Computational Revision of Ecological Process Models

Nima Asgharbeygi, Pat Langley, Stephen Bay
Center for the Study of Language and Information
Stanford University

Kevin Arrigo
Department of Geophysics
Stanford University

Thanks to S. Dzeroski, J. Sanchez, K. Saito, J. Shrager, and L. Todorovski for their contributions to this research, which is funded by the US National Science Foundation.

Data Mining vs. Scientific Discovery

There exist two computational paradigms for discovering explicit knowledge from data.

The *data mining* movement develops computational methods that:

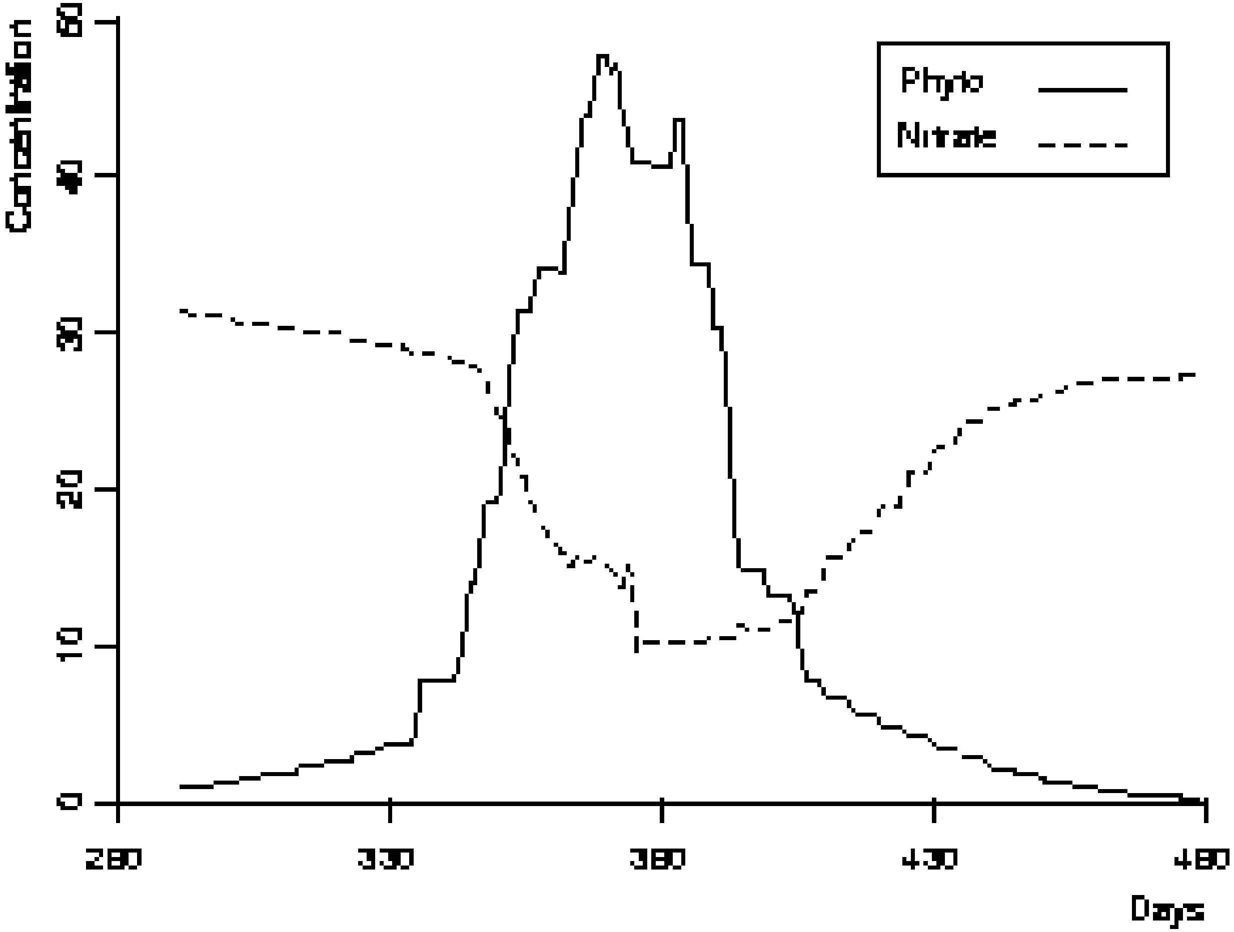
- induce predictive models from large (often business) data sets;
- represent models in notations invented by AI researchers.

In contrast, *computational scientific discovery* focuses on:

- constructing models from (often small) scientific data sets;
- stated in formalisms invented by scientists themselves.

This talk focuses on applications of the second framework to environmental and ecosystem modeling.

Observations from the Ross Sea



A Model of the Ross Sea Ecosystem

model RossSeaEcosystem

variables: phyto, zoo, nitro, residue
observables: phyto, nitro

d[phyto,t,1] = – 0.307 × phyto – 0.495 × zoo + 0.411 × phyto

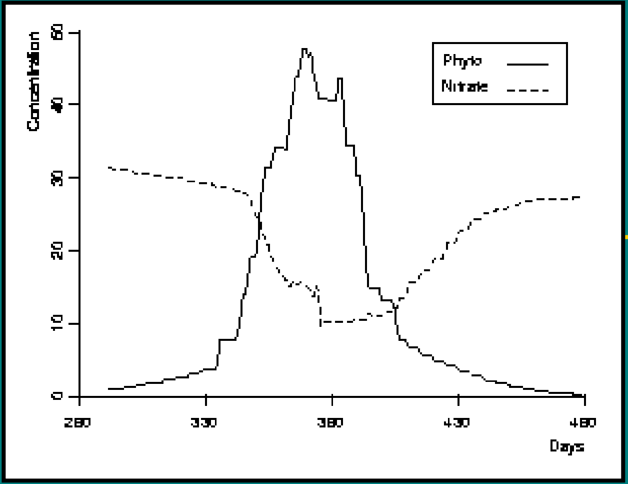
d[zoo,t,1] = – 0.251 × zoo + 0.615 × 0.495 × zoo

