Manuscript: Statistical exploration of gaseous elemental mercury (GEM) measured at Cape Point from 2007 to 2011 (Ref. No.: acp-2014-971), Atmos. Chem. Phys. Discuss., 15, 4025–4053, 2015, doi:10.5194/acpd-15-4025-2015.

Anonymous Referee #1

Received and published: 05 March 2015

We thank Anonymous Referee #1 for the positive review of this paper and recommending that the paper can be published in its current form.

Anonymous Referee #2

Received and published: 04 May 2015

We thank Anonymous Referee #2 for a detailed review of our manuscript. The comments and remarks have been processed and explained in the manuscript, which we believe has gained in clarity and scientific soundness. Below is a point-by-point reply (in blue) to the comments of Referee #2 (in black font).

Venter and co-authors present a statistical exploration of a 4-year record (2007-2011) of gaseous elemental mercury (GEM) concentrations measured at Cape Point, South Africa. Firstly, this paper aims at identifying the origin of high and low mercury concentrations events using a dataset already presented and discussed elsewhere (e.g., Slemr et al., 2013; Slemr et al., 2015) and based on back-trajectories and cluster analysis. Secondly, multi-linear regression analysis was used to predict GEM concentrations from other atmospheric parameters measured at the station. The paper is clearly organized, easy to follow and well written but overall lacks robust statistics. I recommend major revisions.

Major comments

1.Cluster analysis

I generally agree with the comments of B. Denzler and will avoid duplication. Briefly, I would focus the analysis on the extreme data points. Indeed using only two clusters, most of the points lie very close to the 0.904 ng/m3 threshold and results from the two clusters are very close (Fig.4).

Since Referee #2 and B Denzler had similar comments relating to the cluster analysis we addressed these comments together.

We agree with Referee #2 and B Denzler that a two cluster analysis results in points lying very close to the 0.904 ng/m^3 threshold. However, in this study the aim was for statistical analysis of the data to lead to certain deductions/conclusions. Therefore, it was not decided beforehand that a two grouping solution was going to be used. Interpretation of the cluster analysis resulted in identifying two clusters as the best solution with the optimum separation and representation of the data. Although a 5, 6 and 7 cluster solution was obtained with slightly better separation, some of these clusters represented only a small fraction of the total GEM distribution at Cape Point. Clusters representing smaller fractions of the GEM data are considered to be extreme cases that can be evaluated as separate case studies. Studies of this nature can be an important future perspective, which was indicated in the Conclusions as follows:

"Data indicated as extreme events, as indicated by 5, 6 and 7 cluster solutions should be investigated as special case studies."

Furthermore, the 0.904 ng/m^3 threshold is close to, but not the same as the mean and median of the data, which are 0.917 and 0.925 ng/m^3, respectively. This threshold value was statistically determined with cluster analysis and is specific for this dataset. If the dataset was only separated according to mean/median value, extreme data points would have been included in calculations. After cluster analysis data not included in the clusters (that can include extreme data points) was not considered in further processing.

We agree with B Denzler that the use of quartiles would have been a feasible alternative. However, this approach would have required a pre-classification of the data into four groups and using two of these (highest and lowest) in further processing. As mentioned, in this study the aim was for statistical analysis of the data to lead to certain deductions/conclusions and not pre-classification. In our approach cluster analysis was performed that resulted in a 2 cluster optimum solution. Cluttering was dealt with by subtracting the back trajectory analysis of each of the two clusters identified, which resulted in more detailed plots from which two distinct observations could be made, i.e. GEM values associated with ship routes and air masses of continental origin.

2.Trend sign at Cape Point

The authors report a decline in GEM concentrations at Cape Point over the 2007-2011 period. In contrast, Slemr et al. (2015) reported a change in the trend sign at the same station from decreasing mercury concentrations in 1996-2004 to increasing concentrations over the 2007-2013 period. How do the authors explain these contrary conclusions?

The trends identified by Slemr et al. (2015) and in this paper cannot be directly compared. Slemr et al. (2015) utilised pre-processed, i.e. de-trended and de-seasonilised data, while in this study all the data was considered for statistical evaluation. In order to clarify this to the readers the following text was added:

"In contrast, Slemr et al. (2015) reported an increase in GEM concentration at CPT GAW. However, this increase was calculated by utilising pre-processed, i.e. de-trended and de-seasonilised data, which was not the case in this study. Therefore these different approaches cannot be directly compared."

3.Back-trajectories as an alternative tool to distinguish continental/marine GEM contributions

Several studies (e.g., Slemr et al. (2013) and Brunke et al. (2004)) used 222Rn measurements to determine the continental/maritime origin of air masses reaching the Cape Point station. In this paper, the authors used back-trajectories as an alternative tool in order to distinguish continental and marine GEM contributions. The hourly arriving back trajectories were divided into groups according to the time that these air masses had spent over the continent. This work needs to include a more critical discussion of results obtained by both methods and associated uncertainties. According to the authors, the errors accompanying a single trajectory are 15-30% of the trajectory distance travelled. How does it compare with 222Rn measurements uncertainties? Could 222Rn concentrations also be a tool to determine the time spent by a trajectory over the continent? Lacking any of the above, it is not clear to me what new results this paper brings to the topic.

Although ^{222}Rn is considered to be a good tracer for studies of emissions from terrestrial surfaces, according to Jacob et al. (1997), the assumption of a uniform ^{222}Rn emission rate of 1 atom cm^{-2} s^{-1} is accurate to roughly 25% globally, or by a factor of 2 regionally. Therefore the 15 – 30% error associated with back trajectory analysis is in the same range as the uncertainties associated with ^{222}Rn as tracer. Fig. 6 was augmented by adding Fig. 6(b) that indicates the ^{222}Rn concentration range associated with air masses classified by back trajectory analysis. It is evident from comparison between Fig. 6(a) (originally Fig. 6) and Fig. 6(b) that back trajectory analysis provides a more sensitive method of characterising GEM according to time that air masses spent over the continent up to 11 hours, while ^{222}Rn classification only allows separation within three hours that air masses spent over the continent. The text was modified accordingly to:

"However, the ^{222}Rn classification method only allows for the separation of the CPT GAW GEM data into relatively few classes, i.e. marine background, mixed and continentally influenced, while the back trajectory analysis methods provide a more quantified classification based on the length of time that air masses spent over the continent resulting in increased GEM concentrations. It is evident from comparison between Fig. 6(a) and Fig. 6(b) that back trajectory analysis provides a more sensitive method of characterising GEM according to time that air masses spent over the continent up to 11 hours (where GEM concentrations reached a plateau), while ^{222}Rn classification only allows separation within three hours that air masses spent over the continent.The difference in average GEM concentrations between air masses that had spent one hour or less over the continent, i.e. 0.92 ng m^{-3} and air masses that had spent more than 11 hours on the continent, i.e. 1.09 ± 0.150 ng m^{-3}, therefore provides some quantified indication of the possible continental contribution of GEM at CPT GAW. When GEM concentrations were classified according to ^{222}Rn levels, i.e. ^{222}Rn levels > 1000 mBq m^{-3} indicating continentally influenced air masses (Slemr et al., 2013), 50% of the data was greater than 0.92 ng m^{-3}. This value is somewhat lower than the average concentration value determined for air masses spending more than 11 hours over the continent, i.e. 1.09 ng m^{-3}.

According to Jacob et al. (1997), the assumption of a uniform ^{222}Rn emission rate of 1 atom cm^{-2} s^{-1} is accurate to roughly 25% globally, or by a factor of 2 regionally. Therefore the 15 – 30% error associated with back trajectory analysis is in the same range as the uncertainties associated with ^{222}Rn as tracer."

Minor comments

p. 4026

l. 8: please define SA

We have replaced "SA" with "South Africa" in the text as follows:

"...semi-arid interior of South Africa and..."

l. 17-19: "Both measured and MLR calculated data confirm a decline in GEM concentrations at CPT GAW over the period evaluated". See major comment #2.

This was addressed in major comment #2.

p.4028

l. 3-5: Angot et al. (2014) and Slemr et al. (2015) should be included as references in addition to Ebinghaus et al. (2002).

These references were included in the text:

"...for the Southern Hemisphere (Ebinghaus et al. (2002); Slemr et al., 2011; Angot et al. (2014); Slemr et al. (2015))."

Angot et al. (2014) and Slemr et al. (2015) were added to the Bibliography. Ebinghaus et al. (2002) was included in the Bibliography.

p. 4031

l. 24-28: "Eight-day back trajectories with hourly arrival times at an arrival height of 100m (....). An arrival height of 100m was chosen since the orography in HYSPLIT is not very well defined, and therefore lower arrival heights could result in increased error margins". I wonder why the authors used an arrival height of 100m given that measurements are were carried out higher, on the top of a cliff at 230m a.s.l.

The 100m arrival height mentioned here is above ground level of the location from where the back trajectory is calculated and not above sea level. This was clarified in the text:

"...at an arrival height of 100 m (above ground level) were calculated..."

p. 4033

l.21-23: "However, significant differences between these two overlay trajectory maps (: : :) are not that evident". I agree, see major comment #1.

As mentioned at comment #1 subtraction of the back trajectory analysis of each of the two clusters resulted in more detailed maps from which distinct observations could be made. This is also explained in the text:

"Therefore, a third overlay trajectory map (Fig. 4c) was drawn, which represents the difference between the two individual maps, i.e. subtracting the percentage of trajectories passing over each correlating 0.2° x 0.2° grid cell in Fig. 4b from the percentage of trajectories passing over each 0.2° x 0.2° grid cell in Fig. 4a. In Fig. 4c, positive values (red) correspond with areas over which cluster one's (> 0.904 ng m^{-3}) air masses dominated, whereas negative values (dark blue) indicate areas over which air mass movement of cluster two (< 0.904 ng m^{-3}) were dominant. From this map (Fig. 4c), two observations can be made. Firstly, oceanic regions along both the east- and west coast around CPT GAW correspond with air masses mostly related to cluster one (higher GEM values), which could potentially indicate the influence of shipping routes on GEM measured at CPT GAW. Secondly, air masses that had passed over the very sparsely populated semi-arid Karoo region, almost directly to the north of CPT GAW, were mostly associated with cluster two (lower GEM values)."

p. 4035

l. 12-13: “An evident trend is observed in Fig. 6, i.e. an increase of GEM concentrations for air masses that spent more time over the continent”. Air masses spending less than 10 hours over the continent are associated with highly variable GEM concentrations. Is the mean statistically different from one group to another? This should be tested statistically.

The graph presented in Fig. 6(a) is statistical representation of the Hg data with mean, median and quartile values, as well as 99.3% of the data coverage indicated. This is explained by the caption of the figure:

“Figure 6: (a) The statistical distribution of GEM concentrations as a function of time spent over the continent and (b) ^{222}Rn distribution as a function time air masses spent over the continent. The mean is indicated by the black stars, the median by the red line, the 25- and 50 percentile by the blue box and the whiskers indicating 99.3 % data coverage (if a normal distribution is assumed), while the black line connects the mean values to provide an indication of the trend observed”

From the whiskers of the plot it is evident that there are more lower Hg levels associated with air masses that spent shorter periods over the continent. In addition, considering the typical atmospheric Hg concentrations measured at Cape Point, there is also a significant difference between the mean/median values of air mass spending less than 10 hours over the continent.

p. 4036

l. 1-3: “The average marine background GEM concentration for the entire sampling period according to the 222Rn level classification (<350 mBq/m3 –as proposed by Slemr et al. (2013) and Brunke et al. (2004)) was 0.92 ±0.275 ng/m3.” I believe they rather used a 100-250 mBq/m3 threshold. Does it affect the calculated mean marine background GEM concentration?

We thank Reviewer #2 for pointing out this misquote. The threshold value must be 100 mBq/m3 as indicated by Brunke et al. (2004). The reference to Slemr et al . (2013) was removed from the text, since these authors did not consider marine classification. The average marine background GEM was recalculated for the 100 mBq/m3 threshold as 0.89 ± 0.106 ng/m3. The text was changed as follows:

“The average marine background GEM concentration for the entire sampling period according to the ^{222}Rn level classification (< 100 mBq m^{-3} – as proposed by Brunke et al., 2004) was 0.89 ± 0.106 ng m^{-3}, while the average GEM level for air masses that spent one hour or less over the continent (Fig. 6(a)) was 0.92 ± 0.300 ng m^{-3}.”

l. 17-20: “When GEM concentrations were classified according to 222Rn levels, i.e. 222Rn levels > 1200 mBq/m3 indicating continentally influenced air masses ((Slemr et al., 2013) and (Brunke et al., 2004)), 50% of the data was greater than 0.99 ng/m3”. Same as above, Slemr et al. (2013) used a threshold of > 1000 mBq/m3 rather than > 1200 mBq/m3. Does it affect the calculated mean GEM concentration of continentally influenced air masses?

