET and the NN are failing to fit the coarse mode of the size distribution.

To investigate why the coarse mode is poorly fit, we binned the size distributions by aerosol load as reflected by the values of the AOD(440nm) as shown in Figure 5d below:

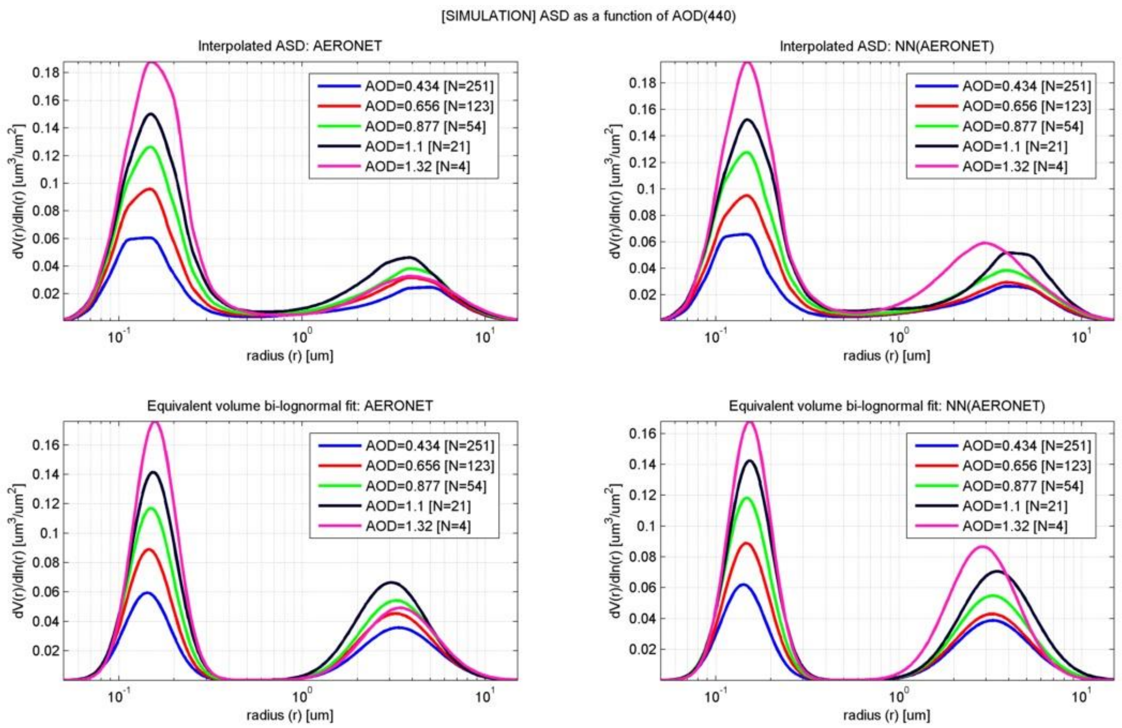


***Figure 5d***: Impact on the shape of the mean size distribution at Mongu resulting from aerosol load as measured by the magnitude of AOD(440m). It is clear that increasing the load significantly affects the fine mode (the left-most peaks) for both AERONET and the NN predictions and has a much lower impact on the coarse mode.

The fact that both AERONET and NN predictions at Mongu poorly modelled the biomass burning coarse suggested that modelling the size distribution with a bi-log-normal fit is incorrect. This inspired us to study the modelling of the size distribution with more modes. Daily-averaged GOCART data was obtained for the following sites of different key aerosol types:

DESERT DUST: Banizoumbou
BIOMASS BURNING: Mongu
URBAN SO2: Washington GSFC
MARINE SEA SALT: Lanai

and was ranked by percentage of aerosol type so as to identify the date when that type was at a maximum. Using this method it was found that the highest percentage of dust contribution to the absorption AOD at Banizoumbou for example, was on the 16th of March 2005. The same was done at Mongu, Washington GSFC and Lanai. The AERONET size distribution on these peak days was extracted and modelled with an increasing number of Gaussians. Using the sum of squared error (SSE), the coefficient of determination (R2) and the root mean square error (RMSE) as measures of the goodness of fit of the multi-modal Gaussians against the raw size distribution data, the optimum number of Gaussians needed to fit the distribution was obtained for each aerosol type as illustrated in Figure 6 below:

