Development of HYSPLIT inverse modeling technique to improve particulate matter (PM2.5) forecasts in the US

Tianfeng Chai\textsuperscript{1,2}, and Ariel Stein\textsuperscript{2}

\textsuperscript{1}: Cooperative Institute for Climate & Satellites – Maryland, University of Maryland
\textsuperscript{2}: NOAA Air Resources Laboratory
Motivation

- Smoke from wildfires has significant negative impacts on public health;
- Many extreme events are caused by PM$_{2.5}$ emitted from large wildfires;
- The PM$_{2.5}$ forecast quality is greatly hindered by the large uncertainties of the wildfire emission estimates;
- We aim to objectively and optimally estimate the fire sources based on the satellite observations of the fire plumes.

A helicopter makes a water bucket drop as it flies through smoky air while fighting a wildfire that flared up in the late afternoon near Omak, Wash., on Thursday (Aug 27, 2015). (Ted S. Warren/AP)

Photo taken around noon 8/29/2015 from the Highway 26 viewpoint northeast of Prairie City. Smoke from the Canyon Creek Fire was heading in northeast direction. http://oregonsmoke.blogspot.com/

Meanwhile since Friday, more than 1,000 firefighters have struggled with a blaze started by lightning in the Chelan, Wash., area, where at least 49 buildings have been destroyed and authorities have issued evacuations that affect some 3,000 people. http://news.discovery.com/earth/weather-extreme-events/will-more-wildfires-combust-our-health-150828.htm
Smoke forecasts with HYSPLIT: Current status

**NOAA NESDIS HMS  Smoke and fire detection**
Incorporates imagery from NOAA and NASA satellites (GOES-West, GOES-East, Terra/Aqua MODIS, AVHRR on NOAA-15/-17/-18)

Provides fire locations, starting time, durations

**USFS’s BlueSky: Estimate PM$_{2.5}$ emissions and plume rise**

Emission aggregated and assumptions are made to for the forecasting period

Smoke column from the HYSPLIT model (blue) and NESDIS Hazardous Mapping System (orange)

Operational since March, 2007
NOAA NAQFC PM\textsubscript{2.5} forecasts with CMAQ

BlueSky emission terms similar to HYSPLIT added to CMAQ PM\textsubscript{2.5} forecasts since January, 2015
PM$_{2.5}$ forecasts with CMAQ: underestimation
NOAA NESDIS GOES Aerosol/Smoke products (GASP)

The GASP product is a retrieval of the Aerosol Optical Depth (AOD) made from the current GOES West/East visible imagery. Satellite measured aerosol optical depth (AOD) is available at a **30-minute interval and 4 km X 4 km** spatial resolution during the sunlit portion of the day. ([http://www.ssd.noaa.gov/PS/FIRE/GASP/gasp.html](http://www.ssd.noaa.gov/PS/FIRE/GASP/gasp.html))

Can we objectively and optimally estimate the fire sources based on the satellite observations of the fire plumes, instead of using BlueSky to estimate the emissions?
Methodology

• An independent HYSPLIT simulation starting at each HMS fire location with given starting time and duration is run with unit source, at several possible release height to generate a *Transfer Coefficient Matrix (TCM)*.

• Source terms are solved by minimizing a cost function built to mostly measure the differences between model predictions and observations, following a general data assimilation approach.

\[
F = \frac{1}{2} \sum_{i=1}^{N} \sum_{k=1}^{Z} \frac{(q_{ik} - q^b_{ik})^2}{\sigma_{ik}^2} + \frac{1}{2} \sum_{m=1}^{M} \frac{(c^h_m - c^o_m)^2}{\epsilon_m^2} + F_{other}
\]
1. Fukushima source term estimation

Ref: Source term estimation using air concentration measurements and a Lagrangian dispersion model—Experiments with pseudo and real cesium-137, T Chai, R Draxler, A Stein – Atmos. Environ., 2015

2. Volcanic ash application - Kasatochi eruption
NASA WORLDVIEW – Corrected reflectance (true color) Aqua & Terra MODIS

8/17/2015

8/18/2015
Twin experiments

- In twin experiments, known wildfire sources are released to generate smoke plumes. Then pseudo-observations (satellite mass loadings) are generated based on the HYSPLIT simulation results;
- With exact solutions available, the inverse algorithm can be fully evaluated.