d[residue,t,1] = 0.307 × phyto + 0.251 × zoo + 0.385 × 0.495 × zoo – 0.005 × residue

d[nitro,t,1] = – 0.098 × 0.411 × phyto + 0.005 × residue

Inductive Revision of Ecosystem Models

observations



model RossSeaEcosystem

variables: phyto, zoo, nitro, residue
observables: phyto, nitro

d[phyto,t,1] = – 0.307 × phyto – 0.495 × zoo
+ 0.411 × phyto

d[zoo,t,1] = – 0.251 × zoo + 0.615 × 0.495 × zoo

d[residue,t,1] = 0.307 × phyto + 0.251 × zoo
+ 0.385 × 0.495 × zoo – 0.005 × residue

d[nitro,t,1] = – 0.098 × 0.411 × phyto + 0.005 × residue

initial model

Revision

revised model

model RossSeaEcosystem

variables: phyto, zoo, nitro, residue
observables: phyto, nitro

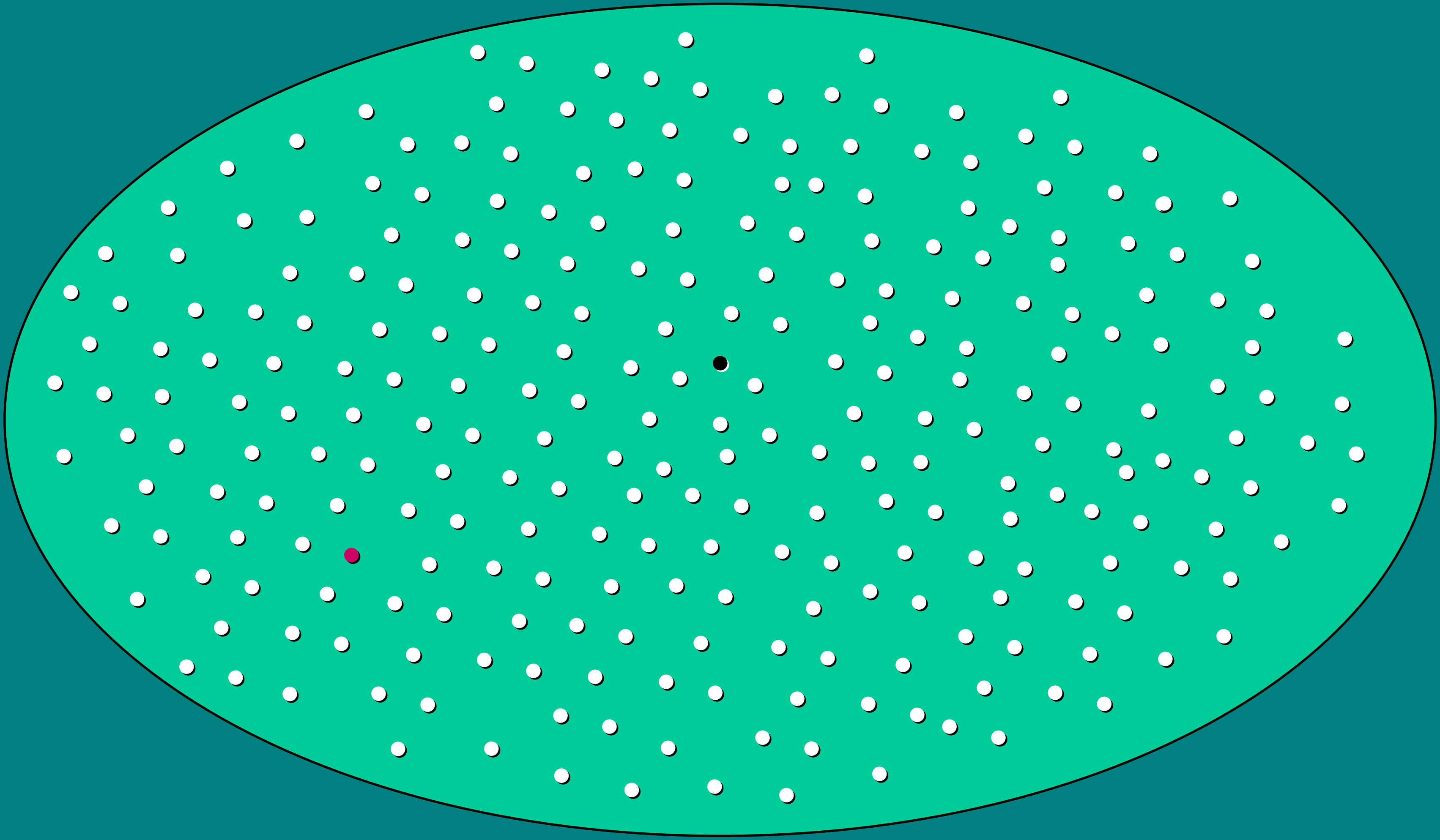
d[phyto,t,1] = – 0.307 × phyto – 0.495 × zoo
+ 0.411 × phyto

d[zoo,t,1] = – 0.251 × zoo + 0.615 × 0.495 × zoo

d[residue,t,1] = 0.307 × phyto + 0.251 × zoo
+ 0.385 × 0.495 × zoo – 0.005 × residue

d[nitro,t,1] = – 0.098 × 0.411 × phyto + 0.005 × residue

A Space of Ecosystem Models



Model revision requires ways to constrain search through this space.

Phytoplankton Loss in the Ross Sea

model RossSeaEcosystem

variables: phyto, zoo, nitro, residue
observables: phyto, nitro

$\boxed{d[\text{phyto},t,1] = -\ 0.307 \times \text{phyto}} -\ 0.495 \times \text{zoo} + 0.411 \times \text{phyto}$

$d[\text{zoo},t,1] = -\ 0.251 \times \text{zoo} + 0.615 \times 0.495 \times \text{zoo}$

$\boxed{d[\text{residue},t,1] = 0.307 \times \text{phyto}} + 0.251 \times \text{zoo} + 0.385 \times 0.495 \times \text{zoo} -\ 0.005 \times \text{residue}$

$d[\text{nitro},t,1] = -\ 0.098 \times 0.411 \times \text{phyto} + 0.005 \times \text{residue}$

Phytoplankton loss is a *process* that affects two variables; no model should include one influence without the other.

