Reducing stray light contamination in the limb profiles of the Ozone Mapper Profiler Suite (OMPS)

John W. Bergman
Computational Physics Inc., Boulder CO

Didier F. Rault
National Aeronautics and Space Administration, Langley VA

John S. Hornstein
Naval Research Laboratory, Washington DC

Jerry D. Lumpe
Computational Physics Inc., Boulder CO

"This is all you really need to know"

We are working on an algorithm to reduce stray light contamination in ozone retrievals by the OMPS limb-profiles (LP).

Our most recent approach uses a regression representation of the instrument in the forward model of the optimal estimation procedure.

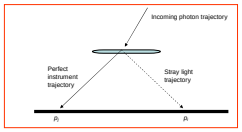
The method has the following benefits:

- There is no explicit inversion of the stray light operator and, so, the retrieval doss not act upon amplified noise
- A regression model can be computationally efficient because:
  - The bulk of the cost is incurred during off-line training
  - The regression exploits large-scale patterns in the fields to reduce the dimension of the stray light operator
- Input to the regression model is output from an idealized radiative transfer (RT) model making the regression model (in principle) perfectly trainable.

Our most recent efforts focus on developing a training strategy for the regression model based on the subdivision of state space into regions over which the RT model is nearly linear.

For this method to be ultimately successful, we will need a reliable characterization of the instrument and a strategy for retraining the regression model to accommodate future changes in the instrument characterization

"The Problem is, stray photons don't go where they're supposed to"



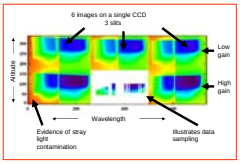
Stray light contamination results from internal scattering within the instrument that causes photons to strike the image plane at a different location than it would otherwise.

It is described via the stray light operator $\mathbf{\Theta}$ whose elements θ_{ij} represent the probability that a photon that would strike the image plane at location p_j in the absence of stray light actually strikes the image plane at location p_i.

Removing stray light amounts to applying the inverse operator $\mathbf{\Theta}^{-1}$ to the measurements $\mathbf{m}$:

- This is analogous to performing a de-convolution, a process that is notoriously unstable and prone to noise amplification.
- The measurements are noisy
- Measurements typically do not describe the entire image plane and, thus, $\mathbf{\Theta}^{-1}\mathbf{m}$ is an inaccurate representation of what the measurements would be without stray light

"We constructed an instrument simulator to investigate stray light contamination in the OMPS-LP"

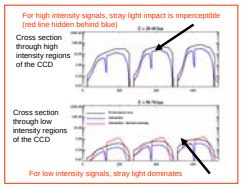


To perform our investigation, we constructed an instrument simulator for the OMPS-LP that includes:
Quantum efficiency, Optical through put, Stray light, Stray light filters, High and low gain images, Two different integrating times, and Noise

Relevant to this study:

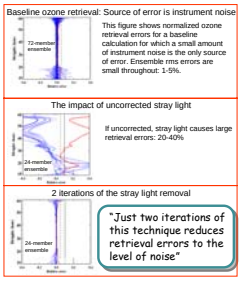
- Evidence for stray light is found in between images where the signal should be zero.
- Multiple images on the same CCD allow cross-image contamination by stray light.
- Less than 20% of the data is utilized – making it difficult to remove stray light via the inverse stray light operator.

"Stray light dominates the signal for low-intensity portions of the CCD, but is a small perturbation to the radiant energy overall"



The figure above shows measurements along cross sections through a high intensity region of the CCD (top panel) and a low intensity region. The black line represents a perfect instrument. The blue shows the impact of attenuating influences, and the red line represents the full instrument model. The difference between the red and blue lines quantifies the impact of stray light. While stray light dominates the signal for low intensity regions, overall it is small effect. Our work exploits the 'smallness' of stray light.

"We first explored an iterative method based on the Taylor expansion of the inverse stray light operator to remove stray light from the measurements."



"Just two iterations of this technique reduces retrieval errors to the level of noise"

"Our iteration method to invert the stray light operator has two principal problems."

The success of the iteration method to invert the stray light operator was very encouraging, but it has two problems.

It is expensive. It requires two applications of an expensive instrument model.

It requires full knowledge of the measurements, which we will not have due to the sub-sampling strategy used to meet data rate constraints.

"We use linear regression to circumvent both problems."

Linear regression can be efficient because the bulk of the expense is incurred during offline training. In addition, regression models are able to exploit large-scale patterns in the radiance and measurement fields, allowing the information in these fields to be stored with a small number of parameters. This latter property makes the regression efficient. It also allows us to remove the impact of nearly all sources of stray light despite severe sub-sampling of the image plane.

