

DATABASE TOOLS FOR ECOLOGICAL DATA INTEGRATION AND SYNTHESIS

Judy Cushing¹, Eda Colom-Melendez², Anne Fiala¹, Michael Finch¹, Nicole Kaplan³, Nalini Nadkarni¹, Ken Ramsey⁴ and Kristin Vanderbilt⁵

¹The Evergreen State College, Olympia WA; ² Luquillo LTER, San Juan Puerto Rico; ³ Shortgrass Steppe LTER, Colorado State Univ., Fort Collins CO; ⁴ Jornada LTER, New Mexico State Univ., Las Cruces NM; ⁵Sevilleta LTER, Univ. of New Mexico, Albuquerque NM

THE CHALLENGE:

1. To synthesize across research sites syntactically disparate, but thematically similar, data.
2. To efficiently perform cross-site synthesis, using new informatics tools that exploit database component technology.
3. To aid analysis of ecological data through visualization tools that take advantage of informatics-processed data.

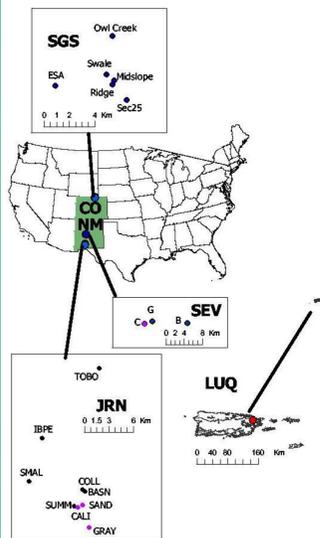
OBJECTIVES:

- To test Canopy Database Project (CDP) prototype software **Databank** (database generator) and **CanopyView** (visualization) - within and beyond forest canopy studies using data from 4 Long-Term Ecological Research (LTER) sites:
- Luquillo – Puerto Rico (Tropical Forest)
 - Jornada Basin – New Mexico (Grassland)
 - Sevilleta – New Mexico (Grassland)
 - Shortgrass Steppe – Colorado (Grassland)

TROPICAL FOREST: To visualize the 3-D upper canopy surface height over time after Hurricane Hugo (1989).

GRASSLANDS/BUSH: To compare net primary productivity (NPP) data across a large landscape by combining field data from three separate projects into a single database.

STUDY SITES



- Creosote Bush
- Grassland
- Tropical Forest

METHODS:

DATA COLLECTION:

Luquillo - Forest

The 16 ha long-term forest dynamics plot is sub-divided into a 5m x 5m grid of points. In 1992, 1994, & 1996 maximum heights of vegetation above each grid point were recorded.

Jornada and Sevilleta – Grassland/Bush

ANPP data were collected in two distinct phases:

- I) 1-m² Quadrat Measurements – Non-destructive measures of plant sizes were recorded for every species in each quadrat 3x/yr (winter, spring, and fall).
- II) Plant Harvests - Plants were harvested from areas adjacent to the quads, and regressions based on these harvests were used to estimate biomass within the quadrats.

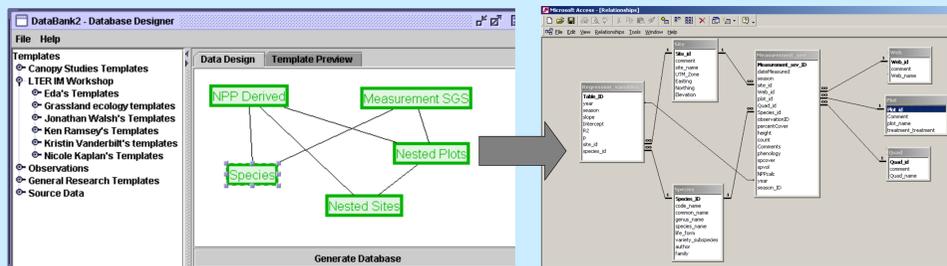
Shortgrass Steppe - Grassland

ANPP data were collected in a single phase 1x/yr at the end of the growing season:

- I) Plant Harvests – The current year's above-ground growth was clipped from plants located within each 0.25-m² quadrat. Clipped samples were weighed by species.

DATABASE CREATION:

We designed **Databank** components, and the system combined them into MS Access database packages and generated Ecological Metadata Language (EML) files:



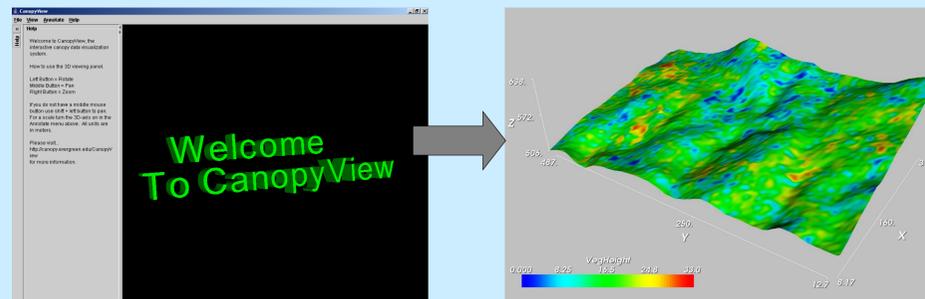
DataBank – Example NPP database design

Generated MS Access Database Tables

Commonalities among data collection methods (nesting of sites) allowed the use of a single extensible database component for 3 sites; thus the 3 data sets could be compared.

