

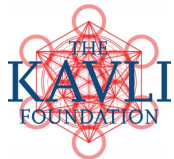
Knowledge-Powered Data Science for Integrated Modeling in Geosciences

Yolanda Gil

Information Sciences Institute
Department of Computer Science
Spatial Sciences Institute

University of Southern California

<http://www.isi.edu/~gil>
@yolandagil
gil@isi.edu



IARPA
BE THE FUTURE



The Future is Here: Knowledge-Powered Data Science

- Focus shifted from data to **models**
 - Model characterization, reuse, and integration
- Need to **incorporate model-centered science knowledge** about phenomena and context
 - Knowledge about **physical, geological, chemical, biological, ecological, and anthropomorphic factors**
 - Knowledge about the user **goals and context**
- This would enable novel forms of reasoning, integrating, visualizing, managing, learning, and discovery with geosciences data

REVIEW ARTICLES

Intelligent Systems for Geosciences: An Essential Research Agenda

By Yolanda Gil, Suzanne A. Pierce, Hassan Babaie, Arindam Banerjee, Kirk Borne, Gary Bust, Michelle Cheatham, Imme Ebert-phoff, Carla Gomes, Mary Hill, John Horel, Leslie Hsu, Jim Kinter, Craig Knoblock, David Krum, Vipin Kumar, Pierre Lermusiaux, Yan Liu, Chris North, Victor Pankrati, Shanan Peters, Beth Plale, Allen Pope, Sai Ravela, Juan Restrepo, Aaron Ridley, Hanan Samet, Shashi Shekhar

Communications of the ACM, January 2019, Vol. 62 No. 1, Pages 76-84

10.1145/3192335

[Comments](#)

VIEW AS:



SHARE:



SIGN IN for Full Access

User Name

Password

» [Forgot Password?](#)

» [Create an ACM Web Account](#)



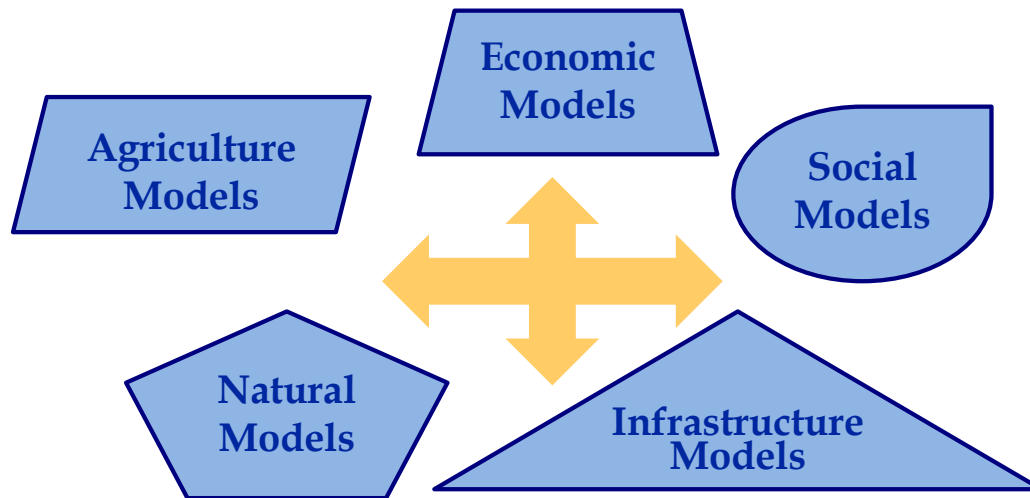
Many aspects of geosciences pose no systems research. Geoscience data is to be uncertain, intermittent, sparse scale. Geosciences processes and object spatiotemporal boundaries. The lack model evaluation, testing, and computing these challenges requires breakthrough transform intelligent systems, while geosciences in turn. Although there beneficial interactions between the geosciences communities,^{4,12} the potential research in intelligent systems for ge



Outline

- The need for integrated modeling in geosciences
- Diversity of models across disciplines
- MINT: knowledge-powered data science for integrated modeling
- Intelligent systems for geosciences

A Grand Challenge for Geospatial Data Science



Integrated Modeling

- Increased demand
- Months or years
- Currently a craft

- Integrating pairs of natural models is very hard
 - Eg, surface water models + ground water models
- Integrating natural and human models is even harder
 - Eg, agriculture + socioeconomic models of human behavior