Grazing in the Ross Sea Ecosystem

model RossSeaEcosystem

variables: phyto, zoo, nitro, residue
observables: phyto, nitro

d[phyto,t,1] = – 0.307 × phyto – 0.495 × zoo + 0.411 × phyto
d[zoo,t,1] = – 0.251 × zoo + 0.615 × 0.495 × zoo
d[residue,t,1] = 0.307 × phyto + 0.251 × zoo + 0.385 × 0.495 × zoo – 0.005 × residue
d[nitro,t,1] = – 0.098 × 0.411 × phyto + 0.005 × residue

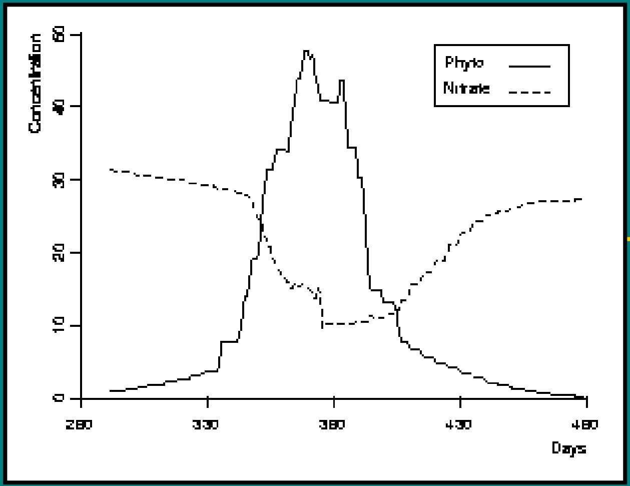
We can view an ecosystem model as a set of processes that provide an alternative way to encode its assumptions.

Process Model of Ross Sea Ecosystem

```
model RossSeaEcosystem

variables: phyto, zoo, nitro, residue
observables: phyto, nitro

process phyto_loss
  equations: d[phyto,t,1] = – 0.307 × phyto
             d[residue,t,1] = 0.307 × phyto

process zoo_loss
  equations: d[zoo,t,1] = – 0.251 × zoo
             d[residue,t,1] = 0.251 × zoo

process zoo_phyto_grazing
  equations: d[zoo,t,1] = 0.615 × 0.495 × zoo
             d[residue,t,1] = 0.385 × 0.495 × zoo
             d[phyto,t,1] = – 0.495 × zoo

process nitro_uptake
  equations: d[phyto,t,1] = 0.411 × phyto
             d[nitro,t,1] = – 0.098 × 0.411 × phyto

process nitro_remineralization;
  equations: d[nitro,t,1] = 0.005 × residue
             d[residue,t,1 ] = – 0.005 × residue
```

Inductive Revision of Process Models

observations



revised model

model RossSeaEcosystem

variables: phyto, zoo, nitro, residue
observables: phyto, nitro

d[phyto,t,1] = – 0.307 × phyto – 0.495 × zoo
+ 0.411 × phyto

d[zoo,t,1] = – 0.251 × zoo + 0.615 × 0.495 × zoo

d[residue,t,1] = 0.307 × phyto + 0.251 × zoo
+ 0.385 × 0.495 × zoo – 0.005 × residue

d[nitro,t,1] = – 0.098 × 0.411 × phyto + 0.005 × residue

Revision

model RossSeaEcosystem

variables: phyto, zoo, nitro, residue
observables: phyto, nitro

d[phyto,t,1] = – 0.307 × phyto – 0.495 × zoo
+ 0.411 × phyto

d[zoo,t,1] = – 0.251 × zoo + 0.615 × 0.495 × zoo

d[residue,t,1] = 0.307 × phyto + 0.251 × zoo
+ 0.385 × 0.495 × zoo – 0.005 × residue

d[nitro,t,1] = – 0.098 × 0.411 × phyto + 0.005 × residue

initial model

process exponential_growth
variables: P {population}
equations: d[P,t] = [0, 1,∞] × P

process logistic_growth
variables: P {population}
equations: d[P,t] = [0, 1, ∞] × P × (1 – P / [0, 1, ∞])

process constant_inflow
variables: I {inorganic_nutrient}
equations: d[I,t] = [0, 1, ∞]

process consumption
variables: P1 {population}, P2 {population}, nutrient_P2
equations: d[P1,t] = [0, 1, ∞] × P1 × nutrient_P2,
d[P2,t] = – [0, 1, ∞] × P1 × nutrient_P2

process no_saturation
variables: P {number}, nutrient_P {number}
equations: nutrient_P = P

process saturation
variables: P {number}, nutrient_P {number}
equations: nutrient_P = P / (P + [0, 1, ∞])

generic processes

Generic Processes for Aquatic Ecosystems

```
generic process exponential_loss
  variables: S{species}, D{detritus}
  parameters: $\alpha$ [0, 1]
  equations:   $d[S,t,1] = -1 \times \alpha \times S$
               $d[D,t,1] = \alpha \times S$

generic process grazing
  variables: S1{species}, S2{species}, D{detritus}
  parameters: $\rho$ [0, 1], $\gamma$ [0, 1]
  equations:   $d[S1,t,1] = \gamma \times \rho \times S1$
               $d[D,t,1] = (1 - \gamma) \times \rho \times S1$
               $d[S2,t,1] = -1 \times \rho \times S1$

generic process nutrient_uptake
  variables: S{species}, N{nutrient}
  parameters: $\tau$ [0, $\infty$], $\beta$ [0, 1], $\mu$ [0, 1]
  conditions:  $N > \tau$
  equations:   $d[S,t,1] = \mu \times S$
               $d[N,t,1] = -1 \times \beta \times \mu \times S$

generic process remineralization
  variables: N{nutrient}, D{detritus}
  parameters: $\pi$ [0, 1]
  equations:   $d[N, t,1] = \pi \times D$
               $d[D, t,1] = -1 \times \pi \times D$

generic process constant_inflow
  variables: N{nutrient}
  parameters: $\nu$ [0, 1]
  equations:   $d[N,t,1] = \nu$
```