VISUALIZATION:

A **Databank**-generated database was used in **CanopyView** to visualize canopy height

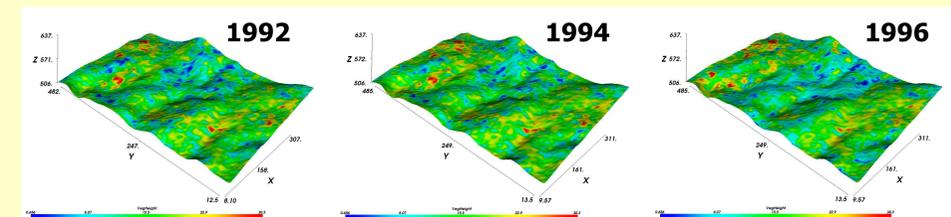


CanopyView - Visualization

Sample Luquillo visualization generated

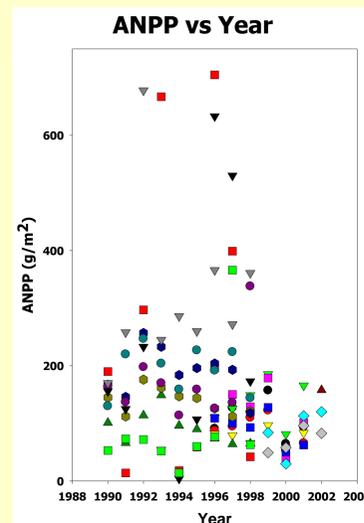
PRELIMINARY RESULTS:

TROPICAL FOREST:



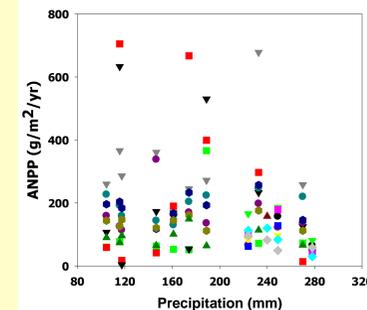
Comparison of Upper Canopy heights 3-, 5-, and 7-yr's Post Hurricane Hugo

GRASSLANDS/BUSH - Sub-site Level:

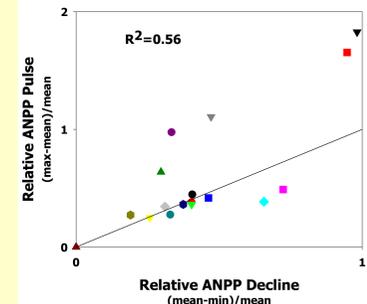


- Grassland**
- ESA
 - MIDSLOPE
 - OWL
 - Ridge
 - SEC25
 - Swale
 - G
 - B
 - BASN
 - IBPE
 - SUMM
 - COLL
 - SMAL
 - TOBO
- Creosote Bush**
- C
 - CALI
 - GRAV
 - SAND

ANPP vs Annual Precipitation



ANPP - Pulse vs Decline



* Types of analysis partially based on Knapp, A.K., and M.D. Smith. 2001. Variation among biomes in temporal dynamics of aboveground primary production. *Science*. 291:481-484.

CONCLUSIONS:

- Data visualizations of the Luquillo forest provided an effective alternate analysis of canopy height data -- illustrating decreased upper canopy height after 1994.
- Data set integration using **Databank** allowed comparison of NPP across a larger landscape than was possible within individual LTER sites. Such informatics tools offer promise for cross-site research and data synthesis.
- This collaboration among LTER Information Managers, Ecology Researchers, and Computer Scientists provided insights into designing database components and informatics tools for ecologists, and improving the user interface of existing software prototypes.
- Analysis of syntactically-incomparable datasets that represent related concepts can be facilitated with conceptual design techniques and **Canopy Database Project** tools.

FUTURE WORK:

1. Further analyze these particular cross-site NPP field data and statistical aggregates (in particular with respect to missing data).
2. Generalize data analysis methodology for other cross-site NPP measures.
3. Describe our derived data products in ecological metadata formats, which could provide a model for describing other ecological syntheses.

INTRODUCTION:

The Canopy Database Project (CDP):

- Develops informatics tools for forest canopy scientists.
- Documents and publishes datasets that demonstrate use of these tools.
- Characterizes, visualizes, and formalizes (in informatics terms) fundamental structures of the canopy.
- Generalizes the tools to be applied to the larger discipline of ecology.

CDP Informatics Tools include:

- **Databank** -- generates ecology databases using a library of design components.
- **CanopyView** -- creates data visualizations from **Databank** databases for analysis.
- **Big Canopy Database** -- provides research reference information to canopy researcher (canopy.evergreen.edu/bcd).

Long-Term Ecological Research (LTER) Sites: The Test Bed

Tropical Forest – Canopy Height:

Luquillo is located in an area susceptible to hurricane disturbance, and time series data were recorded post-Hurricane-Hugo, which hit the site in 1989. These data on maximum height over time are amenable to visualization for examining dynamic patterns of disturbance.

Grasslands/Bush - Above-Ground Net Primary Production (ANPP):

Jornada, Sevilleta, and Shortgrass Steppe have similar, but not directly comparable, NPP data, which were chosen for an ecological synthesis exercise. Combining data from 3 studies is enabling us to examine ANPP patterns across the larger landscape.

ANPP is:

- The change in plant mass including any losses to death and decomposition over a given period of time (measured in g/m²/time interval).
- A fundamental ecological variable that measures rates of carbon consumption and fixation.
- Important in understanding energy flow at a community level and spatial and temporal responses of the community to a range of ecological processes.

ACKNOWLEDGEMENTS:



Databank: canopy.evergreen.edu/databank **Canopyview:** canopy.evergreen.edu/canopyview
For more information contact us c/o: fialaa@evergreen.edu