"The efficiency of the regression model allows us to explore a variety of stray light removal methods"

Ozone retrieval from the OMPS-LP is essentially a two step process. A data processing step (e.g., producing an SDR), with operator $\mathbf{Q}$ that produces a data vector $\mathbf{y}$ from measurements $\mathbf{m}$

$$\mathbf{y} = \mathbf{Q}(\mathbf{m})$$

And the ozone retrieval that inverts the relationship

$$\mathbf{y} = \mathbf{F}(\mathbf{x})$$

using the optimal estimation procedure of Rodgers (2000), where $\mathbf{x}$ is the state vector to be retrieved (i.e., the vertical profile of ozone number density) and $\mathbf{F}$ is the forward model.

Within this simple view of the retrieval process, we have two choices for implementing stray light removal.

Method 1: We can include stray light removal in the data processing step. In this case, $\mathbf{Q}$ is a regression representation of the inverse instrument operator and $\mathbf{F}$ is a radiative transfer model.

Method 2: We can include the effects of stray light in the forward model of the optimal estimation. In this case, $\mathbf{Q}$ is essentially the identity, and $\mathbf{F}$ combines radiative transfer with a regression representation of the instrument.

A side issue: "The fact that a regression model is a tangent linear model is the key to the our ability to place a regression representation of the instrument into the forward model"

Optimal estimation inverts the relationship

$$\mathbf{y} = \mathbf{F}(\mathbf{x})$$

via the tangent linear model, or kernel, $\mathbf{K}$

$$\mathbf{y} = \mathbf{K}\mathbf{x};\ K_{ij} = \frac{\partial y_i}{\partial x_j}$$

For a state vector of dimension n and data vector of dimension m calculating, determining $\mathbf{K}$ requires enough runs of the forward model to determine $n \times m$ derivatives, which for large n and m can be expensive.

However, since a regression model is itself a tangent linear model, updating the radiative transfer kernel $\mathbf{K}_{rad}$ to include the instrument kernel $\mathbf{K}_{inst}$ only requires multiplying $\mathbf{K}_{rad}$ by a matrix of regression coefficients – *which are pre-calculated*.

For example, consider the radiance vector $\mathbf{r}$

$$\mathbf{r} = \mathbf{K}_{rad}\mathbf{x}$$

So,

$$\mathbf{y} = \mathbf{K}\mathbf{x} = \mathbf{K}_{inst}\mathbf{r} = \mathbf{K}_{inst}\mathbf{K}_{rad}\mathbf{x}$$

or

$$\mathbf{K} = \mathbf{K}_{inst}\mathbf{K}_{rad}$$

Regression training: "The regression is performed for each measurement vector element"

For the case of method 2: We use regression to calculate the measurement vector $\mathbf{m}$ from the radiance vector $\mathbf{r}$.

$$\mathbf{m} = \mathbf{m}_0 + \mathbf{m}'$$
$$= \mathbf{m}_0 + \boldsymbol{\beta} + \mathbf{A}\mathbf{r}';\ \mathbf{r} = \mathbf{r}_0 + \mathbf{r}'$$

To determine the coefficients $\boldsymbol{\beta}$ and $\mathbf{A}$, we utilize a training set of measurement/radiance pairs $\mathbf{m}(t)$, $\mathbf{r}(t)$; $t = 1,T$ and minimize the least squares error for each vector element m_i in the set of equations

$$m_i'(t) = m_{0i} + \beta_i + \sum_j a_{ij} r_j'(t)$$

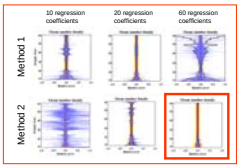
This process leads to the following set of equations

$$m_i' \cdot r_k' = \sum_j a_{ij} r_j' \cdot r_k'$$
$$\overline{m}_i = m_{0i} + \beta_i$$

where; $f_i(t) = \overline{T}_i + T_i'$

$$\overline{T} = \frac{1}{T}\sum_t f(t)$$
$$f \cdot g = \frac{1}{T}\sum_t f(t)g(t)$$

"In our experiments, the method that places a regression representation of the instrument into the forward model of the optimal estimation procedure performs best"



This figure compares noise amplification by the two stray light removal techniques for different number of regression coefficients used.

Note that while we can improve the performance of method 2 by increasing the number of coefficients used in the regression model, improvements using method 1 are limited by the increased noise amplification that accompanies such an increase.

Method 2 also outperforms method 1 in experiments that simulate the impact of pixel failure, pixel saturation, errors in the instrument characterization, and errors in the assumed properties of the atmosphere.

