scoring

model performance

Stefan Kinne
MPI-Meteorology, Hamburg

initial thoughts (1)

- why scoring ?
  - deficiencies in current model assessments
    - ***subjective:*** model performance evaluations are often tied to sub-sets (of often favorable) data … also visual inspection of plots can be misleading
    - ***limited:*** model evaluation are often tied to one or at best a few variables
  - need to ‘quantify’ model performance
    - ability to track impacts of changes in modeling
    - ability to diagnose deficiencies in modeling
    - … while satisfying demand for a ‘single score’

initial thoughts (2)

- how to score?
  - not every variable is meaningful
    - properties could be result of off-setting errors
      - total AOD value vs component AOD values
      - annual average vs monthly averages vs daily
      - global average vs regional average vs local
  - not every statistical tool is meaningful
    - statistical applications could be misleading
      - impact of outliers (… on average and std. dev)
      - variability only matters, if there IS variability
    - different stat. applications could be redundant

initial thoughts (3)

… due to a misrepresentation … there is also **a danger in (combined single) scores**!

thus, several aspects need to be addressed:

- what are the properties to be tested ?
- how accurate is the (data) reference?
- what are the (smallest) relevant scales?
- what are the relevant statistical methods?
  - bias *(is it larger or smaller?)*
  - spatial variability *(test spatial pattern)*
  - temporal variability *(test temporal change)*

bias

- use value ranks !
  - apply to all valid data-pairs
  - throw all data in a single array
  - rank all elements by value
  - sum all ranks associated with data D (=13)
  - sum all ranks associated with reference R (=8)
  - determine bias: bias= D-R/(D+R) (= 5/21)
    - (bias can be positive or negative)
  - determine the bias score S_B = 1-abs(bias)
  - reduce the bias error "abs(bias)" , if the range of data is smaller than the average

Ref./Data
0.2 / 0.3
0.4 / 0.6
0.5 / 0.7

→

0.2	1.
0.3	2.
0.4	3.
0.5	4.
0.6	5.
0.7	6.

variability

- use rank correlation
  - apply to all valid data-pairs
  - determine Spearman correlation coefficient C

  - determine the variability score $S_V = 1 - (C+1)/2$
  - reduce the variability error "(C+1)/2 " , if the range of data is smaller than the average

  - apply for spatial variability
  - apply for temporal variability

overall score

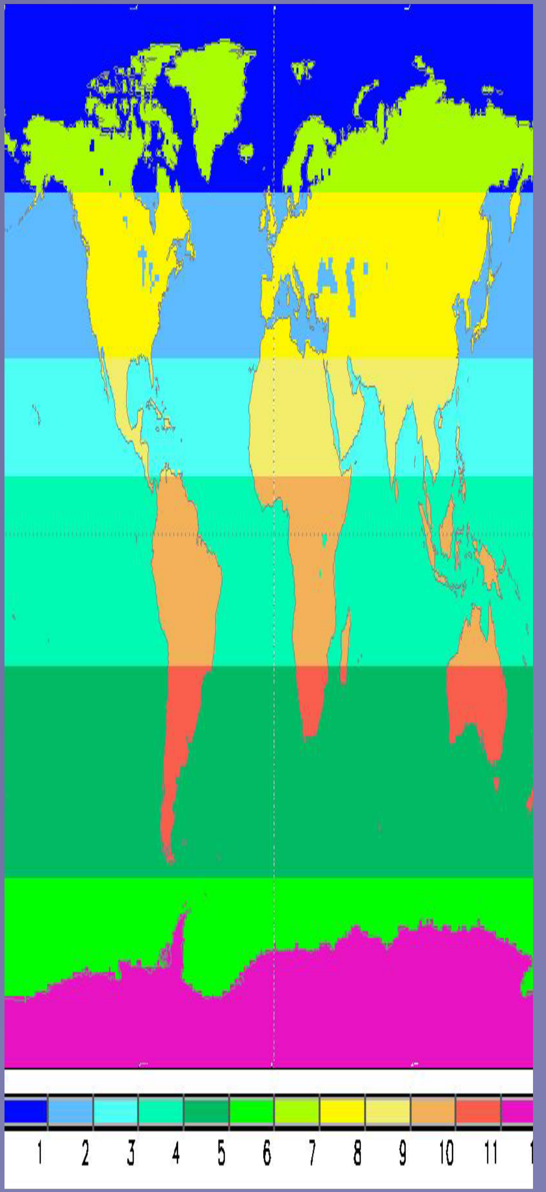
- as bias and variability (sub-) scores *[1-0 range]* are better the closer to 1 *('1.0' is perfect)* ...

- sub-scores are multiplied $S = S_B * S_{V,s} * S_{V,t}$

- scoring procedure
  - pick regions
  - at the smallest temporal scale for each region
    - determine bias and spatial variability score
  - advance to larger temporal (up to annual) scales
    - combine (add) bias and variability scores
    - determine a temporal score (with regional median)
  - advance to larger spatial (up to global) scale
    - weigh according to surface area of region

an example

- stratify globe into regions
  - 6 land and 6 ocean zones ⇨
- decide on resolution
  - 2D, 1x1deg lat/lon
- pick the low temporal scale
  - monthly data
- select a property
  - total AOD
- select a reference
  - AERONET gridded data

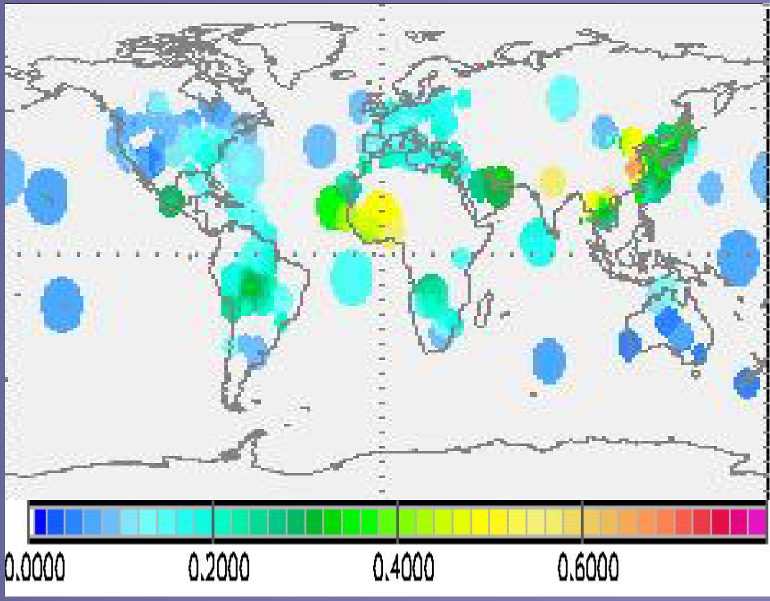


the reference

… for 1x1 gridded data

- **all sites are scored on**
  - **Quality**
  - **Regional representation**
    - **site-data of good sites can be extended to adjacent pixels (dot size)**
    - **more ‘objective’ satellite data-based scores for the regional representation are desirable.**

… note, there are gridded data for all AERONET properties



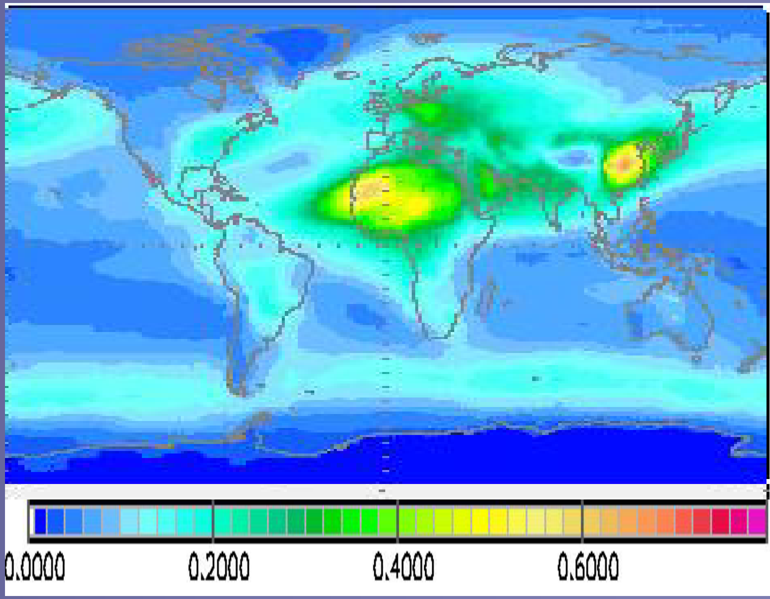
annual AOD averages based on all available CIMEL sunphotometer samples (1996-2007)

(here dots are expanded for better visibility, as data are more sparse)

the data

- all (18) model* data are interpolated to 1x1 deg
- the model median ⇨
- let the game begin …

*match, impact, lsce, loa, gfdl, sprintars, gocart, mirage, cam, oslo-ctm, ncar, ulaq, giss, ham-echam5, grantour, oslo-gcm, echam4, canada



annual model median based on median com-posites for monthly Fields based on 18 different component aerosol module results

the essence

- the ‘median model’ scores better than any individual model on an global annual basis
  - global: 0.63 (models: 0.60 to 0.45) neg.
  - ocean: 0.64 (models: 0.62 to 0.39) neg.
  - contin: 0.62 (models: 0.61 to 0.44) neg.
- the ‘median model’ is NEVER the best on a regional or monthly basis
- global annual ‘median model’ sub-scores indicate issues with spatial variability (.75) and seasonality (.85); bias error is small.