9 fire locations, constant releases each day for 2 days from 6Z on 8/17/15, at 1500m or 2000m, MET: gdas1

Hourly, at 0.5°x0.5° resolution
Case 1:

- All 48 hourly non-zero mass loadings are assumed retrieved accurately;
- $\ln(q_{ik})$ as control variables to avoid negative emission results;
- First guess is $10^5$kg/hr at all location/height;
- Background term effect is minimized with extremely large uncertainty given;
- Observations uncertainties at 10% $M + 0.003$kg/m²;
- Minimization stops after cost function reduced to be $10^{-6}F_{init}$

**Case 1 result**

**Actual sources**
**Source term error statistics**

<table>
<thead>
<tr>
<th>Source term</th>
<th>MAE (kg/hr)</th>
<th>Normalized MAE</th>
<th>RMSE (kg/hr)</th>
<th>Normalized RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 1</td>
<td>534.9</td>
<td>0.77%</td>
<td>841.4</td>
<td>1.21%</td>
</tr>
<tr>
<td>Day 2</td>
<td>1760.5</td>
<td>2.53%</td>
<td>3332.5</td>
<td>4.78%</td>
</tr>
<tr>
<td><strong>Case 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 1</td>
<td>1985.8</td>
<td>2.85%</td>
<td>3310.2</td>
<td>4.75%</td>
</tr>
<tr>
<td>Day 2</td>
<td>1393.0</td>
<td>2.00%</td>
<td>2943.2</td>
<td>4.22%</td>
</tr>
</tbody>
</table>

---

48 hr observations → 24 hr observations

Case 2: Remove 1st day observations from Case 1
### Spatial coverage

<table>
<thead>
<tr>
<th>Case</th>
<th>Day 2 observations, w/o Region A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 3</td>
<td>Day 2 observations, w/o Region A</td>
</tr>
<tr>
<td>Case 4</td>
<td>Day 2 observations, w/o Region B</td>
</tr>
<tr>
<td>Case 5</td>
<td>Day 2 observations, w/o Region C</td>
</tr>
</tbody>
</table>

### Source term error statistics

<table>
<thead>
<tr>
<th></th>
<th>Source term</th>
<th>MAE (kg/hr)</th>
<th>Normalized MAE</th>
<th>RMSE (kg/hr)</th>
<th>Normalized RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 3</td>
<td>Day 1</td>
<td>606.4</td>
<td>0.87%</td>
<td>1156.3</td>
<td>1.66%</td>
</tr>
<tr>
<td></td>
<td>Day 2</td>
<td>301.2</td>
<td>0.43%</td>
<td>573.4</td>
<td>0.82%</td>
</tr>
<tr>
<td>Case 4</td>
<td>Day 1</td>
<td>23834.6</td>
<td>34.21%</td>
<td>32157.9</td>
<td>46.16%</td>
</tr>
<tr>
<td></td>
<td>Day 2</td>
<td>66177.5</td>
<td>94.99%</td>
<td>78653.3</td>
<td>112.90%</td>
</tr>
<tr>
<td>Case 5</td>
<td>Day 1</td>
<td>3974.9</td>
<td>5.71%</td>
<td>8803.3</td>
<td>12.64%</td>
</tr>
<tr>
<td></td>
<td>Day 2</td>
<td>3400.6</td>
<td>4.88%</td>
<td>10663.2</td>
<td>15.31%</td>
</tr>
</tbody>
</table>
Observation errors

- 2\textsuperscript{nd} day observations only;
- without spatial blocking;
- Gaussian-distributed errors are added to pseudo observations.

### Source term error statistics

<table>
<thead>
<tr>
<th>Case</th>
<th>Source term</th>
<th>MAE (kg/hr)</th>
<th>Normalized MAE</th>
<th>RMSE (kg/hr)</th>
<th>Normalized RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Day 1</td>
<td>1448.6</td>
<td>2.08%</td>
<td>2259.7</td>
<td>3.24%</td>
</tr>
<tr>
<td></td>
<td>Day 2</td>
<td>4418.7</td>
<td>6.34%</td>
<td>11121.8</td>
<td>15.96%</td>
</tr>
<tr>
<td>7</td>
<td>Day 1</td>
<td>2105.8</td>
<td>3.02%</td>
<td>3954.7</td>
<td>5.68%</td>
</tr>
<tr>
<td>10%</td>
<td>Day 2</td>
<td>8567.9</td>
<td>12.30%</td>
<td>22884.4</td>
<td>32.84%</td>
</tr>
<tr>
<td>20%</td>
<td>Day 1</td>
<td>6227.6</td>
<td>8.94%</td>
<td>12047.1</td>
<td>17.29%</td>
</tr>
<tr>
<td>50%</td>
<td>Day 2</td>
<td>22034.5</td>
<td>31.63%</td>
<td>67298.2</td>
<td>96.60%</td>
</tr>
<tr>
<td>9</td>
<td>Day 1</td>
<td>10104.2</td>
<td>14.50%</td>
<td>19759.9</td>
<td>28.36%</td>
</tr>
<tr>
<td>100%</td>
<td>Day 2</td>
<td>34560.9</td>
<td>49.61%</td>
<td>131203.9</td>
<td>188.33%</td>
</tr>
</tbody>
</table>
Summary

• Wildfire emission inversion system has been built based on HYSPLIT model, its TCM, and a cost function;
• With pseudo observations generated using HYSPLIT model simulations (twin experiments), true emissions (release height and emission rate) can be recovered;
• First day emission sources are easier to estimate than the second day emissions;
• Spatial coverage of satellite retrieval is important;
• 100% satellite retrieval errors resulted in 17.3%/28.4% normalized MAE/RMSE errors of first day emission rates;
• The system will be further tested before implementation with real observations.
Thank you!
Twin experiments