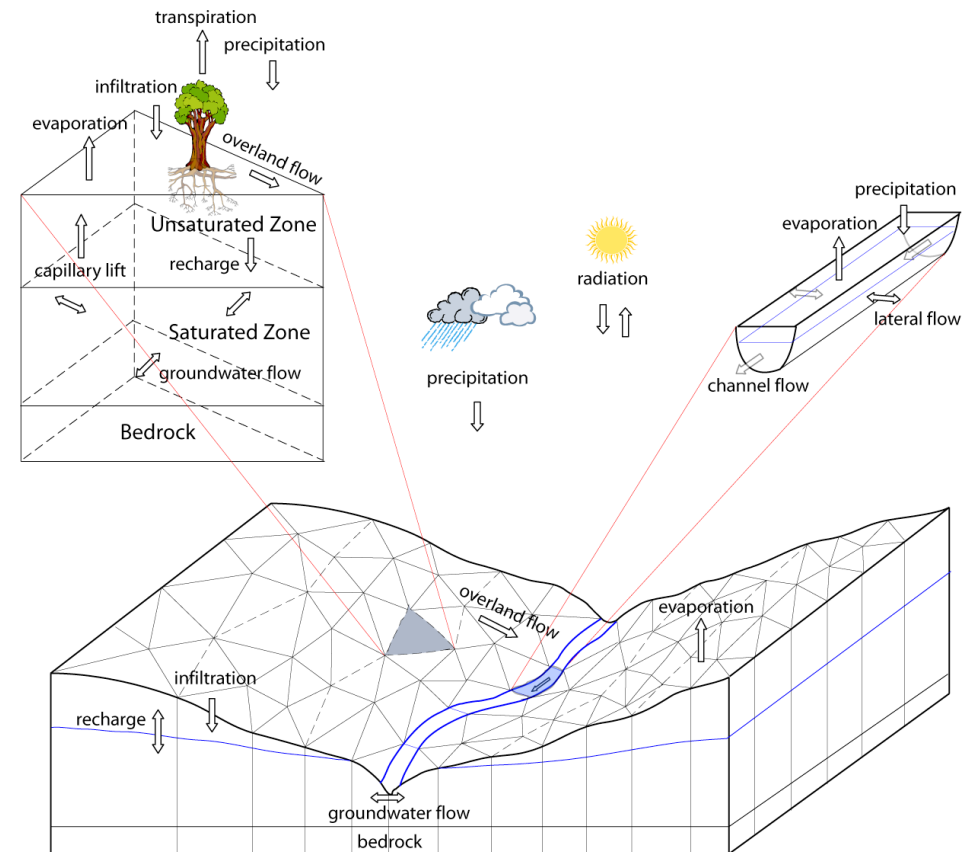
Outline

- The need for integrated modeling in geosciences
- Diversity of models across disciplines
- MINT: knowledge-powered data science for integrated modeling
- Intelligent systems for geosciences

