ICAP-Ensemble Meeting

Climatological-Persistence Models (CLIPERs)

Jeffrey Reid[1], Jianglong Zhang[2], Buck Sampson[1], James Hansen[1], and **Walter Sessions[3]**

[1]NRL Monterey

[2]University of North Dakota

[3]CSC

May, 2011

jeffrey.reid@nrlmry.navy.mil/ 1 (831) 656-4725

http://www.nrlmry.navy.mil/aerosol/

http://www.nrlmry.navy.mil/flambe/

Overview

- **Why a CLIPER?**
- **CLIPER basics**
- **Single site examples**
- **Global findings**
- **The way ahead for ICAP?**

Why a CLIPER?

- The last ICAP meeting highlighted the need to uniform baselines to assess forecast skill be adopted across centers.
- Comparison of forecasts to persistence is an unappealing baseline: The environment is always changing.
- Climatology is an equally unappealing baseline: We are NWP after all.
- So combine them both- Construct a “parametric forecast” based on combining current observations (persistence) and climatologies… the “CLIPER”
- Bottom line: We need an agreed upon ICAP baseline to start cooperation in verification!

CLIPER Basics

- CLIPERs may be new to aerosol or atmospheric constituent forecasting, but are commonly used in NWP and seasonal forecasting.
- CLIPERs often take the role as anything from a statistical baseline, to a full blown parametric model. For example, there are solid ENSO CLIPERs that often beat many forecast models.
- Generally, the CLIPER constructed with non-NWP input (e.g., observations and best available climatologies).
- In its simplest form, a CLIPER is an exponential decay from observation to climatology:

$$y(t) = C_o + C_1 \cdot e^{-t/\tau_y}$$

y is the forecast variable at time t
C_o is climatology
C_1 is perturbation from climatology
τ_y is the e-fold decay of y

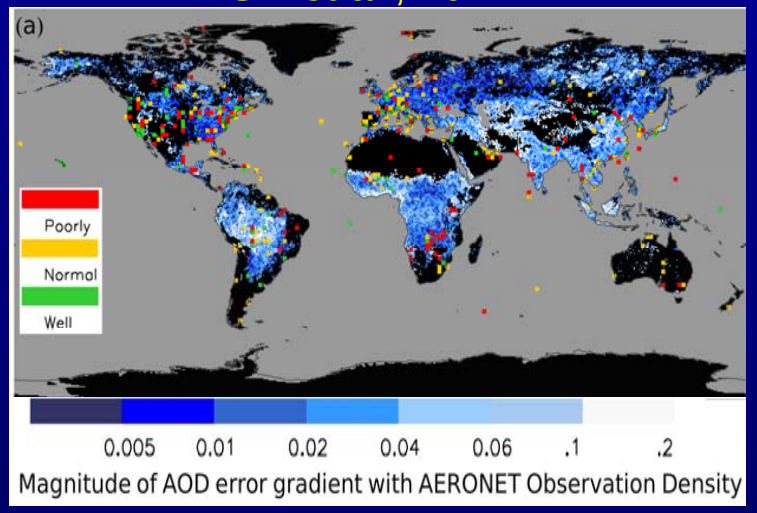
Things to keep in mind with a baseline CLIPER

- REMEMBER: The CLIPER is a statistical baseline, not a forecast model. Well, at least for us…..
- Even so, CLIPERs are often referred to as the "No-skill model." Generally in *mature* modeling environments, if you can't beat the CLIPER, then your model is said to be without skill. This is not really true because often the forecast has qualitative value (e.g., you will get whacked by a dust storm in 4 days).
- What should you expect? At some critical point all models loose to climatology. Similarly, for good observations, the model will loose to the CLIPER at t=0. The object is to increase the ground in between.
- If RMSE is the norm, it may beat all of the models for a while.
- There is plenty of room to ratchet things up on the modelers (add a tendency term, smart climatologies, etc…). But lets be nice for now.
- The first questions in developing a CLIPER "How good are your observations and climatology?" A CLIPER has to verify too….