Similar to the previous comment the threshold was changed to > 1000 mBq/m3 and Brunke et al., (2004) was removed from the text. The median continental GEM was recalculated for the 1000 mBq/m3 threshold as 0.92 ± ng/m3. The text was changed as follows:

“When GEM concentrations were classified according to ^{222}Rn levels, i.e. ^{222}Rn levels > 1000 mBq m^{-3} indicating continentally influenced air masses (Slemr et al., 2013), 50% of the data was greater than 0.92 ng m^{-3}.”

p. 4037

l. 7-9: “Minimization of the RSME was attained when the number of independent variables included in the optimum solution of the equation was increased to eight, and had a RMSE of 0.1205”. Values of RMSE are very close to each other. How do you know if the small difference is statistically significant?

We agree with Reviewer #2 that this was not explained adequately. The text was changed to clarify as follows:

“...if the optimum MLR solution contained more independent variables. The optimised RMSE was attained when the number of independent variables included in the optimum solution of the equation was increased to eight, and had an RMSE of 0.1205. The measure of optimisation was taken as at least 1% contribution to the overall reduction of RMSE. Table 2 indicates the identity of the independent parameters determined for each of the optimum MLR solutions.”

p.4039

l.14: “a slight decrease of GEM concentrations at CPT GAW over the evaluated period”. Please see major comment #2.

This was addressed in major comment #2.

p.4040

l.2-3: “such analyses could be used as an alternative tool to distinguish between continental and marine GEM contributions”. Please see major comment #3.

This was addressed in major comment #3.

Figure 1: It is hard to see anthropogenic point sources. Please consider using different colors.

We agree that it is difficult to see the different types of sources. On this scale it is also difficult to improve differentiation, since point sources spatially overlap in certain areas. This implies that the use of different markers and/or colours would not improve the legibility. However, our intention was to indicate the concentrations of major point sources in this region of southern Africa and not necessarily the different types. Therefore we have grouped the three different source types to be represented with the same marker.

Table 1: What about the eight-, nine- and ten-cluster solutions?

As explained in our response to major comment #1, a two cluster solution was considered to be the optimum separation and representation of the Cape Point Hg data. Fig. 2 was augmented to only indicate 7 clusters to avoid confusion as pointed out by Reviewer #1.

Table 2: Please define WGS.

WGS was defined in the text on p 4037 l. 27.

B. Denzler

Received and published: 24 March 2015

We thank B. Denzler for a detailed review of our manuscript. The comments and remarks have been processed and explained in the manuscript, which we believe has gained in clarity and scientific soundness. Below is a point-by-point reply (in blue) to the comments of B. Denzler (in black font).

Summary

In the article the authors analyze a gaseous elemental mercury (GEM) time series from the Cape Point Global Atmosphere Watch (CPT GAW) station ranging from 2007 until 2012. Different statistical methods and back-trajectory analysis were applied to identify the origin of high and low mercury concentrations. Furthermore, multiple linear regression (MLR) was used to predict mercury concentrations at CPT GAW from trace gas concentration and other atmospheric parameters. The regression was also used to gain insight into the relation of the parameters with mercury concentrations.

General impression

I regard the measurement series especially at this location in the southern hemisphere as highly important. Therefore, the analysis of this series is of great interest. Generally I would argue that the methods used to either identify source regions or estimating GEM concentrations are not suitable and not well enough applied to draw concise conclusions. I encourage the authors to reconsider their methods before resubmitting the manuscript. The data-set is highly interesting and worth being published.

Major comments

1. Cluster analysis

The cluster analysis was used to distinguish between high an low concentrations. Strangely only two clusters were formed. The authors justify this choice with a high silhouette number for two clusters (fig. 2) and significant amount of GEM data. But since only two groups are formed I would assume the separation at 0.904 ng/m3 amounts about to the mean or median concentration, which could serve as a separation equally good.

The problem with using only two clusters is visible when looking at the source region analysis. Here they compare all the values above 0.904 ng/m3 with the ones below. Yet most of the measurement points lie very close to this line and are certainly not containing much valuable information and still dominate the plots (fig. 4). I would argue that using quantiles on their data and comparing for example data below the first quartile (low concentrations) with data points above the third quartile (high concentrations) would result in much more detailed plots. It would focuse the analysis on the extreme data points, not on the majority of data points lying in the middle. I therefore question the cluster analysis as the adequate method in this case.

Since Referee #2 and B Denzler had similar comments relating to the cluster analysis we addressed these comments together.

We agree with Referee #2 and B Denzler that a two cluster analysis results in points lying very close to the 0.904 ng/m^3 threshold. However, in this study the aim was for statistical analysis of the data to lead to certain deductions/conclusions. Therefore, it was not decided beforehand that a two grouping solution was going to be used. Interpretation of the cluster analysis resulted in identifying two clusters as the best solution with the optimum separation and representation of the data. Although a 5, 6 and 7 cluster solution was obtained with slightly better separation, some of these clusters represented only a small fraction of the total GEM distribution at Cape Point. Clusters representing smaller fractions of the GEM data are considered to be extreme cases that can be evaluated as separate case studies. Studies of this nature can be an important future perspective, which was indicated in the Conclusions as follows:

"Data indicated as extreme events, as indicated by 5, 6 and 7 cluster solutions should be investigated as special case studies."

Furthermore, the 0.904 ng/m^3 threshold is close to, but not the same as the mean and median of the data, which are 0.917 and 0.925 ng/m^3, respectively. This threshold value was statistically determined with cluster analysis and is specific for this dataset. If the dataset was only separated according to mean/median value, extreme data points would have been included in calculations. After cluster analysis data not included in the clusters (that can include extreme data points) was not considered in further processing.

We agree with B Denzler that the use of quartiles would have been a feasible alternative. However, this approach would have required a pre-classification of the data into four groups and using two of these (highest and lowest) in further processing. As mentioned, in this study the aim was for statistical analysis of the data to lead to certain deductions/conclusions and not pre-classification. In our approach cluster analysis was performed that resulted in a 2 cluster optimum solution. Cluttering was dealt with by subtracting the back trajectory analysis of each of the two clusters identified, which resulted in more detailed plots from which two distinct observations could be made, i.e. GEM values associated with ship routes and air masses of continental origin.

2. Multiple linear regression

When looking at part 3.3. I question that multiple linear regression (MLR) has been adequately applied here. Since the root mean square error (RMSE) is always decreasing with increasing variables, the choice of eight variables for the MLR comes at random. The choice of number of variables must be made according to a criterion which penalizes an MLR with many variables (expl. Akaike

Information Criterion (AIC)). However, the relationship they obtain from the MLR can also be obtained by simply doing individual linear regression of the chosen parameters with the GEM measurements.

We agree with B Denzler that this was not explained adequately. The text was changed to clarify as follows:

"...if the optimum MLR solution contained more independent variables. The optimised RMSE was attained when the number of independent variables included in the optimum solution of the equation was increased to eight, and had an RMSE of 0.1205. The measure of optimisation was taken as at least 1% contribution to the overall reduction of RMSE. Table 2 indicates the identity of the independent parameters determined for each of the optimum MLR solutions."

Furthermore, MLR is a mathematical calculation where one dependent variable is related to a number of independent variables simultaneously, which is based on nearest Euclidian distances calculated for each data point entry. Therefore we consider MLR to be more advanced calculation compared to simple linear regression of individual parameters.

3. Conclusion

In Section 4 the authors present a summary of their work and an outlook, but the conclusions are missing. It is not clear what processes; ships or cities are responsible for GEM emissions.

We thank B Denzler for pointing this out. We have rewritten the Conclusions and also added future perspectives in which it is clarified why specific sources could not be identified in this statistical evaluation of GEM data. The Conclusion section was changed as follows:

"As far as the authors could assess, this is the first study that has evaluated continuous high resolution GEM data of CPT GAW with different statistical analysis techniques. Cluster analysis on the dataset indicated that the GEM data could be divided into two clusters, separated at atmospheric concentrations of 0.904 ng m^{3}. Trajectory analyses of the individual clusters, as well as the differences between these clusters, indicated that shipping around Cape Point could be a significant source of GEM. In contrast, low GEM concentrations originated from the southern oceanic background and terrestrial areas with very low anthropogenic activities/population density. Correlation of the time that back trajectories spent over the African continent and GEM concentration, proved that such analyses could be used as an alternative tool to distinguish between continental and marine GEM contributions.

It was also demonstrated that MLR analysis could be used to determine an equation that can be used to predict GEM at CPT GAW. Moreover, this equation provided some insight into the complex nature of GEM chemistry. Lastly, the evaluation of both continuously measured and calculated (with the determined MLR Eq. 1) GEM concentrations seem to indicate a decline in GEM concentrations over the period evaluated in this paper. It remains to be seen whether this decline continues, which would reflect a positive response to global Hg emission reductions, or if it is only part of a longer-term cycle with a temporary decline.

From this statistical study of continuous GEM measurement at Cape Point additional research questions and/or perspectives were identified. Data indicated as extreme events, as indicated by 5, 6 and 7 cluster solutions should be investigated as special case studies. Further research quantifying the contribution of shipping should be undertaken, not only for the southern African region, but also for other busy shipping routes. In addition, source apportionment should be conducted in order quantify the contribution of specific sources."

Minor comments

Some acronyms are not defined in the manuscript, or too late

We have rectified this matter.

US and British English is not used consistently.

We have read through the manuscript and could not find any inconsistencies relating to the UK English used in the text. However, if considered necessary, we would appreciate it if the reviewer could point out specific instances of inconsistent use of UK English.

Fig. 6, showing the GEM concentration against duration above the ocean in the same plot would be interesting

We thank B Denzler for this suggestion. However, marine air masses that contribute to most of the air masses at Cape Point are considered to be representative of the Southern Hemispheric background levels of Hg, which remains relatively constant over a 24 hour period.

on fig. 5 it would be interesting to enlarge the point of interest; South Africa

We agree that an enlargement of South Africa in this figure would be interesting. However, this image was obtained from the Automated Mutual-Assistance Vessel Rescue System (AMVER) website, sponsored by the United States Coast Guard. Enlargement of the South African region resulted in an image with poor quality. Since we do not have access to detailed data of shipping routes, it is unfortunately not possible for us to construct a figure with higher resolution.

a figure of the whole GEM series with lower resolution and plotted mean concentration would be interesting

Although a figure of this nature would be very interesting we believe that the two current figures (Fig. 3 and 8) representing GEM trends for the entire sampling period in the manuscript adequately present the complete GEM data series. Another figure of this type could be considered repetitive representation of the dataset.

on p.4037, l.27, WD (wind direction) is not mentioned as a parameter included on the MLR (eq. 1). However, eq. 1 does not include WD.(?)

We excluded the text:

"...WD the wind direction in degrees..."

from the manuscript, since it was not included as a parameter in the equation obtained with MLR. Furthermore, based on this comment we also noticed that UVb was also mentioned in the text, but not in the equation. This text:

"and UVb the ultraviolet radiation in minimum erythema dose (MED) units"

was also excluded.

Statistical exploration of gaseous elemental mercury (GEM) measured at Cape Point from 2007 to 2011

A.D. Venter[1], J.P. Beukes[1], P.G. van Zyl[*1], E.-G. Brunke[2], C. Labuschagne[2], F. Slemr[3], R. Ebinghaus[4] and H. Kock[4]

[1]{Unit for Environmental Sciences and Management, North-West University, Potchefstroom, South Africa}

[2]{South African Weather Service, P.O. Box 320, Stellenbosch 7599, South Africa}

[3]{Max-Planck-Institut fur Chemie, Hahn-Meitner-Weg 1, 55128 Mainz, Germany}

[4]{Helmholtz-Zentrum Geesthacht, Institute of Coastal Research, Max-Planck-Strasse, 21502 Geesthacht, Germany}

Correspondence to: P.G. van Zyl (pieter.vanzyl@nwu.ac.za)

Abstract

The authors evaluated continuous high resolution gaseous elemental mercury (GEM) data from the Cape Point Global Atmosphere Watch (CPT GAW) station with different statistical analysis techniques. GEM data was evaluated by cluster analysis and the results indicated that two clusters, separated at 0.904 ng m^{-3}, existed. The air mass history for the two-cluster solution was investigated by means of back-trajectory analysis. The air mass back-trajectory net result showed lower GEM concentrations originating from the sparsely populated semi-arid interior of South Africa ~~SA~~ and the marine environment, whereas higher GEM concentrations originated predominately along the coast of SA that most likely coincide with trade routes and industrial activities in urban areas along the coast. Considering the net result from the air mass back-trajectories, it is evident that not all low GEM concentrations are from marine origin, and similarly, not all high GEM concentrations have a terrestrial origin. Equations were developed by means of multi-linear regression (MLR) analysis that allowed for the estimation/prediction of atmospheric GEM concentrations from other atmospheric parameters measured at the CPT GAW station. These equations also provided some insight into the relation and interaction of GEM with other atmospheric parameters. Both measured

and MLR calculated data confirm a decline in GEM concentrations at CPT GAW over the period evaluated.

