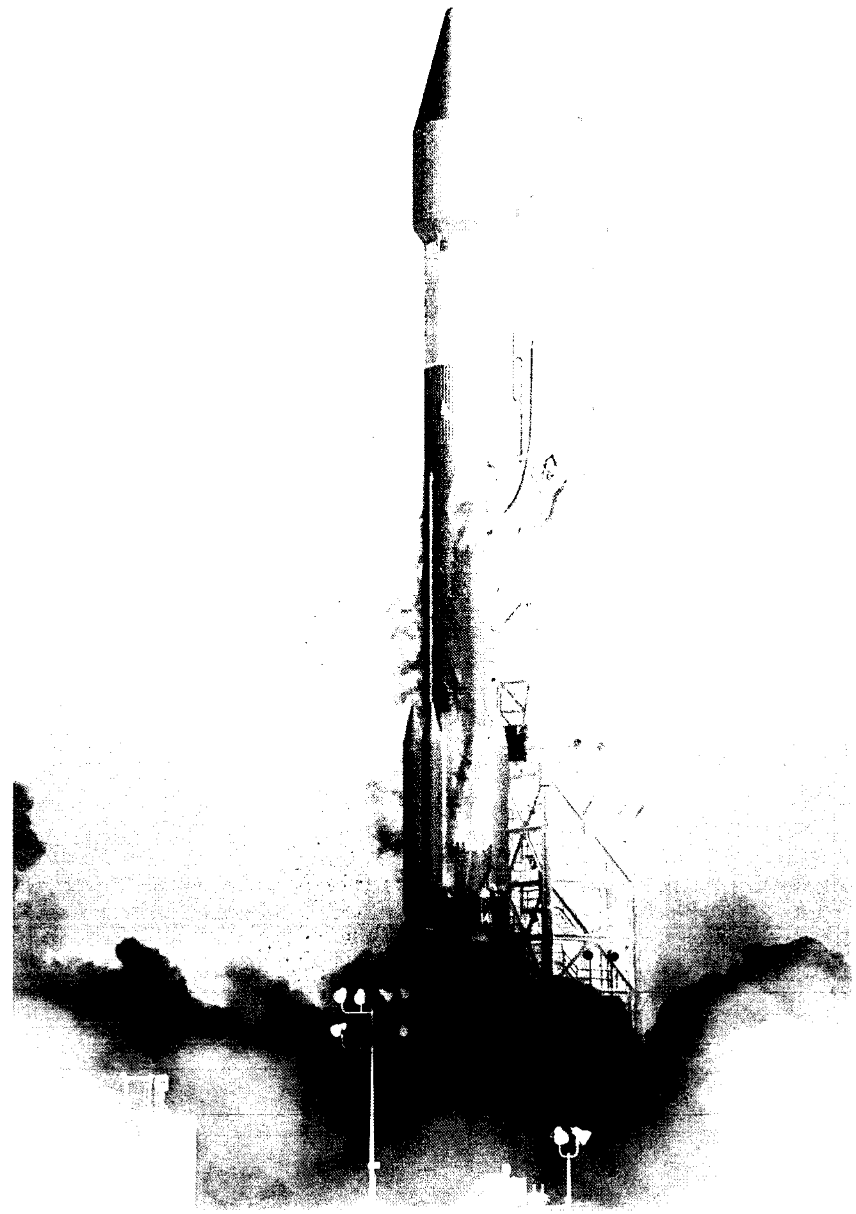
Data Fusion for Earth Science Remote Sensing

Amy Braverman
Jet Propulsion Laboratory,
California Institute of Technology

Mail Stop 126-347
4800 Oak Grove Drive
Pasadena, CA 91109-8099



Launch of the Terra satellite, December 19, 1999

Outline

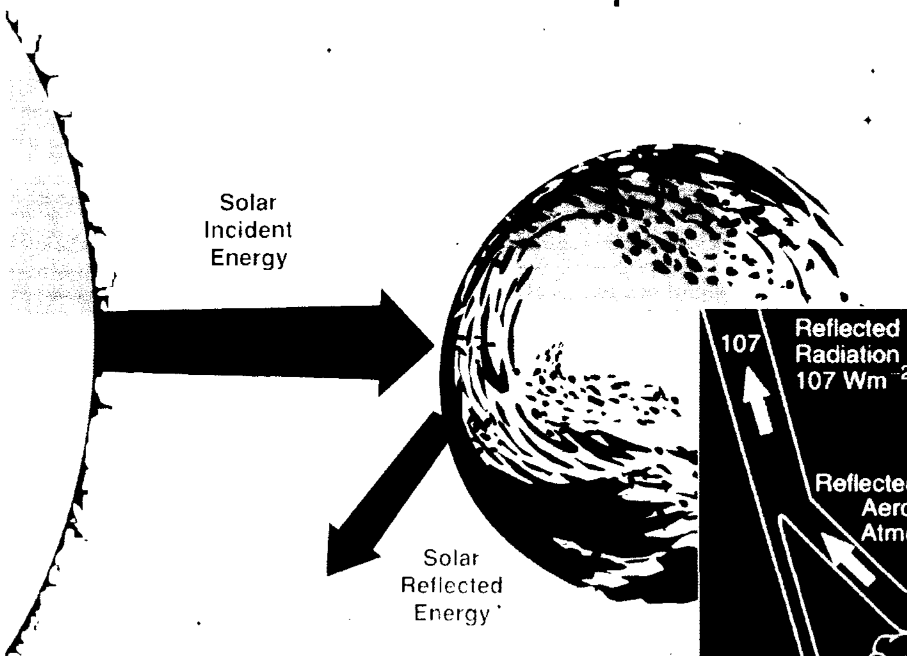
Data collected from multiple sensors on Terra.