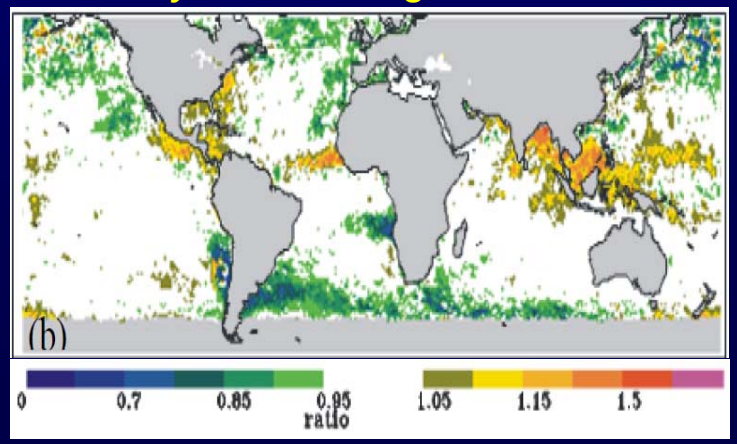
Data Considerations

- The AOT CLIPER can be global, based on satellite data, or at specific AERONET sites.
- The benefit of AERONET is it is the best available data with measurement uncertainties much lower than the models. The problem with AERONET is representativeness.
- For satellite data, we get global coverage, but climatologies have big errors.
- In both cases we expect contextual or sampling biases in the climatologies as well as when the CLIPER can be run. For both AERONET and satellite data there is some clear sky bias. For satellite data , there are also regional biases as well as magnitude biases.
- For regional air quality CLIPERs, it is easier to deal wit these issues.

Bias Gradient w/ AERONET locations
Shi et al., 2011



Clear sky bias, Zhang and Reid, 2009



Single Site Simple Example

Autocorrelation times from NAAPS reanalysis

	GSFC Garden Variety Pollution	Rome Pollution & Dust	Shirahama Complex Mixtures	Cape Verde Dust	Alta Floresta Seasonal Burning
Annual	4.25	1.75	1.5	2.75	33

Winter	0.5	1	1	2.25	2
Spring	1	0.5	0.5	3.5	1
Summer	1.5	1.5	1.75	1.25	8
Fall	1.5	2	1	2	10

But each location has its own spin, e.g.:

Rome: Summertime has a long t for pollution (4 days), but short for dust(2).

Shirahama: Spring pollution 0.75 days, dust (1.5)

AERONET? Raw Alta Floresta is 5 days…

First Go at Global CLIPER

- Calculated annual seaonal and monthly values.
- Performed Autocorrelation every 1x1 degree with NAAPS analysis.
- Kept the simple form of:

$$y(t) = C_o + C_1 \cdot e^{-t/\tau_y}$$

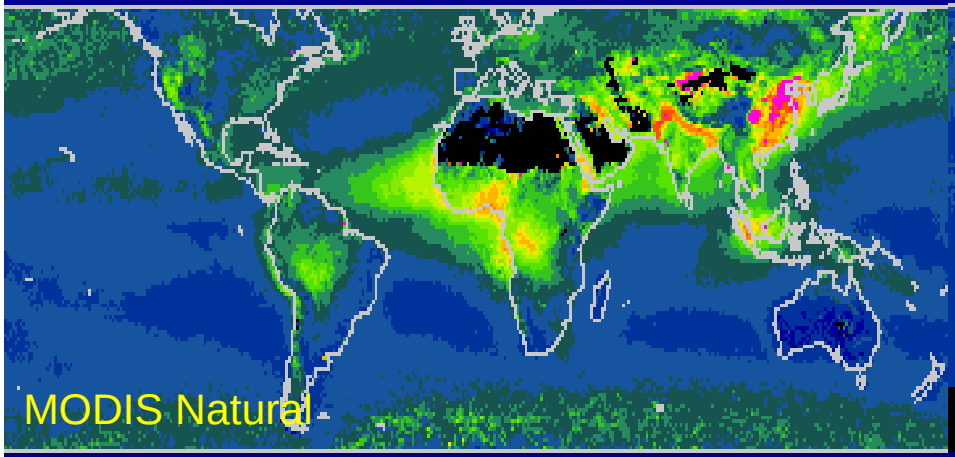
- Decided to stop short of a tendency term
- Now we need a climatology….

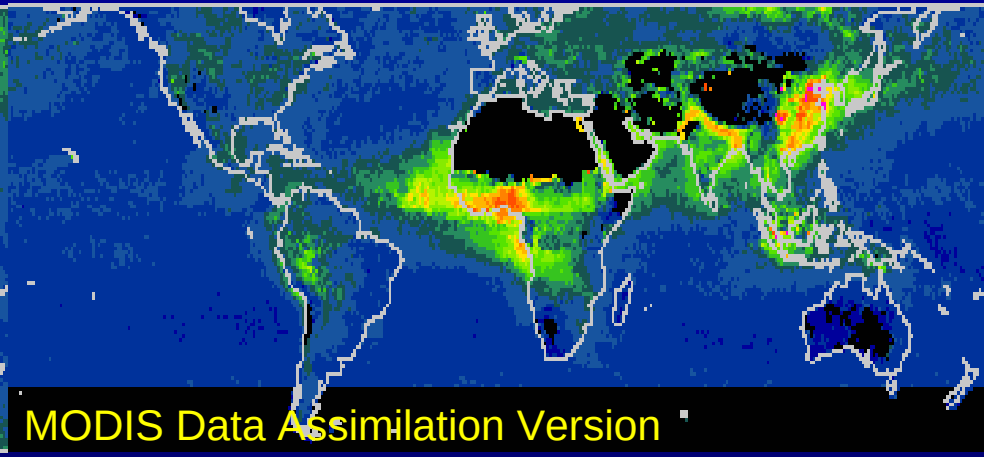
CLIPER Global AOT Climatology?