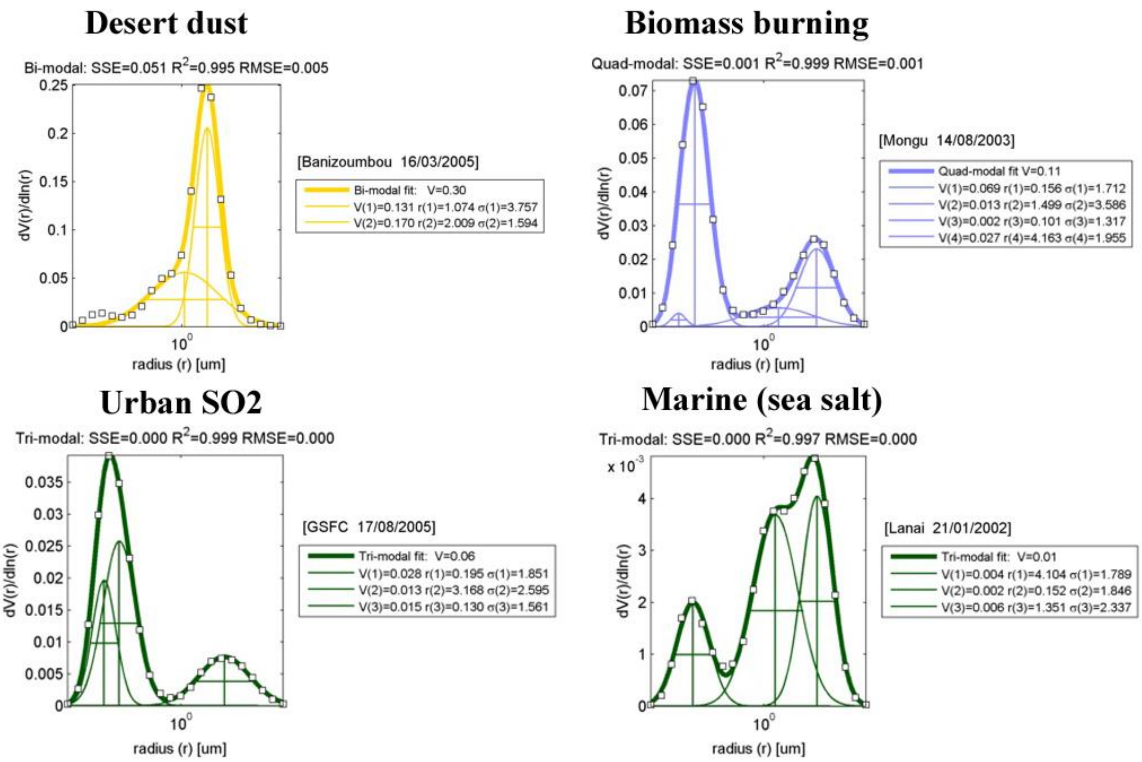



***Figure 6***: Gaussian mixture model fits to the raw AERONET size distribution at a desert dust, biomass burning, urban SO2 and marine site on the day of purest aerosol type presence. Desert dust was best modelled with 2 modes but the fine mode peak was not fit. The biomass burning site was best fit with 4 modes (essentially a combination of two fine modes, a coarse mode and an intermediate mode). The urban SO2 site and the marine site were both best modelled with three modes. The urban SO2 site involved a combination of two nearby fine modes while the marine site involved the overlap of two mid-coarse modes – the characteristic mode around 1 micron for sea salt is evident as one of them.

This approach offered a partial explanation into why the fine mode of desert dust is difficult to model with trained NN and also why the coarse mode of biomass burning aerosol is difficult to fit. The results of this analysis and the development of the multi-modal technique using Gaussian mixture models is being written up into a paper and will soon be submitted for publication.

The outcome of Tasks A1-3 and B1-B2 led to satisfaction of Objectives 1-4 and provision of Deliverable 2 and Deliverable 4 of the project.

> ***Deliverable 2***: Verification and uncertainty analysis demonstrating the ability of neural networks to extrapolate retrievals to regions of similar aerosol type
>
> ***Deliverable 4***: Verification and uncertainty analysis demonstrating the ability of neural networks to use satellite inputs to extrapolate retrievals at new site locations

The results of the desert dust study have been written up and the fellow is in the process of submitting a paper that describes AEROMAP's core methodology and constitutes an assessment report on the function approximation ability and extrapolation power of the NNs.

> ***Milestone 1***: Assessment report on the function approximation ability and extrapolation power of the NNs

***TASK C1 (months 9-12): Cluster analysis of global AERONET sites***

AEROMAP has already identified a new technique for assessing aerosol types based on using GOCART 3-hourly measurements of aerosol type contribution to extinction AOD as described above  This is novel and a paper is in preparation on the methodology developed here. As we describe below, this was no accident but instead resulted from the failure of cluster analysis to offer a way forward. Early in month 3 of the project during the site selection process for task A3, AEROMAP repeated the cluster analysis used by Omar et al (2005) on AERONET data up until 2002, but applied the method to the complete AERONET record spanning the years 1996-2011 to assess how cluster centres have changed in the 10 years since that study was performed. The Omar study identified 6 clusters and this was also the conclusion of our study on the full dataset as shown in figures 5a and 5b below:

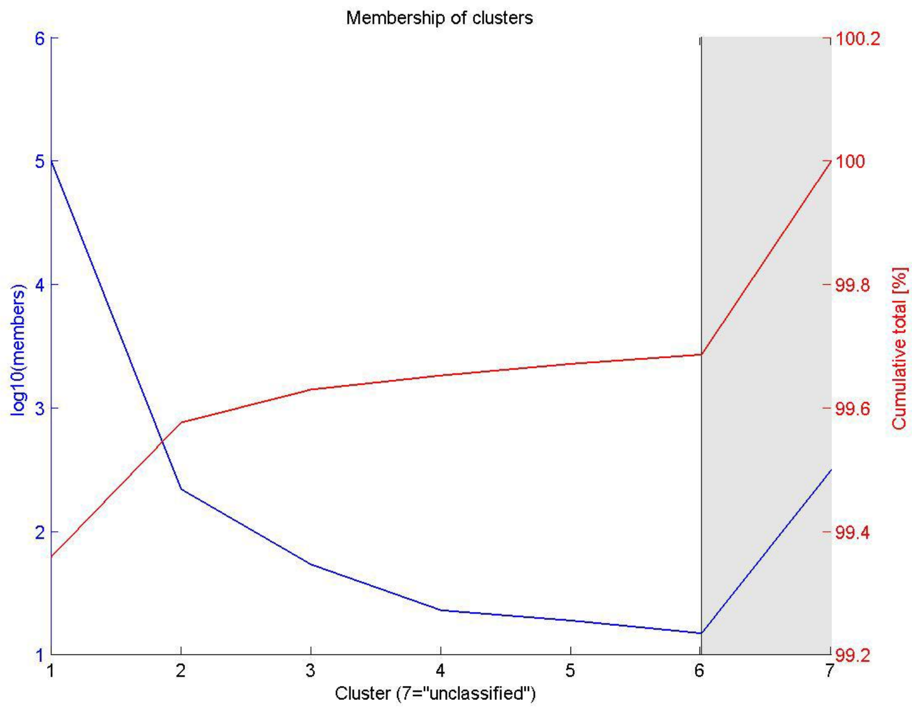



***Figure 5a***: Cluster analysis on the complete AERONET Level Version 2.0 inversions product from 1996-2011 showing that 6 aerosol types (clusters) account for approximately 99.7% of the data