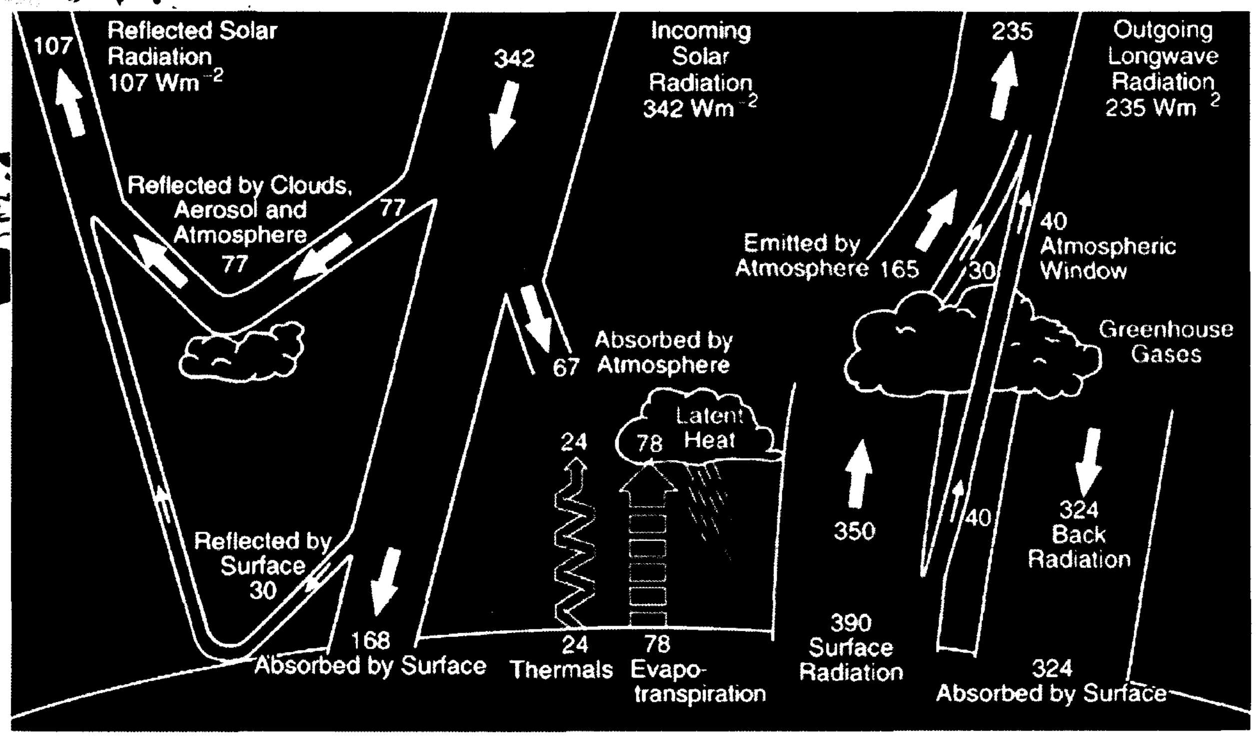
- Introduction: NASA's Earth science mission.
- The Earth Observing System.
- Satellite data acquisition and processing.
- The challenge of data fusion.
- A framework.
- Conclusion: your ideas?