1 Introduction

Mercury (Hg) is a volatile metal emitted into the atmosphere, naturally or through anthropogenic activities, such as the combustion and processing of fossil fuels (Pirrone et al., 2010). It can be transported over large distances in the atmosphere due to its low reactivity and solubility. The removal of Hg from the atmosphere is facilitated mainly through wet deposition. This occurs when Hg is oxidised to less volatile and more soluble compounds (Lin et al., 2006). The deposited aqueous Hg compounds can then be converted into methylated Hg, which is a potent toxin for humans and animals. Methylated Hg bio-accumulates in the aquatic food chain and may lead to the build-up of high concentrations in predatory fish that can pose serious health risks to people and animals that depend on a fish diet (Mergler et al., 2007). These negative environmental impacts of Hg have led to an increase in atmospheric Hg research (Lindberg et al., 2007).

Current emission inventories and models indicate that anthropogenic emissions are the largest source of atmospheric Hg (2880 t yr^{-1}), followed by marine (2680 t yr^{-1}) and terrestrial (1850 t yr^{-1}) sources (Mason, 2009; Pirrone et al., 2010). Industrial coal combustion processes, which include electricity generation, petrochemical plants and gasification processes, are considered to be the major anthropogenic sources of atmospheric Hg (Laudal et al., 2000; Wagner, 2001). It is estimated that coal combustion accounts for approximately a third of anthropogenic Hg emissions in the United States of America (USA) (Laudal et al., 2000). Other main sources of anthropogenic Hg emissions include non-ferrous metal production, gold refining, cement production and other combustion sources. The US Environmental Protection Agency (US-EPA) introduced the Clean Air Mercury Rule in March 2005 enforcing the capping of mercury emissions from new and existing coal-fired power plants. The USA and European Union (EU) were among the first to regulate Hg pollution, and it is widely expected that this could significantly influence the way in which South Africa adopts Hg control legislation. In 2013, the Minamata Treaty was signed by South Africa and 98 other countries to protect human health and the environment from anthropogenic emissions and releases of elemental Hg and Hg compounds. It is expected that the Minamata Treaty will have far-reaching implications for South Africa, since it is globally considered to be the 6th largest emitter of mercury, emitting ~50 t yr^{-1} due to the reliance on fossil fuels (Scott & Mdluli, 2012).

The global uncertainty associated with anthropogenic Hg emissions is considered to be ± 30%, while the uncertainties associated with emissions from oceans and terrestrial surfaces are ± 50% (Lin et al., 2006; Lindberg et al., 2007). Long-term monitoring of Hg is important to reduce these uncertainties associated with Hg emissions from different sources, as well as to provide important information relating to the oxidation mechanism of atmospheric Hg (Slemr et al., 2008, Slemr et al., 2013). Although atmospheric Hg is monitored extensively in the Northern Hemisphere, few studies have been published in peer-reviewed literature for the Southern Hemisphere (Ebinghaus et al. (2002); Slemr et al., 2011; Angot et al. (2014); Slemr et al. (2015))~~(Slemr et al., 2011)~~. The German Antarctic research station measured total gaseous Hg (TGM) from January 2000 to January 2001 (Ebinghaus et al., 2002). Gaseous elemental Hg (GEM) measurements have been conducted at the Cape Point Global Atmosphere Watch (CPT GAW) atmospheric monitoring station in South Africa since 1995 and have yielded several publications on long-term trends, depletion events, seasonal cycles and flux rates (Baker et al., 2002; Slemr et al., 2008; Brunke et al., 2010). From 1995 until 2004, approximately 200 three-hour GEM samples have been collected each year with a manual double amalgamation technique (Slemr et al., 2008) at the CPT GAW station, while continuous high-resolution Hg measurements commenced in 2007.

In this paper, a combination of different statistical techniques was applied to continuous high-resolution Hg data collected between March 2007 and December 2011, as well as back-trajectory analyses that were performed in order to provide greater insight into the source regions of GEM at the CPT GAW station. This approach is different from previous studies of GEM measured at the CPT GAW station, where ^{222}Rn measurements were used to determine the origin of air masses, i.e. maritime or continental (Brunke et al., 2004; Slemr et al., 2013). The relationship between GEM and other atmospheric parameters measured at the CPT GAW station was also determined statistically in order to establish whether GEM levels could be estimated or predicted from these parameters.

2 Experimental

2.1 Site description

The CPT GAW station (34º21' S, 18º29' E) is located in a nature reserve approximately 60 km south of Cape Town at the southernmost tip of the Cape Peninsula on the top of a coastal cliff at 230m above sea level. Its location within a regional context is indicated in Fig. 1. The site experiences moderate temperatures with dry summers during which occasional biomass burning events in the surrounding area occur and increased precipitation during the winter (Slemr et al., 2013). Fig. 1 is a composite map indicating population density and major point sources (Lourens et al., 2011; 2012; Venter et al., 2012) in South Africa. Population density data was obtained from the Socioeconomic Data and Applications Center (SEDAC) – a data centre in the Earth Observing System Data and Information System (EOSDIS) of NASA. As is evident from Fig. 1, the industrial hub of South Africa that is concentrated around the Johannesburg-Pretoria megacity conurbation (in the Gauteng Province, GP) and a relatively high density of large point sources (e.g. coal-fired power plant, petrochemical operations, metallurgical smelters) located in parts of the Mpumalanga, North West and Free State provinces that border on GP, are somewhat remote from the CPT GAW station. The nearest large anthropogenic point or area sources that could impact the CPT GAW station directly include the Cape Town metropolitan area, industries associated with the conurbation and an iron smelter on the west coast at Saldanha.

Insert Fig. 1

2.2 Sampling

Continuous Hg measurements were conducted with a Tekran 2537A vapour-phase mercury analyzer (Tekran Inc., Toronto, Canada) with a 15-min time resolution. A description of the instrument, as well as its calibration and maintenance were presented by Slemr et al. (2008 and 2013). The analyser is capable of measuring low Hg concentrations typically measured at background locations (Ebinghaus et al., 1999), with a TGM detection limit of ~0.05 ng m^{-3}. Under the prevailing atmospheric conditions at the CPT GAW station, and due to the presence of hygroscopic sea salt aerosols, it can be assumed that the reactive gaseous mercury (RGM) will be adsorbed by the inlet tubing and the aerosol filter allowing for the exclusive measurement of atmospheric GEM (Brunke et al., 2010). The 15-min GEM data have been converted to 30-min averages in order to correlate the results with other trace gas and meteorological measurements at CPT GAW. As described by Brunke et al. (2004), the ^{222}Rn detector was designed by the Australian Nuclear Scientific & Technology Organisation (ANSTO) and was partially constructed locally. Trace gas measurements are drawn from the top of a 30 m high mast located on the instrument deck of the laboratory. Several trace gases (carbon monoxide, CO; carbon dioxide, CO_2; methane, CH_4; ozone, O_3; nitrogen dioxide, NO_2), and halocarbons are measured at CPT GAW on a continuous basis. Of these, CO, CO_2, CH_4 and O_3 are used in this paper. Detailed instrumental descriptions of these instruments and meteorological parameters measured can be found in previous publications (Brunke et al., 1990, 2004, 2010).

2.3 Statistical analysis

In contrast to previously published atmospheric measurements of the CPT GAW station, the dataset used in this paper was not de-trended (Brunke et al., 2004). The complete dataset (March 2007 to December 2011) was evaluated statistically in order to determine the influence and interaction of GEM and other atmospheric species on a temporal and spatial basis. The statistical techniques applied are subsequently discussed.

During cluster analysis, GEM data was processed in MATLAB by making use of the k-means function from the clustering toolbox. K-means clustering is an iterative partitioning method that uses squared Euclidean distances to partition a data matrix into k clusters. The centroid of each cluster is the mean of the points in that cluster. K-means uses a two-phase iterative algorithm to minimise the sum of point-to-centroid distances, summed over all k clusters. All GEM data considered, i.e. 41 499 measurements, were subjected to five consecutive iterations using cluster analysis. Silhouette numbers obtained from the clustering were further considered to quantify the separation of the clusters. Silhouette numbers are a measure of how close individual points in a specific cluster are to points in the neighbouring clusters. The silhouette numbers range from +1, indicating points that are very distant from neighbouring clusters, through 0, indicating points that are not distinctly in one cluster or another, to -1, indicating points that are probably assigned to the wrong cluster.

Multi-linear regression (MLR) analysis models the relationship between two or more explanatory independent variables and a dependant variable by fitting a linear equation to the observed data. The MLR equation obtained can be utilised to calculate values for the dependent variable. In this study, GEM values were considered the dependent variable, while 27 other ancillary data parameters (such as gaseous species concentrations and meteorological data) were considered the explanatory independent variables. MLR was used to determine the optimum combination of independent variables to derive an equation that could be used to predict GEM concentrations, while root mean square error (RMSE) was used to compare the calculated values with the measured values.

2.4 Back-trajectory analysis

The origin of air masses arriving at CPT GAW was determined by making use of NCEP (National Centers for Environmental Prediction) GDAS (Global Data Assimilation System) data obtained from NOAA ARL (Air Resources Laboratory of National Oceanic and Atmospheric Administration, ftp://arlftp.arlhq.noaa.gov/pub/archives/gdas1/). Individual eight-day back trajectories with hourly arrival times at an arrival height of 100 m (above ground level) were calculated with the Hybrid Single Particle Lagrangian Integrated Trajectory Model (HYSPLIT) 4.820. An arrival height of 100 m was chosen since the orography in HYSPLIT is not very well defined, and therefore lower arrival heights could result in increased error margins on individual trajectory calculations. Considering the above, 24 back trajectories for each day were obtained for the entire sampling period. Individual back trajectories generated in HYSPLIT (24/day x 365 days x ~5 years) were superimposed and further analysed in MATLAB to form overlay back trajectory maps. In these overlay back trajectory maps, a colour code indicates the percentage of trajectories passing over 0.2° x 0.2° grid cells, with red being the highest percentage. Therefore, such images indicate the main areas over which air masses arriving at CPT GAW had passed. Similar overlay back trajectory analyses have previously been used in other non-GEM related atmospheric studies, e.g. Venter et al. (2012), Lourens et al. (2011) and Vakkari et al. (2013).

3 Results

3.1 Cluster analysis

Cluster analysis of GEM data was performed to statistically determine the optimal number of clusters that the data could be divided into. The cluster analysis results were assessed by calculating the average silhouette numbers of each cluster solution. Figure 2 shows the average silhouette numbers for the various cluster solutions. According to these results, dividing the GEM data into two clusters yielded an average silhouette number of 0.638, while three or four cluster divisions yielded lower average silhouette numbers. The division of the GEM data into five to ten clusters yielded higher average silhouette numbers, which indicate better separation between clusters. In Table 1, the percentage GEM data distribution (columns) of each cluster solution (rows) is presented, and indicates that cluster solutions containing more than four clusters did not contain statistically significant amounts of GEM data in each cluster, e.g. the five-, six- and seven-cluster solutions all had clusters with ≤ 0.05% of the data. Therefore, the two-cluster solution was chosen as the optimal representation for this study. The GEM data distribution in the two clusters for the entire sampling period is presented in Fig. 3. According to the clustering approach applied, separation between the two clusters was at a GEM concentration of 0.904 ng m^{-3}. Fig. 3 also indicates a small number of data points with GEM concentrations > ~2 ng m^{-3} that could be investigated as possible plume events, or special case studies. However, these possible higher concentration case studies were not considered in this paper, but may be an important topic to consider in future studies.