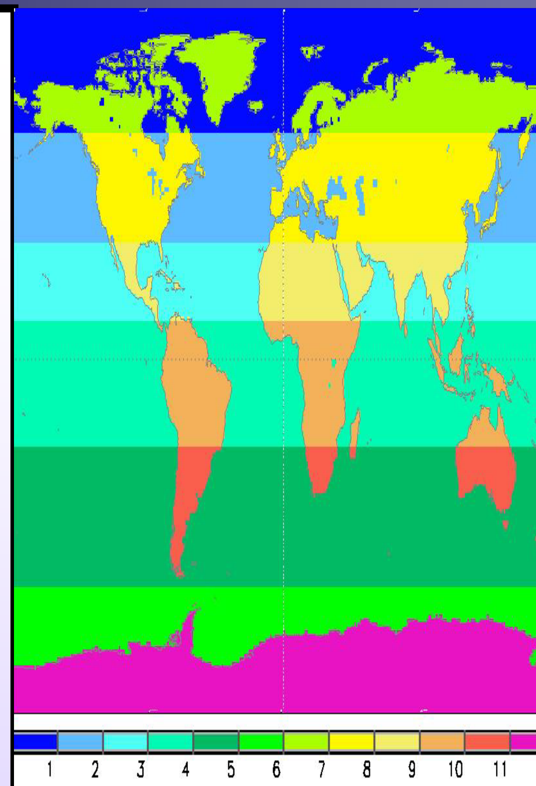
the ‘oscars’

- the 8 better scoring models underestimate AOD … compared to AERONET
- best overall scores (score > 0.56) are by
  - match, impact, lsce, loa, gfdl and sprintars
- best seasonal scores by
  - ocean: echam4
  - continent: cam
- best spatial variability scores by
  - ocean: sprintars
  - continent: impact, gocart

more regional scoring detail

- .36 (.56 impact)
- .67 (.75 mirage)
- .79 (.80 cam)
- .56 (.70 sprinta)
- .35 (.48 match)
- ***ocean scores***

median (best model)



- .30 (.52, ncar)
- .71 (.74 oslo, t)
- .58 (.69, cam)
- .53 (.68, cam)
- .52 (.65, gfdl)
- ***land scores***

median (best model)

scores are far from perfect *(even excl. polar regions)*
dust and biomass regions have major deductions
northern mid-latitudes score best
southern oceans score poor *(few data though)*

summary

- this it just one of many ways to score
  - it seems to work though, as score for all sub-scores improved for MODIS coll 5 vs coll 4
- it provides one overall score … while still providing scoring at spatial and temporal sub-scales (for detailed diagnostics)
- global and sub-scores of many other properties (beyond AOD) could be combined for more adequate scoring
  - good total AOD scores may have resulted from ‘tuning’

back to the questions

- what are observables to be tested ?
  - total AOD (AERONET)
  - fine-mode ($r<0.5\mu m$) AOD (AERONET)
  - absorption AOD (AERONET)
  - AOD above 678 (?) mb (Lidar-networks)
- how accurate is the (data) reference?
  - reduce error weight with incr. uncertainty
- what are the (smallest) relevant scales?
  - 1x1, monthly sufficient …or daily?
- what are the relevant statistical methods?

New Data

AERONET news

- AERONET data have been reprocessed
  - monthly/daily data: at ftp ftp-projects.zmaw.de
    cd aerocom/aeronet/grd_stat808
  - also statistics … just for satellite overpasses
  - special PDF statistics have been developed for comparisons to NOAA ground obs. for
    - single scattering albedo
    - Angstrom parameter

    investigated at
  - Bondville, Illinois
  - Cart site, Oklahoma
  - Mauna Loa, Hawaii (high altitude)

column vs ground

- SSA
  - near surface fall/winter values at cont. sites are less absorbing, if absorption is strongest
  - near surface absorption at M. Loa is stronger
  - column absorption at cont. sites is stronger at intermediate AOD values
- Angstrom
  - near surface particles are much smaller at cont. sites, especially in winter and spring
  - better agreement at Mauna Loa but much more diversity at surface compared to column

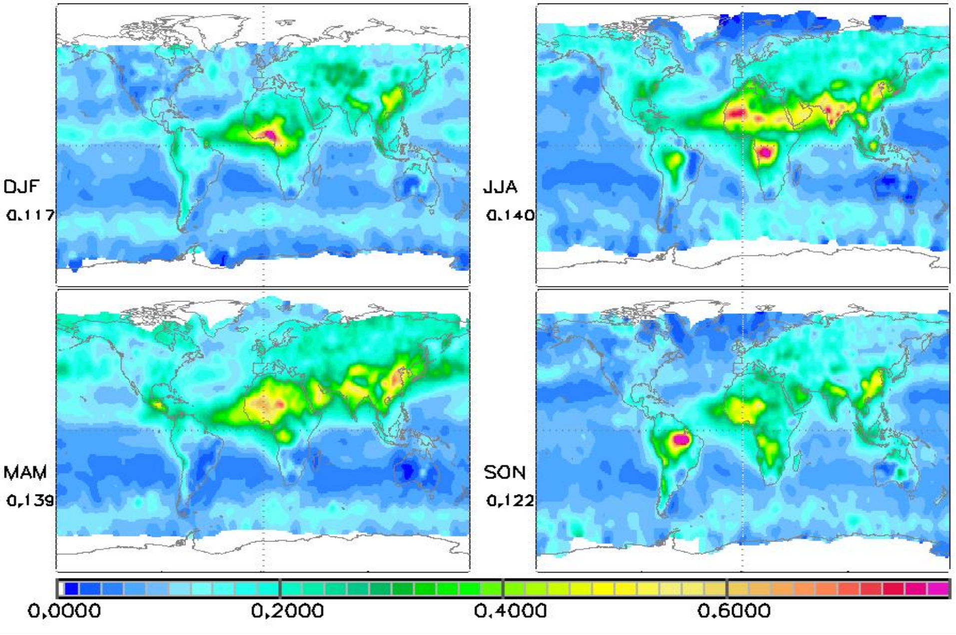
Lidar data

- EARLINET, NIES, MPL-net … network data
  - examining and scoring the aerosol vertical distribution is still an underexplored activity
  - suggest a rather simple diagnostics (e.g. AOD above vs below a threshold altitude

satellite data

- aerosol
  - updates to MODIS, MISR, POLDER, OMI …
  - GlobAER products (G.Thomas)
  - development of a new satellite composite based solely on remote sensing data
    - seasonal AOD fields

ftp ftp-projects.zmaw.de
cd aerocom/climatology/
satellite_aod/gocompo03.nc



- clouds
  - Calispo / Cloudsat

request to all

- the JPL CloudSat group requests your input: what of their products at what vert. resolution … would help us ?
  - I requested global and annual … but what properties to pick: overlap, LWC, IWC, … ?
  - a good treatment of clouds is an essential aspect to get aerosol processing right!

- note, Jay Mace already has provided a 3/6/10/14/20km CALIPSO/CloudSat product on cloud statistics for addressing cloud (cover) overlap

extras

data of interest

- aerosol column properties
  - attenuation *(total direct loss* ⇨ *AOD)*
  - absorption *(loss frac. not absorbed* ⇨ *SSA)*
  - size *(if scattered … how?* ⇨ *P, g)*
- vertical distribution
  - aerosol *AND* clouds
- environmental properties
  - clouds *(impact on aerosol)*
    - …surface properties, ambient water vapor
- anthropogenic fraction

column attenuation

- aerosol optical depth
  - a component mixture
    - sulfate
    - organic carbon
    - black carbon (soot)
    - sea-salt
    - dust
  - component weights differs by region
  - ~0.13 is the global ann. average at λ=.55μm

.

annual maps (.55μm) ⇨

total

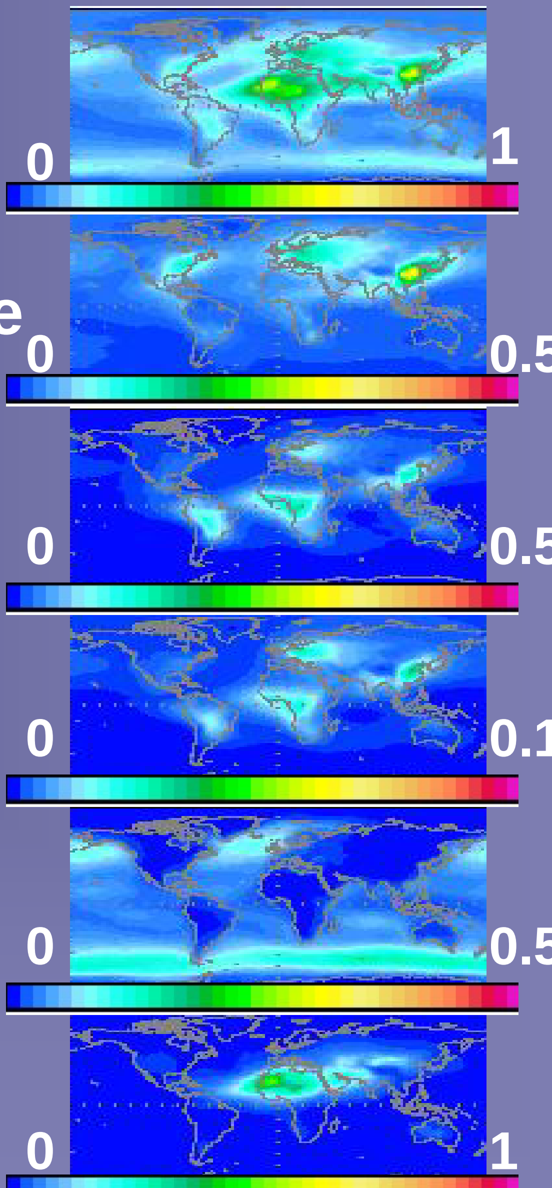
sulfate

org.c

soot

salt

dust

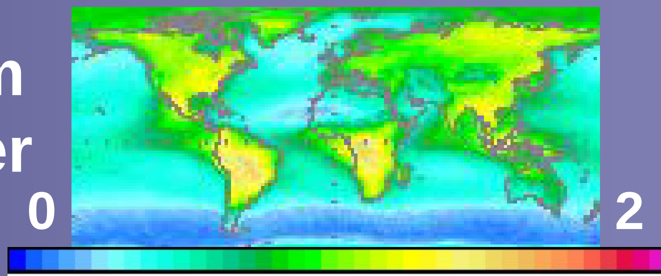


column absorption and size

- **size**
  - **AOD** spec dep → **AP**
    - **AP<1: larger sizes**
    - **AP>1: smaller sizes**
  - **fine mode (r <.5μm) AOD fraction**
- **absorption**
  - **single scatt alb**
    - **absorb potential**
  - **absorption- AOD**
    - **eff. absorption**

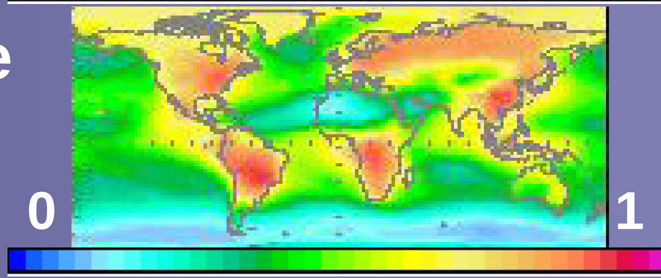
ann. maps (.55μm) ⇨

Angstrom parameter



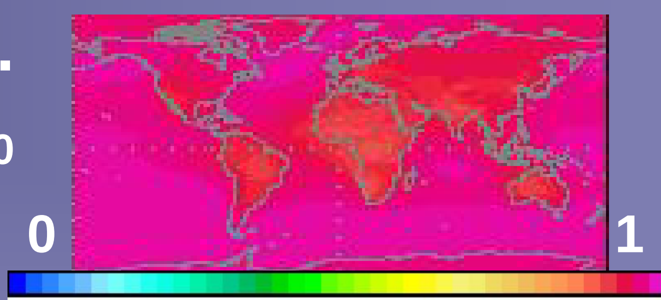
0 2

fine-mode fraction



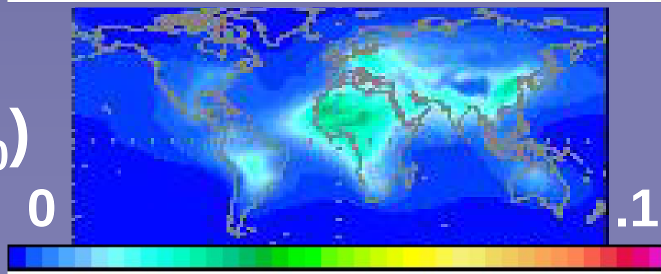
0 1

single sc. albedo ω_0



0 1

abs- AOD aod*(1- ω_0)