Introduction

NASA's Earth science mission: *To understand and protect our home planet*

Earth Radiation Components

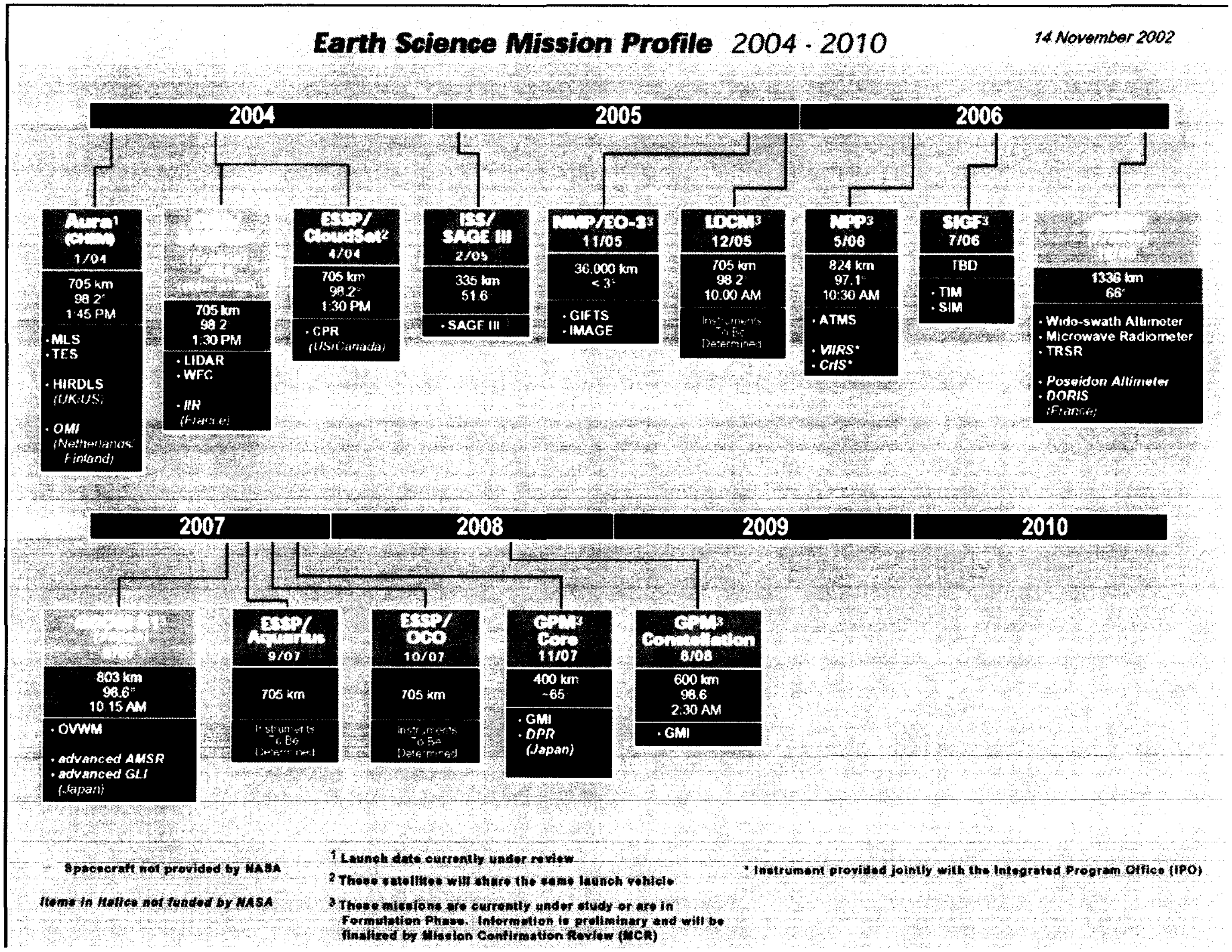




Complex system with many "forcings" and feedbacks:

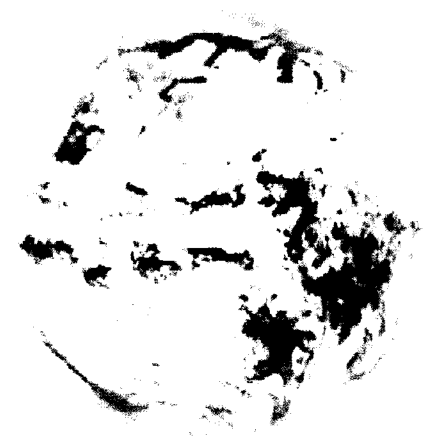
- atmosphere
- ocean
- land
- biosphere
- cryosphere

Earth Observing System



Synergistic Data Sources

- Terra satellite instruments:
  - Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)
  - Clouds and the Earth's Radiant Energy Systems (CERES)
  - Multi-angle Imaging SpectroRadiometer (MISR)
  - Measuring of Pollution in the Troposphere (MOPITT)
  - Moderate Resolution Spectrometer (MODIS)



- Aqua satellite instruments:
  - Atmospheric Infrared Sounder (AIRS)
  - Advance Microwave Radiometer for EOS (AMSR-E)
  - Advanced Microwave Sounding Unit (AMSU)
  - Clouds and the Earth's Radiant Energy Systems (CERES)
  - Humidity Sounder Brazil (HSB)
  - Moderate Resolution Spectrometer (MODIS)
- Aura satellite instruments:
  - High Resolution Dynamics Limb Sounder (HiRDLS)
  - Microwave Limb Sounder (MLS)
  - Ozone Measuring Instrument (OMI)
  - Tropospheric Emission Spectrometer (TES)

CloudSat

Cloud-Aerosol Lidar and Infrared

Pathfinder Satellite Observation (CALIPSO)

Orbiting Carbon Observatory (OCO)