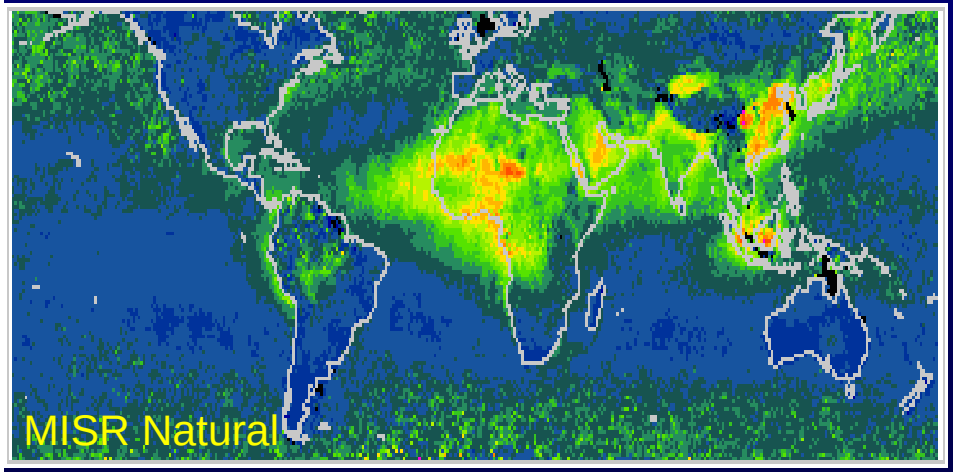
Sorry, even the climate folks can't get it right.
We need to construct our own.

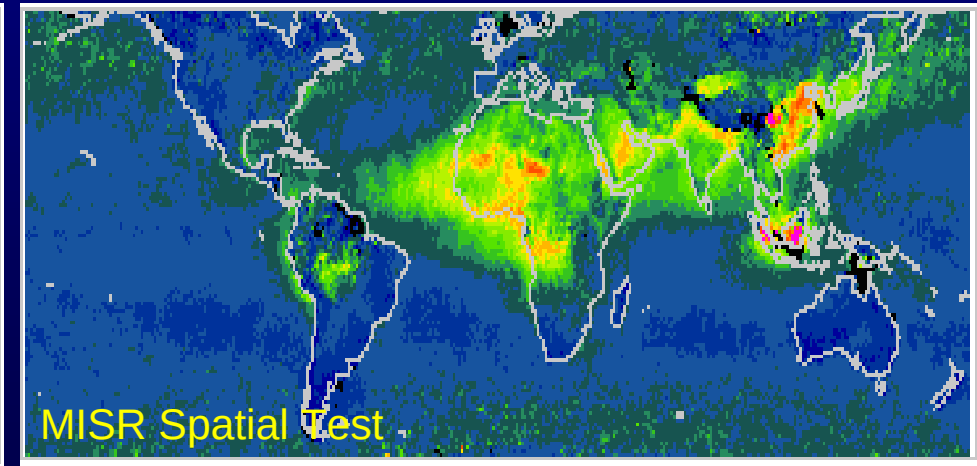
- We want to be independent from NWP models and the data they commonly assimilate.
- Plus, we want reasonable coverage with as little subjective interference from us.
- MISR is a pretty good start for a climatological baseline. But, there are three issues:
  - MISR has a well known 20-30% low bias for AOT>0.5
  - MISR retrievals stop at a lower AOT than MODIS. Hard to get data above 1.
  - Like MODIS, MISR has a cloud mask problem, particularly in the southern 40's
- But, for a climatology these biases are probably not too bad over most of the globe. In the future, we may fuse MISR with MODIS at some latitudes.

MODIS and MISR Climatologies 2006









AOT (Green)