0 .1

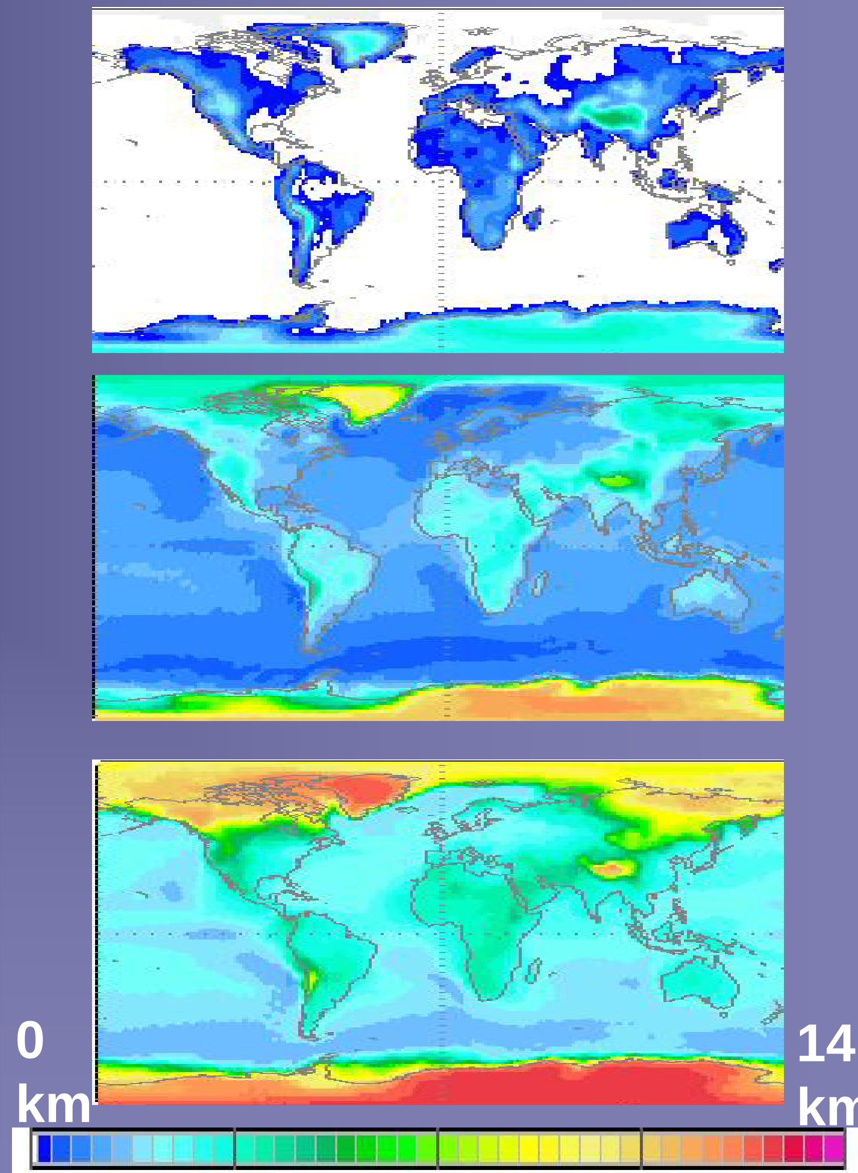
aerosol altitude

- **0% PDF 0.3 km**

  surface altitude
  - **100 % AOD above**

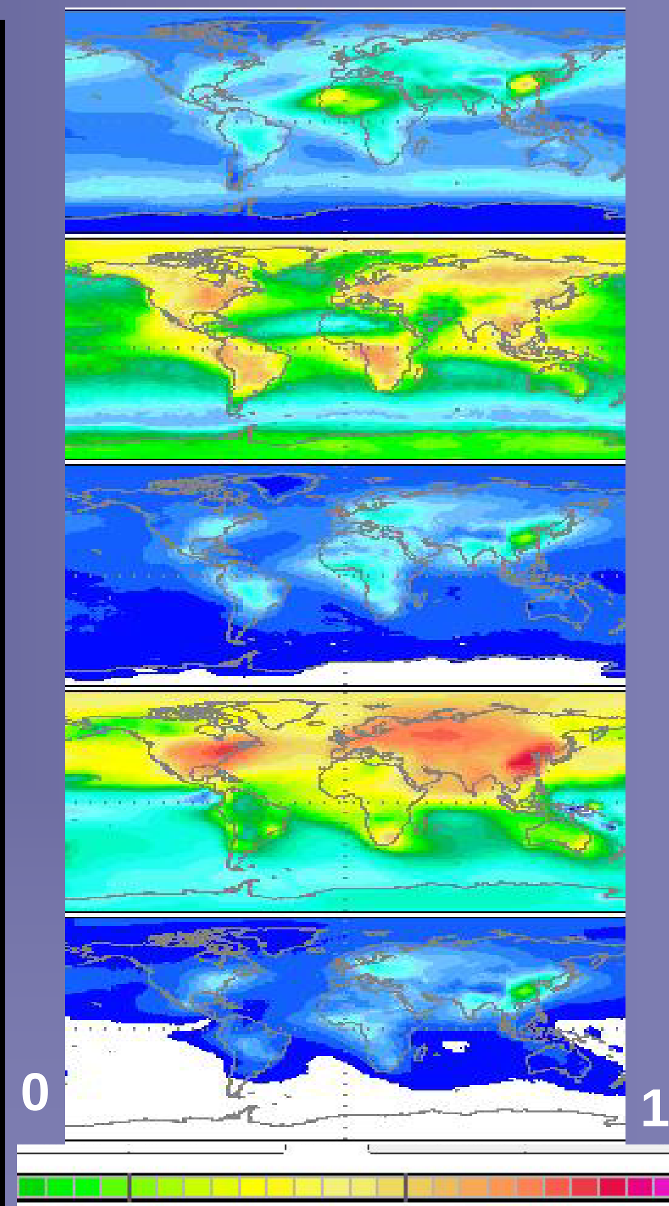
- **50% PDF 1.7 km**
  - **50 % AOD below**
  - **50 % AOD above**

- **90% PDF 3.4 km**
  - **10 % AOD above**
  - **90 % AOD below**



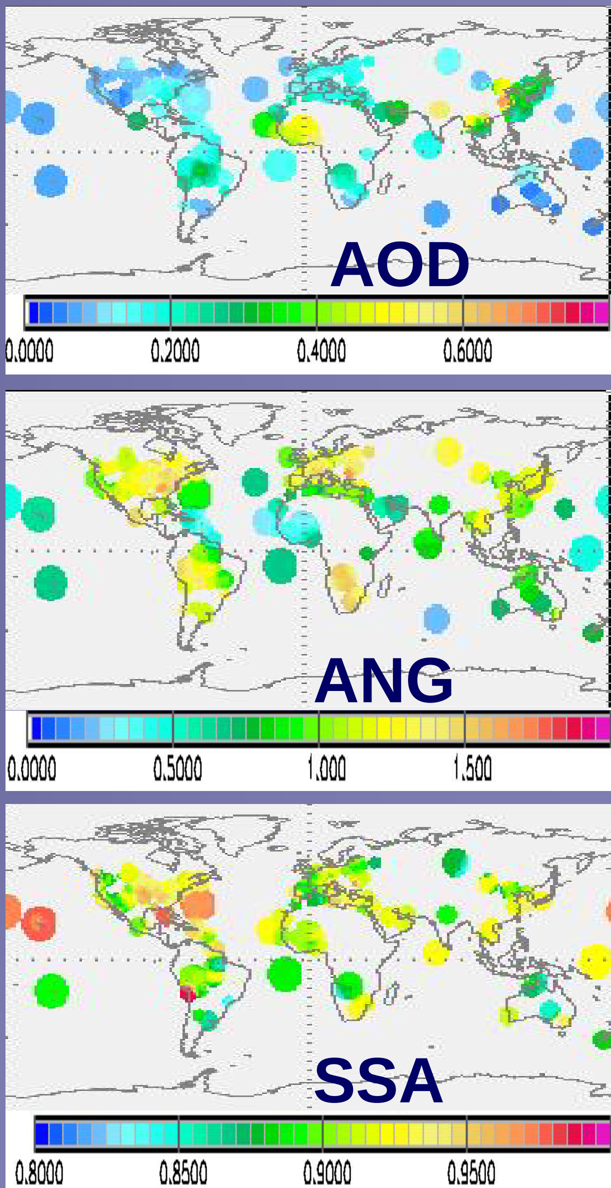
anthropogenic fraction

- ○ **.13**
  - **total AOD** ***(aerosol opt. depth)***
- ○ **.46**
  - **AOD fine / total ratio**
- ○ **.06**
  - **fine mode AOD** ***(r < 0.5μm)***
- ○ **.48**
  - **AOD anthr.fine / total fine ratio from global modeling**
- ○ **.04**
  - **anthropogenic AOD**