"In hindsight, the performance of this method should be of no surprise, placing a regression representation of the instrument into the forward model to remove stray light has a number of beneficial properties"

Placing the regression model into the forward model brings stray light removal under the umbrella of the optimal estimation routine. This regularizes stray light inversion as it does inversion of radiative transfer.

In addition, input to the regression model is output from an idealized RT model. This has two important consequences:

1. Input to the regression model is noise free, so the regression model does not directly amplify noise.

2. It is, in principle, possible to know a *priori* the full range of output from the RT model and, thus, the regression model is perfectly trainable. The sub-sampling of measurements does not adversely affect our ability to remove stray light.

"So, to complete our work: we need to develop a training for the regression model, develop a strategy for retraining to accommodate future instrument changes, and a reliable characterization of the instrument"

"The major obstacle for a training strategy to overcome is the non-linearity of radiative transfer"

To fully train the regression model, the training data (sets of radiance output from the idealized RT model paired with the corresponding measurements) should span the input and output spaces of the instrument.

One factor that makes our regression approach appealing is that the input space to the regression model is constrained by the output of an idealized RT calculation, which constrains the amount of training data required.

To exploit that factor, we must know a *priori* what that restricted radiance space is in order to train the regression model. We do not have that information.

We do, however, know the constraints on the state space that is input to the RT model.

IF the RT model was linear, developing a training strategy (i.e., deciding on an appropriate set of training data) would be simple.

We could construct radiance/measurement pairs from a set of basis state vectors that span the input space of the RT model.

But, the RT model is not linear …

"The strategy we are pursuing is to divide the state space into linear sub-spaces over which the RT calculation is approximately linear"

To help us develop and test this strategy, we first consider a state space for which only ozone concentration and temperature vary. All other aspects of the atmospheric/viewing space are held fixed: solar angle, viewing angle, surface reflectivity, aerosol properties, …

We investigate the properties of the RT calculation over this state space – investigating the performance of a linearized calculation to determined the number of linear sub-spaces required to span the entire state space.

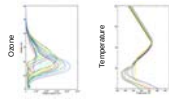
We construct basis temperature and ozone number density profiles for each sub-space and calculate training radiance/measurement data from those basis sets.

Our current basis vectors are calculated using sinusoidal perturbations to the state vector at the center of the linear sub-space.

$$T^{min}(z) = T_0 + 4.0\cos\left(\frac{2\pi k z}{z_c}\right)$$
$$T^{max}(z) = T_0 + 4.0\sin\left(\frac{2\pi k z}{z_c}\right)$$
$$O^{min}(z) = O_0\left[1 + 0.5\cos\left(\frac{2\pi k z}{z_c}\right)\right]$$
$$O^{max}(z) = O_0\left[1 + 0.5\sin\left(\frac{2\pi k z}{z_c}\right)\right]$$

"To test our method, we constructed 2 sets of verification data"

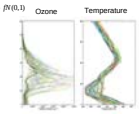
The first set is considered to be a necessary, but probably not sufficient test of the method. We use it because it is constructed with the same standard ozone and temperature profiles (Wellemeyer et al., JGR, 1997) used to test operational codes.



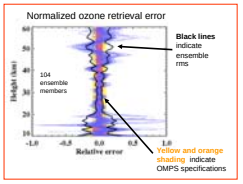
The second set is considered a more rigorous test and uses profiles with Markovian perturbations to the standard profiles.

$$A(z) = A(z-1)\left(1.0 - \frac{1.0}{Z}\right) - fN(0,1)$$
$$O(z) = \overline{O}(z)(1 + A(z))$$
$$T(z) = \overline{T}(z) + 10.0A(z)$$

De-correlation heights:
$Z = 5, 7.5, 10$ km
Noise amplitude:
$f = 0.3$



"We first tested the strategy with state vectors for which ozone number density is the only varying quantity."

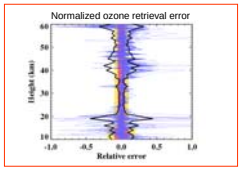


This figure displays normalized retrieval errors for the case that regression training only incorporates ozone variability. There are systematic errors (note the *wavy* pattern) that we hope to learn how to alleviate – but the overall performance is encouraging.

These results are encouraging, but this is a very simplified test

"Then we added temperature variability with similar success"

This figure displays normalized retrieval errors for the case that regression training incorporates both ozone and temperature variability. There are systematic errors are no worse that the previous case with ozone variability only.



Conclusions

- Stray light contamination is a small, linear perturbation (for the OMPS limb profiler)
- Regression is an inexpensive way to handle small, linear perturbations
- Incorporating the regression model into the forward model protects the retrieval from noise amplification
- Requires an accurate characterization of the instrument
- Requires a careful training strategy with many subtle issues to resolve