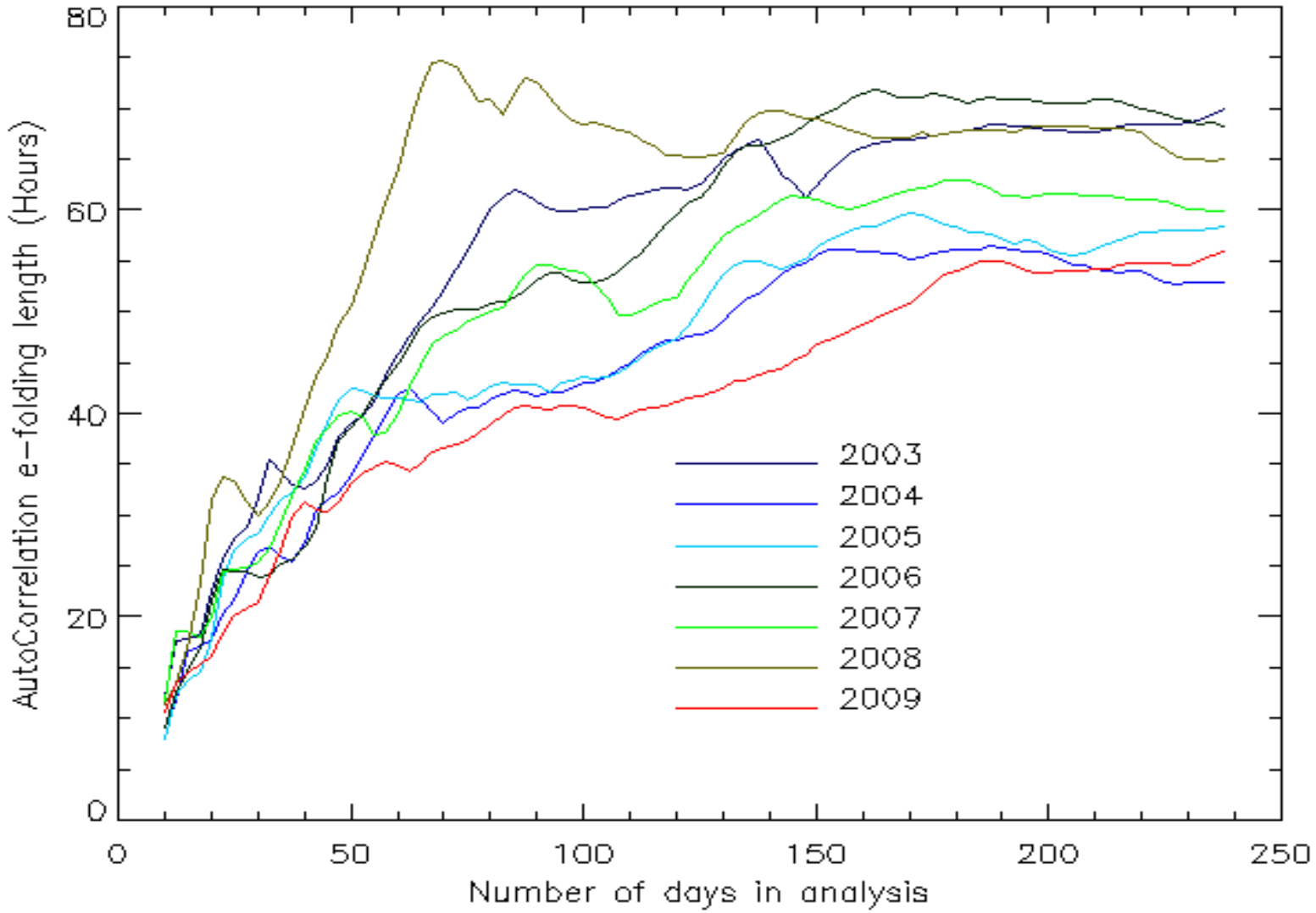
Where do we objectively get τ_y?
Lots of ways, all different mathematically

- Autocorrelation: Simple statistic based on observation or reanalysis. But, this is sensitive to sample size
- Relaxation time: Progressively average observation O from t(0) to t(x) for each observation. Define τ_y then when O is within e^{-1}. This is probably an overestimate τ_y
- Relaxation time-B: Group by O amplitude, and local climatology. Then evaluate when $\tau_i(\overline{x})$ reaches within e^{-1}.of the climatology. This is probably an underestimate τ_y
- Spectral Analysis: Are we over analyzing this?
- Brute force/Moores law: For each bin, optimize RMSE for an ensemble of τ_y.