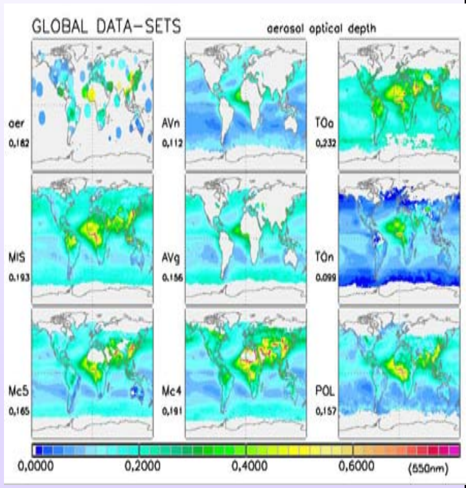
AERONET the reference

- **advantages**
  - **transmission measurement**
  - **all properties, consistently**
    - **direct sun: AOD, Angstrom**
    - **sky radiances: *also* absorption, size-distribution and shape**
- **disadvantages**
  - **local … though connected**
  - **lower conf. on absorption**
- **action**
  - **grid monthly statistics with**
    - **site scores for regionality**
    - **site scores for accuracy**



satellite AOD fields

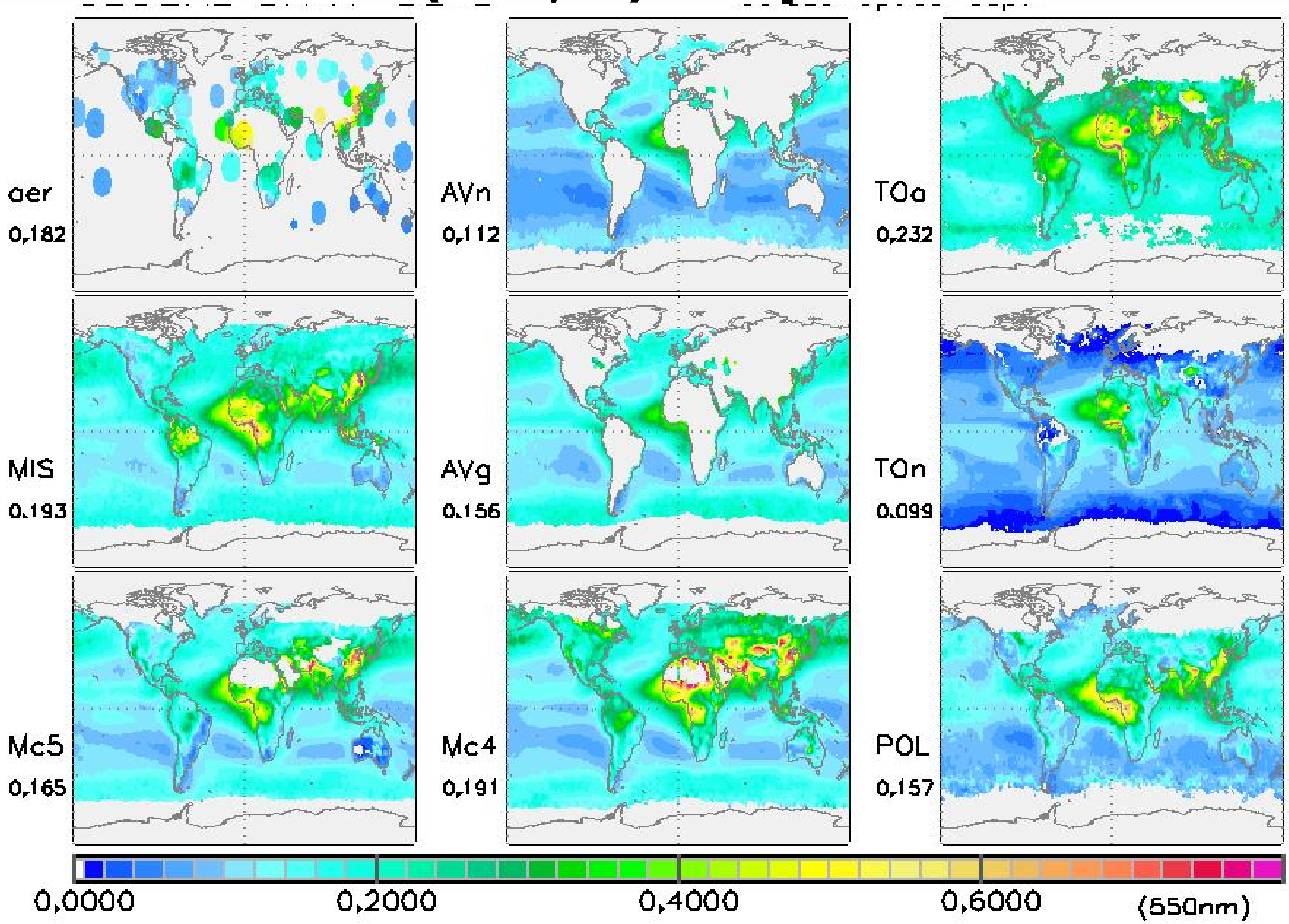
- multi-annual AOD 550nm maps
  - **MIS** MISR (2000-2005)
  - **Mc5** MODIS coll. 5, AQUA +TERRA (2000-2005)
  - **Mc4** MODIS coll. 4, AQUA +TERRA (2000-2005)
  - **AVn** AVHRR NOAA (1981-1990)
  - **AVg** AVHRR GACP (1984-2001)
  - **TOo** TOMS - old p (1979-2001)
  - **TOn** TOMS - new p (1979-2001)
  - **POL** POLDER (1987, 2002)



  - time periods with enhanced stratospheric aerosol loading (e.g. after El Chichon or after Mt.Pinatubo volc. eruptions) are excluded from these averages.

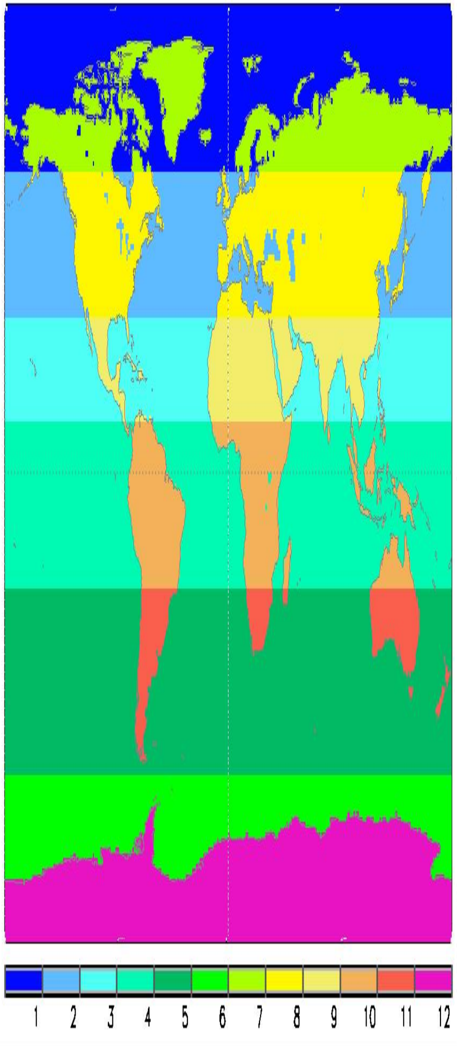
satellite AOD fields

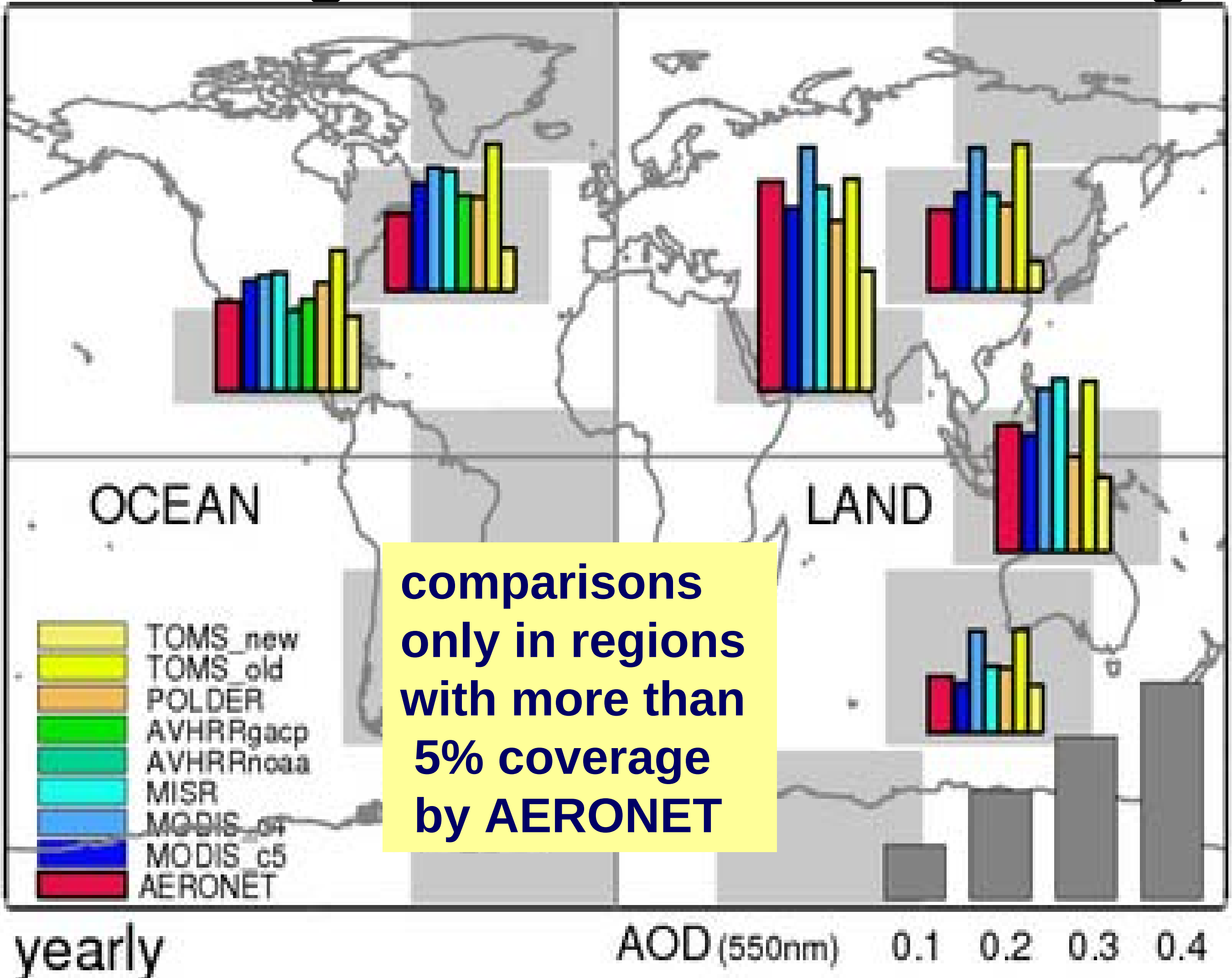
- multi-annual AOD (0.55μm) maps
  - MIS
  - Mc5
  - Mc4
  - AVn
  - AVg
  - TOo
  - TOn
  - POL



regional comparisons

- comparison averages over ocean/land reg.