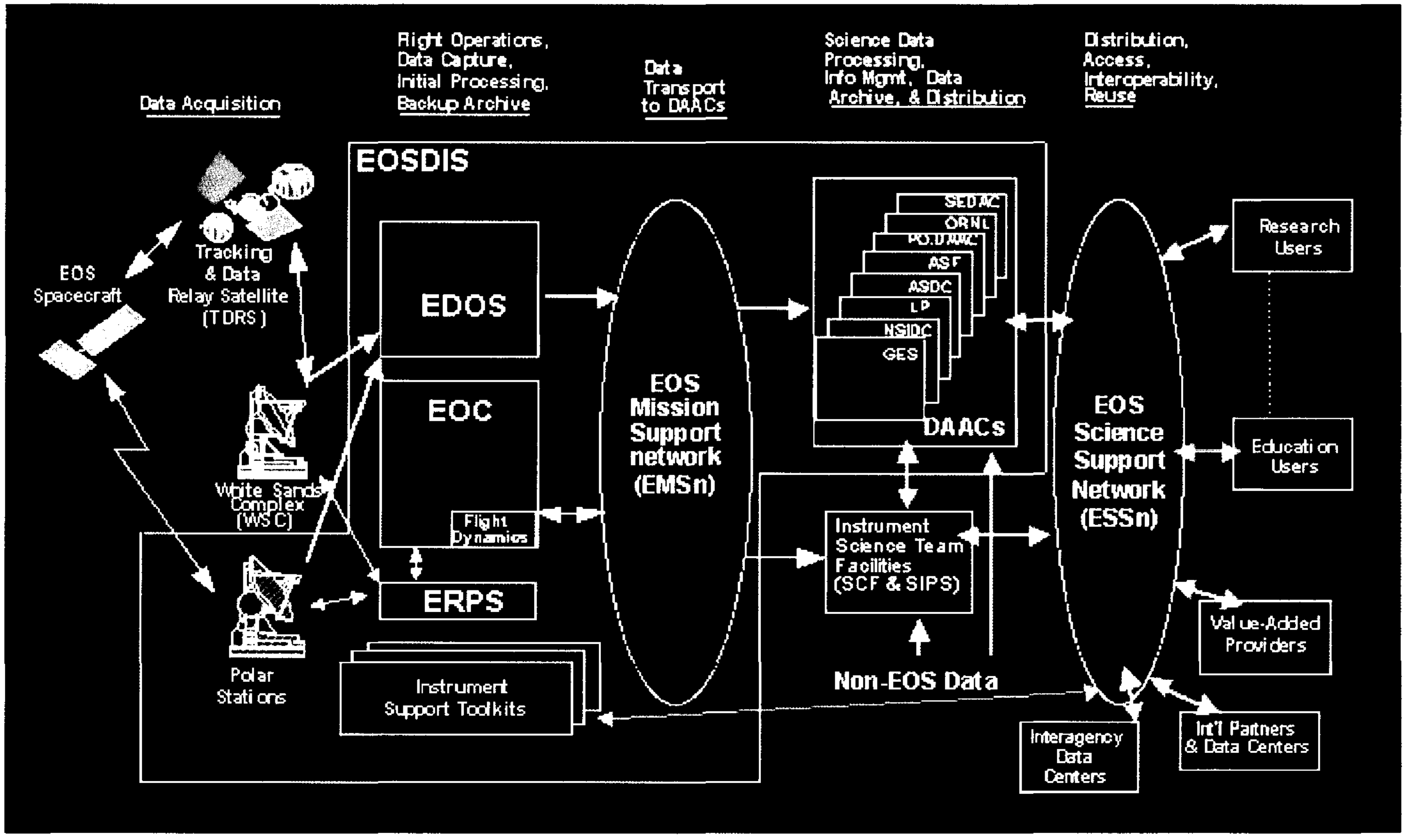
PARASOL

Technically not part of EOS

Data Acquisition

Data downlinked by "granule".

No delay allowed- the firehose is on!



National Aeronautics and Space Administration
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Data Processing

Level 0
Reconstructed, unprocessed instrument/payload data at full resolution; any and all communications artifacts, e.g., synchronization frames, communications headers, duplicate data removed.

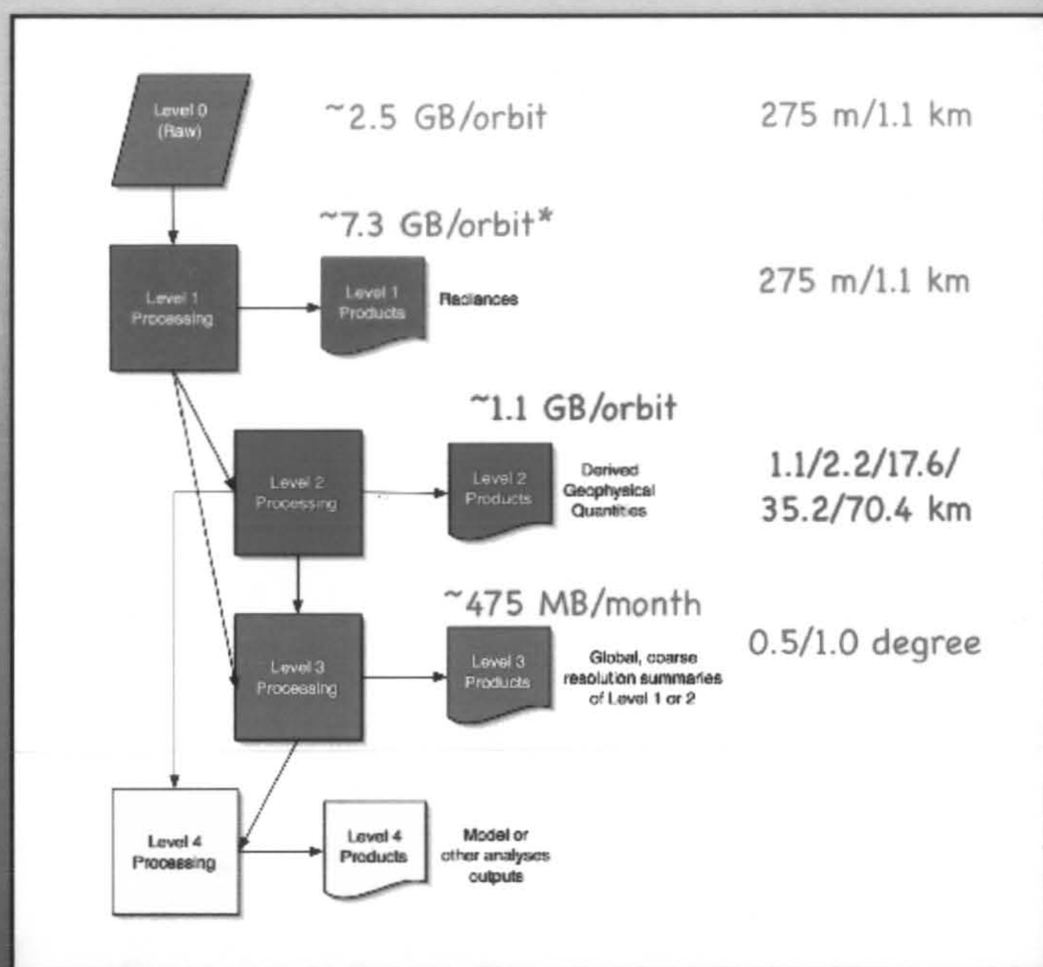
Level 1A
Reconstructed, unprocessed instrument data at full resolution, time-referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and georeferencing parameters, e.g., platform ephemeris, computed and appended but not applied to the Level 0 data.