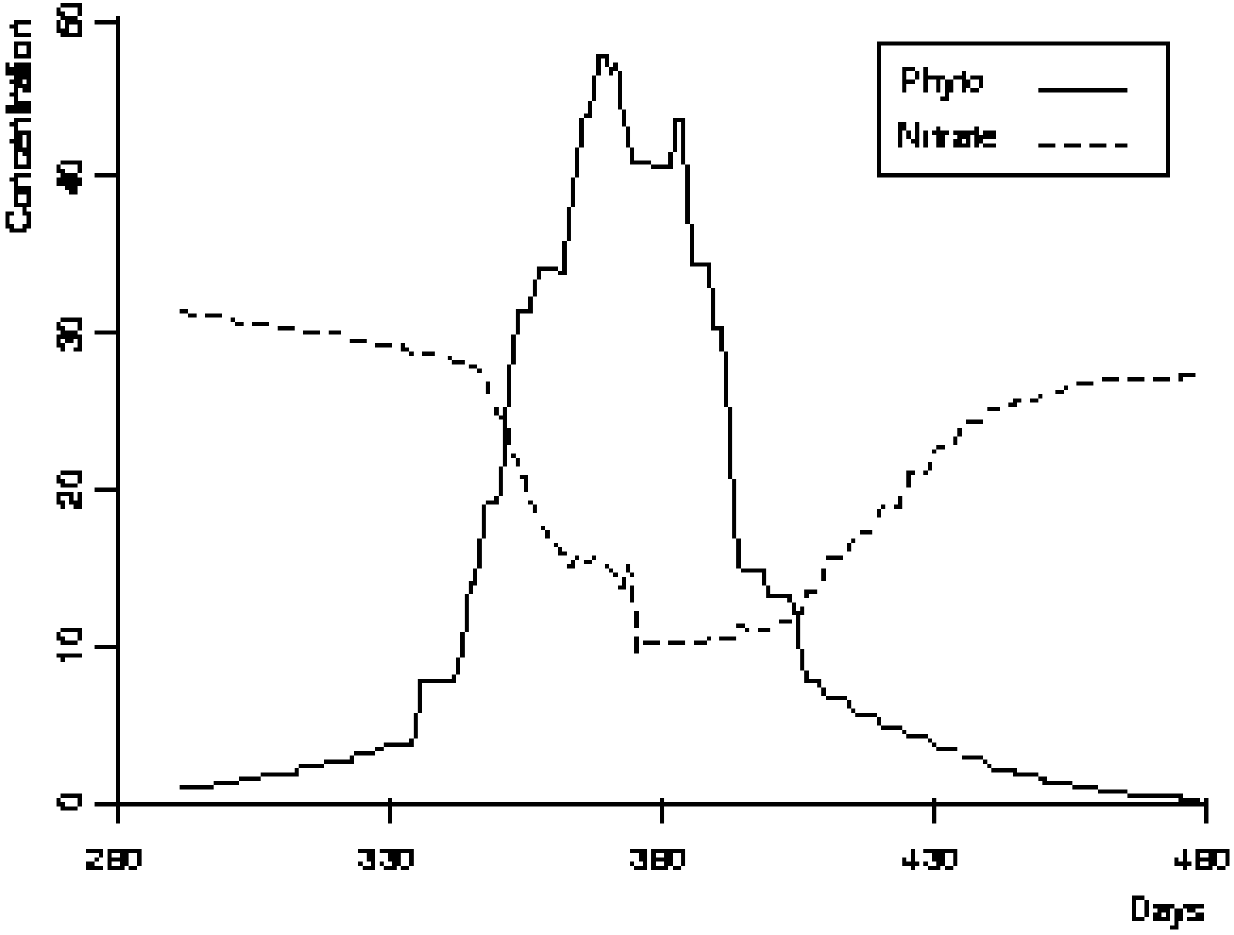
A Method for Process Model Revision

We have implemented RPM, an algorithm that revises an initial process model in four main stages:

1. Find all ways to instantiate available generic processes with specific variables, subject to type constraints;
2. Generate candidate model structures by deleting the current processes and adding new ones, subject to complexity limits;
3. For each generic model, carry out search through parameter space to find good coefficients [difficult];
4. Return a list of revised models ordered by their overall scores.

The evaluation metric can be squared error or description length based on error and distance from the initial model.

Observations from the Ross Sea



Revised Model of Ross Sea Ecosystem

model RossSeaEcosystem

variables: phyto, zoo, nitro, residue, light, G, growth_rate, nitro_rate, light_rate
observables: phyto, nitro, light

d[phyto,t,1] = – 0.307 × phyto *– G × zoo* + growth_rate × phyto

d[zoo,t,1] = 0.615 × G × zoo

d[residue,t,1] = 0.307 × phyto *+ 0.385 × G × zoo – 0.083 × residue*

d[nitro,t,1] = – 1 × n_to_c × growth_rate × phyto *+ 0.083 × n_to_c × residue*

G = 0.415 × (1 – exp(– 1 × 0.27 × phyto)