comparisons only in regions with more than 5% coverage by AERONET

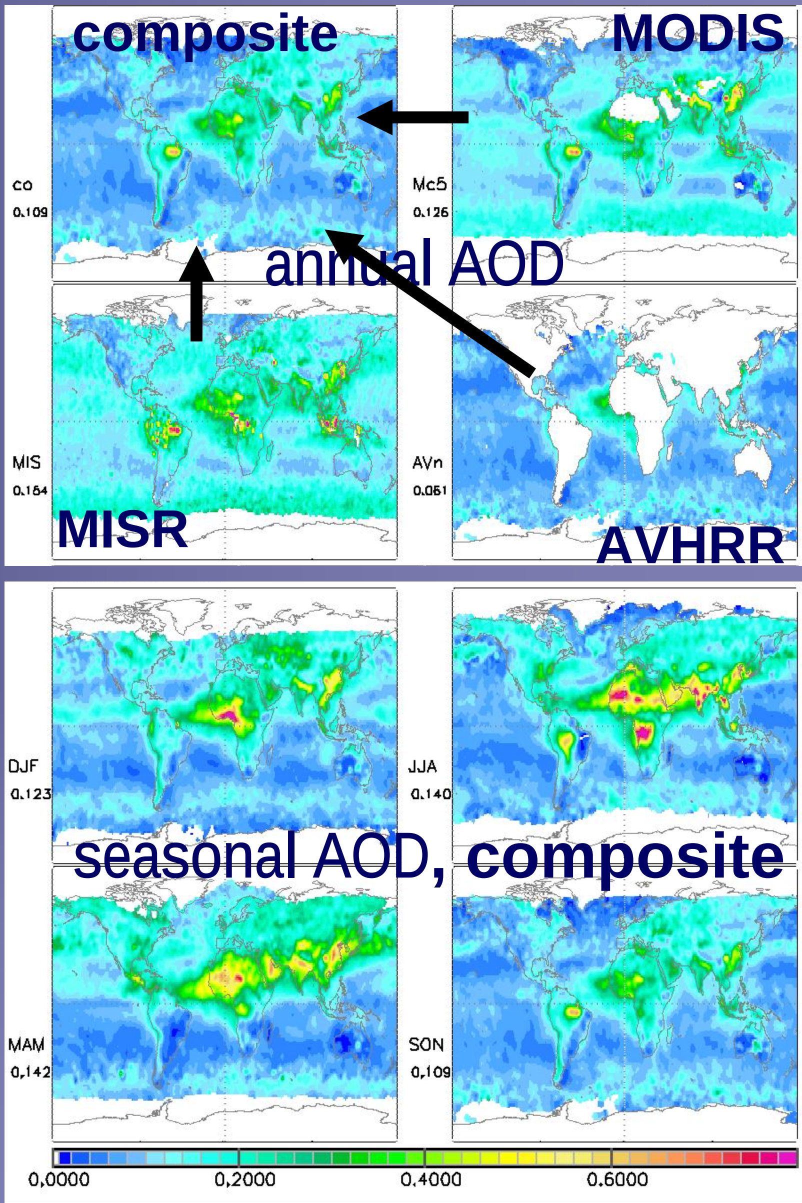
what data to recommend ?

- each satellite set has regional strength …
  … and weaknesses
- create an AOD composite combining strengths:
- selection based on *'objective'* rank scoring
  - involving all data-pairs in the region
- total score is composed of sub-scores for
  - bias
  - regional variability
  - seasonality

the AOD composite

for each region …

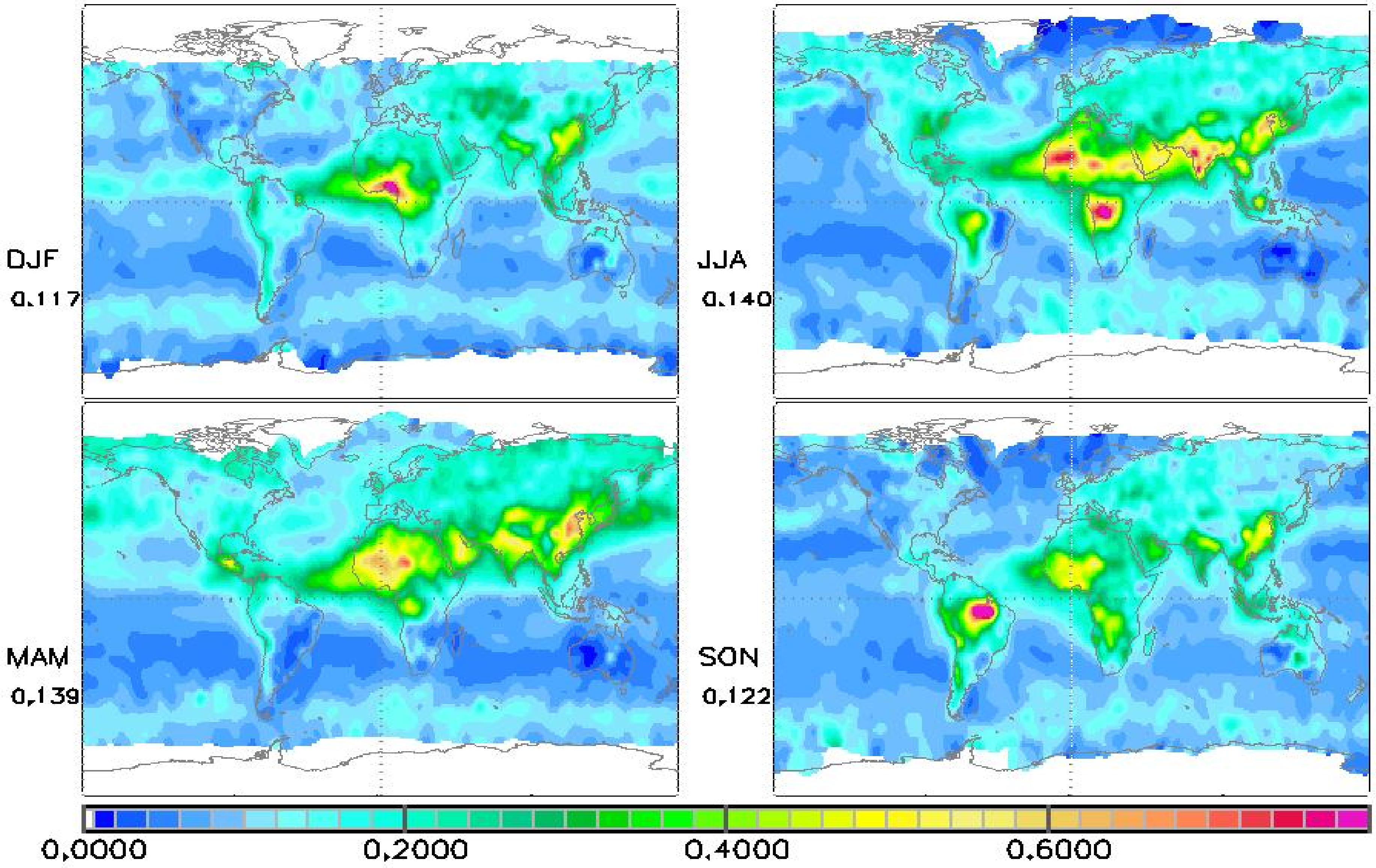
- **score vs. AERONET**
  - **bias**
  - **regional variability**
  - **the seasonality**
- **pick satellite data with highest overall score**
  - ocean: **AVHRR,POLDER**
  - land: **MISR**(NH), **MODIS**(SH)
- **create a composite**
  -



can do better: merging !

- even the best scoring satellite AOD retrievals are far from perfect (vs AERONET)
  - e.g. satellite AOD data are larger
- merge AERONET data into sat-composite
  - 1. regrid AERONET data *(use site scores)*
  - 2. identify grid-points with data pairs
  - 3. extend local grid-points ratios globally with decaying weights *(separately for land and ocean)*
  - 4. establish grid-point (weight) domains
  - 5. apply ratios (of global map) in the grid-point domains to the background field (composite)

the unbiased AOD composite



still …

- the ‘unbiased composite’ is probably one of the better global aerosol ‘data’ products

- but …
  - it only covers AOD (a single aerosol prop.)
  - assumptions to other (aerosol /environment) properties in sat-retrievals lack consistency
  - there are regions of no data (e.g. polar/desert)
  - there may be sampling biases

global modeling

- advantages
  - all aerosol properties are provided
  - consistency among aerosol properties
  - complete (no temporal or spatial data gaps)
- drawbacks
  - many processes (lack of transparency)
  - some tuning to (global annual) constraints
- compromise: median of 20 global models

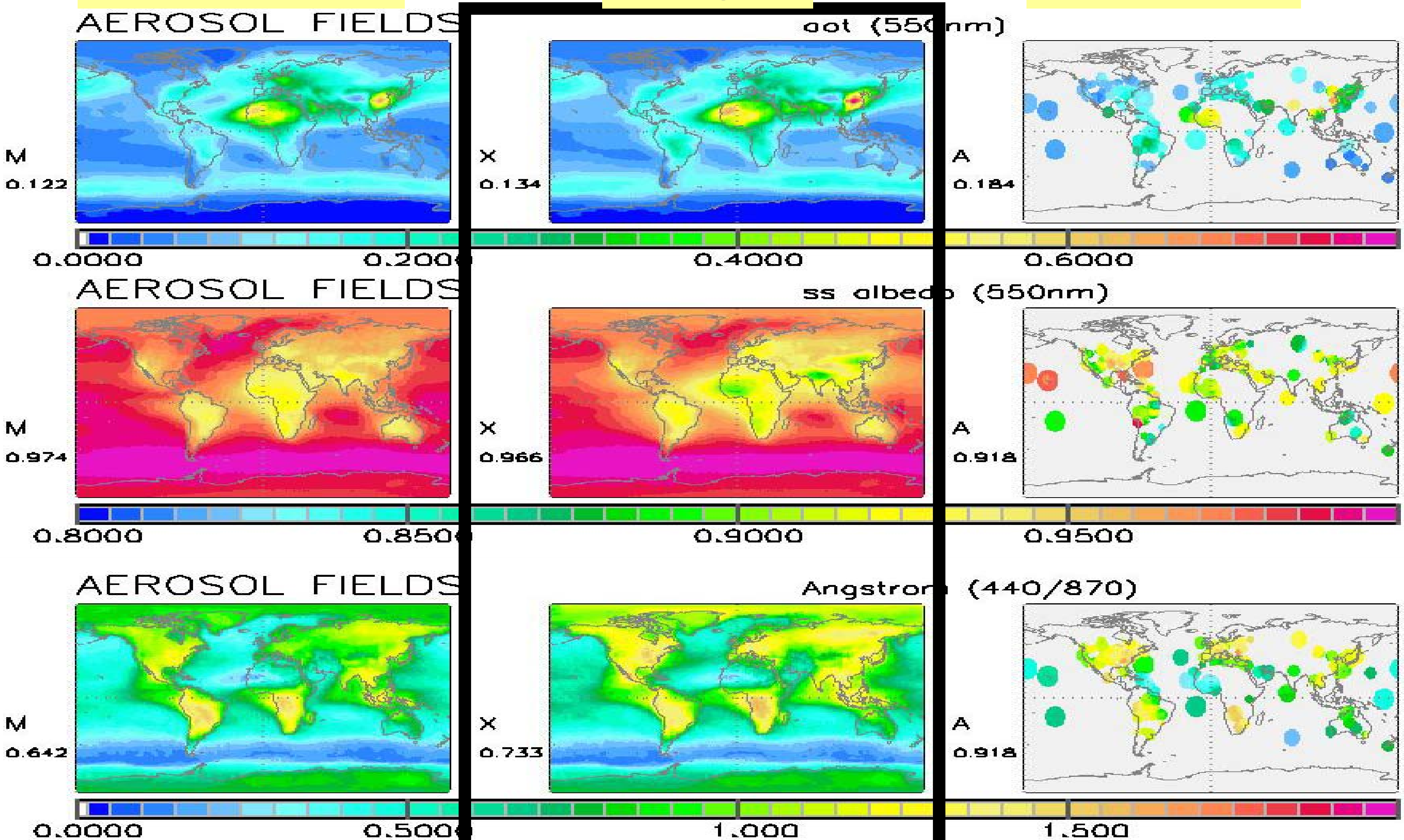
  \+ central (typical) model behavior (no extremes)