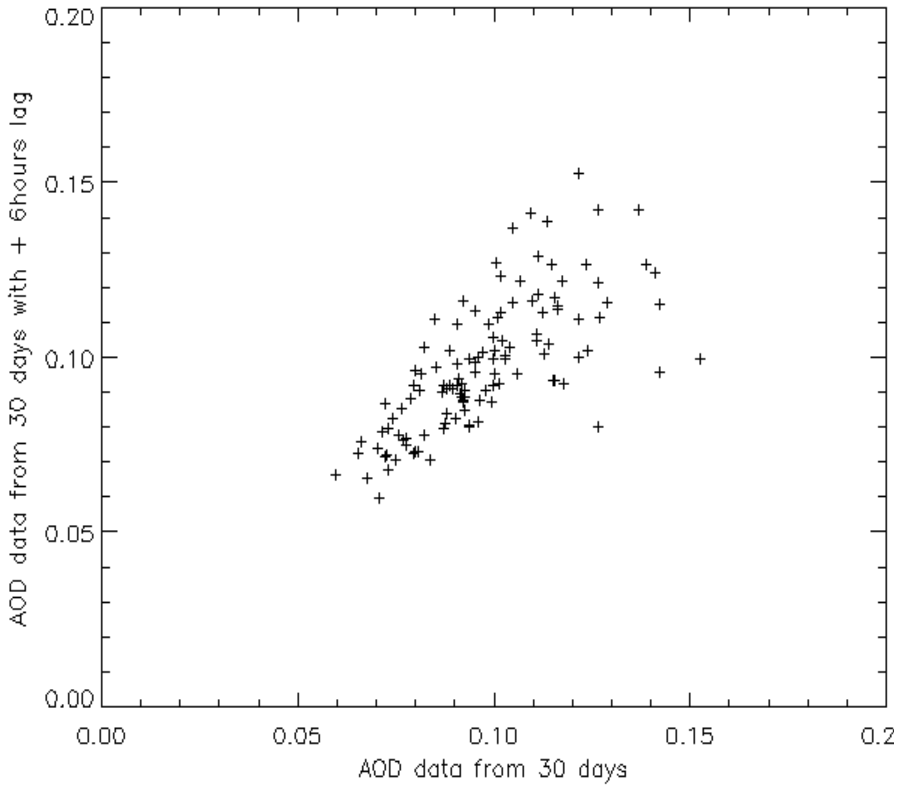
- But, remember we need to verify the CLIPER too!

Autocorrelation sensitivities

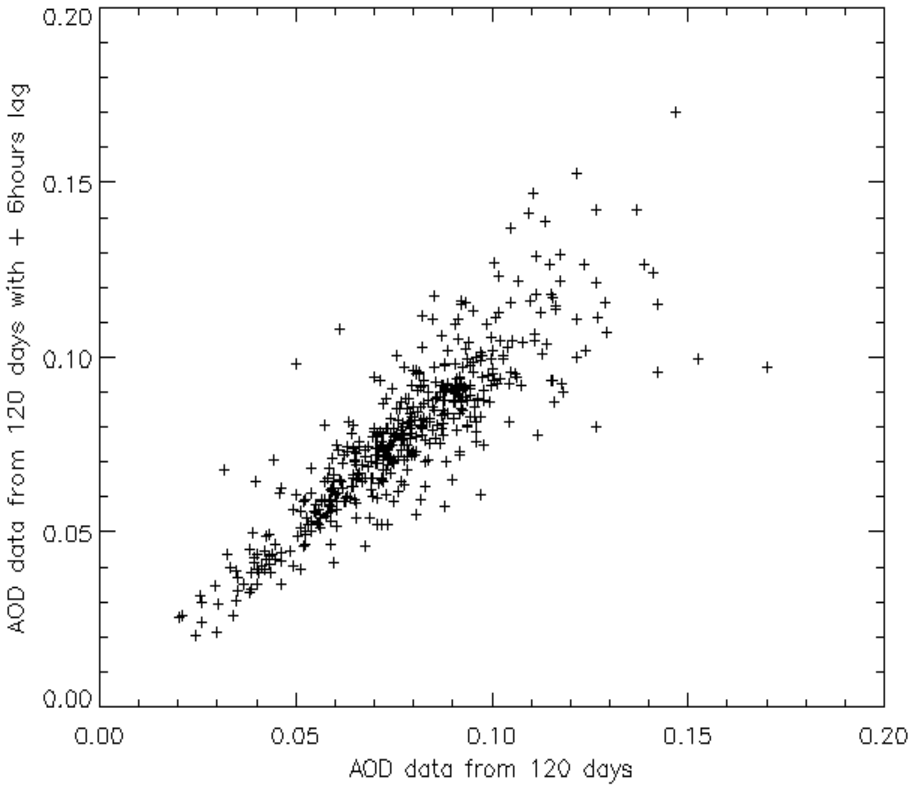
Autocorrelation increases with analysis time as we expand dynamic range and include longer wavelength features.



Example: Scatter plot of AOD data vs. 6 hour lag of AOD data from NAAPS (random sample from May 01, 2007)