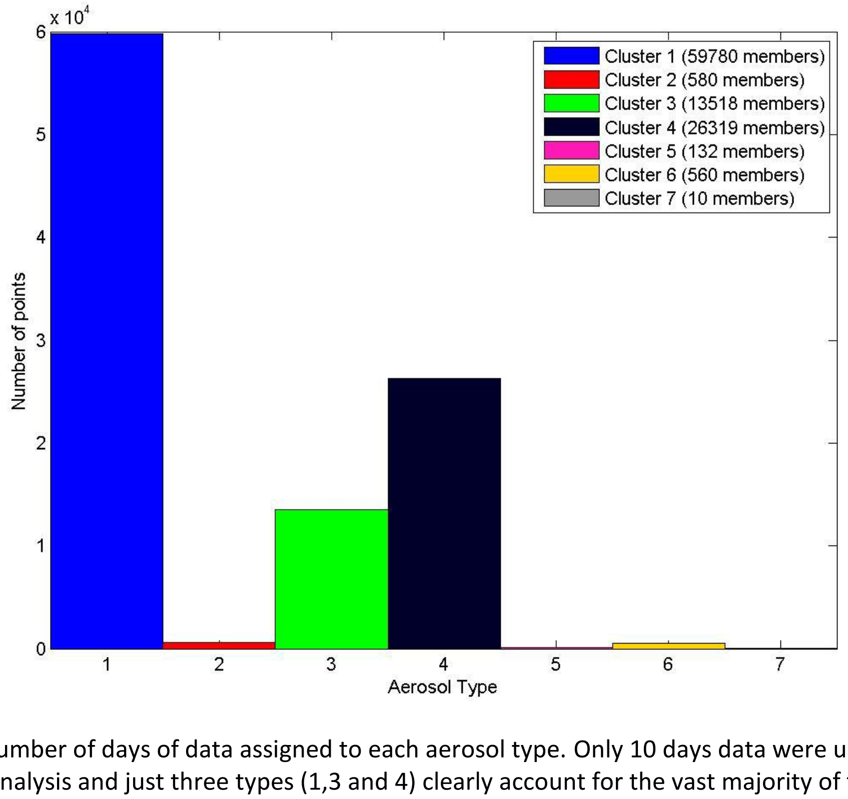



***Figure 5b***: The number of days of data assigned to each aerosol type. Only 10 days data were unclassified by the cluster analysis and just three types (1,3 and 4) clearly account for the vast majority of the data.

However, just three types were seen to dominate. Analysis of their optical and microphysical properties allowed us to identify them as desert dust (type 1), biomass burning (type 4) and urban SO2 (type 3). It is clear that the method failed to identify the vast source of global aerosol arising from sea salt and this raised alarm bells. It was realised that there is a logistical problem with applying cluster analysis to AERONET Level 2 (and Level 1.5) inversion data. As mentioned earlier under Task A1, AERONET complete records of inversions are only provided when certain meteorological conditions are met such as AOD>0.4. Ocean aerosol has been found to have loads much lower than this around 0.1-0.15 and these sites do not have inversion data in the database entries. Indeed, analysis of the EXCEL file ranking sites by complete records of inversion data revealed this selection effect and led to the absence of such sites from the list. As such, it was decided that cluster analysis applied to inversion data led to a dead end and a different approach to site selection should be adopted. This provided the motivation for the approach using GOCART 3-hourly data as a viable alternative. AEROMAP still has one remaining obstacle to overcome with regard to aerosol classification. While robust and reliable neural networks have been trained for desert dust, biomass burning and urban SO2 aerosols, AEROMAP still needs to find a way to model ocean aerosol in the absence of full inversion data. In this regard, the expertise of Dr Alexandra Tsekeri (please see section 2.3.1 for a justification of her involvement in the project) will be paramount. Dr Tsekeri has performed extensive work on the assessment of marine aerosol with AERONET and LIDAR data and a collaboration has begun to assess suitable AERONET data and satellite inputs that can be used to model marine aerosol. This is an active line of current enquiry of the project and key to the development of a NN for marine pixels. Furthermore, it will be necessary to investigate how to model and predict the aerosol optical and microphysical parameters for mixtures of aerosol (e.g. 50% desert dust and 50% urban SO2) – i.e. given satellite data at a pixel on the Earth surface, how should the individual networks for desert dust and urban SO2 be combined? This is still an open question and is the current focus of the project.

***TASK C2 (months 9-15): Pilot studies***

To date, the trained and validated dust, biomass burning and urban SO2 neural networks obtained in Phases A and B have been tested and validated at the following sites:

- DESERT DUST: Banizoumbou (Sahara), Solar Village (Arabian Peninsula) and Dalanzadgad (Gobi desert)
- BIOMASS BURNING: Alta Floresta (Amazon) and Mongu (Central Africa)
- URBAN SO2: Washington GSFC (USA) and Moscow MSU-MO

AEROMAP is in the process of testing the networks at other key locations important for their socio-economic impact:

a) Urban industrial sites: Belsk (Russia) and Taihu (China)
b) European heritage cities: Rome (Italy) and Athens (Greece)
c) Polluted marine sites: Sicily (Italy) and Crete (Greece)

Other important site-pairs such as volcanic ash and radiation cloud-affected regions will also be identified and investigated. Any discrepancies found will be seriously investigated and used to provide information of the feasibility of using neural networks to extrapolate aerosol parameter values globally. Furthermore, the site pairs Rome and Athens (urban industrial sites) and Sicily and Crete (polluted marine sites) have been chosen for the potential impact of the findings on other aspects of the European research effort on environmental quality. The archaeological importance of Rome and Athens means that national (Italian and Greek) and European policy-makers have an interest in the impact for example of desert dust or urban pollution on historical sites. Crete and Sicily are islands having a micro-economy dominated by tourism, fishing and agriculture. A correct aerosol assessment at these locations can have a significant local impact.

***TASK E1 (months 0-12): To design a project website, dissemination and public outreach activities***

This task corresponds to Phase E (months 1-24): Website design, maintenance and public engagement and is an activity that continues during the full duration of the project. Phase E1 (months 0-9) was dedicated to the design and development of the project website and is now complete. Please also kindly refer to section 2.3.6 for details. The project website:

1. provides a non-technical description of climate change and the role of the global aerosol system with links to non-technical resources and multimedia
2. provides a non-technical description of the project
3. describes in detail the project participants and has prominent links to the host institute
4. provides transparency by allowing public access to the project proposal description of participants, and financials (where appropriate)
5. shows the progress of the project on a timeline of the work plan and indicates project milestones and actual deliverables
6. provides open access to the dissemination of results in peer-reviewed publications, conference proceedings, posters and Power-point presentations
7. provides contact details for submission of public queries on the project that will be answered by email
8. provides dedicated pages for events organized by the fellow in relation to the project
9. contains an education gateway complete with links to video animations, approachable feature articles in the popular science press, and lecture courses as part of the Open Course Ware initiative

As part of the next phase of the project he will code the data portal which will house the real-time monitor. Furthermore, the website will hopefully increase its traffic following the issuing of press releases outlining project findings and results. Any media generated by this initiative including press cuttings, radio interviews or podcasts, and news video clips will be added to the Media page of the website. The fellow is already an active contributor to Wikipedia and will write a dedicated entry on AEROMAP. A Twitter account has already been opened having the hash-tag: _AEROMAP and the fellow is in the process of creating a Facebook Group for the project. Two important gateways have also been added to the project website to facilitate interaction with non-specialist visitors and educationists and to provide them with informative content.