  \+ no data gaps ⇨ ideal background fields

  \- *not necessary consistent anymore*

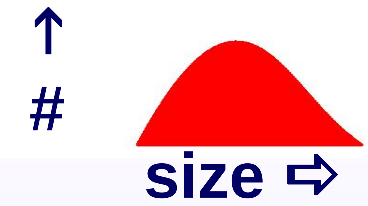
(mid-) visible optical properties

MODEL med → **merged** ← **AERONET**

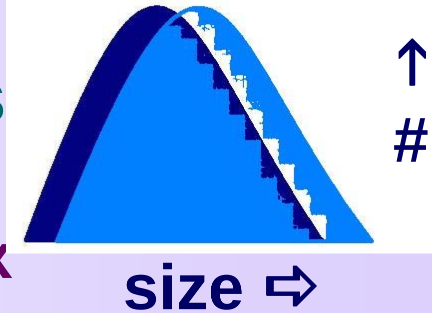


spectral extension (1)

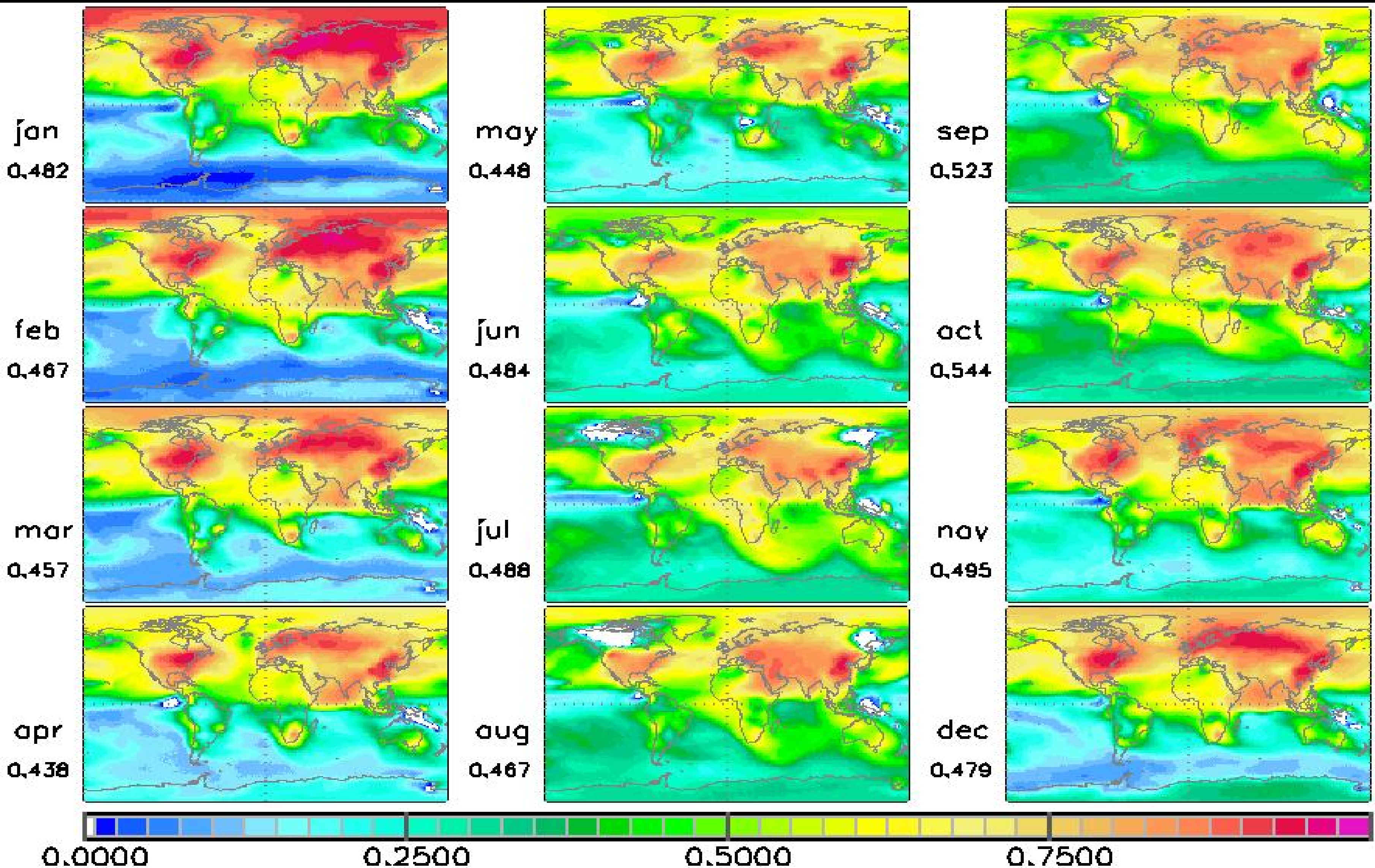
- assume a bi-modal distribution
  - corase mode (radii > 0.5μm)
  - fine mode (radii < 0.5μm)



- prescribe coarse mode single scatt. prop
  - dust (+size) or sea-salt (mix?) based on SSA
  - sizes > 1μm ⇨ Angstrom: Avis,coarse = 0.0
- set the fine mode Angstrom parameter
  - Avis,fine = 2.3 for dry conditions
  - Avis,fine = 1.7 for wet conditions
    - (scaled) low cloud cover fraction of cloud climatology as wetness index



Aff anthrop fine mode AOD fraction

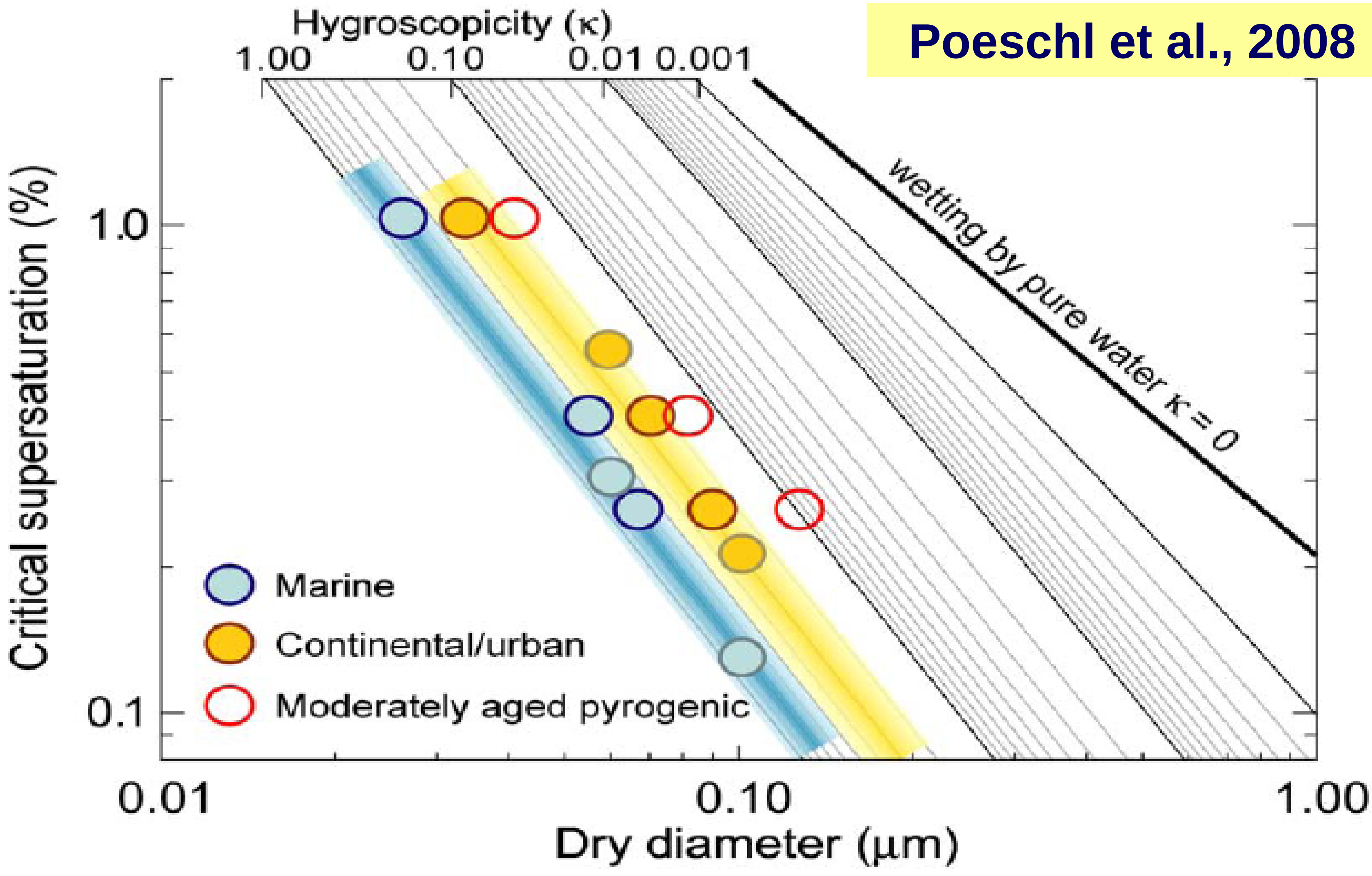


CCN conc. / enhancements

- aerosol can influence the hydrolo.cycle
  - cloud micro/macro-physics (more droplets)
  - precipitation (arguments for less and more)
- aerosol to act as CCN depends on
  - supersaturation (updraft)
  - aerosol particle size
  - hygroscopicity
- hygroscopicity seems well constrained
  - effective hygroscopic factors κ cluster at
  - 0.3 +/- 0.1 over continental regions
  - 0.7 +/- 0.2 over marine region

crit.size / hygrosc. / supersat.