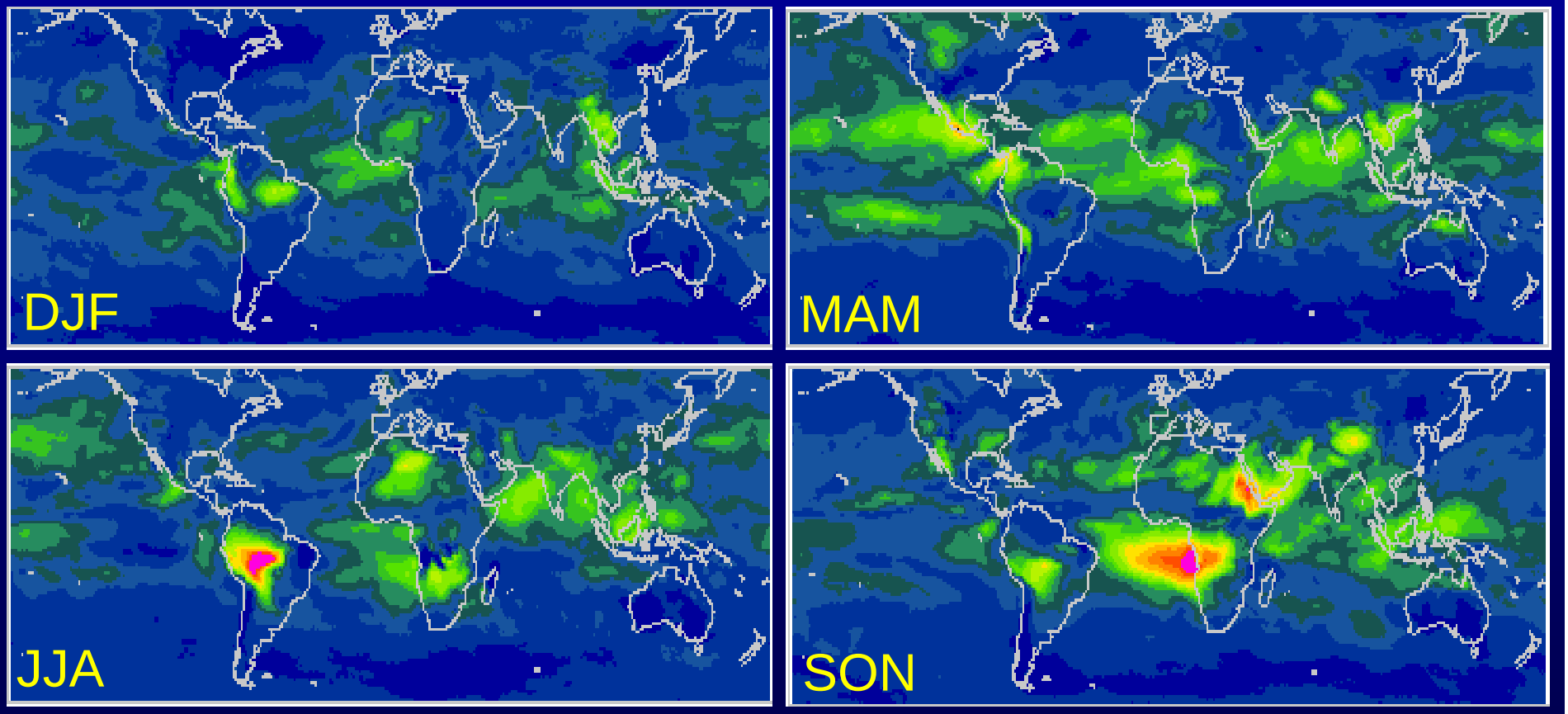
30 day's data were included in the analysis