As the project progresses into phases E2 and E3, the plan of outreach mass media activities and public engagement will be implemented. Some progress has already been made in this direction. In particular, the fellow is making arrangements for a public talk entitled "Marie-Curie and atmospheric chemistry" which will pay tribute to the accomplishments of Marie-Curie and which will describe her legacy in the form of Marie-Curie actions, in particular, the fellowship for the proposed project and its focus on atmospheric chemistry. This will take place in the visitor centre of NOA and which is open to the public one evening per month. It accommodates up to 125 visitors and is oriented to ages 8 years and above. The fellow will arrange with the organiser, Dr Nikolaos Matsopoulos, to assign a date for this and a second date dedicated to a public debate (to be introduced and chaired by the fellow) on the subject of climate change and pollution. It is hoped that these events will increase awareness of the important role played by Marie-Curie actions in facilitating European research and mobility, the role of scientific research on society, the project AEROMAP, the fellow, the host institute, and work currently being funded and undertaken in the ERA on the global aerosol system.

The fellow has disseminated the results of the project via participation in the large (>500 participant) European Aerosol Conference in Granada, Spain, 2-7 September 2012 and via presentation of a digital poster session geared to the general public at the European Science Open Forum in Dublin, Ireland, 11-15 July 2012. Two research articles have been prepared for submission to peer-review in academic journals and he has designed and developed a project website/portal containing gateways for the interested non-scientist, for researchers and for educationalists. AEROMAP also had a prominent role in the host institute ACP Group's poster and stand at the FP7-funded Researcher's Night at the National Hellenic Research Foundation and the fellow is arranging a venue for a public talk on the role of aerosol research on understanding climate change, the environment, public health and the general quality of life. The success of this programme of science communication activities will lead to satisfaction of Objective 9.

## 2.3 Project management during the period

### 2.3.1 Management tasks and achievements

The host institute and in particular the SIC, has successfully provided research supervision, expertise and has facilitated collaborations essential for the advanced training needs of the fellow. Following provision of a comfortable workspace in the ACP Group at IERSD-NOA and incorporation of the fellow into the group and its website, the SIC helped the fellow smoothly through the process of registering with the Greek tax office, obtaining health insurance, arranging payment and generally settling in to his new environment. Necessary technical equipment (namely the purchase of a state-of-the-art desktop PC for the project) and software (namely MATLAB) was made an immediate priority and quickly acquired. On the technical side, the SIC assisted the fellow in accessing and obtaining satellite remote sensing and ground-based datasets and showed his great experience in this regard – allowing the fellow to get up to speed rapidly. As a result, the first neural networks were trained within the first 4 months of the start date and gave the fellow the opportunity to meet the call for abstracts at the European Aerosol Conference 2012 where the first results of the project were presented. In addition, the SIC early on capitalised on his collaboration network and opened up a collaboration with the University of Ioannina where Dr Antonios Gkikas had recently developed a more precise technique for sampling MODIS pixel data. Furthermore, the SIC placed the fellow in contact with an expert on airborne aerosol measurement and retrieval, Dr Alexandra Tsekeri, who has recently completed her doctorate at the Colombia University of New York. In particular, Dr Tsekeri has facilitated the fellow with aerosol typing criteria, access to CALYPSO satellite aerosol databases (and locally-developed tools and products) in order to better assess focused data to be used by the project, particularly with regard to pilot studies. Her experience of handling measurement errors has helped arm the fellow with the knowledge base needed for developing precise neural network results from satellite data whose uncertainties can be quantified. The SIC has also taken steps to ensure that the algorithms, computational tools and data to be developed and acquired during this project will be made available at the project website/portal and incorporated into the Greek GEO Office data centre. Technologies developed by AEROMAP will also be included in the National Bank of Environmental Observation which is coordinated by NOA. In particular, the Greek GEO Office which since 2008, has initiated an inventory of meteorological and atmospheric data coming from the Public Sector, Research Institutes and Universities, ensures that its national portal harmonizes and shares data among producers, providers and policy-makers in the field of environmental observation. This database will be enhanced with the climatological data collected during the operation of AEROMAP which is expected to assist the identification of gaps between available and required Earth observations. AEROMAP's databases will be linked to existing Earth Observation portals and online resources and the portal will enable users to access and download data, products, algorithms and tools. The invitation of the fellow to be part of NOA and to lead the project is bringing international expertise and interdisciplinary research to the host institute and helped raise the profile of the fellow. Finally, the host institute's ground-based Atmospheric Remote Sensing Station (ARSS) is the federated site of AERONET that monitors aerosol pollution over the city of Athens. As a result, the fellow as a member of the ACP Group is in direct contact with scientists that are expert in aerosol measurement, online data provision and database management. This skill base will prove to be invaluable as the project progresses into Phase D and the development of the real-time monitor. The fellow has also been provided with access to UV radiative-transfer codes and to the European UV database should the need arise and time permit the assessment of atmospheric radiative-budgets. In summary, the general scientific arrangements required for the success of this project have been comprehensively provided to the fellow including: a) access to satellite aerosol databases, b) access to ground-based aerosol retrieval algorithms and techniques, c) access to a high-tech AERONET station, d) access to PCs, computing power, resources and software for the coding and running of neural networks, and image processing and web development software, and e) access to expertise in the field via direct interaction with or via the collaborative networks of APC group members. The wealth of experience that the host institute has in accommodating visiting scientists and implementing projects of this scale, means that AEROMAP has progressed smoothly and is unlikely to face never-before encountered administrative issues. The regular personal mentoring by the SIC and regular group meetings have enabled the fellow to flourish.