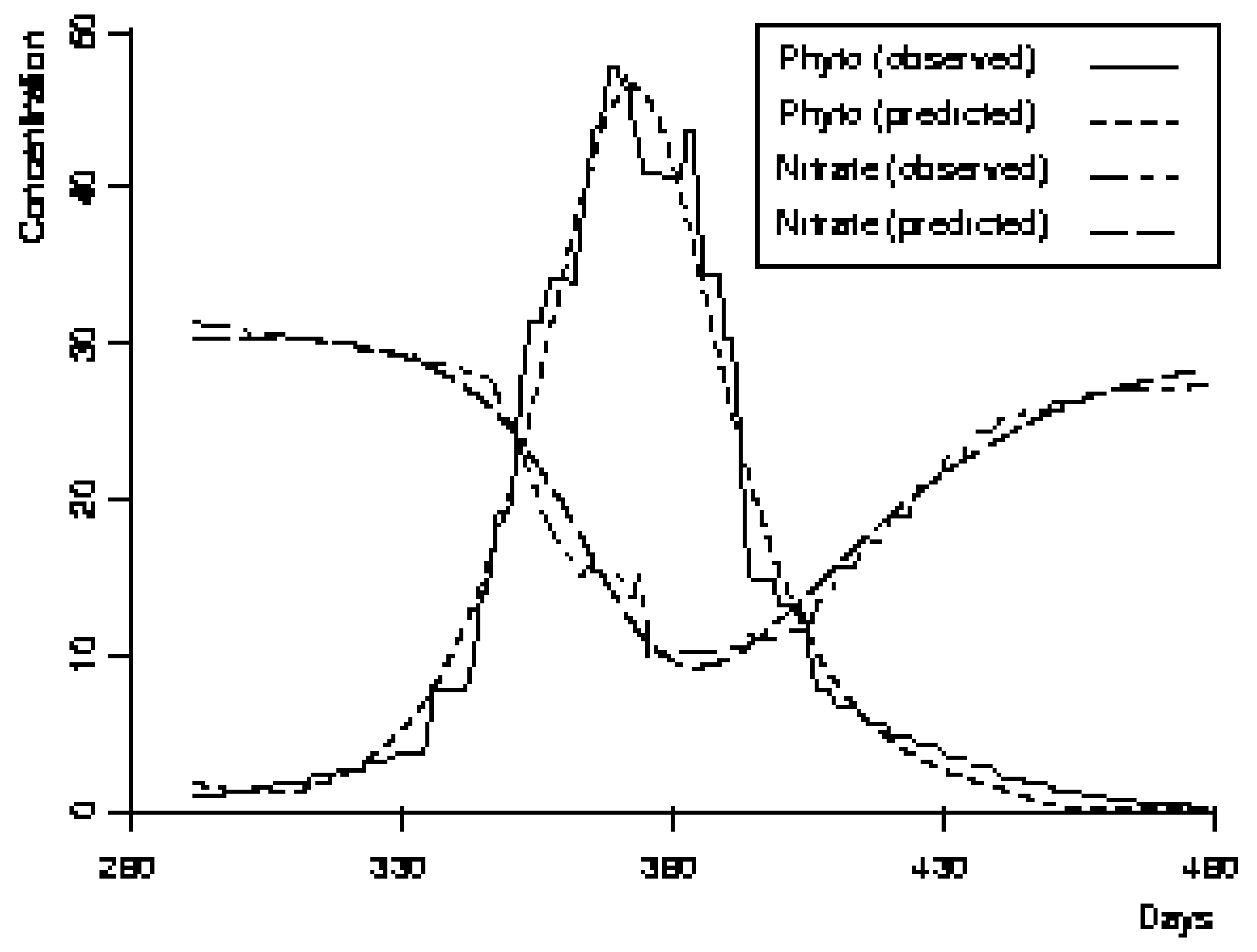
growth_rate = r_max × min(nitro_rate, light_rate)

nitro_rate = nitro / (nitro + 4.33)

light_rate = light / (light + 11.67)

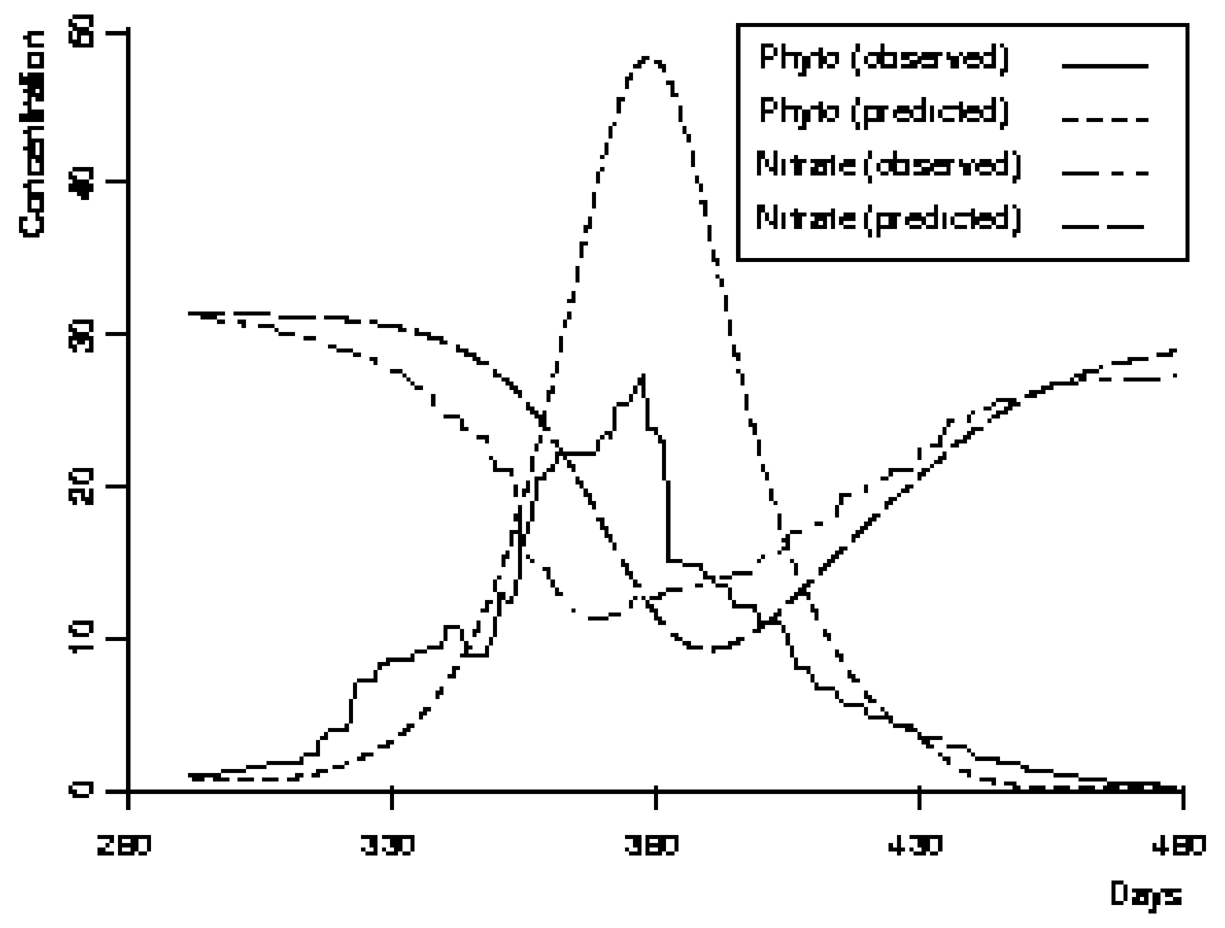
n_to_c = 0.251, r_max = 0.194, remin_rate = 0.0676