Insert Fig. 2

Insert Table 1

Insert Fig. 3

3.2 Source region analysis

In order to identify possible sources and/or source regions, the two clusters identified were further investigated by means of associating the arrival times of calculated back trajectories with the GEM data in the two different clusters. This was performed by correlating the GEM concentrations measured during the half hour before and the half hour after the hourly arrival time of the calculated back trajectory. Therefore, all the calculated back trajectories could be divided as being associated with one of the two GEM data clusters. The normalised (to 100 percent) overlay back trajectory analysis map associated with GEM data cluster one, i.e. GEM concentration > 0.904 ng m^{-3}, is presented in Fig. 4a, while the overlay back trajectory analysis associated with cluster two, i.e. GEM concentration < 0.904 ng m^{-3}, is shown in Fig. 4b. From both these overlay back trajectory maps, it is evident that CPT GAW is mainly influenced by air masses from the south west that had passed mainly over the marine background. It also seems as if the air masses associated with GEM concentrations

< 0.904 ng m^{-3} (cluster two, Fig. 4b) have a slightly larger fetch region over the ocean than the air masses associated with GEM concentrations > 0.904 ng m^{-3} (cluster one, Fig. 4a). However, significant differences between these two overlay trajectory maps, which could indicate possible differences in sources/source regions for the two GEM clusters, are not that evident. Therefore, a third overlay trajectory map (Fig. 4c) was drawn, which represents the difference between the two individual maps, i.e. subtracting the percentage of trajectories passing over each correlating 0.2° x 0.2° grid cell in Fig. 4b from the percentage of trajectories passing over each 0.2° x 0.2° grid cell in Fig. 4a. In Fig. 4c, positive values (red) correspond with areas over which cluster one's (> 0.904 ng m^{-3}) air masses dominated, whereas negative values (dark blue) indicate areas over which air mass movement of cluster two (< 0.904 ng m^{-3}) were dominant. From this map (Fig. 4c), two observations can be made. Firstly, oceanic regions along both the east- and west coast around CPT GAW correspond with air masses mostly related to cluster one (higher GEM values), which could potentially indicate the influence of shipping routes on GEM measured at CPT GAW. Secondly, air masses that had passed over the very sparsely populated semi-arid Karoo region, almost directly to the north of CPT GAW, were mostly associated with cluster two (lower GEM values). The afore-mentioned differences in source regions for GEM did not seem to be associated with back trajectory heights, e.g. subsiding air masses from the free troposphere. It was found that the mean height of trajectories resulting in low and high GEM concentrations were almost identical, i.e. 1 178 m and 1 104 m, respectively. A slight difference was observed for the mean air mass back trajectory maximum height, with the lower and higher GEM concentrations peaking at 2 785 m and 2 654 m, respectively. Similarly, the heights of trajectories passing over land (five hours or more) were investigated. The mean heights of continental trajectories from cluster one (high GEM) and cluster two (low GEM) were 1 490 m and 1 632 m, and the mean maximum heights of continental trajectories were 3 352 m and 3 599 m, respectively.

Insert Fig.4

The possibility of shipping routes contributing to the high GEM concentrations observed around the coast of South Africa was further investigated. From the Automated Mutual-Assistance Vessel Rescue System (AMVER) website, sponsored by the United States Coast

Guard, Fig. 5 was obtained. In Fig. 5, the daily ship locations, derived from an average of 4 634 ships per day, for July 2011 are presented in a density plot. The July 2011 timeframe was chosen since it was within the GEM data analysis period. The shipping route density plot indicated in Fig. 5 also closely correlates with NO_x shipping tracks characterised by satellite observations (Skjølsvik et al., 2000, Richter et al., 2011). The similarities of shipping route densities and NO_x shipping tracks, with the areas indicating additional contributions to higher GEM values (as indicated in Fig. 4c), provide credibility to the postulation that shipping around South Africa contributes meaningfully to GEM measured at CPT GAW.

Insert Fig.5

Slemr et al. (2013) and Brunke et al. (2004) demonstrated how ^{222}Rn could be used to identify lower GEM concentrations that were associated with air masses passing over the marine background and higher GEM levels that were associated with a continental influence. In an effort to gain further insight into the difference between continental and marine background GEM concentrations at CPT GAW, the hourly arriving back trajectories calculated for the entire sampling period were divided into groups according to the time that these air masses had spent over the African continent. The statistical distribution of GEM concentrations and the time that air masses spent over the continent are presented in Fig. 6.

Insert Fig. 6

An evident trend is observed in Fig. 6, i.e. an increase of GEM concentrations for air masses that spent more time over the continent. As expected, back trajectories of air masses spending shorter times over the continent were on average associated with the lowest GEM concentrations. However, these groups also had the largest spread in GEM concentrations (±0.6 ng m^{-3}). This can at least partially be explained by uncertainties in the back trajectory calculations. During these calculations, a single hourly position of the air mass was determined for each of its 96 hourly backward positions. This relatively weak temporal resolution implies that an air mass could, for instance, have passed over the relatively nearby Cape Town conurbation, without being registered as spending any time over the continent.

Additionally, the accuracy of trajectories depends on the quality of the underlying meteorological data used during the calculations. Considering the afore-mentioned limitations, the errors accompanying a single trajectory are currently estimated to be 15 to 30% of the trajectory distance travelled (Vakkari et al. 2013; Stohl 1998; Riddle et al. 2006). It is therefore possible that a trajectory could have spent a short period over the continent, without being calculated as such. Notwithstanding the possible errors associated with the calculation of individual back trajectories, the large number of trajectories calculated in this study (24/day x 365 days x ~5 years) ensures that the data obtained is statistically representative. The average marine background GEM concentration for the entire sampling period according to the ^{222}Rn level classification (< 100 mBq m^{-3} – as proposed by Brunke et al., 2004) was 0.89 ± 0.106 ng m^{-3}, while the average GEM level for air masses th one hour or less over the continent (Fig. 6(a)) was 0.92 ± 0.300 ng m^{-3}.~~The average marine background GEM concentration for the entire sampling period according to the ^{222}Rn level classification (< 350 mBq m^{-3} – as proposed by Slemr et al., 2013 and Brunke et al., 2004) was 0.92 ± 0.275 ng m^{-3}, while the average GEM level for air masses that spent one hour or less over the continent (Fig. 6) was 0.92 ± 0.300 ng m^{-3}.~~ This indicates that the classification of air masses arriving at the CPT GAW station with back trajectory analysis correlates well with air mass classification according to ^{222}Rn levels. However, the ^{222}Rn classification method only allows for the separation of the CPT GAW GEM data into relatively few classes, i.e. marine background, mixed and continentally influenced, while the back trajectory analysis methods provide a more quantified classification based on the length of time that air masses spent over the continent resulting in increased GEM concentrations. It is evident from comparison between Fig. 6(a) and Fig. 6(b) that back trajectory analysis provides a more sensitive method of characterising GEM according to time that air masses spent over the continent up to 11 hours (where GEM concentrations reached a plateau), while ^{222}Rn classification only allows separation within three hours that air masses spent over the continent. The difference in average GEM concentrations between air masses that had spent one hour or less over the continent, i.e. 0.92 ng m^{-3} and air masses that had spent more than 11 hours on the continent, i.e. 1.09 ± 0.150 ng m^{-3}, therefore provides some quantified indication of the possible continental contribution of GEM at CPT GAW. This value is somewhat lower than the average concentration value determined for air masses spending more than 11 hours over the continent, i.e. 1.09 ng m^{-3}.

According to Jacob et al. (1997), the assumption of a uniform ^{222}Rn emission rate of 1 atom cm^{-2} s^{-1} is accurate to roughly 25% globally, or by a factor of 2 regionally. Therefore the 15 – 30% error associated with back trajectory analysis is in the same range as the uncertainties associated with ^{222}Rn as tracer.However, the ^{222}Rn classification method only allows for the separation of the CPT GAW GEM data into relatively few classes, i.e. marine background, mixed and continentally influenced, while the back trajectory analysis methods provide a more quantified classification based on the length of time that air masses spent over the continent resulting in increased GEM concentrations.

Additionally, from Fig. 6, it is evident that the GEM concentrations reach a plateau for air masses that spent more than 10 hours over the continent. The difference in average GEM concentrations between air masses that had spent one hour or less over the continent, i.e. 0.92 ng m^{-3} and air masses that had spent more than 10 hours on the continent, i.e. 1.09 ± 0.150 ng m^{-3}, therefore provides some quantified indication of the possible continental contribution of GEM at CPT GAW. When GEM concentrations were classified according to ^{222}Rn levels, i.e. ^{222}Rn levels > 1200 mBq m^{-3} indicating continentally influenced air masses (Slemr et al., 2013; Brunke et al., 2004), 50% of the data was greater than 0.99 ng m^{-3}. This value is somewhat lower than the average concentration value determined for air masses spending more than 10 hours over the continent, i.e. 1.09 ng m^{-3}.

3.3 Relationship of GEM with other parameters

In an effort to determine relationships between atmospheric GEM concentrations and other atmospheric parameters measured at the CPT GAW station, as well as to establish whether GEM levels could be estimated or predicted from these parameters, multi-linear regression (MLR) analysis was conducted. In Fig. 7, the RMSE difference between the calculated and measured GEM values, as a function of the number of independent variables included in the optimum MLR solution, is presented. The linear equation containing only a single optimum independent variable, which was determined as absolute humidity, had an RMSE of ~0.1250, while the RMSE was lowered to ~0.1231 if an MLR equation containing the optimum combination of two independent variables, i.e. absolute humidity and O_3, was calculated. The RMSE difference between the experimental and calculated GEM values could further be reduced if the optimum MLR solution contained more independent variables. Minimisation of the RMSE was attained when the number of independent variables included in the optimum solution of the equation was increased to eight, and had an RMSE of 0.1205. The measure of optimisation was taken as at least 1% contribution to the overall reduction of RMSE. Table 2 indicates the identity of the independent parameters determined for each of the optimum MLR solutions.

Insert Fig. 7

Insert Table 2

The inclusion of more independent variables in the MLR solution did not significantly reduce the RMSE, and this can be seen in Fig. 7. A new MLR equation was determined with every addition of independent variables to determine the optimum variable combinations. This implies that new constants were calculated for all independent variables in each new optimum variable combination. Therefore, the afore-mentioned reductions in RMSE observed with an increase in the number of independent variables included in the optimum solution was an overall reduction of RMSE resulting from the increased number of independent variables included in the combination and not due to the contribution of a single (or two) independent variable(s). The identity and constants associated with the independent variables in the identified optimum MLR solution to predict GEM, i.e. the dependant variable, are provided in Eq. 1:

$$GEM = -1.2308 + 1.492\times10^{-3} AbsH - 3.790\times10^{-3} O_3 + 6.220\times10^{-4} CO + 1.3630\times10^{-3} P + 6.280\times10^{-3} T + 2.180\times10^{-4} CH_4 - 5.530\times10^{-6} Rn - 1.110\times10^{-3} WGS \quad (1)$$

with GEM in ng.m^{-3}, gaseous species (CO, CH_4, O_3 and CO_2) in ppb, T in ºC, WGS indicating wind gust speed in km/h, ~~WD the wind direction in degrees,~~ AbsH the absolute humidity in g.m^{-3} and~~,~~ P the ambient pressure in hectopascal (hPa)~~ and UVb the ultraviolet radiation in minimum erythema dose (MED) units~~. In Eq. (1), independent variables associated with positive constants indicate that an increase in these parameters would statistically lead to an increase in atmospheric GEM, whereas the increase in independent variables associated with negative constants would statistically lead to lower GEM. Although the MLR equations cannot be used to explicitly derive the origin and/or reaction mechanistic information about GEM at CPT GAW, it could be used to provide some insight.

The positive constants associated with CO and CH_4 could indicate that higher GEM can be attributed to anthropogenic emissions such as fossil fuel (e.g. shipping) and household combustion, as well as natural biomass burning observed during pollution events (Brunke et al. 2010). Higher O_3 leads to higher hydroxyl ($^{\bullet}OH$) radical concentrations, therefore the possible negative constant associated with O_3. As discussed in Lau et al. (2012) and the references therein, GEM may be oxidised by $^{\bullet}OH$, nitrate ($NO_3^{\bullet}$) or halogen ($X^{\bullet}$) radicals. Gierens et al. (2014) recently indicated that $^{\bullet}OH$ concentrations reach a peak around midday in the interior of SA, and since $^{\bullet}OH$ has a lifetime of ~1 second, its diurnal variation will therefore follow the diurnal variation of the UV radiation with wavelength capable to photolyse O_3 to O^1D, and therefore achieving peak GEM oxidation during midday. Additionally, the negative constant associated with O_3 in Eq. (2) could also indicate aged air masses, in which GEM decreased (e.g. by oxidation and deposition). The photochemically-driven oxidation of GEM results in the formation of gaseous oxidised mercury (GOM). Particulate bound mercury (PBM) and GOM typically reach diurnal minima before sunrise and maxima in the afternoon (Lau et al., 2012). It has been suggested that an abundant halogen radical ($X^{\bullet}$) concentration present in the marine environment may lead to higher GOM concentrations (Mao & Talbot 2012). The photochemically-driven oxidation of GEM to GOM in summer depletes GEM levels during midday when solar radiation, O_3 levels and atmospheric halogens produced by sea spray are the most intense and would therefore explain negative constants in Eq. (1). This signifies the complex nature of the interaction between chemical species and the physical environment in the atmosphere.