***TASK: To help the fellow acquire competencies in the field of environmental modelling***

The expert training provided to the fellow by the SIC during frequent (weekly) project meetings has meant that a steep learning curve was achieved quickly in this rapidly developing research area. The fellow now has a high level the technical background knowledge in the field of global aerosol characterization and measurement. This is evidenced by the three papers in preparation that have arisen as by-products of the fellow's initial research into site variability, aerosol type and the mathematical details of the aerosol size distribution. The very positive climate of collaboration engendered at the host institute has allowed the fellow to be well integrated in the ACP Group. An early outcome of this has been the introduction of the fellow to the wider geophysical, atmospheric chemistry and physics community. This directly led to the fellow being invited to participate in the peer review of research articles in key journals in these fields. Already, the fellow is acquiring strong competencies that will establish him as an independent scientist in the field of atmospheric remote sensing from satellites and aerosol related research. In particular, after just 12 months, he is about to submit 4 research articles which will substantially enhance his publication record. The early achievements of the project are testament to the successful collaboration between the fellow and the SIC and the transfer of knowledge of between the fields of data mining and aerosol science. In addition, the application of the methodology developed in objectives 1-3 and the wealth of new findings expected to be delivered as a result of implementing the real-time monitor, will give added weight to the presence of the fellow in a newly opened field and stimulate future cooperation with the host institute and other foreign research groups in the field of environmental research. Finally, the scientific experience of the SIC and the host institute will assure the effective exploitation of the results as the project progresses into its second year.

***TASK: To contribute to the career development and re-establishment of the fellow***

One of the principal goals of the fellowship in terms of the medium to long-term development of the fellow's career is to enable the fellow to achieve the level of principal investigator as a medium term goal, and to help him secure a permanent research or academic position in the field of environmental modelling in atmospheric physics as a long-term goal. The project funding from a Marie Curie IEF has provided the mobility that enabled the fellow to travel to an expert host where he has been able to interact and collaborate with world-leading researchers in the field. As a result, he is obtaining competence in the fields of aerosol science and measurement techniques and environmental modelling in atmospheric physics. As such, the fellowship is supporting his transition from an applied mathematician applying advanced methods to several fields, to an experienced scientist in one field. The fellowship, in addition to helping focus the fellow's skills in one field, is also supporting the fellow's transition to research independence. The fellow has already taken a leading role in terms of the preparation of the proposal work plan and performing this work under the training supervision of the SIC. This has led to significant new results and research publications which will strengthen the fellow's CV. Furthermore, the fellow is developing the skills required to manage a project of this scale and importance which will be valuable assets in the future when he applies for more senior scientific positions. Working at the host institute has exposed the fellow to the working of satellite-based instruments, to global monitoring techniques and ground-based atmospheric information analysis. His involvement with the latest methodologies used to obtain aerosol and solar irradiance information retrievals from MODIS and AERONET has provided him with in-depth knowledge of contemporary measurement techniques, retrieval methodologies, uncertainties and limitations of atmospheric remote sensing from space and ground stations. Furthermore, the machine learning techniques he has developed and is starting to apply are supporting the much needed transition from local to global scale modelling and characterization of aerosols. The successful completion of the project will enrich the fellow's profile and help him establish his professional career.

***TASK: To contribute to European excellence and competitiveness***

AEROMAP is contributing to one of the most important climate change parameters - the global characterization of aerosols. As such, it will significantly raise the profile of the ERA for researchers in this field. The published results will contribute to European excellence and competitiveness by demonstrating that European centres like the host institute NOA are pioneering new methods to exploit satellite remote sensing to characterize aerosols globally. Furthermore, the use of cutting edge mathematical/computational methods like neural networks will help to create a pole of attraction to scientists from outside the ERA who will be interested in learning such techniques so as to apply them in their own fields of research. The cross-fertilisation of knowledge that this engenders will raise the status of European science as originator of such methodologies and will hopefully attract more talent to the ERA. As the project progresses towards the production of global maps and the implementation of the real-time monitor and alerting service, the project website will mature into a showcase for the project, the fellow, the host institute and Marie Curie actions in general. It is hoped that the project website will also provide a unique data portal to global aerosol characteristics provided nowhere else. Research findings facilitated by AEROMAP may not only help reduce the uncertainty in the effect of aerosols on global climate change but it is likely that this can have a significant impact on environmental policy-making decisions may have worldwide repercussions that will further turn eyes towards the ERA. Until now, it has not been possible to exploit the full-Earth coverage provided by satellite remote sensing in order to globally-characterise aerosols via their microphysical properties. As a result Europe has followed the example of other continents in building instruments and investing in ground stations federated to AERONET. The methodology adopted to extrapolate from local AERONET sites to non-existent site locations is now starting to yield accurate deductions of aerosol optical and microphysical parameters on the daily timescale. While it is clear that the best way to validate neural network-derived retrievals has been with co-located AERONET data, one direct long-term structuring effect of AEROMAP is that its products will allow for determination of the best sites to locate such stations. In this regard, the development by the fellow of the Coefficient of Variability [paper in preparation] is highly relevant. Furthermore, small scale but high impact projects like AEROMAP will help to ensure that Europe can meet the challenges of the 21st century, where the quality of the environment affects everyday life and the impact of aerosols on climate change, air quality and business (e.g. aviation) is high in the public conscience. Europe is strongly affected by inflows of dust from the Sahara and the Arabian Peninsula, biomass burning from wildfires like those in Moldova, Northern Greece, Spain and the coastline region of Portugal. This provided the motivation for the choice of pilot studies by the fellow in his paper on multiple-mode aerosol identification via the size distribution.

### 2.3.2 Problems that have occurred and how they were solved or solutions envisioned

It is important to say that this project is very ambitious. Early results at the end of Phase A revealed that the proposed use of only 3 MODIS satellite AOD measurements to obtain optical and microphysical parameters was only capable of retrieving mean values and was not going to be sufficient for the goal of real-time monitoring. However, the change of coarse to include two further inputs which are available to satellites – the columnar water from MODIS and the absorption AOD from the OMI satellite was well founded and brought dividends. Networks were trained for three principal aerosol types (dust, biomass burning and urban SO2) with 5 rather than 3 inputs, and were shown in Phase B to be able to retrieve the required aerosol parameters to a satisfactory level of accuracy at the daily timescale required for a real-time monitor. More careful work will still need to be done to be able to model ocean aerosol and to devise a method to combine networks trained for individual aerosol types into composite models of aerosol mixtures. This is of course, work in progress, and the fellow strongly believes that the expertise of the SIC, the host institute and its strong collaboration network will lead to solutions to these challenges in the months to come.