Initial Results on Ross Sea Training Data



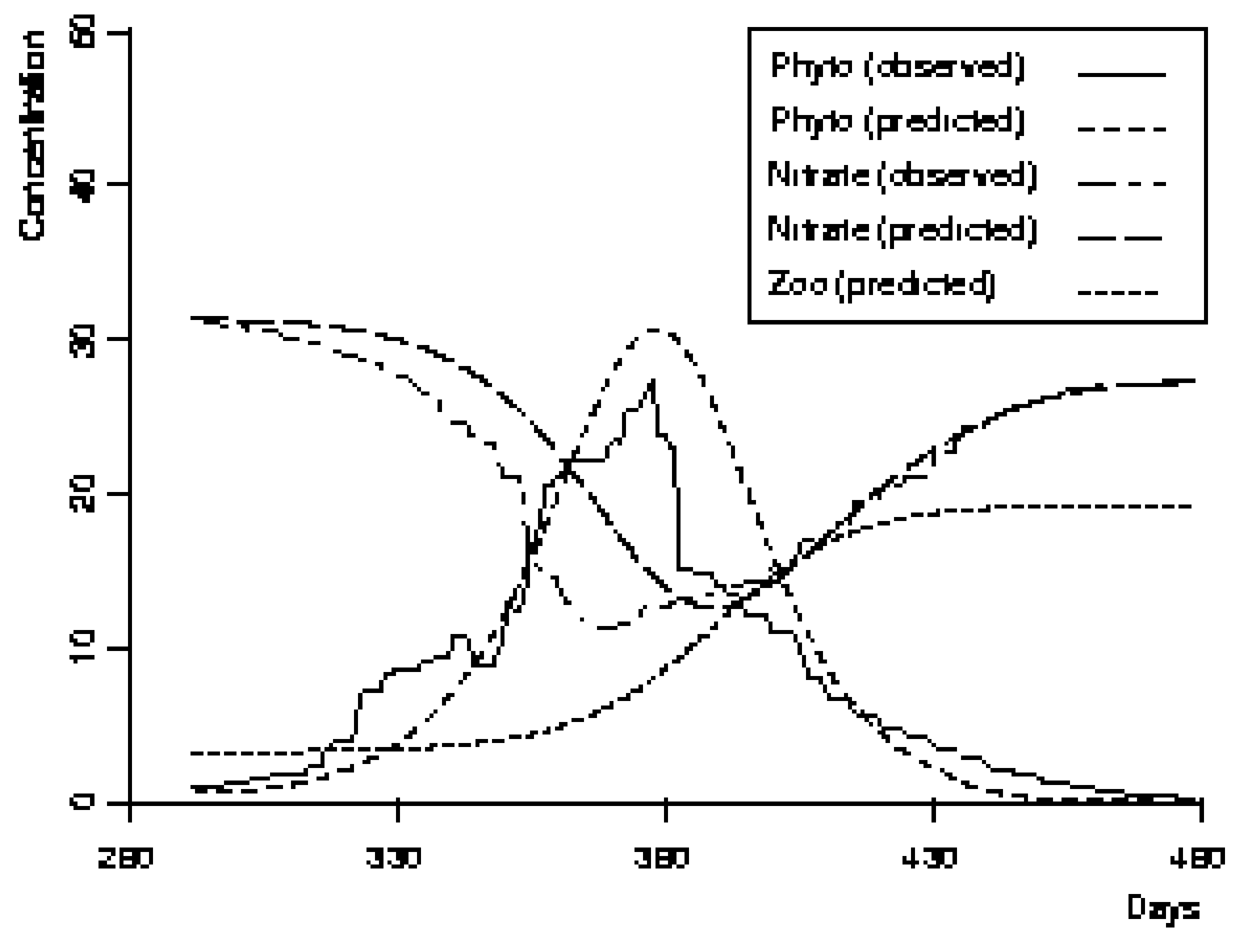
The best revised model reproduces the observations quite well.