Poeschl et al., 2008



climatology application

- knowing …
  - supersaturation
  - aerosol concontration (assumed AOD profile)
  - effective hygroscopicity factors (κ)
  - ambient temperature
- … the critical radius can be determined:
- CCN (by definition) are
  - all particles of the coarse size mode
  - those particles of the accumulation mode, whose radii excced the critical radius

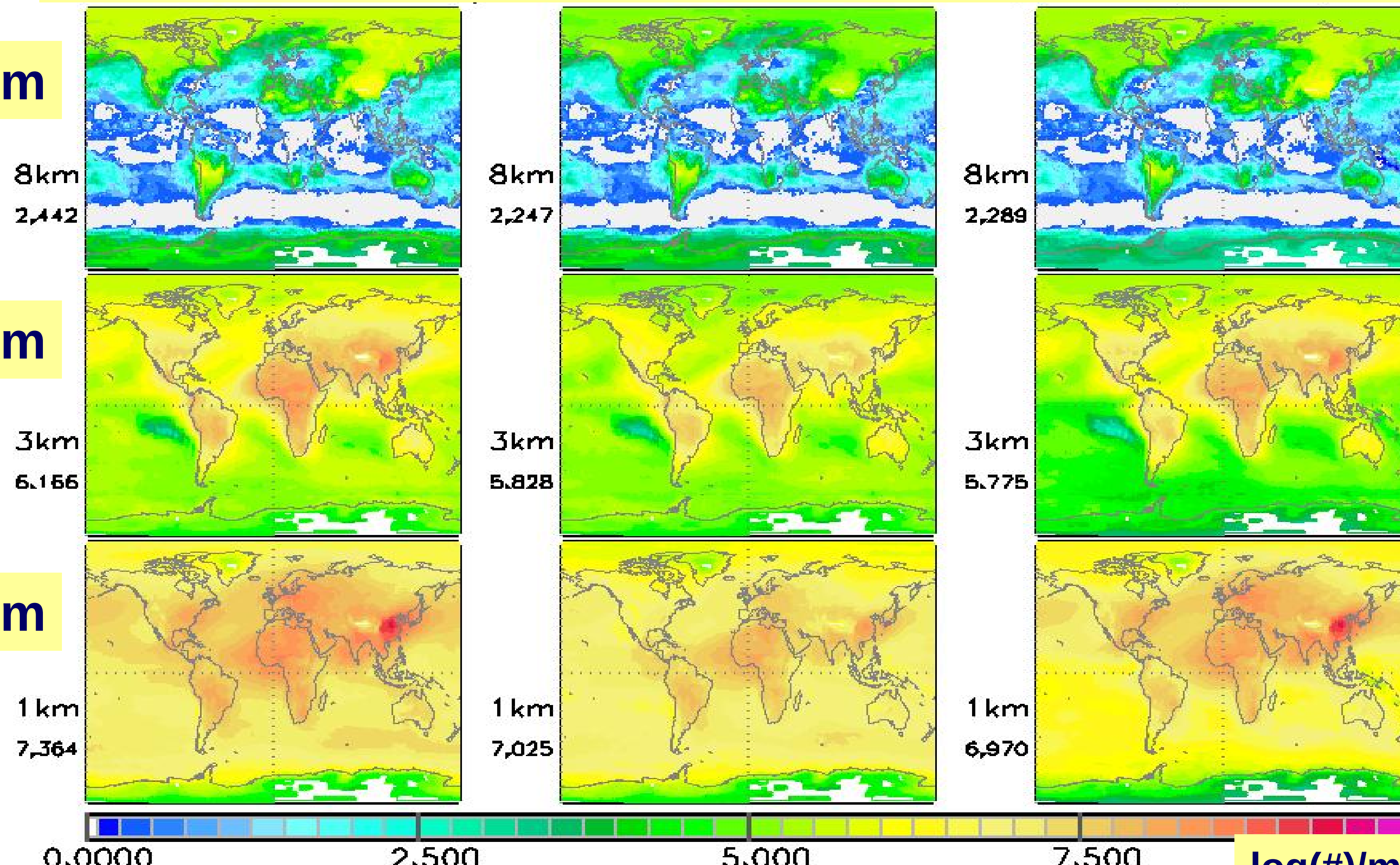
CCN at 0.1% supersaturation

total aerosol | **natural aerosol** | **anthropogenic**

8 km

3 km

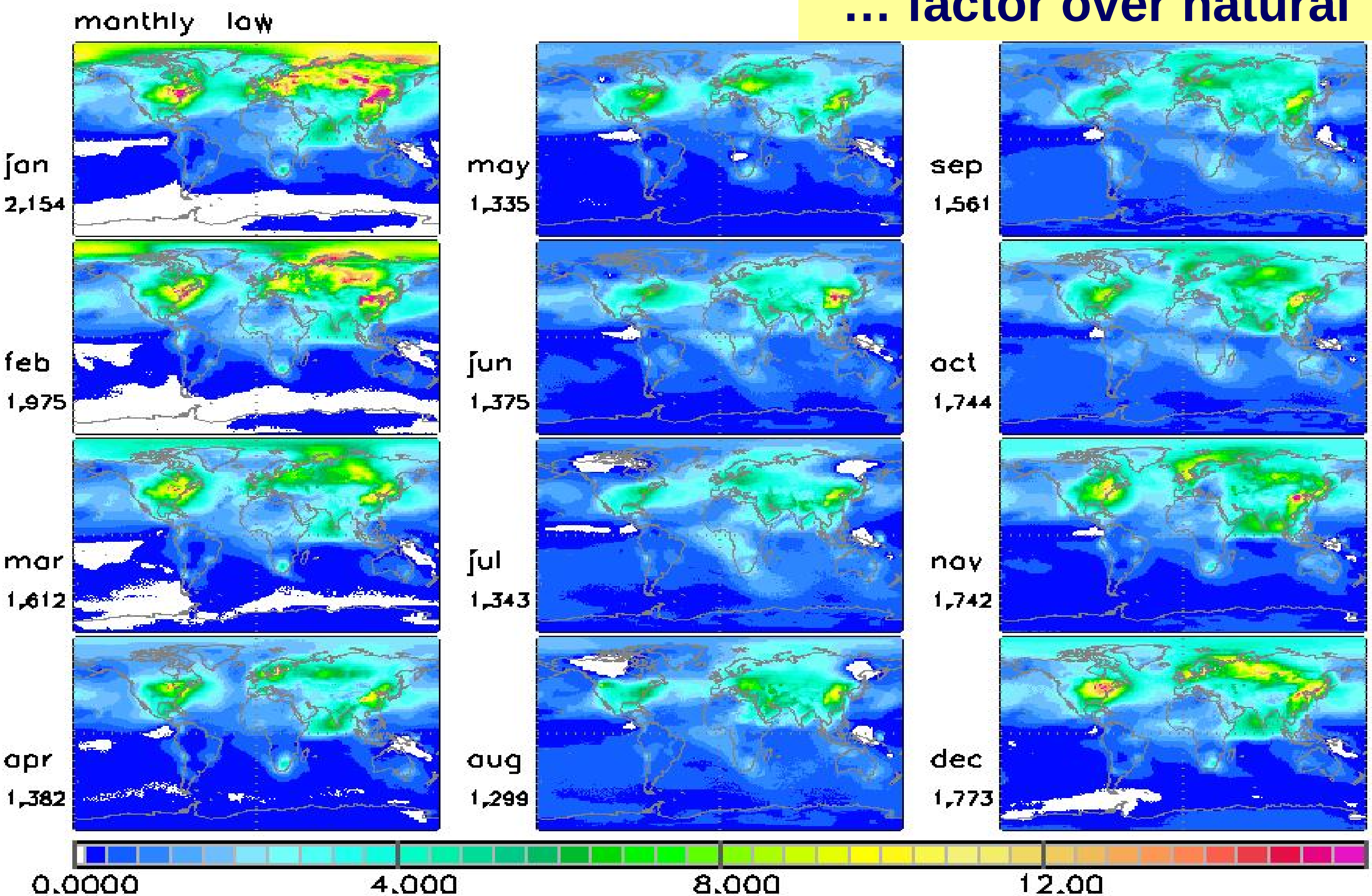
1 km



log(#)/m3

anthrop. CCN enhancement

… factor over natural

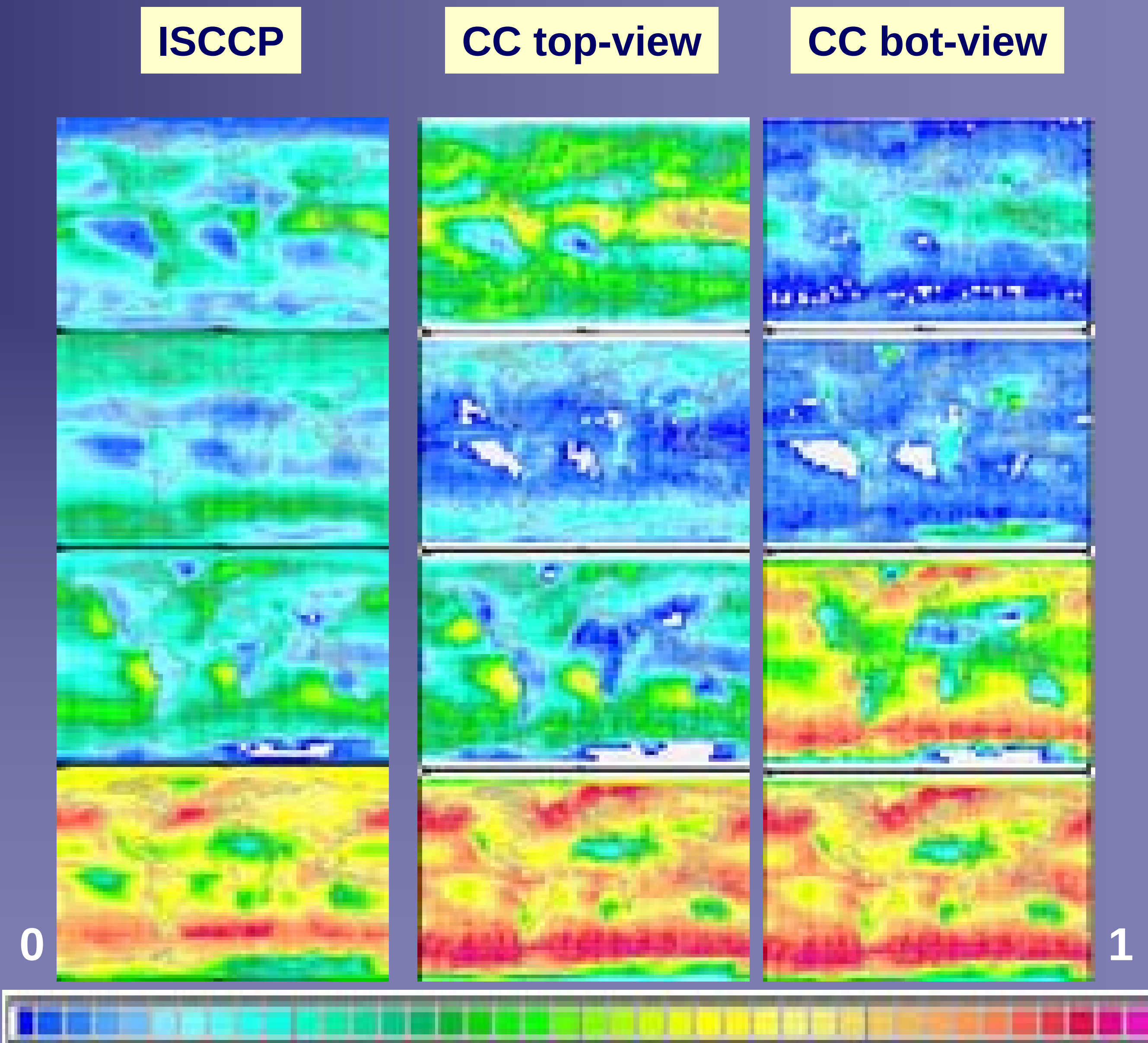


place-holder

- placeholder for aerosol properties in ‘faster’ simulations of with reduced (aerosol) complexity
- currently implemented for testing in ECHAM5 and ECMWF global models
- data are available via anonymous ftp
  - ftp ftp-projects,zmaw.de
  - cd aerocom/climatology/ band_30