### 2.3.3 List of key project meetings, dates and venues

- Euro-Science Open Forum 2012, Dublin, Ireland: 11-15/07/2012.
- European Aerosol Conference 2012, Granada Spain: 2-7/09/2012.
- APC Group Seminar: 26/09/2012.
- *Researchers Night 2012* at the National Hellenic Research Foundation, Athens: 28/09/2012.

- APC Group Seminar: 01/04/2013.

### 2.3.4 Project planning and status

A project meeting is held once a week between the research fellow and the scientist in charge. These meetings have proven invaluable in that:

1. they allowed communication to the SIC from the fellow of the progress of the work
2. they allowed the fellow to learn from the SIC who highlights the significance of new publications in the field of atmospheric physics relevant to the project
3. they functioned as efficient brainstorming sessions for new ideas to advance the work and to overcome technical problems
4. they allowed the SIC to help the fellow in managing the workload efficiently
5. they allowed the SIC to assist the fellow with administrative paperwork related to routine administrative matters such as tax declarations.

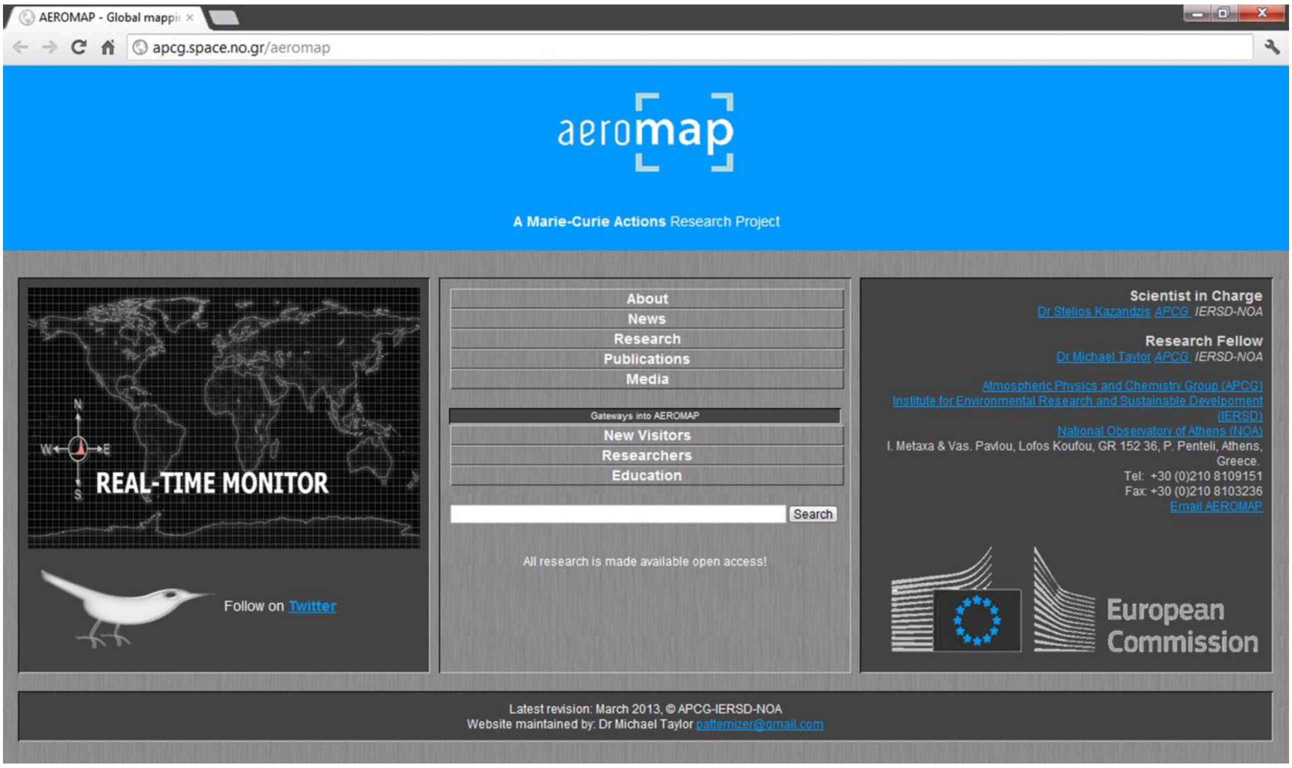
The project is progressing smoothly and it is envisaged that all remaining project deliverables will be accomplished in good time.

### 2.3.5 Impact of possible deviations from the planned milestones and deliverables

It is envisaged that all of the stated project deliverables will be met. Indeed, as this report hopes to show, bi-products have been created as a result of novel problem-solving strategies and from new collaborations the fellow is building with the environmental modelling community.

### 2.3.6 Development of the Project website

As a result of completing Task E1, an extensive project website has been designed and produced and is housed at the host institute's website: http://apcg.space.noa.gr/aeromap. The home page is shown below:




The website is structured around 5 core pages:

1. The *About* page contains a brief abstract outlining the project and lists the project's objectives, expected results, phases and milestones.
2. The *News* page presents a chronological list of events such as conference attendances of the fellow and participation in outreach activities, and highlights important landmarks such as the completion of project phases and research studies. Readers can be kept up to date of changes by following @_AEROMAP on Twitter.
3. The *Research* page lists the lines of investigation being undertaken and the research methods being implemented by AEROMAP.
4. The *Publications* page listed all open-access peer-review publications resulting from the project and presents details of materials resulting from conference participation such as posters, abstracts and proceedings. This page will also host two project bulletins relating to the mid-term report and the final report.
5. The *Media* page is reserved for press coverage of the project and will be used to house multimedia to include: i) an image gallery of graphical abstracts and published images, ii) Power-point presentations and talks and iii) posters produced by AEROMAP.