Initial Results on Ross Sea Test Data



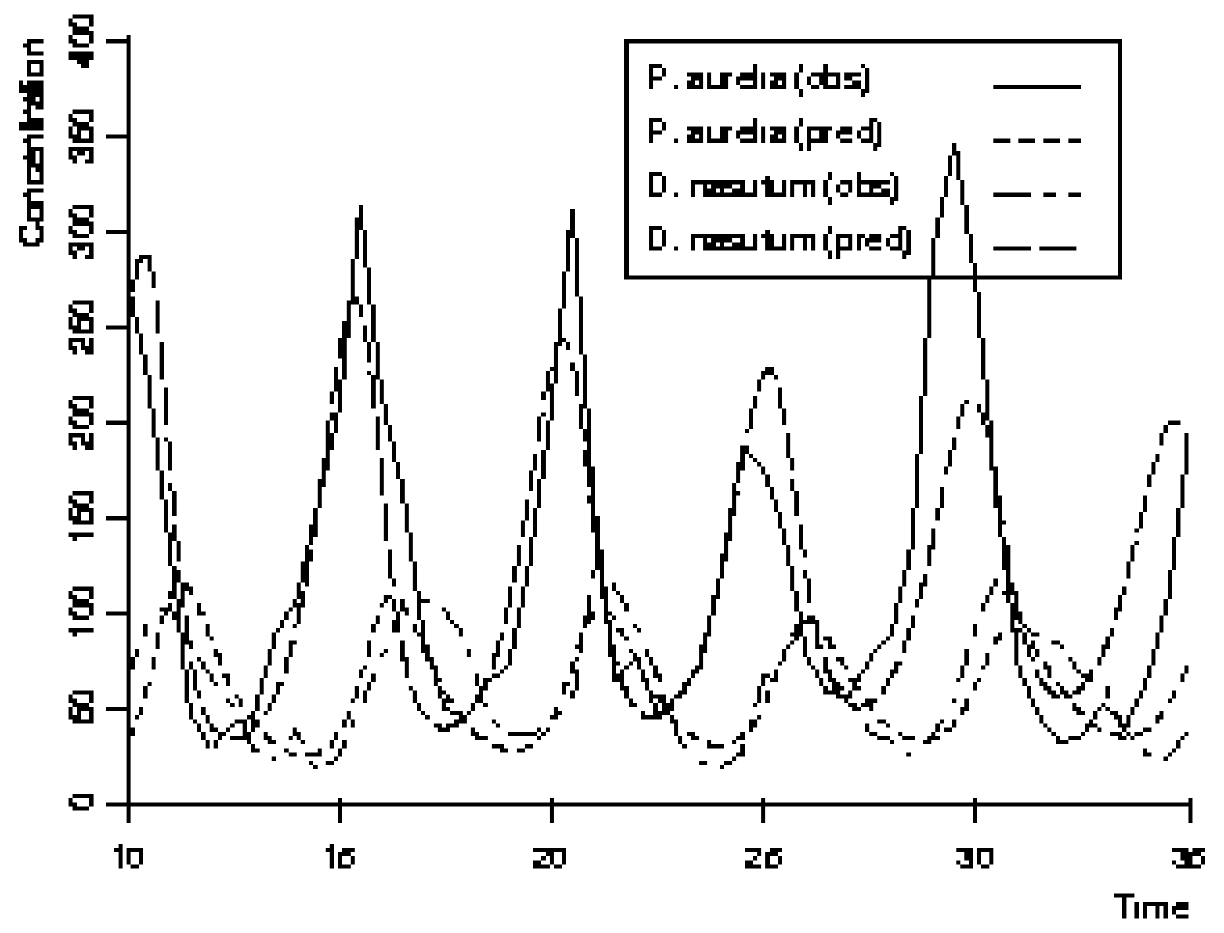
But the model predicts nearly the same behavior for both years.

Revised Results on Ross Sea Test Data

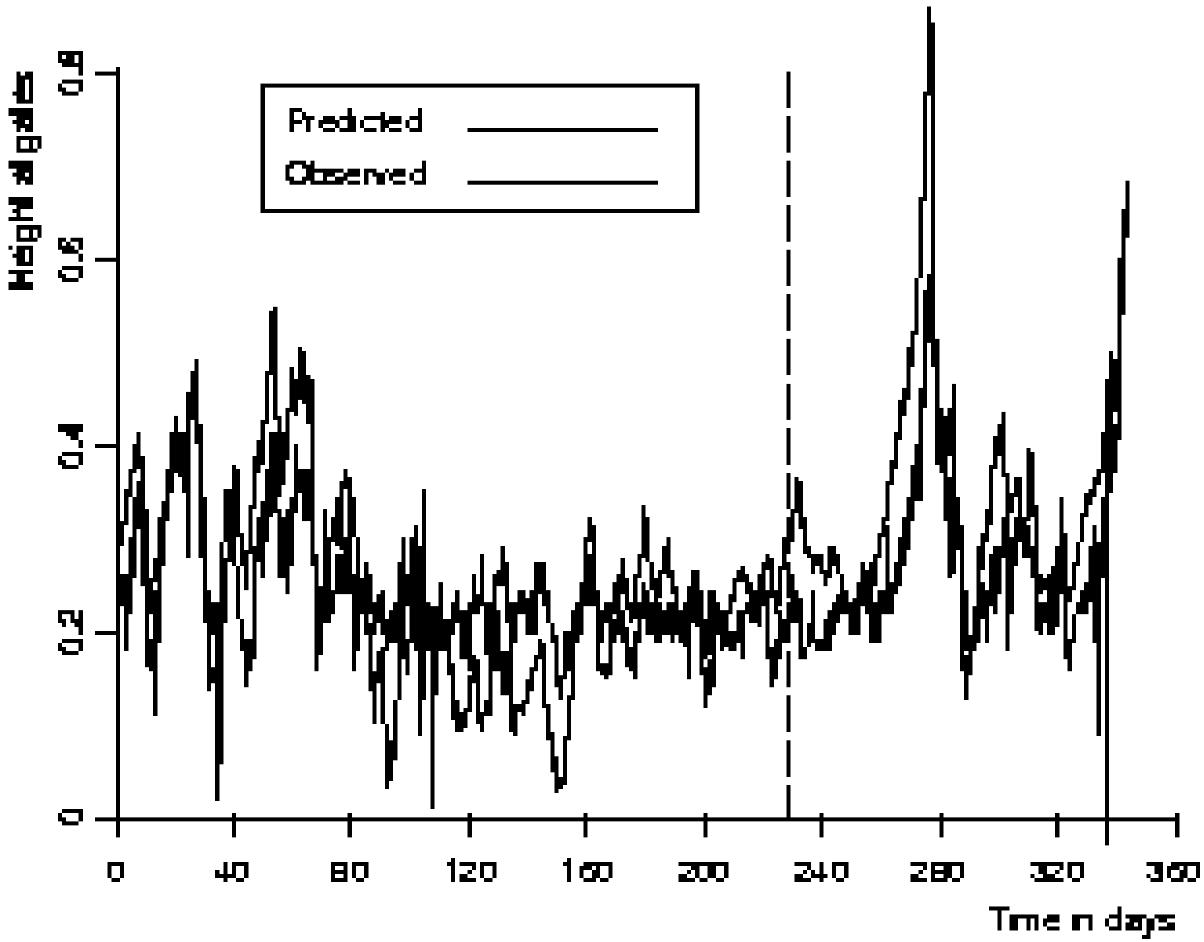


Refitting initial values for zooplankton gives better generalization.