120 day's data were included in the analysis

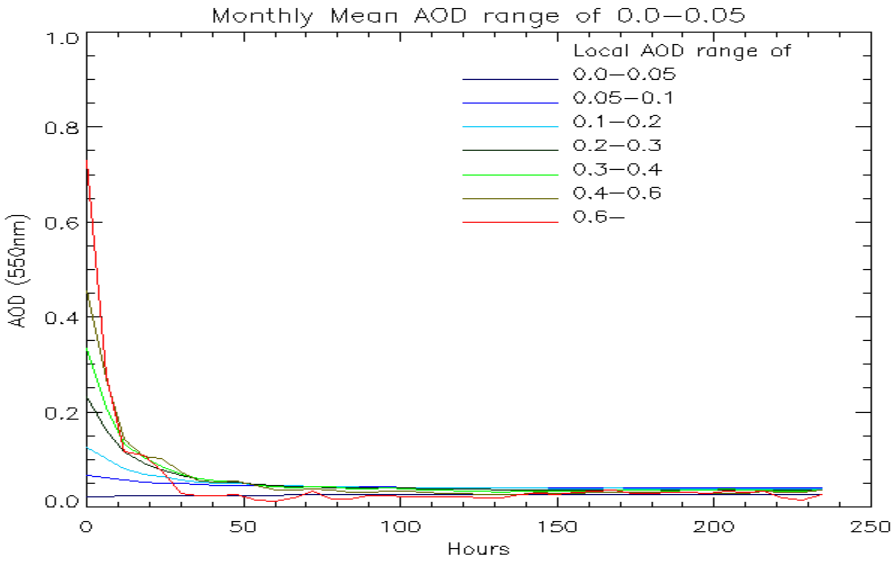
Seasonal Global Autocorrelation Times

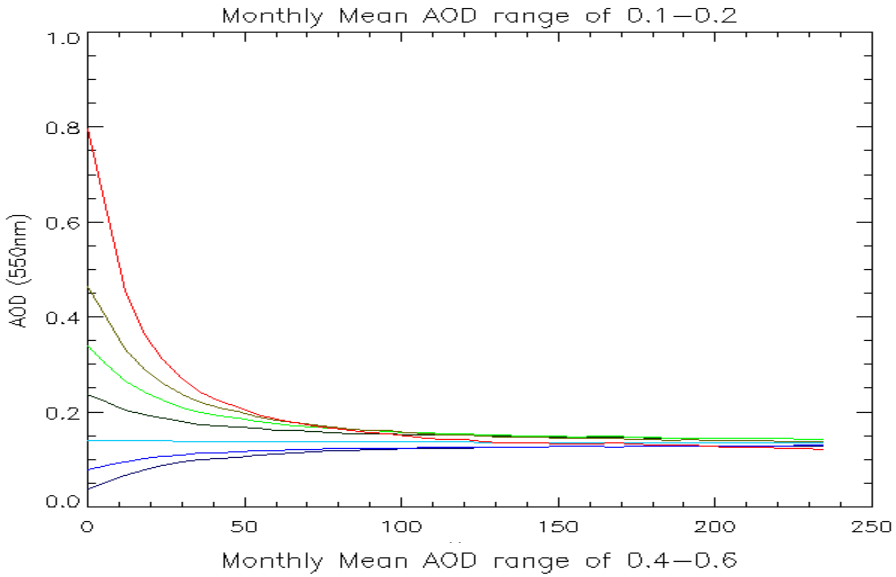
Calculated each year, then averaged the years

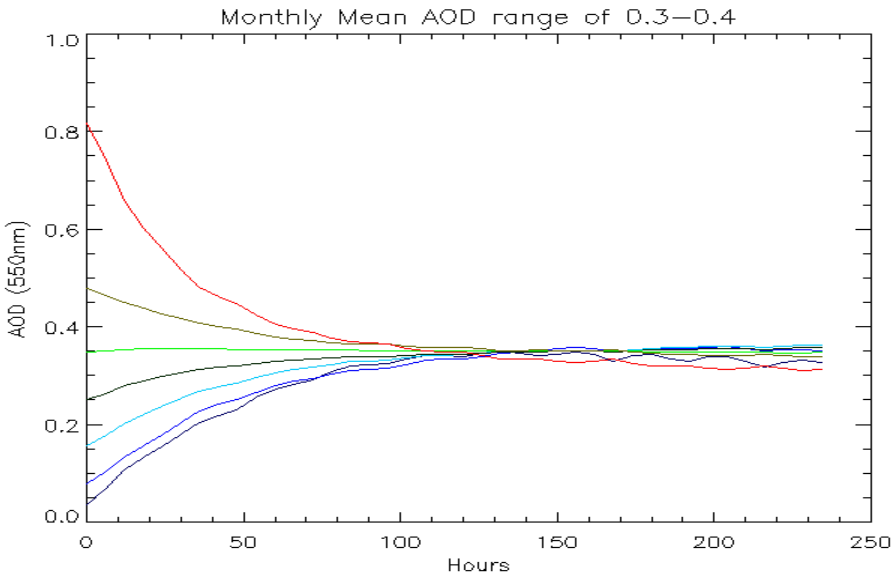


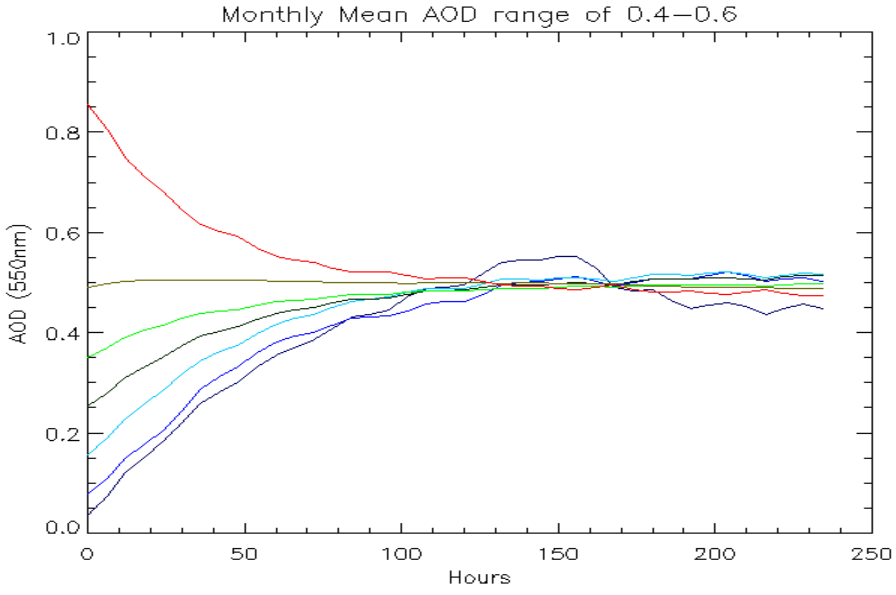
Autocorrelation Time (hours)

