Geosci. Instrum. Method. Data Syst. Discuss., 4, C333–C336, 2015
www.geosci-instrum-method-data-syst-discuss.net/4/C333/2015/

Geoscientific Instrumentation Methods and Data Systems
Discussions

Interactive comment on "Designing optimal greenhouse gas observing networks that consider performance and cost" *by* D. D. Lucas et al.

D. D. Lucas et al.

ddlucas@alum.mit.edu

Received and published: 23 March 2015

Response to Anonymous Referee #1

We appreciate the feedback from Anonymous Referee #1. The most significant issues identified by the referee involve improving our discussion of the impacts of our assumptions on our results. The motivation for improving this discussion is to better understand how the prototype GHG observing network presented in the manuscript could be extended and used for a real-world network design. In particular, the reviewer is interested in the assumptions about the space and time dependence of noise, measurement frequencies at the stations, and the cost function. Each of these

points is addressed in more detail below and will be included in our *Summary and conclusions* in our revised manuscript.

In addition to our responses and revisions, we also note that we are releasing a data set associated with our simulations and analysis to a public domain data repository and ftp site. This data set will allow other researchers to use different inversion and optimization methods, and to test different cost functions, noise characteristics, and other assumptions about the analysis. The final details about the data release will be included in the revised manuscript, but tentatively the data will be accessible at the UCI Machine Learning repository under the name "Greenhouse Gas Observing Network" (http://archive.ics.uci.edu/ml/) and through the LLNL Green Data Oasis ftp site (ftp://gdo148.ucllnl.org/pub/ghg_data.tgz).

Last, we also thank the reviewer for identifying the typo on Page 713, line 20 in the manuscript.

Space and time dependence of noise. As noted in Sect. 2.2 in our manuscript, the noise added to the time series differs from one location to another because different random seeds are used at each site. However, the noise amplitude is assumed to be constant in space and time (i.e., a fixed value of ρ is used) and the noise is not spatially correlated. These assumptions are consistent with i.i.d. noise (independent and identically distributed), which is often a reasonable starting point for statistical analysis. In practice, however, the GHG time series noise may be spatially correlated and may have an amplitude that scales with mixing ratio. To properly characterize these features, we would need to analyze actual observations, rather than model simulations, of HFC-134a or a comparable tracer, but only limited observations are available (e.g., at Trinidad Head and Scripps). In the absence of extensive observations, we can only surmise about the impacts of relaxing our noise assumptions. By including spatially correlated noise, we expect that the optimization scheme would penalize stations that are close to each other because neighboring grid cells would experience similar fluctuations. However, the spatial correlation length scale is also expected to be relatively small (e.g., less than 10–20 km) because California has rough surface features and complex topography. The net effect of including spatially correlated noise on our analysis is therefore anticipated to be minor. By relaxing our constant noise amplitude assumption, on the other hand, we anticipate that the uncertainty in the inferred emissions of large emitting regions would increase (i.e., the error bars for regions 7 and 12 in Fig. 6), which would drive the optimization scheme to increase the preference for stations nearby those regions.

Same measurement frequency for all stations. Our assumption of using the same measurement frequency at all of the stations in the network was a matter of convenience, though relaxing this assumption would not be difficult. 12 design variables, instead of 7, could have been used to optimize the location and frequency of each station. The computational time required to design the network, however, would increase with the number of design variables. We expect that such a change would result in a network with stations that collect measurements relatively more frequently in locations that are far from important sources (e.g., regions 1 and 6) than locations that are nearby (e.g., regions 7 and 12). The data set that we will be releasing to the public repository can be used by other researchers to test independent measurement frequencies of 1, 2, 3 and 4 samples per day at different locations.

Cost function. As noted in the manuscript, the cost function used in the network design is highly idealistic. The form and structure of the cost function was chosen to help illustrate the notion of competing objectives (performance versus cost) and impart convexity to the Pareto frontier. The authors have more expertise on the performance side of network design than the cost side, so it is difficult for us to extrapolate our

results to situations involving more detailed cost models. Collaborative efforts with other experts are needed to build out a comprehensive cost model and better explore the impacts of different cost decisions on network design. We also invite researchers in the community to apply different cost models to the data set we will be releasing.

Interactive comment on Geosci. Instrum. Method. Data Syst. Discuss., 4, 705, 2014.