In addition to providing some insight into the origin and/or reactions of GEM at CPT GAW, it would also be useful to predict GEM at this site with the MLR Eq. (1). In Fig. 8, the measured time series for atmospheric GEM at CPT GAW are presented (blue markers) and compared to the GEM values calculated with Eq. (1) (red markers). Although slight differences are observed, the MLR equation predicts atmospheric GEM concentrations relatively well, with the exception of very high and low levels. Therefore, the MLR Eq. (1) could be used to predict GEM concentrations at CPT GAW if actual measurements thereof are not available.

Insert Fig. 8

Although not indicated Fig. 8 (to prevent cluttering of the graph), linear fitting of the actual continuous measured and calculated (with Eq. 1) GEM concentrations indicated negative slopes of -5.579×10^{-6} and -1.391×10^{-5}, respectively. This indicates a slight decrease of GEM concentrations at CPT GAW over the evaluated period. Brunke et al. (2010) previously also reported a decrease in GEM at CPT GAW from 1995 to 2009, but this reported decline only include approximately two years of continuous measurement. The decline observed for the longer time series reported here provides additional support to the observed decline. In contrast, Slemr et al. (2015) reported an increase in GEM concentration at CPT GAW. However, this increase was calculated by utilising pre-processed, i.e. de-trended and de-seasonilised data, which was not the case in this study. Therefore these different approaches cannot be directly compared.

4 Conclusions

As far as the authors could assess, this is the first study that has evaluated continuous high resolution GEM data of CPT GAW with different statistical analysis techniques. Cluster analysis on the dataset indicated that the GEM data could be divided into two clusters, separated at atmospheric concentrations of 0.904 ng m^{3}. Trajectory analyses of the individual clusters, as well as the differences between these clusters, indicated that shipping around Cape Point could be a significant source of GEM. In contrast, low GEM concentrations originated from the southern oceanic background and terrestrial areas with very low anthropogenic activities/population density. Correlation of the time that back trajectories spent over the African continent and GEM concentration, proved that such analyses could be used as an alternative tool to distinguish between continental and marine GEM contributions.

It was also demonstrated that MLR analysis could be used to determine an equation that can be used to predict GEM at CPT GAW. Moreover, this equation provided some insight into the complex nature of GEM chemistry. Lastly, the evaluation of both continuously measured and calculated (with the determined MLR Eq. 1) GEM concentrations seem to indicate a decline in GEM concentrations over the period evaluated in this paper. It remains to be seen whether this decline continues, which would reflect a positive response to global Hg emission reductions, or if it is only part of a longer-term cycle with a temporary decline.

From this statistical study of continuous GEM measurement at Cape Point additional research questions and/or perspectives were identified. Data indicated as extreme events, as indicated by 5, 6 and 7 cluster solutions should be investigated as special case studies. Further research quantifying the contribution of shipping should be undertaken, not only for the southern African region, but also for other busy shipping routes. In addition, source apportionment should be conducted in order quantify the contribution of specific sources.~~As far as the authors could assess, this is the first study that has evaluated continuous high resolution GEM data of CPT GAW with different statistical analysis techniques. GEM data was firstly evaluated by conducting cluster analysis on the dataset. This indicated that the GEM data could be divided into two clusters, separated at atmospheric concentrations of 0.904 ng m^{-3}. Trajectory analyses of the individual clusters, as well as the differences between these clusters, indicated that shipping around Cape Point could be a significant source of GEM. Further research quantifying the contribution of shipping should be undertaken, not only for the southern African region, but also for other busy shipping routes. Correlation of the time that back trajectories spent over the African continent and GEM concentration, proved that such analyses could be used as an alternative tool to distinguish between continental and marine GEM contributions. It was also demonstrated that MLR analysis could be used to determine an equation that can be used to predict GEM at CPT GAW. Moreover, this equation provided some insight into the complex nature of GEM chemistry. Lastly, the evaluation of both continuously measured and calculated (with the determined MLR Eq. 1) GEM concentrations seems to indicate a decline in GEM concentrations over the period evaluated in this paper. It remains to be seen whether this decline continues, which would reflect a positive response to global Hg emission reductions, or if it is only part of a longer-term cycle with a temporary decline.~~

Acknowledgements

The authors would like to thank the South African Weather Service for making available the GEM data measured at CPT GAW. The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF.

5 Bibliography

United States Coast Guard. 2014. *AMVER*. [ONLINE] Available at: http://www.amver.com/. [Accessed 17 November 14].

Angot, H., Barret, M., Magand, O., Ramonet, M., and Dommergue, A.: A 2-year record of atmospheric mercury species at a background Southern Hemisphere station on Amsterdam Island, Atmospheric Chemistry and Physics 14, 11461-11473, 2014.

Baker, PGL, Brunke, E-G, Slemr, F & Crouch, AM (2002), 'Atmospheric mercury measurements at Cape Point, South Africa', *Atmospheric Environment*, no. 36, pp. 2459-2465.

Brunke, E-G, C., L, Slemr, F, Ebinghaus, R & Kock, H (2010), 'Atmospheric Mercury Measurements at Cape Point, South Africa', *Clean Air Journal*, vol. 18, no. 1, pp. 17-21.

Brunke, E-G, Labuschagne, C, Ebinghaus, R, Kock, HH & Slemr, F (2010), 'Gaseous elemental mercury depletion events observed at Cape Point during 2007-2008', *Atmospheric Chemistry Physics*, no. 10, pp. 1121-1131.

Brunke, E-G, Labuschagne, C, Parker, B, Scheel, HE & Whittlestone, S (2004), 'Baseline airmass selection at Cape Point, South Africa: application of 222Rn and other filter criteria to CO2', *Atmospheric Environment*, no. 38, pp. 5693-5702. doi:10.1016/j.atmosenv.2004.04.024

Brunke, E-G, Scheel, HE & Seiler, W (1990), 'Trends of tropospheric CO, N2O and CH4 as observed at Cape Point, South Africa', *Atmospheric Environment*, no. 24A, pp. 585-595.

Ebinghaus, R, Jennings, SG, Schroeder, WH, Berg, T, Donaghy, T, Guentzel, J, Kenny, C, Kock, HH, Kvietkus, K, Landing, W, Muhleck, T, Munthe, J, Prestbo, EM, Schnee-berger, D, Slemr, F, Sommar, J, Urba, A, Wallschlager, D & Xiao, Z (1999), 'International field intercomparison measurements of atmospheric mercury species at Mace Head, Ireland', *Atmospheric Environment*, no. 33, pp. 3063-3073.

Ebinghaus, R, Kock, HH, Temme, C, Einax, JW, Lowe, AG, Richter, A, Burrows, JP & Schroeder, WH (2002), ' Antarctic springtime depletion of atmospheric mercury ', *Environmental Science and Technology*, no. 36, pp. 1238-1244.

Gierens, RT, Laakso, L, Mogensen, D, Vakkari, V, Beukes, JP, Van Zyl, PG, Hakola, H, Guenther, A, Pienaar, JJ & Boy, M (2014), 'Modelling new particle formation events in the South African savannah', *South African Journal of Science*, vol. 110, Art. #2013-0108, 12 pages. http://dx.doi.org/10.1590/sajs.2014/20130108

Laudal, DL, Pavlish, JH, Graves, J & Stockdill, D (2000), 'Mercury mass balances: A case study of two North Dakota Power Plants', *Journal of the Air & Waste management Association*, no. 50, pp. 1798-1804.

Lau, X, Talbot, R, Castro, M, Perry, K & Luke, W (2012), 'Seasonal and diurnal variations of atmospheric mercury across the US determined from AMNet monitoring data', *Atmospheric Chemistry and Physics*, no. 12, pp. 10569-10582. doi:10.5194/acp-12-10569-2012

Lindberg, S, Bullock, R, Ebinghaus, R, Engstrom, D, Feng, X, Fitzgerald, W, Pirrone, N, Prestbo, E & Seigneur, C (2007), 'A synthesis of progress and uncertainties in attributing the sources of mercury in deposition ', *Ambio*, no. 36, pp. 19-32.

Lin, C-J, Pongprueksa, P, Lindberg, SE, Pehkonen, SO, Byun, D & Jang, C (2006), 'Scientific uncertainties in atmospheric mercury models I: Model science evaluation', *Atmospheric Environment*, no. 40, pp. 2911-2928.

Lourens, ASM, Beukes, JP, van Zyl, PG, Fourie, GD, Burger, JW, Pienaar, JJ, Read, CE & Jordaan, JH (2011), 'Spatial and Temporal assessment of Gaseous Pollutants in the Mpumalanga Highveld of South Africa', *South African Journal of Science*, vol. 107, no. 1/2, p. 8. Art. #269, 8 pages. DOI: 10.4102/sajs.v107i1/2.269

Lourens, ASM, Butler, TM, Beukes, JP, van Zyl, PG, Beirle, S & Wagner, T (2012), 'Re-evaluating the NO2 hotspot over the South African Highveld', *South African Journal of Science*, vol. 108, no. 9/10. Art. #1146, 6 pages, http://dx.doi.org/10.4102/sajs.nv108i11/12.1146

Mao, H & Talbot, R (2012), 'Speciated mercury at marine, coastal, and inland sites in New England - Part 1: Temporal variability', *Atmospheric Chemistry and Physics*, no. 12, pp. 5099-5112.

Mason, RP (2009), *Mercury emissions from natural processes and their importance in the global mercury cycle*, Springer, New York, USA.

Mergler, D, Anderson, HA, Chan, LHM, Mahaffey, KR, Murray, M, Sakamoto, M & Stern, AH (2007), 'Methylmercury exposure and health effects in humans: A worldwide concern', *Ambio*, no. 36, pp. 3-11.

Pirrone, N, Cinnirella, S, Feng, X, Finkelman, RB, Friedli, HR, Leaner, J, Mason, R, Mukherjee, AB, Stracher, GB, Streets, DG & Telmer, K (2010), 'Global mercury emissions to the atmosphere from anthropogenic and natural sources', *Atmospheric Chemistry and Physics*, no. 10, pp. 5951-5964.

Riddle, EE, Voss, PB, Stohl, A, Holcomb, D, Maczka, D, Washburn, K & Talbot, RW (2006), 'Trajectory model validation using newly developed altitude-controlled balloons during the International Consortium for Atmospheric Research on Transport and Transformations 2004 campaign', *Journal of Geophysical Research*, vol. 111.

Richter, A, Begoin, M, Hilbol, A & Burrows, JP 2011, 'An improved NO2 retrieval for the GOME-2 satellite instrument', *Atmospheric Measurement Techniques*, no. 4, p. 1147–1159.

Scott, GM & Mdluli, TN (2012), 'The Minamata Treaty / Protocol: Potential Implications for South Africa', *Clean Air Journal*, vol. 22, no. 2, pp. 17-19.

Skjølsvik, KO, Andersen, AB, Corbett, JJ & Skjelvik, JM 2000, *Study of Greenhouse Gas Emissions from Ships*, 2nd edn, Norwegian Marine Technology Research Institute, Trondheim.

Slemr, F, Brunke, E-G, Ebinghaus, R & Kuss, J (2011), 'Worldwide trend of atmospheric mercury since 1995', *Atmospheric Chemistry and Physics*, no. 11, pp. 4779-4787. doi:10.5194/acp-11-4779-2011

Slemr, F, Brunke, E-G, Labuschagne, C & Ebinghaus, R (2008), 'Total gaseous mercury concentrations at the Cape Point GAW station and their seasonality', *Geophysical Research Letters*, no. 35, pp. L11807.

Slemr, F, Brunke, E-G, Whittlestone, S, Zahorowski, W, Ebinghaus, R, Kock, HH & Labuschagne, C (2013), '222Rn-calibrated mercury fluxes from terrestrial surface of southern Africa', *Atmospheric Chemistry and Physics*, no. 13, pp. 6421-6428. doi:10.5194/acp-13-6421-2013

Slemr, F., Angot, H., Dommergue, A., Magand, O., Barret, M., Weigelt, A., Ebinghaus, R., Brunke, E.-G., Pfaffhuber, K. A., Edwards, G., Howard, D., Powell, J., Keywood, M., and Wang, F.: Comparison of mercury concentrations measured at several sites in the Southern Hemisphere, Atmospheric Chemistry and Physics, 15, 3125-3133, 2015.