A Representative Hydrology Model

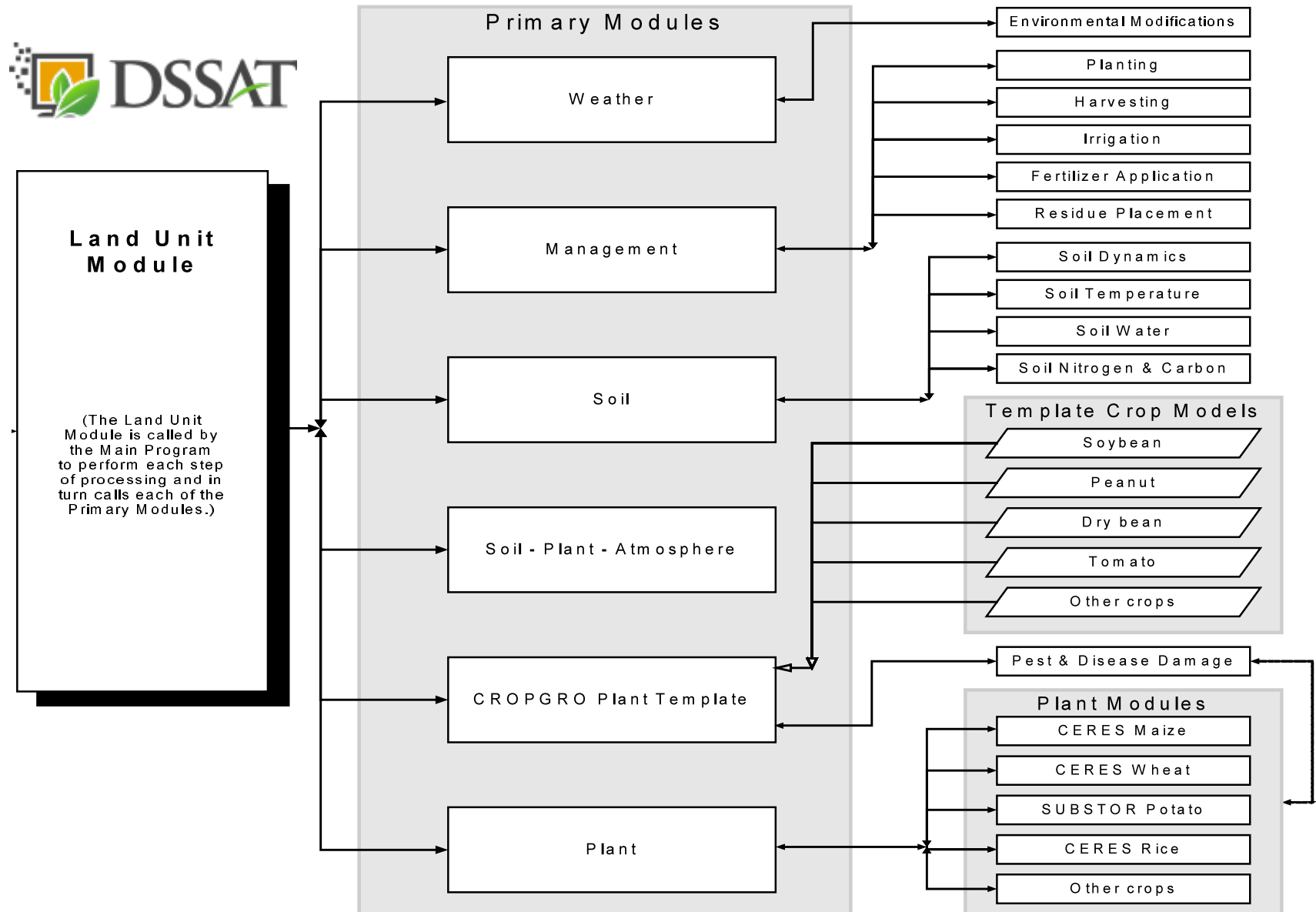
■ PIHM hydrology model

- Needs data on *essential terrestrial variables*: slope, vegetation, etc.
- Generates discharge, flooding

