Twinkle, Twinkle Data the Real Star
How I Wonder Where You Are
(And how good are you and where you should go?)

Craig C. Douglas

University of Wyoming and Yale University
University of Kentucky and Texas A&M

In cooperation with Anthony Vodaceck, Guan Qin, Robert Lodder, and Mauricio Kritz in particular
plus all of the workshop speakers and participants

Supported in part by the National Science Foundation
Typical DDDAS

Schematic showing connection of components via data tools, models, and sensors.
DDDAS Entails the Ability to

- dynamically incorporate data into an executing application and involves the ability of the application to dynamically steer the measurement process.
  - Such capabilities promise more accurate analysis and prediction, more effective measurements, more precise controls, and more reliable outcomes.
  - Incorporation of dynamic inputs offers the promise of computational models that more accurately describe real-world complex systems.
    - Enables the development of applications that intelligently adapt to evolving conditions and that infer new knowledge in ways that are not predetermined by initialization parameters or static data.
Data Center

• **Data acquisition tools** process the incoming data
  – **Retrieval**: from sensors or force sensors to produce
  – **Extraction** of useful data from large inputs
  – **Conversion** to a common set of internal formats
  – **Quality control** is especially important since it allows rejection of data, recollection while it is still inexpensive, and determination of relative errors
  – **Store** data if it is needed again later
  – **Notify** applications that need the data for running processes or continue simulations that need new data

• **Security** has to be maintained throughout the entire process.
Multidisciplinary Research at Multiple Sites and Different Employers

- Data sets are typically owned by a particular site, who stores it in some format, and keeps it...
  - public on a web site: Access is easy unless there are unacceptable legal restrictions placed on the data’s use.
  - private: Access is problematic to the rest of the partners.
  - pay per view: expensive and not practical for most academics unless a government agency pays the expenses over a long time with unlimited viewing.
  - in between: Access may be inconsistent. Worst of all possibilities and most common. Legal limbo.
Most Important Parts of Cooperative Modeling

• Access to the data.
  – Ability to access the data over long periods of time, even if the incoming and outgoing formats change over time.

• Adapt to personnel changes over time.
  – Team must get along instead of constant battles.
    • Death to jihadist big egotists who claim credit for everything and stab in the back their colleagues.

• Adapt to changing computational methods, models, and locations/types of computers.
Data Formats

• In the U.S., NASA keeps a data center just for water quality and analysis of estuaries.
  – There are ~1000 contributors (nightly). There are ~1000 formats.
  – NASA converts all incoming data to a common format for storage. It can deliver the data back to users in any of the ~1000 formats.

• In many academic data collections, the formats change with every new graduate student.

• Interoperability is essential.
Computing Platforms of Interest to Us

• Laptops
• PC’s
• Clusters and SMPs
  – Small, large, and Amazon-scale forests of boxes
• Hybrids
  – GP GPUs (i.e., with 32 and 64 bit IEEE arithmetic)
  – Roadrunner-like atrocities
Data Sources of Interest to Us

• MODIS
• NOOA and NASA
• U.S. Forest Service
• World meteorological organization
• ???
General Purpose Data

- Climate, atmospheric, ocean, weather, …
- Chemical species, temperature, solar absorption, …
- Size and/or quantity
- Movement (or flow) rates and directions
- Quality of data
- ???
Sensor Types of Interest to Us

- Imaging
  - Satellite, airplane, on ground, in or under water
  - (Hyperspectral) spectrometer
    - Various infrared waves (short–long and targeted ones)
    - Visible, ultraviolet, …

- Chemical measurements
  - From satellite, air, water, ground, …

- Physical and biological measurements
  - Tree trunk diameter, …

- Rovers with sensors

- Integrated sensing and processing (ISP)

- Nanosensors
Sensor Placement and Quantity of Interest to Us

• Optimize number and location of sensors
  – Too many cause stability problems, too few cause accuracy problems
  – Dynamic quantity and locations

• Allow sensors to move through changes in the environment (e.g., tree growth)
Sensor Communications

- Via satellite from a planet
- Wired or wireless
  - IP style: WiFi and IPOR (over radio)
  - Specialized transmission (from Mars)
  - Radar
- Cell phone
- Sneakernet
- Hybrid: network of the sensors that communicate with some other method to the world
Powering the Sensors

• Standard electrical methods
  – Nearby power station
  – Solar
  – Water

• Flow of some media
  – For nanobugs underground?
How Do We Construct a Virtual Data Center for Cooperative Research?

- Internal data storage
  - U.S. has one metadata format: Federal geographic data committee (FGDC.gov)
  - LBA’s format
  - NASA format
  - HDF

- Getting access
  - Brazil: Amazonal
  - U.S.: Soil moisture, water quality
How Do We Construct a *Virtual* Data Center for Cooperative Research?

- **Statistics**
  - What to do with missing or awful data
  - Ecology: pattern based modeling (confirm model and data to control errors)
  - “Data without knowing the error in it is worthless.”
  - How to measure and store information about things

- **Data clearinghouse concept**
  - Combine with metadata to find data reliably
How Do We Construct a *Virtual* Data Center for Cooperative Research?

• Geographic simplification
  – Start with *one* area, e.g., Amazonal region
  – Maybe two areas, but not more until at least one works
Summary of Speaker Sensor Types

Raw Notes
Types of Sensors (by speaker)

• Darema
  – Too many to count with 60 application areas funded through grant agencies

• Vodacek
  – Imaging spectrometer (airplane): visual and SWIR+MWIR+LWIR
  – Satellite Worldview-2, 8 spectral bands, 800 MB/picture, 500 GB/orbit

• Almiron
  – Small ISP with wireless communication
Types of Sensors (by speaker)

• Mandel
  – visual and SWIR+MWIR+LWIR and on ground temperature+chemical species and weather stations

• Costas
  – Measurements taken by hand (tree circumference) and satellite data and climate data (real-time)

• Qin
  – In well sensors cable connected and nanobugs
Types of Sensors (by speaker)

- **Barbosa**
  - MRIW imaging spectrometer and hyperspectral imager, currently in situ, but will be automated.
  - Radar images
  - Buoys with sensors and communication to INPE through satellites with hourly data - data public

- **da Silva Dias**
  - Airplane and ground photos, chemical (aerosol, CO$_2$, CO, PM$_{2.5}$, NO, NO$_2$, O$_3$) sensors
  - LWR+SRR
  - Africa ↔ South America interaction
  - Photochemistry (ozone)
Types of Sensors (by speaker)

- Dias
  - Current meters in situ

- Yamasoe
  - PAR sensors on 65m towers above canopy with sensors at 47m

- Efendiev
  - Well sensors (measure pressure and the total flow of oil/gas) and satellite imaging, soil moisture sensors
Types of Sensors (by speaker)

• Campos
  – Satellite images and sensors on buoys, wharfs, etc. and standard ROMS style data collectors

• Kritz
  – Satellite images, canopy level sensors

• Lodder
  – Hyperspectral sensors, IR, and Mars style rover and SSSI with ISP
Types of Sensors (by speaker)

• Carbonel
  – Wind driven forcing function - buoys or satellite data should be sufficient

• Douglas
  – Summary of all speakers’ types (hopefully)