Stohl, A (1998), 'Computation, accuracy and application of trajectories – a review and bibliography', *Atmospheric Environment*, no. 32, pp. 947-966.

Vakkari, V, Beukes, JP, Laakso, H, Mabaso, D, Pienaar, JJ, Kulmala, M & Laakso, L (2013), 'Long-term observations of aerosol size distributions in semi-clean and polluted savannah in South Africa', *Atmospheric Chemistry and Physics*, no. 13, pp. 1751-1770. doi:10.5194/acp-13-1751-2013

Venter, AD, Vakkari, V, Beukes, JP, van Zyl, PG, Laakso, H, Mabaso, D, Tiitta, P, Josipovic, M, Kulmala, M, Pienaar, JJ & Laakso, L (2012), 'An air quality assessment in the industrialized western Bushveld Igneous Complex, South Africa', *South African Journal of Science*, vol. 108, Art. #1059, 10 pages, http:// dx.doi.org/10.4102/sajs.v108i9/10.1059

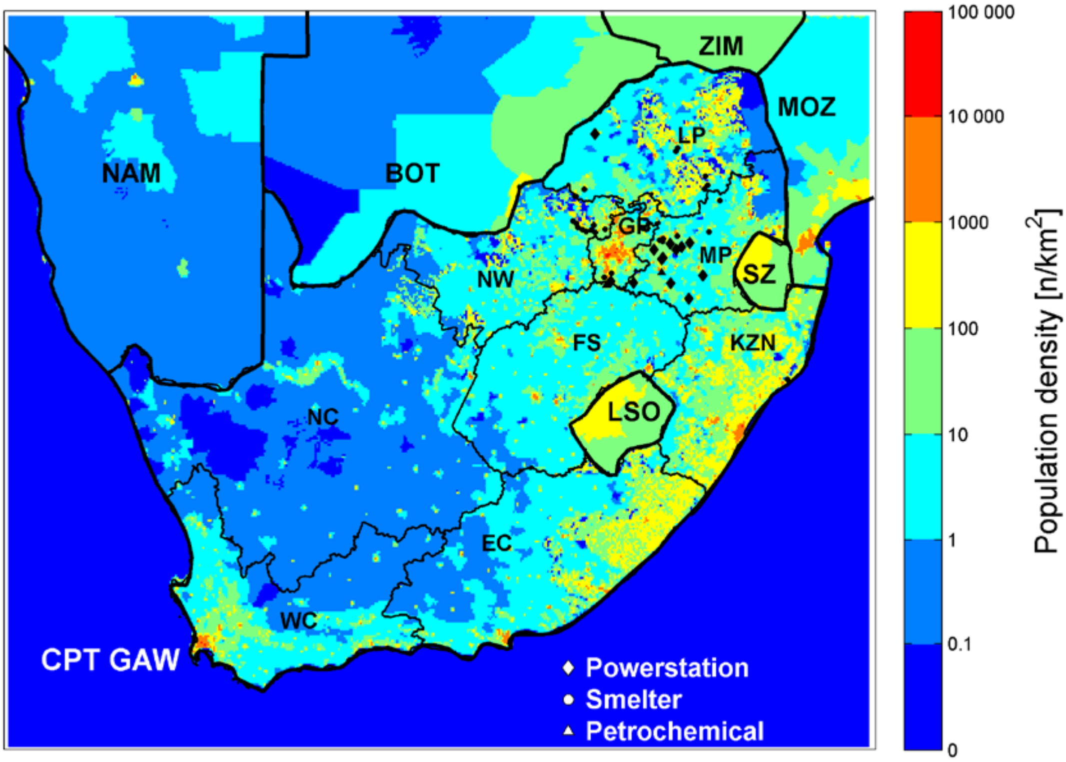
Wagner, NJ (2001), 'Trace Elements in coal, their analysis and environmental impact.' Literature Survey, SASOL R&D, Sasolburg.

Table 1: The percentage GEM data distribution observed for each cluster solution

	% GEM distribution in each cluster						
Cluster solutions	**1**	**2**	**3**	**4**	**5**	**6**	**7**
2	36.79	49.07					
3	3.35	41.34	41.18				
4	1.78	28.54	16.02	39.52			
5	18.37	0.05	27.27	38.47	1.70		
6	11.43	31.36	1.36	27.29	0.05	14.37	
7	1.51	1.23	29.62	11.34	16.76	25.38	0.02

Table 2: The overall identity of independent variables during the determination of the optimum combination of independent variables for GEM calculation utilising the entire dataset

No of independent variables	**MLR solutions for all GEM values**							
1	AbsH							
2	AbsH	O_3						
3	AbsH	O_3	CO					
4	AbsH	O_3	CO	P				
5	AbsH	O_3	CO	P	T			
6	AbsH	O_3	CO	P	T	CH_4		
7	AbsH	O_3	CO	P	T	CH_4	Rn	
8	AbsH	O_3	CO	P	T	CH_4	Rn	WGS


ZIM
MOZ
NAM
BOT
LP
GP
MP
SZ
NW
FS
KZN
LSO
NC
EC
WC
CPT GAW
Powerstation
Smelter
Petrochemical
100 000
10 000
1000
100
10
1
0.1
0
Population density [n/km2]

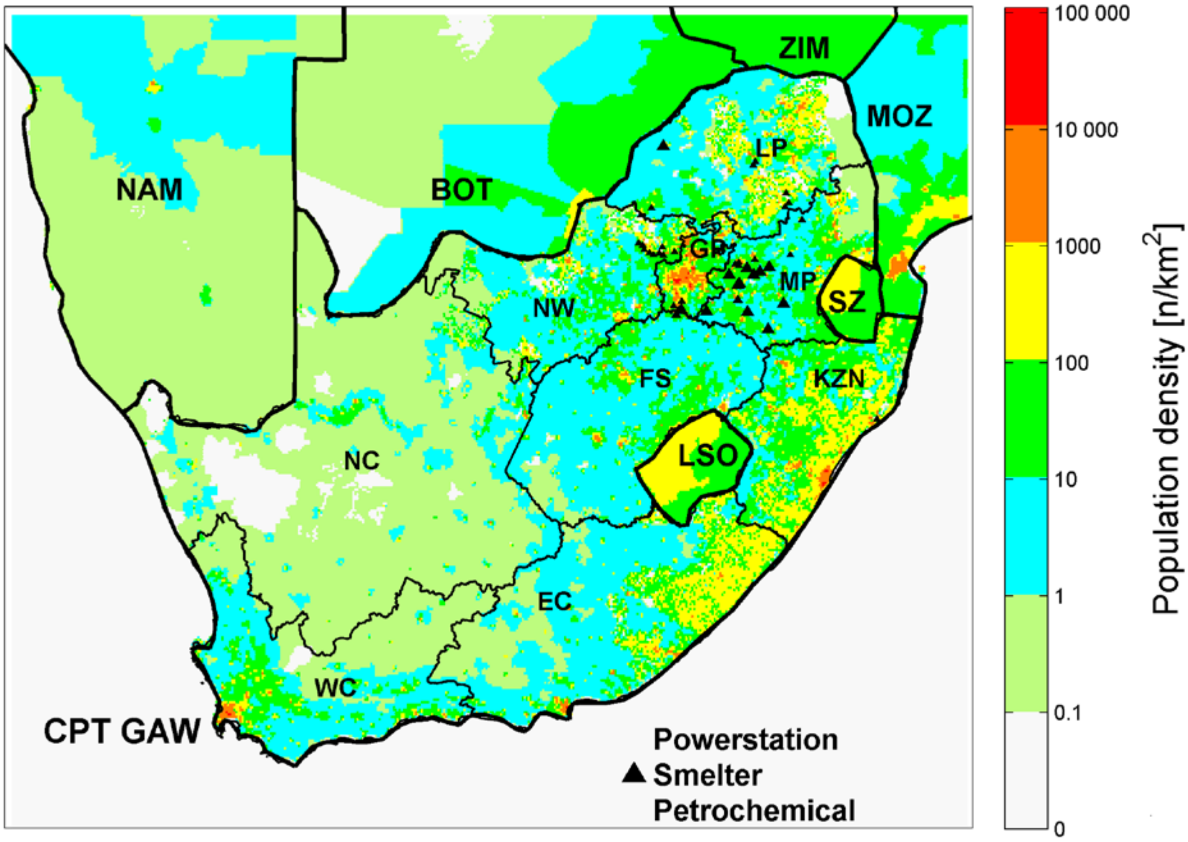

ZIM
MOZ
NAM
BOT
LP
GP
MP
SZ
NW
FS
KZN
LSO
NC
EC
WC
CPT GAW
Powerstation
Smelter
Petrochemical
100 000
10 000
1000
100
10
1
0.1
0
Population density [n/km2]

Figure 1: Position of CPT GAW within a regional context. The population density (people per km^2) provides an indication of the possible location of anthropogenic pollutions sources, while the location of large anthropogenic point sources (e.g. coal-fired power stations, metallurgical smelters and petrochemical plants, adapted from Venter et al., 2012 and Lourens et al., 2011 and 2012) is also indicated. NAM = Namibia, BOT = Botswana, ZIM = Zimbabwe, MOZ = Mozambique, SZ = Swaziland, LSO = Lesotho, WC = Western Cape, EC = Eastern Cape, NC = Northern Cape, NW = North West, FS = Free State, KZN = KwaZulu-Natal, GP = Gauteng, MP = Mpumalanga and LP = Limpopo

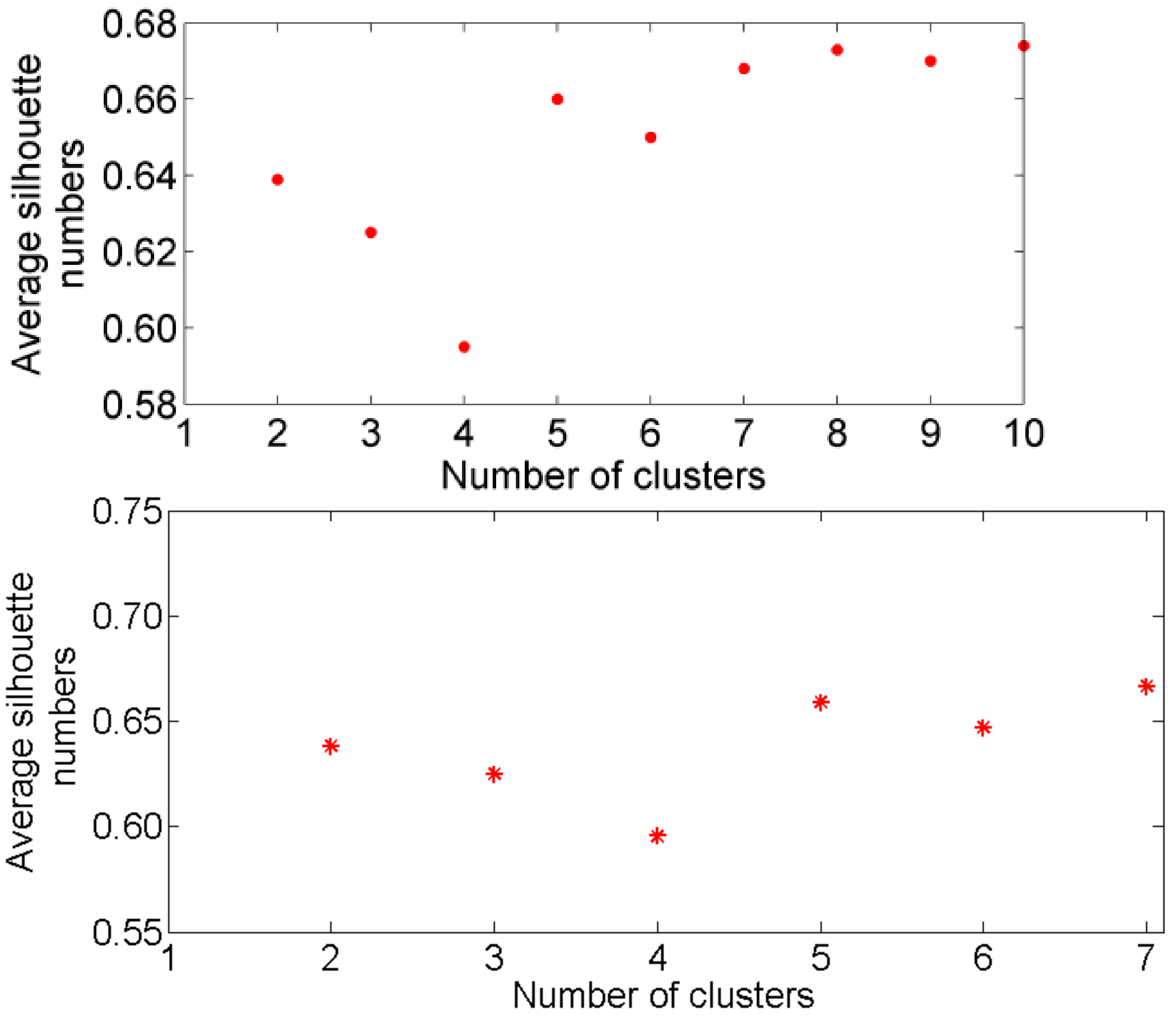


Figure 2: Average silhouette numbers for the various cluster solutions. An increase in silhouette numbers indicates that individual sub-clusters are better separated