Results on Data from Protist Study



Results on Data from Rinkobing Fjord



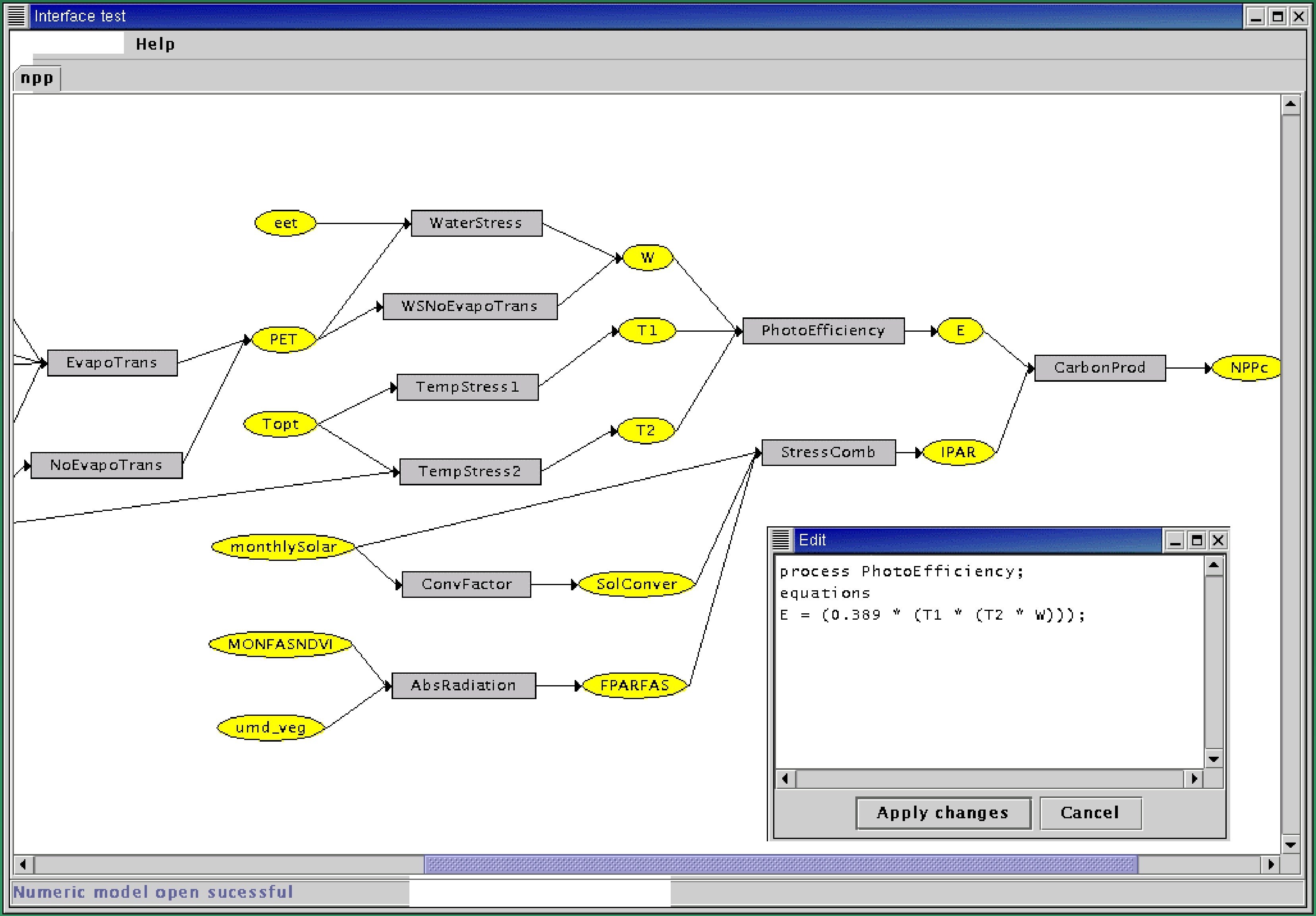
Interfacing with Scientists

Because few scientists want to be replaced, we are developing PROMETHEUS, an interactive environment that lets users:

- specify a quantitative process model of the target system;
- display and edit the model’s structure and details graphically;
- simulate the model’s behavior over time and situations;
- compare the model’s predicted behavior to observations;
- invoke a revision module in response to detected anomalies.

The environment offers computational assistance in forming and evaluating models but lets the user retain control.

Viewing and Editing a Process Model



```
process PhotoEfficiency;
equations
E = (0.389 * (T1 * (T2 * W)));
```

Numeric model open sucessful

Intellectual Influences

Our approach to computational discovery incorporates ideas from many traditions:

- computational scientific discovery (e.g., Langley et al., 1983);
- theory revision in machine learning (e.g., Towell, 1991);
- qualitative physics and simulation (e.g., Forbus, 1984);
- languages for scientific simulation (e.g., STELLA, MATLAB);
- interactive tools for data analysis (e.g., Schneiderman, 2001).

Our work combines ideas from machine learning, AI, programming languages, and human-computer interaction.

Directions for Future Research

Despite our progress to date, we need further work in order to:

- produce additional results on other ecosystem modeling tasks
- develop improved methods for fitting model parameters
- implement heuristic methods for searching the structure space
- utilize knowledge of subsystems to further constrain search
- augment the modeling environment to make it more usable

Process modeling has great potential to aid model development in environmental science.

Contributions of the Research

In summary, our work on computational discovery has produced:

- a new formalism for representing scientific process models;
- an encoding for background knowledge as generic processes;
- a method for revising process models using time-series data;
- an interactive environment for model construction/utilization.

We have demonstrated this approach to model revision on both ecosystem modeling and an environmental domain.

The PROMETHEUS modeling/revision environment is available at:

http://www.isle.org/process.html

End of Presentation