τ_y from Relaxation Time
Higher Climatological AOT, longer decay





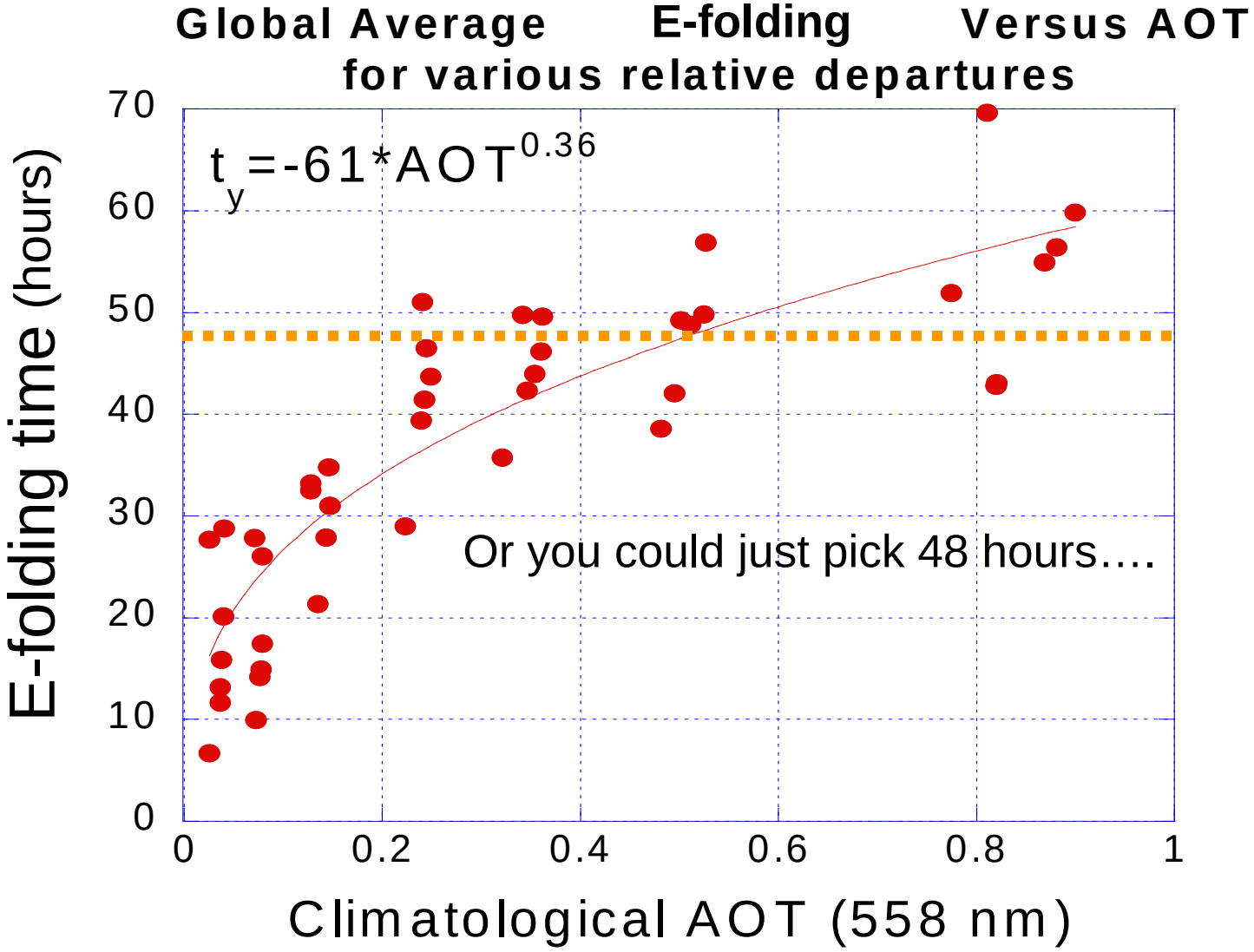




Average convergence times:

Higher the mean AOT, the longer the time.

Clean areas don't have long lasting departures.



The way Ahead for ICAP?

- This CLIPER is simple, but still requires a little bit of thought.
- If everyone is in agreement, as a preliminary test we should code this up in Python for ICAP distribution.
- For now I suggest we stick with corrected MISR, although we may want to blend in MODIS in the high latitudes. Open for suggestion…
- There is not enough contiguous data to do a data only autocorrelation number, although we have isolated segments. Options are:

  a) Create a multi-model analysis and use that for the study, or average the different model values (Easy and transparent).

  b) Base on AERONET and then use the multi-model ensemble to spread the information (Complex and convoluted)
- I suggest a).
- Everything we show here should be paralleled for surface (PM) and layers (lidar). We are going to start a spectral analysis study.