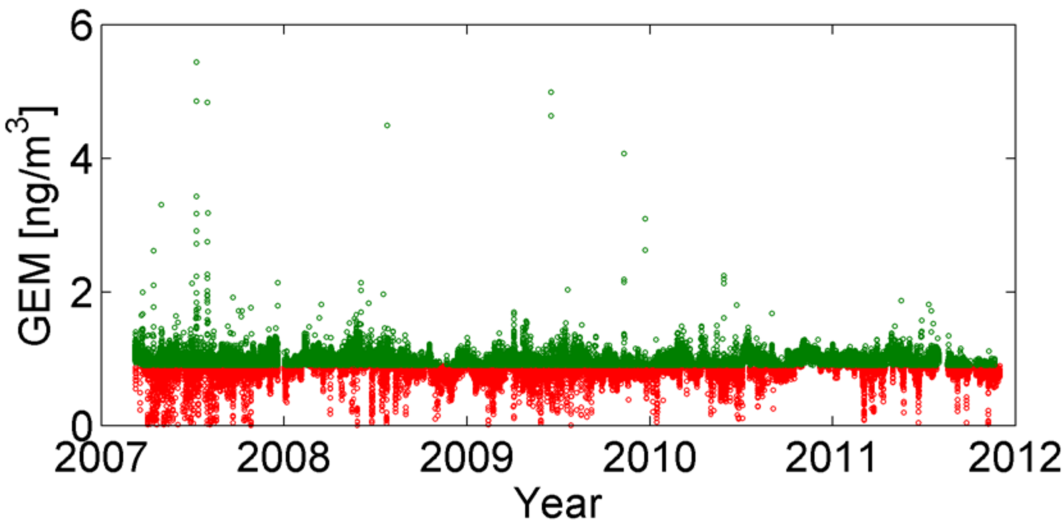
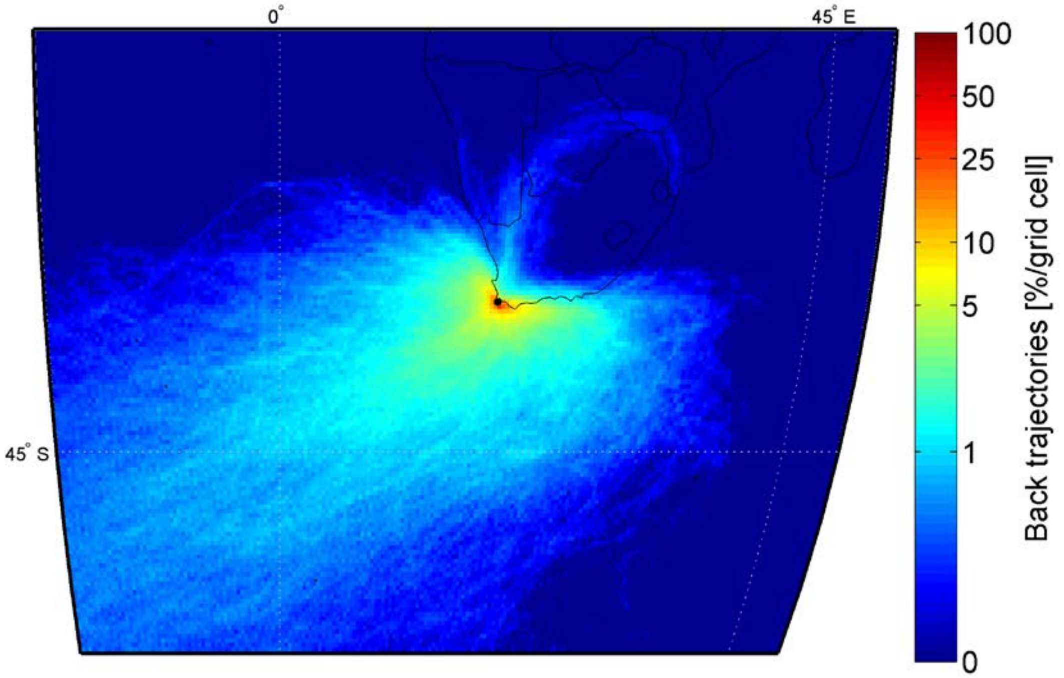
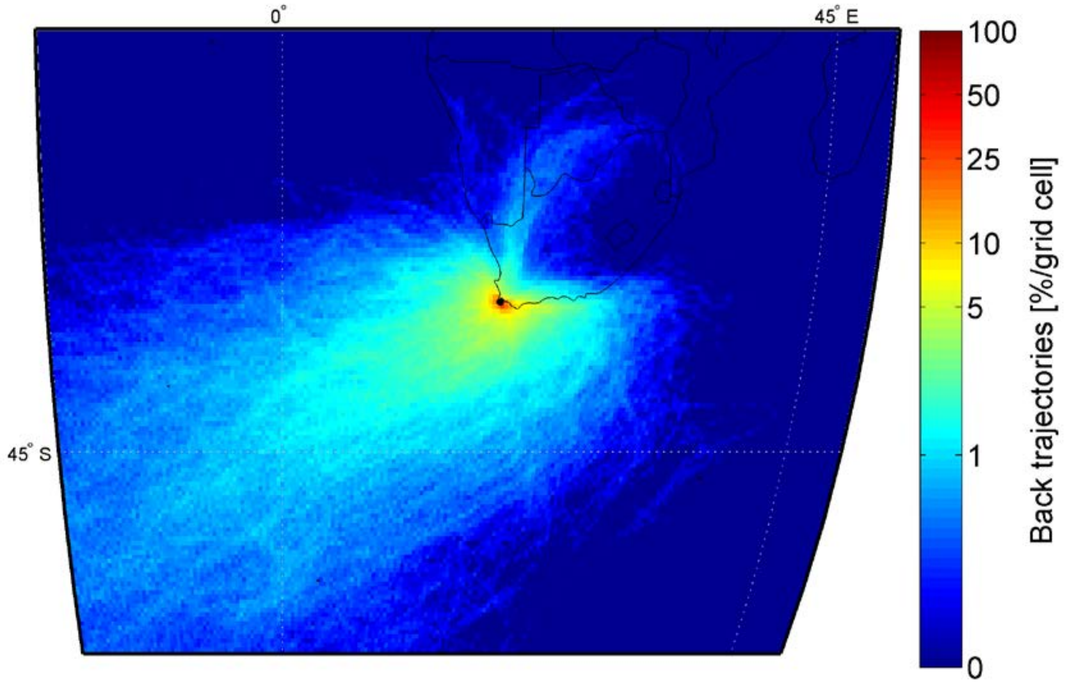


Figure 3: A scatter plot of GEM concentrations over the entire sampling period indicating the two main clusters. According to the clustering applied, division between the two clusters was at a GEM concentration of 0.904 ng m^{-3}

(a)


0°
45° E
45° S
100
50
25
10
5
1
0
Back trajectories [%/grid cell]

(b)


0°
45° E
45° S
100
50
25
10
5
1
0
Back trajectories [%/grid cell]

(c)

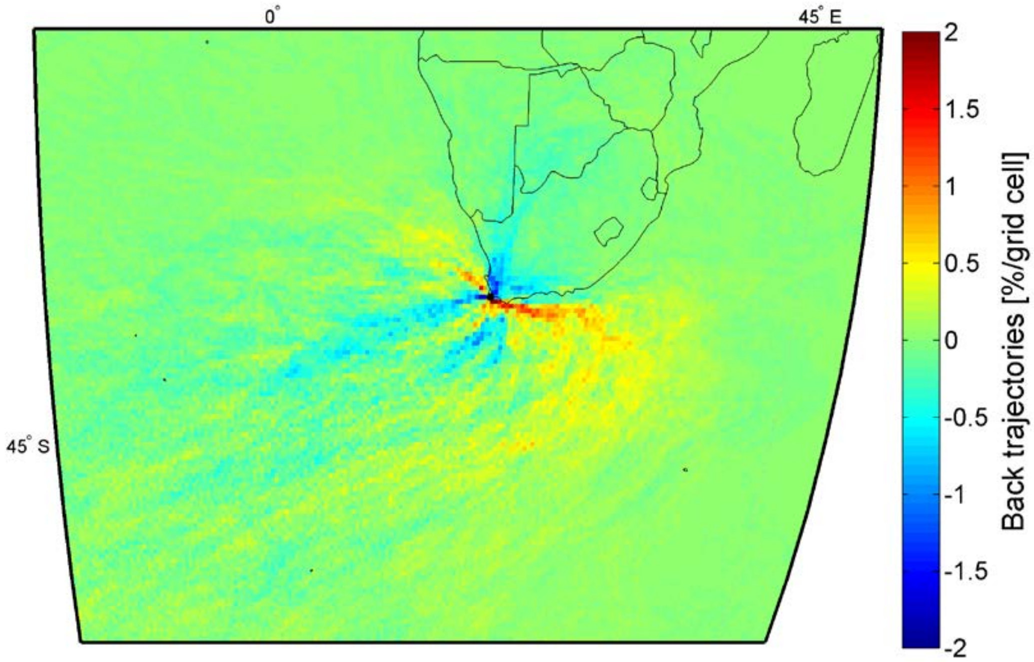


Figure 4: Normalised back trajectory analysis map, i.e. hourly arriving eight-day back trajectories with 100 m arrival height overlaid with MATLAB and normalised to percentage for the entire sampling period for, (a) cluster one, i.e. GEM concentration > 0.904 ng/m^3, (b) for cluster two, i.e. GEM concentration < 0.904 ng/m^3 and (c) the difference between the two individual maps, i.e. percentage of trajectories passing over each correlating 0.2° x 0.2° grid cells in Fig. 4b subtracted from the percentage of trajectories passing over each 0.2° x 0.2° grid cell in Fig. 4a. The colour bar indicates the percentage of trajectories passing over each grid cell