cloud fraction – ISCCP / Cloudsat-C

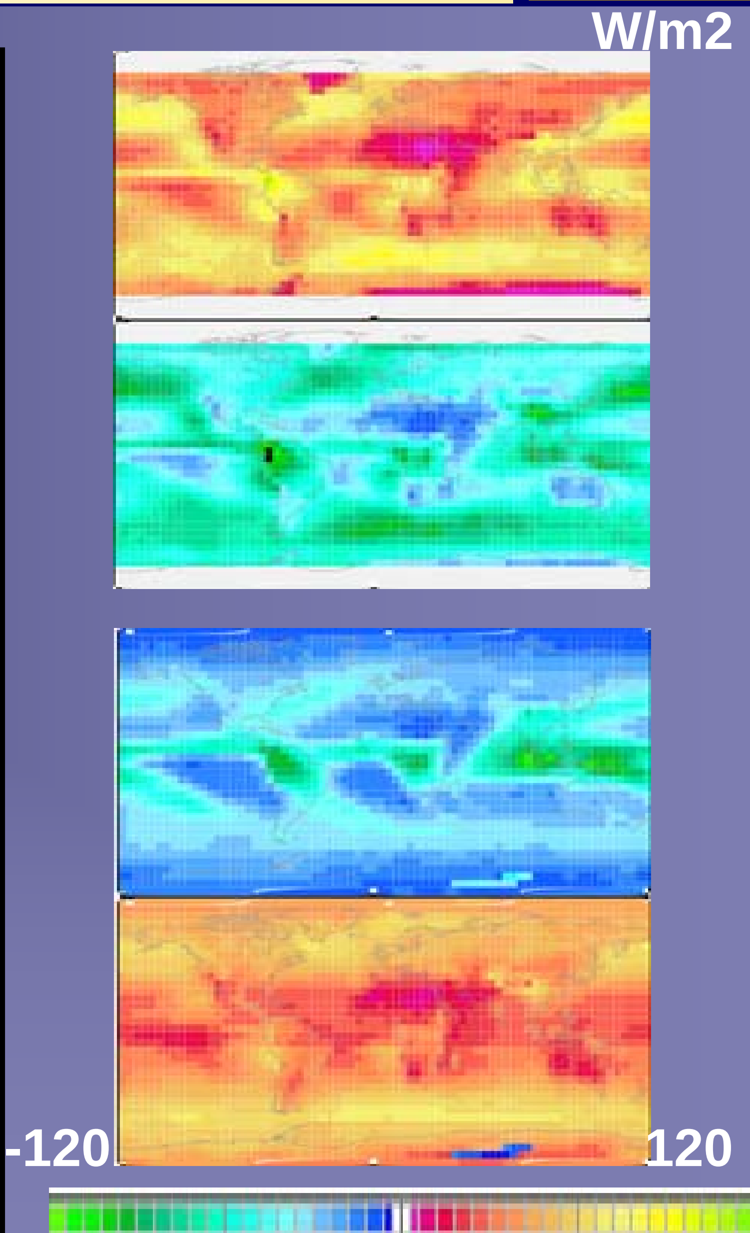
- High
  - .214 ISC
    - .412 top
    - .127 bot
- Mid
  - .192 ISC
    - .086 top
    - .084 top
- Low
  - .263 ISC
    - .249 top
    - .543 bot
- Total
  - .668 ISC
    - .741 top
    - .741 bot



IPCC median cloud effects

- + 47 W/m2 (+/- 12)
  - on solar UP flux at ToA
- - 54 W/m2 (+/- 17)
  - on solar DN flux at surf.
- - 36 W/m2 (+/- 3)
  - on IR UP flux at ToA
- + 38 W/m2 (+/- 5)
  - on IR DN flux at surface

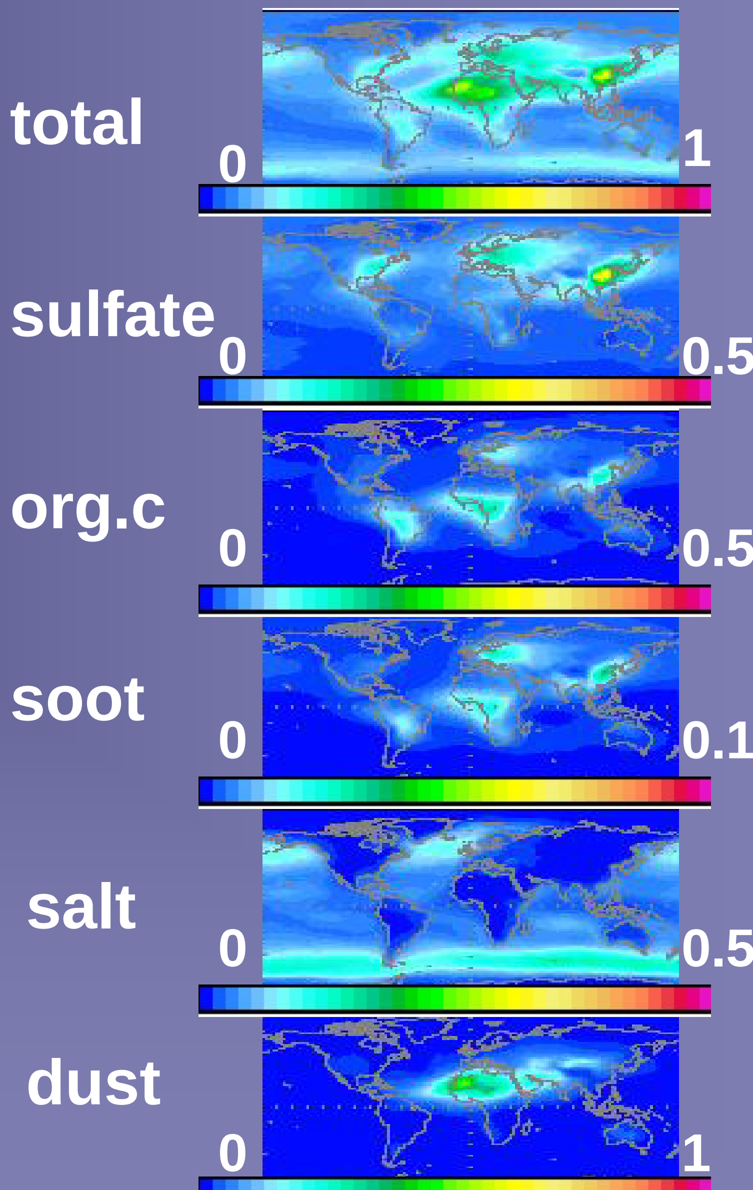
ToA – top of atmosphere



AOD – aerosol column attenuation

- **aerosol optical depth**
  - **a component mixture**
    - **sulfate**
    - **organic carbon**
    - **black carbon (soot)**
    - **sea-salt**
    - **dust**
  - **component weights differs by region**
  - **~0.13** **is the global ann. average at λ=.55μm**

annual maps (.55μm) ⇨

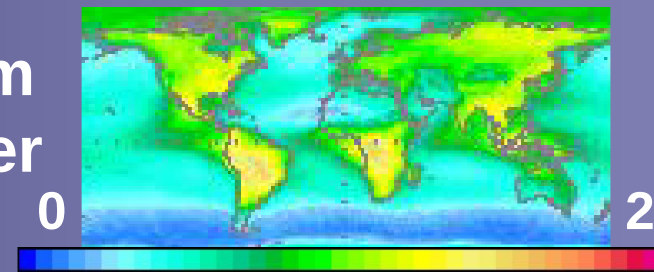


column absorption and size

- **size**
  - **AOD** spec dep → **AP**
    - **AP<1: larger sizes**
    - **AP>1: smaller sizes**
  - **fine mode (r <.5μm) AOD fraction**
- **absorption**
  - **single scatt alb**
    - **absorb potential**
  - **absorption- AOD**
    - **eff. absorption**

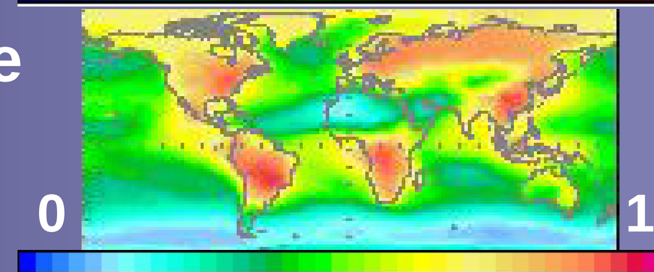
ann. maps (.55μm) ⇨

Angstrom parameter



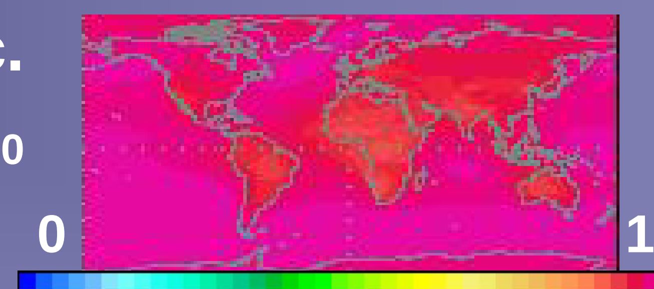
0 2

fine-mode fraction



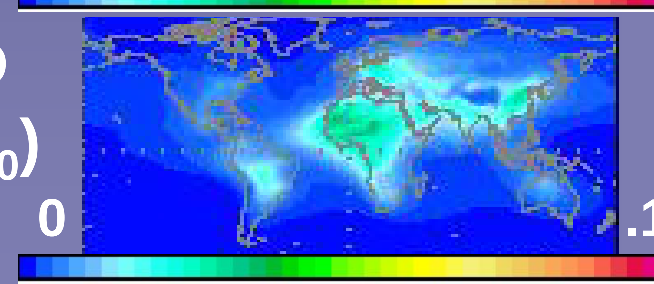
0 1

single sc. albedo ω_0



0 1

abs- AOD aod*(1- ω_0)



0 .1