Level 1B
Level 1A data that have been processed to sensor units (not all instruments will have a Level 1B equivalent).

Level 2
Derived geophysical variables at the same resolution and location as the Level 1 source data.

Level 3
Variables mapped on uniform space-time grid scales, usually with some completeness and consistency.

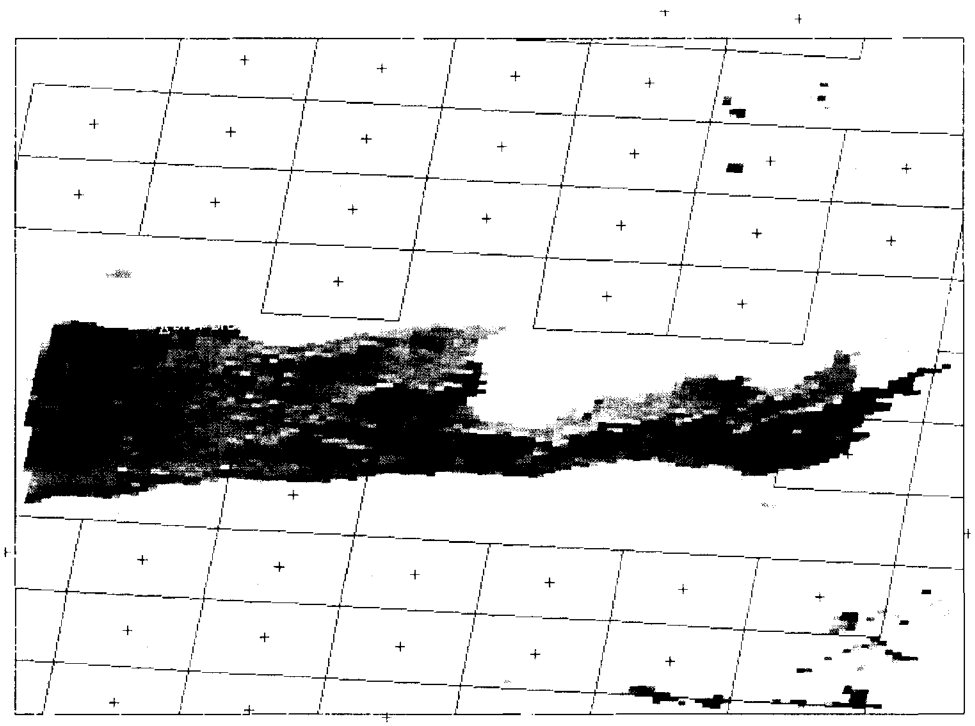
Level 4
Model output or results from analyses of lower level data, e.g., variables derived from multiple measurements.



The Challenge

Example:

MISR and MODIS on Terra



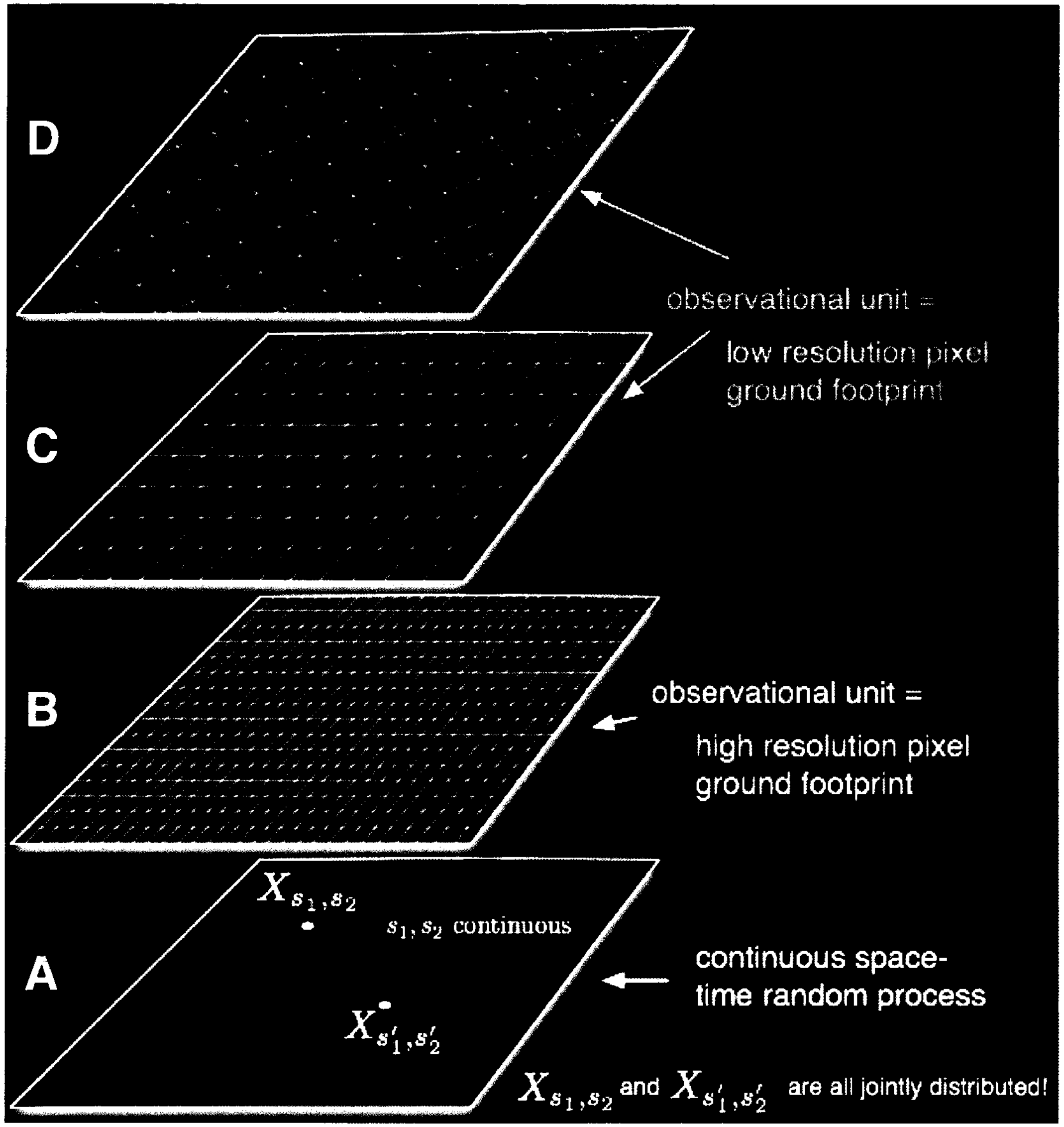
Sampling patterns of level 2 aerosol data sets; MISR (yellow pixels) and MODIS (blue pixels).

Different resolutions (MISR = 17.6 km, MODIS = 10 km), offset and mis-aligned grids => irregular overlap pattern.

The Challenge

- What we wish we had:
  - MISR and MODIS measurements on the same, very high resolution grid, at the same time, and with the same measurement error characteristics.
  - We could then estimate the joint distribution of MISR and MODIS empirically, and use it to make inferences about the underlying physical process. (*We still need a statistical model for the physical process to accomplish this.*)
  - Applies just as well whether MISR and MODIS are measuring "the same thing" or "different things".
- What we do have:
  - MISR and MODIS measurements on different, coarse resolutions (spatial aggregates) overlapping grids at different times*, and with different measurement error characteristics. (* In the simple example on the previous slide, we assume same times.)
- Working definition of data fusion: to infer the joint distribution of two quantities measured on different (but related) observational units.

A Framework



The same measurements taken at different resolutions measure “different” things.

-- thunderstorms vs. hurricanes

Different "windows" on the same underlying phenomena?