In addition, 3 gateways allow routes for users of differing levels of experience and interest:

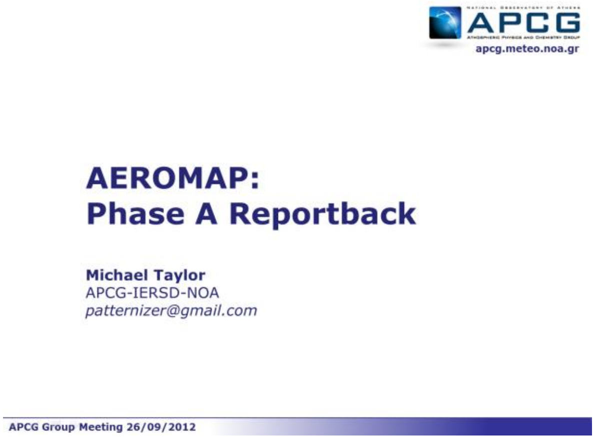
A. The ***New Visitors Gateway*** links to the *About* and *Media* pages and includes a pictorial guide to the project to help first-time users quickly understand what the project is about.
B. The ***Researchers' Gateway*** links to the *About*, *Research* and *Publications* pages for those with technical understanding of atmospheric aerosols to quickly find up-to-date results from the project. In order to provide a vehicle for attracting new researchers in the field to AEROMAP, the gateway also houses a careful chosen and comprehensive links to data portals for the aerosol community. It is envisioned that researchers will visit this page regularly due to its functionality as a portal to data portals. The gateway also houses links to key publications in aerosol science and the empirical techniques used by AEROMAP. Finally, MATLAB code produced by the fellow to assist researchers to load and parse large data files, is made available at this gateway open source.
C. Last but not least, the ***Education Gateway*** presents links to beautiful animated videos of monthly global maps of aerosol optical depth, size and the distribution of fires, chlorophyll, clouds and carbon monoxide. The gateway also links to important reports like those of the International Panel on Climate Change and open-access, online feature articles in magazines like Science and Nature so that readers can get an approachable but accurate overview of the subject. Finally, the gateway links to Open Course Ware online lecture courses on Earth, atmospheric and planetary science and series of graduate lectures on remote sensing, clouds, aerosols and climate.

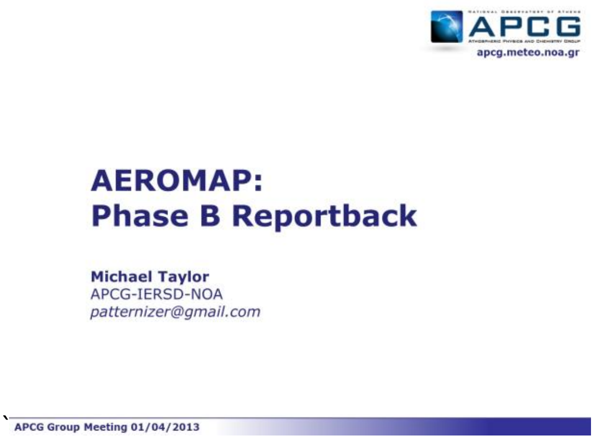
The *Home* page of the project, in addition to providing contact details, also allows experienced users to go directly to the real-time monitor by clicking on the icon once it is ready for implementation. With regard to maintenance of the website, following each weekly project meeting with the SIC, the fellow will be responsible for making changes to the content and for updating the website.

| ***Deliverable* 5**: The project website/portal: http://apcg.space.noa.gr/aeromap |
| --- |

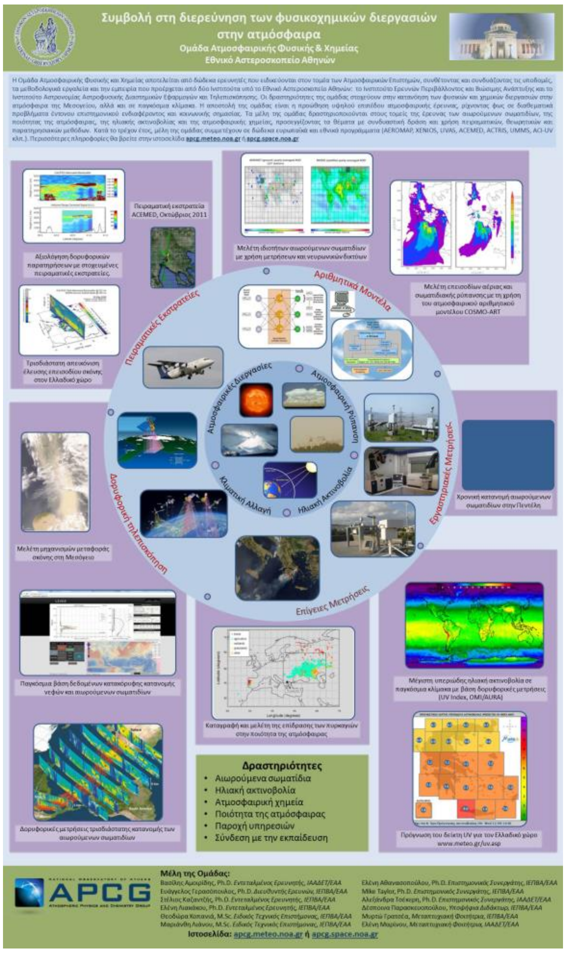
**2.3.7 Training activities**

One of the main training objectives of the project is to aid the transition of the fellow as an applied mathematics to principal investigator level in the field of environmental modelling so that he will have acquired the competences to ensure tenure in this area. The fellow has designed, managed and performed the research in tasks A1-A3, B1-B2 and E1 competently and in good time. The results have led to satisfaction of the project objectives for these phases and have produced bi-products. The fellow has been exposed to a broad array of contemporary methods, techniques and analyses in the field of aerosol science while at the host institute including: i) adaptation to the technical basics of ground-based aerosol instrumentation, measurement, retrieval and databases, ii) adaptation to satellite remote sensing retrieval techniques and databases, iii) the development of algorithms for advanced analysis and have substantially increased the fellow's knowledge of climate change and aerosol science. New collaborations in the field of environmental modelling have been established and have led to co-authorship on an important new article in the Journal of Atmospheric Measurement Techniques. Through presentation of his results at two international conferences, the fellow has helped to introduce neural networks to a community with limited experience in the field. This transfer of knowledge will continue with publication of an article being submitted that outlines the methodology applied in tasks A and B. Furthermore, the fellow participated in bi-monthly meetings of the APC Group at IERSD-NOA. In particular, the he has given two Power-point presentations providing report-backs on the results of Phase A and Phase B and assessed the progress of the work performed in each Phase. These report-backs mark important milestones in the progression of the project and have helped ensure its smooth running and timeliness. The Power-point presentations are available at high resolution on the *Media* page of the project website.