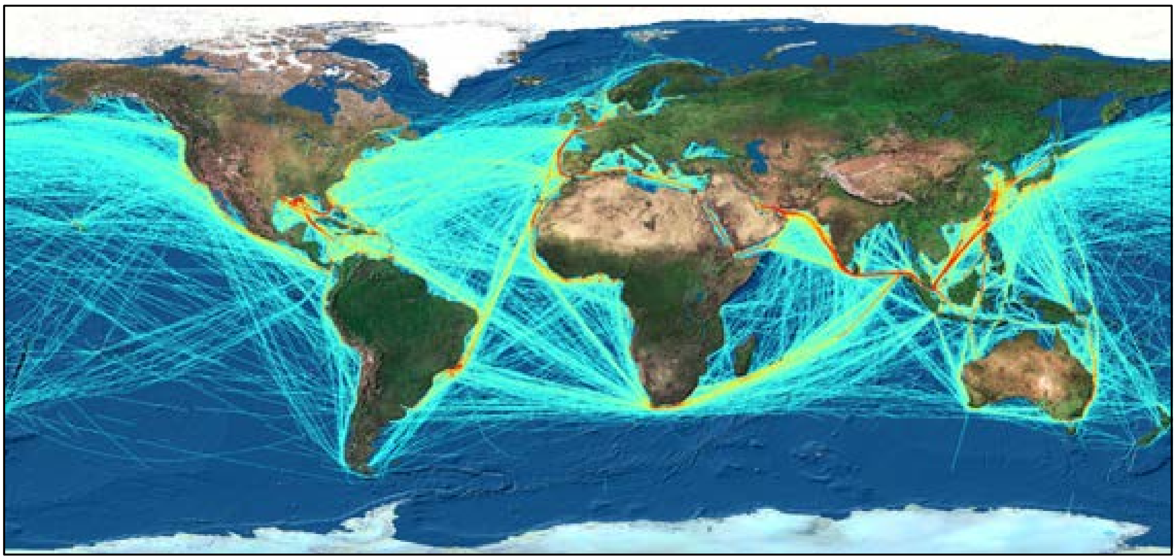
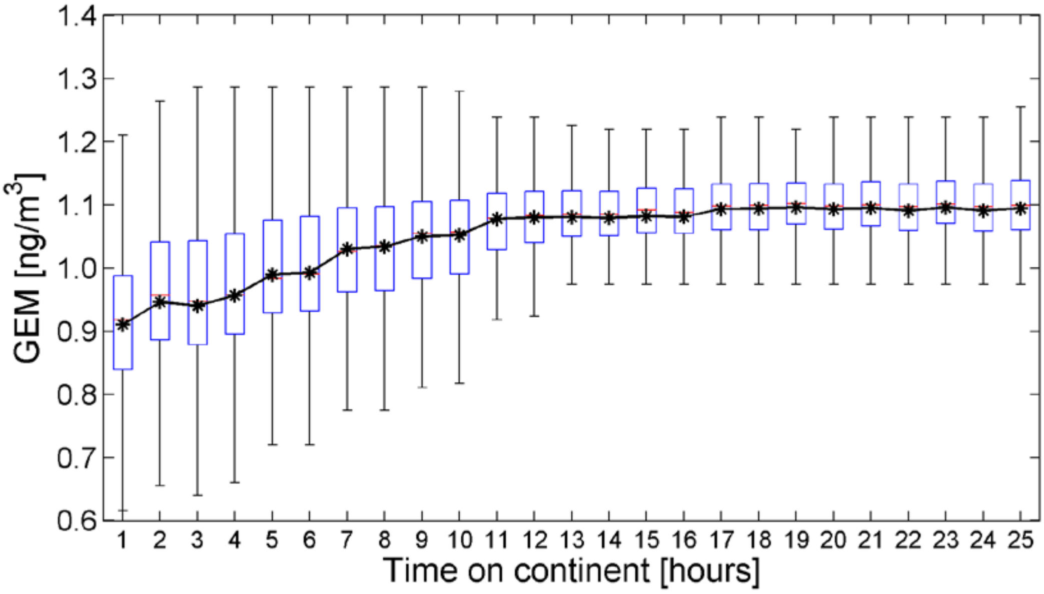
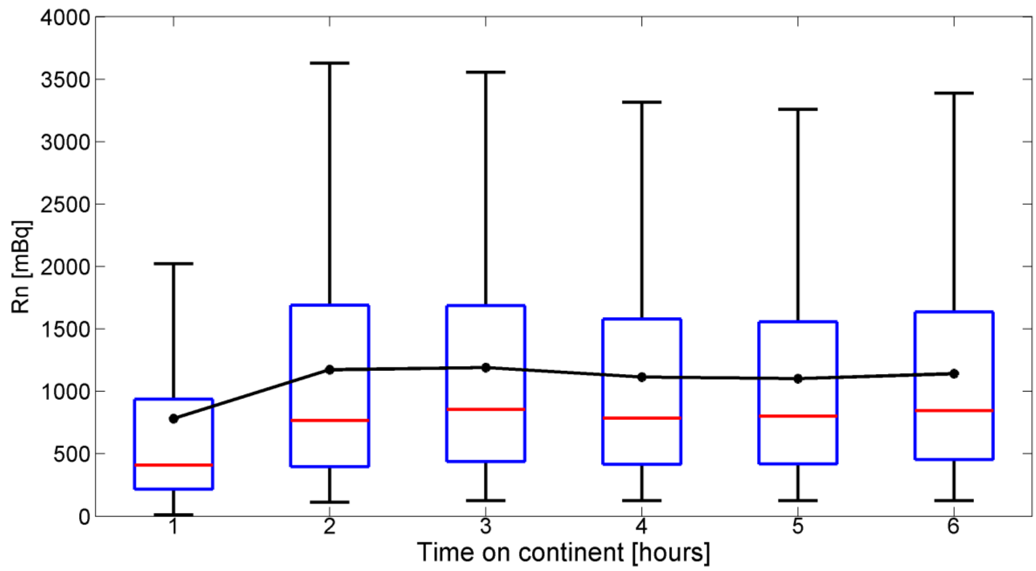


Figure 5: Monthly density plot for the total number of ships registered with Automated Mutual-Assistance Vessel Rescue System (AMVER) for July 2011. AMVER is sponsored by the United States Coast Guard and makes use of the global ship reporting system used worldwide by search and rescue authorities. The ship density plot is compiled from a 2011 average of 4 634 ships per day (United States Coast Guard, 2014)



(a)



(b)

Figure 6: (a) The statistical distribution of GEM concentrations as a function of time spent over the continent and (b) ^{222}Rn distribution as a function time air masses spent over the continent. The mean is indicated by the black stars, the median by the red line, the 25- and 50 percentile by the blue box and the whiskers indicating 99.3 % data coverage (if a normal distribution is assumed), while the black line connects the mean values to provide an indication of the trend observed~~Figure 6: The statistical distribution of GEM concentrations as a function of time spent over the continent. The mean is indicated by the black stars, the~~

~~median by the red line, the 25- and 50 percentile by the blue box and the whiskers indicating 99.3 % data coverage (if a normal distribution is assumed), while the black line connects the mean values to provide an indication of the trend observed~~

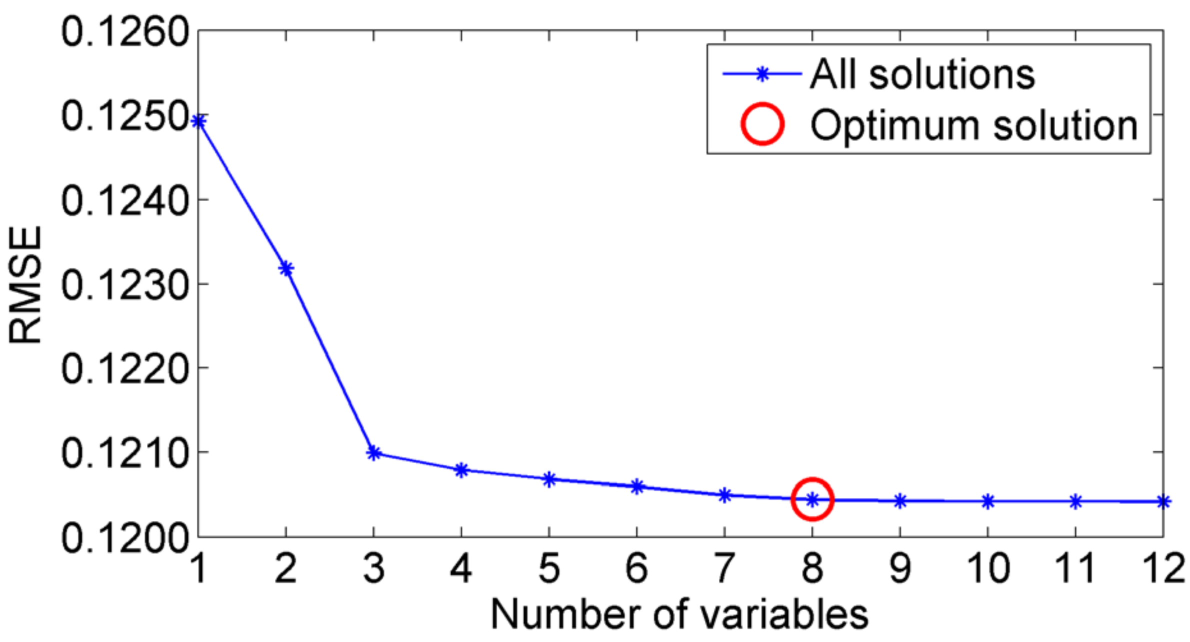
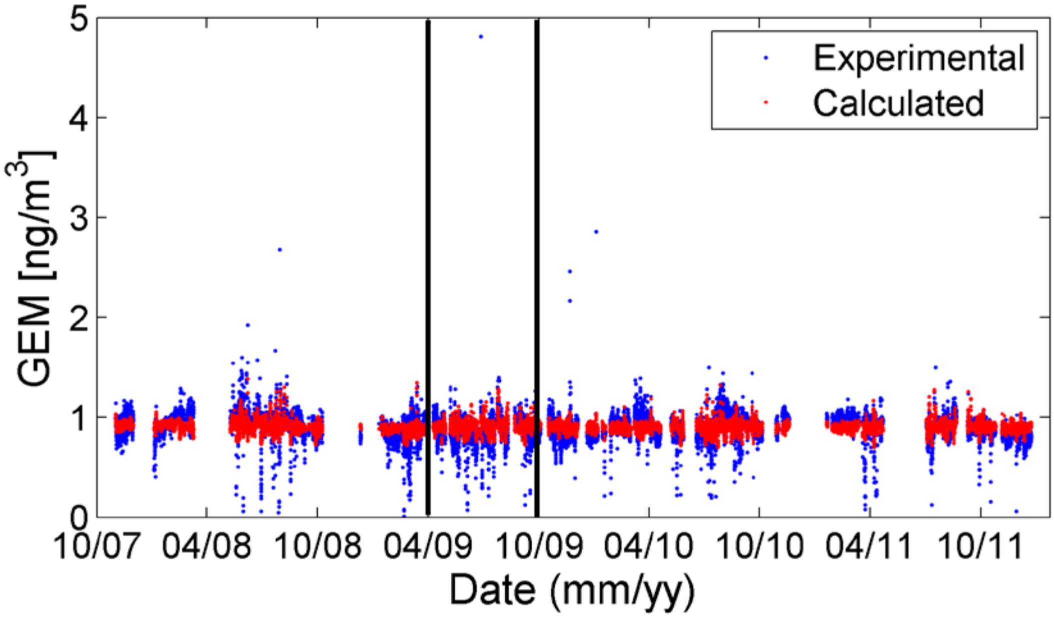


Figure 7: Determination of the optimum combination of independent variables to include in the MLR equation to calculate the dependant variable, i.e. GEM concentration (2007-2011). The root mean square error (RMSE) difference between the calculated and actual GEM concentrations indicated that the inclusion of 8 parameters in the MLR solution was the optimum

(a)



(b)

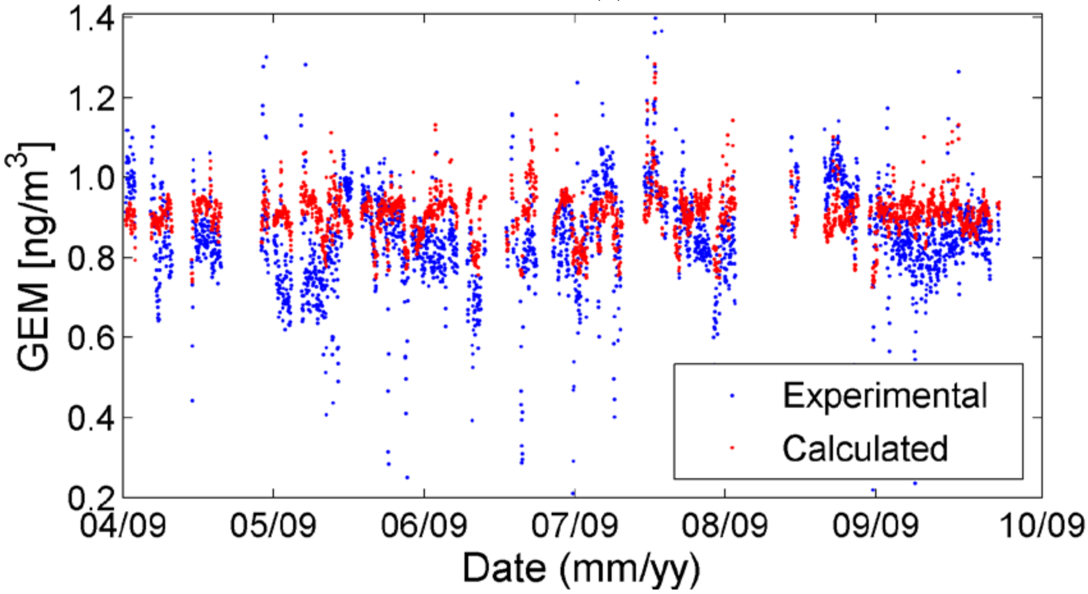


Figure 8: (a) Measured GEM (blue) and calculated GEM concentrations using the MLR Eq. 1 (red) for the entire sampling period. The two vertical black lines in Fig. 8a indicate a period that was enlarged in Fig. 8b to indicate more detailed differences between the measured and calculated GEM concentrations