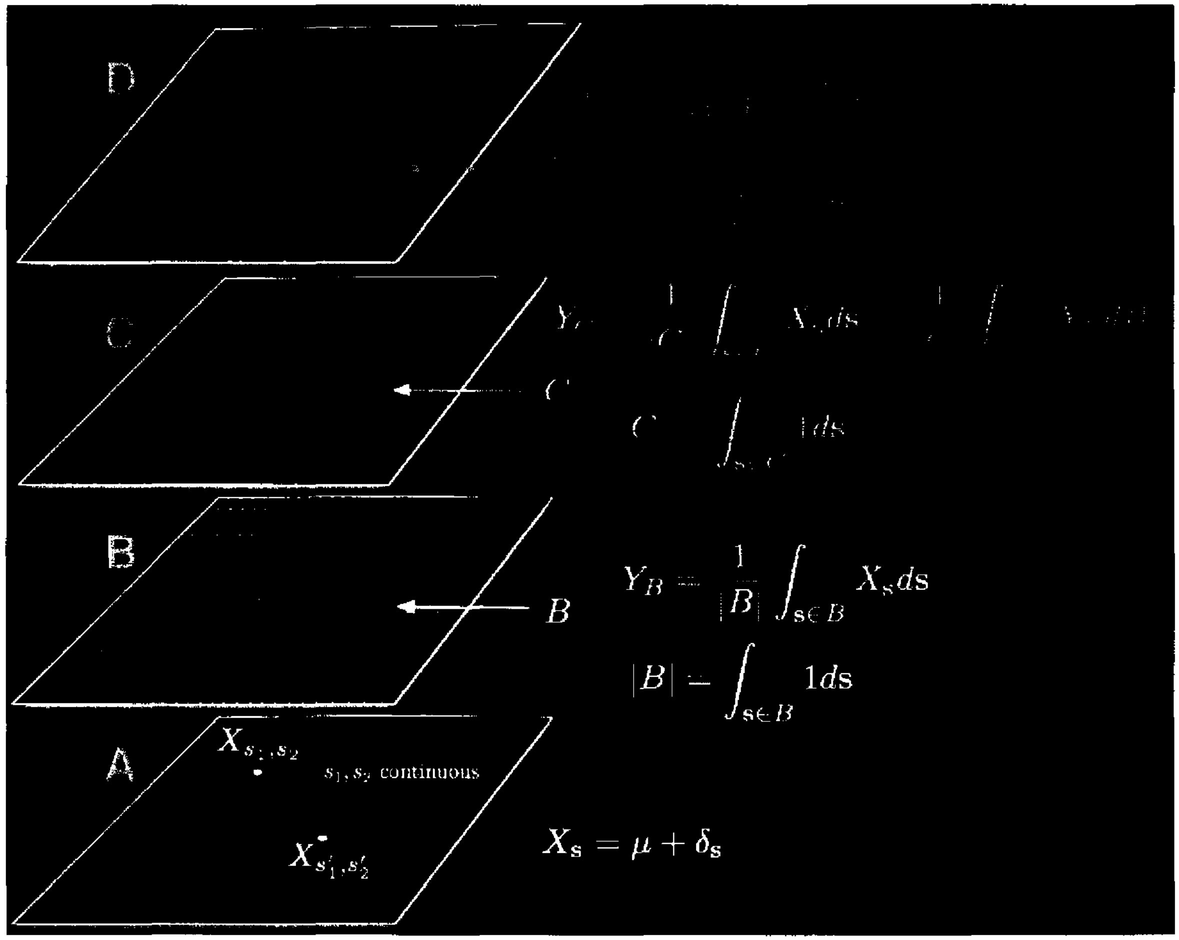
-- cloud fraction

Suppose you observe B. How to get to C (nested: easy)? D (non-nested/ misaligned: hard)? A (very hard).

-- go B to C by aggregating over observational units
-- go to B to A using geostatistical models and methods
-- how to go from C (or B) to D? Via A?

A Spatial Statistics Framework

Statistical model of data generating mechanism at multiple resolutions



Process model:

$$X_{\mathbf{s}} = \mu + \delta_{\mathbf{s}}$$

Data model:

$$Y_B^* = Y_B + \epsilon_B$$
$$= \frac{1}{|B|} \int_{\mathbf{s} \in B} (\mu + \delta_{\mathbf{s}}) d\mathbf{s} + \epsilon_B$$

μ = large scale base value

$\delta_{\mathbf{s}}$ = local perturbation

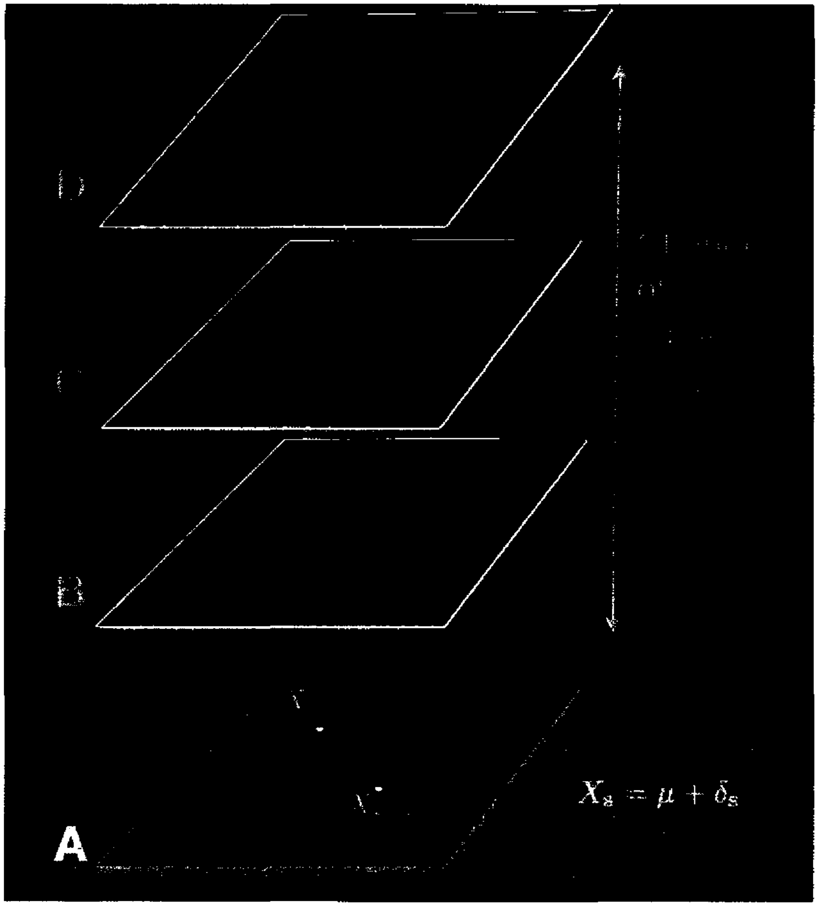
ϵ_B = measurement error at resolution B

μ (or $\mu(\mathbf{s})$) may be all or part deterministic

$Cov(\delta_{\mathbf{s}}, \delta_{\mathbf{s}'})$ specifies spatial covariance

A Spatial Statistics Framework

Can also understand relationships between B, C, D by:



1) Estimating spatial covariance by applying geostatistical methods to blocks (pixels):

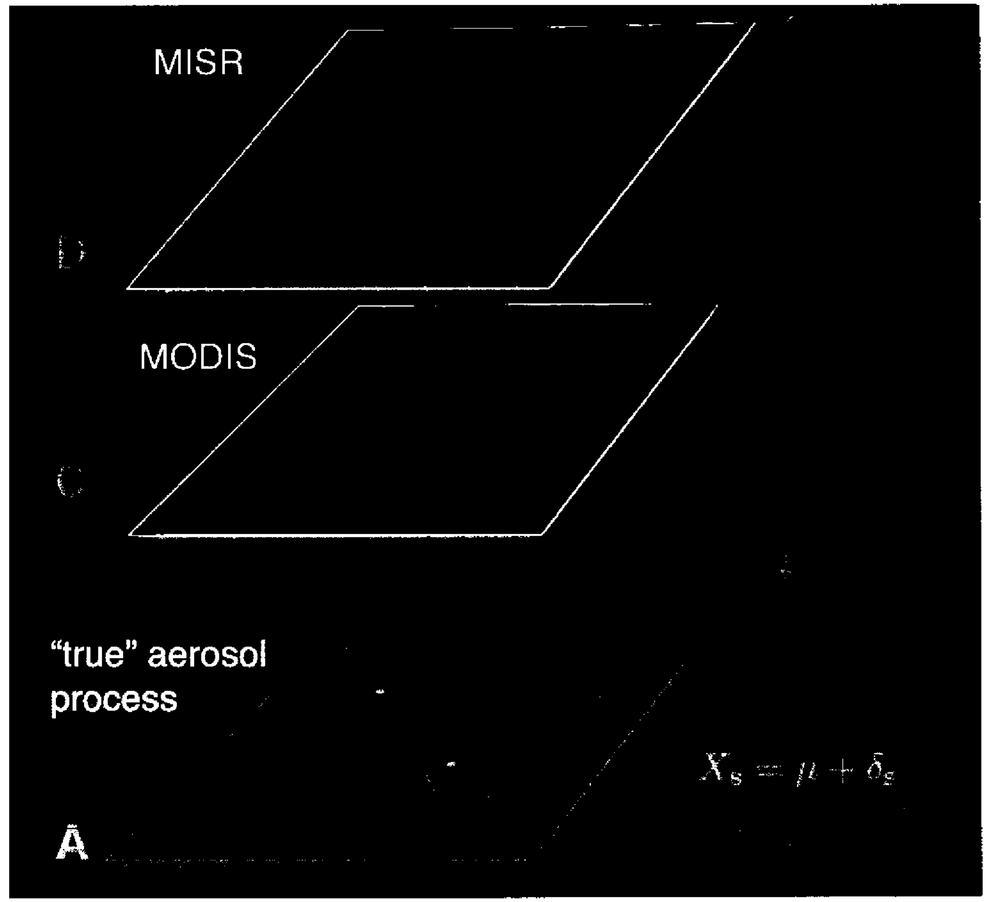
estimate $Cov(\delta_s, \delta_{s'})$ from the data as a function of distance and direction and use optimal statistical interpolation to smooth/estimate (kriging).

2) Using a Markov Random Field approach:

specify $Cov(\delta_s, \delta_{s'})$ as a function of neighborhood structure, i.e. the conditional distribution of δ_s given all $s' \neq s$ depends only on the neighbors of s .

A Framework

The objective of data fusion in the MISR/MODIS case is to combine data sets of the forms C (or B) and D but for what purpose?



Does this mean (1) we want to use information provided by C and D to make inferences about the A process...

or

does this mean (2) we want to use information provided by C and D to make inferences about the spatial process on areal units defined by the superposition of D on C?

What are the implications of deciding one way or the other?

In other words....

As the size of the observational units in C and D (pixels) gets smaller and smaller (2) approaches (1).

Since both MISR and MODIS measure related quantities we learn more and more about their joint distribution at the level of A as (2) approaches (1). (If they measure the same quantity, we expect their joint distribution to have a linear regression!)

Conclusions....

- More questions:
  - How does this framework relate to or connect with other frameworks for data fusion? Joint Directors Laboratories "levels" of fusion, for example.
  - Many data fusion methods seek to obtain correct classifications of pixels or objects in images. What are the spatial-statistical models that underlie those methods and how are they related to this problem?
  - Co-registration is not data fusion because co-registration does not address change of support issues. In what settings has co-registration been used as the basis of subsequent analyses such that this distinction matters?
  - What aspects of this spatial statistical framework are transferable and relevant in other domains? Biology? Defense? Medical imaging? Marketing?

End of File