In addition, the fellow's work was included in the poster portraying the activities of the APC Group at the 2012 FP7-funded Researcher's Night (28/09/2012) at the National Hellenic Research Foundation where he played a pivotal role liasing with the general public. The posters below are available at high resolution on the *Media* page of the project website.


The high international standing of members of the ACP Group has provided the fellow with the opportunity to become a reviewer and to perform formal peer-reviews of research articles for two key journals in the field: the *Journal of Geophysical Research* and the *Journal of Atmospheric Chemistry and Physics*. This opportunity arose as a direct consequence of the great deal of in-house expertise of the ACP group members who are internationally recognized through their editorships on scientific journals, their invited talks at international conferences, invited review papers in top journals, high citation rates, and through their coordination of EU-funded research projects. The fellow, by working at the host institute, is in contact with broad scientific expertise in the fields of aerosol science, climate change, satellite remote sensing and ground-based instrumentation, measurement and data analysis, and this has already brought dividends in the co-authoring by the fellow of an article on aerosol microphysics retrieval with colleagues from the Physics Instrumentation Centre in Moscow and the World Radiation Centre at Davos (see below).

**2.3.8 Co-operation with other projects/programmes**

The SIC has included the fellow as a researcher in 3 submitted proposals for research projects following the completion of AEROMAP in March 2014. The fellow played an active role in the writing and editing of these proposals. In addition, techniques developed during the course of the first 12 months of the project meant that the fellow could contribute results to an important article on aerosol microphysical retrievals from the World Meteorological Organisation's precision filter radiometer. The article is the result of international collaboration including colleagues from the Physics Instrumentation Centre in Moscow and the World Radiation Centre at Davos:

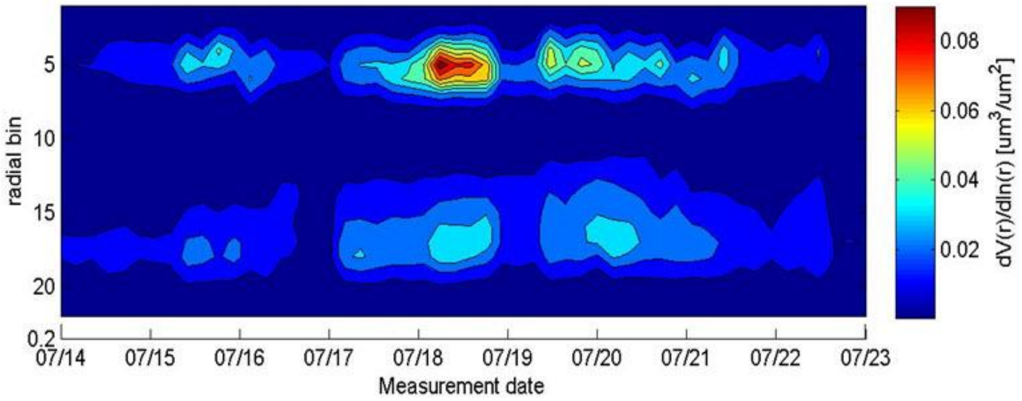



Figure 4: Aerosol size distribution measurements for the experiment period.

S. Kazadzis, I. Veselovskii, V. Amiridis, J. Gröbner, A. Suvorina, S. Nyeki, N. Kouremeti, M. Taylor, A. Tsekeri (2013) *Aerosol microphysical retrievals from WMO's Precision Filter Radiometer (PFR) direct solar radiation measurements*. **J Atmos. Meas. Tech.** (paper in preparation).

It is important to mention that two Greek experts in remote sensing: Dr Alexandra Tsekeri and Dr Antonios Gkikas have made a significant contribution to some of the project's deliverables and have been invited by the SIC to be co-authors on two of the project's publications. In particular, while initial testing of the neural networks were performed by the fellow using standard MODIS satellite data obtained from the Giovanni-MAPSS server, new and more precise methods for sampling data pixels developed by Dr Antonios Gkikas have helped improve the performance of the neural networks. In addition, Dr Alexandra Tsekeri who is an expert on airborne remote sensing instruments and aerosol retrieval has greatly informed the fellow of the limitations of various measurement approaches and their inherent uncertainties.

# 3 Deliverables

## 3.1 Deliverables

The deliverables due in this reporting period include:

***D1***: Identification of the functional relationship between AERONET direct sun AOD data and AERONET retrieval data

***D2***: Verification and uncertainty analysis of the ability of neural networks to extrapolate retrievals to regions of similar aerosol type

***D3***: Identification of the mathematical correlation mapping between MODIS satellite AOD data and AERONET direct sun AOD data

***D4***: Verification and uncertainty analysis of the ability of neural networks to extrapolate retrievals to regions where no sites exist,

***D5***: Creation of a project website/portal

## 3.2 Bi-products of the research performed

In addition, to the main core novelty of AEROMAP – the development of neural networks to invert satellite data and retrieve aerosol optical and microphysical properties, the fellow has also developed other important techniques in the course of completing Phases A and B:

1. A coefficient of variability for ranking global AERONET sites corresponding to different types of aerosol has been developed. This will allow researchers to be able to objectively and easily select geographical locations having high or low variability in terms of different types of aerosol such as desert dust, biomass burning products and urban SO2 pollution [paper in preparation]
2. An algorithm for identifying and characterising multiple-modes in the size distribution has been developed and tested with success on pilot studies such as a severe desert dust storm, the Beijing urban brown cloud, the outbreak of fires in the Africa Savannah and the eruption of volcanic ash from Iceland. The algorithm lifts the constraint of current research to studies of only two aerosol modes (fine particles and coarse particles) and allows researchers to monitor the temporal changes in the size distribution that result from the influx of aerosol during extreme conditions [paper prepared for submission]
3. A method for performing uncertainty analysis based on measuring averages at different timescales. This technique is central to understanding the performance of the neural networks and is given prominence in the core paper describing the methodology used by AEROMAP [prepared for submission]

Publications describing these new techniques are direct by-products of the work being performed to achieve the project objectives and are potential additions to the list of deliverables to be reported on in the final report.