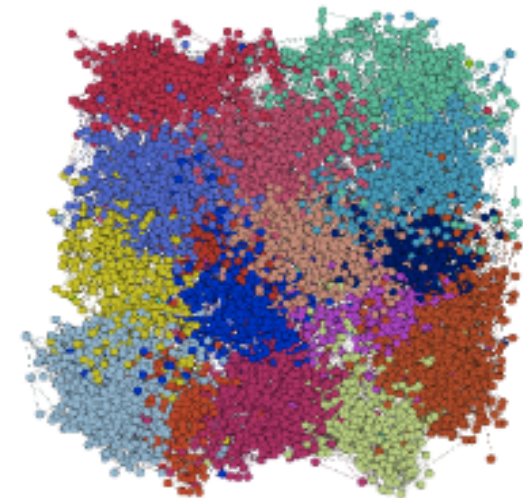
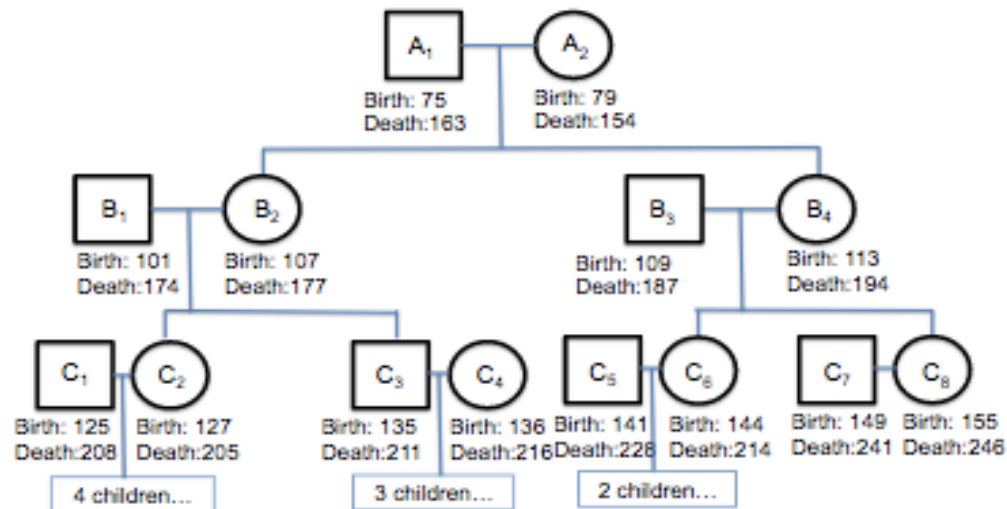
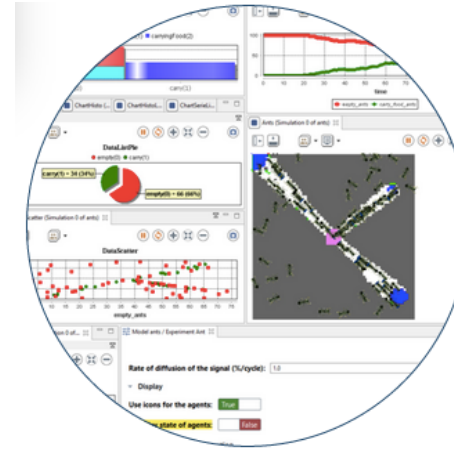
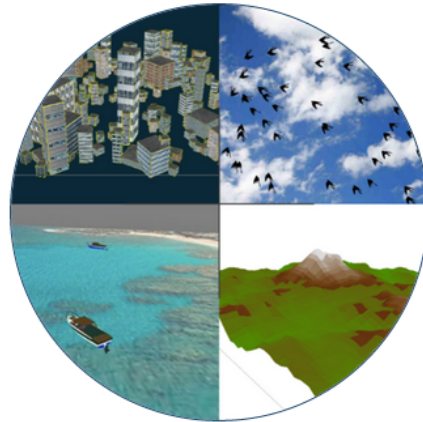


Process	Governing equation/model	Original governing equations	Semi-discrete form ODEs
Channel Routing	St. Venant Equation	$\frac{\partial h}{\partial t} + \frac{\partial(hv)}{\partial x} = q$	$\left(\frac{dh}{dt} = P_s - \sum Q_R + \sum Q_{in} + Q_{out} - E_s \right)_i$
Overland Flow	St. Venant Equation	$\frac{\partial h}{\partial t} + \frac{\partial(hv)}{\partial x} + \frac{\partial(hv)}{\partial y} = q$	$\left(\frac{dh}{dt} = P_s - I - E_s - Q_{in} + \sum_{j=1}^J Q_s^j \right)_i$
Unsaturated Flow	Richard Equation	$C(\psi) \frac{\partial \psi}{\partial t} = \nabla \cdot (K(\psi) \nabla(\psi + Z))$	$\left(\frac{d\psi}{dt} = I - q^e - ET_s \right)_i$
Groundwater Flow	Richard Equation	$C(\psi) \frac{\partial \psi}{\partial t} = \nabla \cdot (K(\psi) \nabla(\psi + Z))$	$\left(\frac{d\psi}{dt} = q^a + \sum_{j=1}^J Q_s^j - Q_i + Q_{ex} \right)_i$
Interception	Bucket Model	$\frac{dS_L}{dt} = P - E_i - P_o$	$\left(\frac{dS_L}{dt} = P - E_i - P_o \right)_i$
Snow melt	ISNOBAL	$\frac{dS_{snow}}{dt} = P - E_{snow} - \Delta w$	$\left(\frac{dS_{snow}}{dt} = P - E_{snow} - \Delta w \right)_i$
Evapotranspiration	Pennman-Monteith Method	$ET_o = \frac{\Delta(R_n - G) + \rho_a C_p \frac{(e_s - e_a)}{r_a}}{\Delta + \gamma(1 + \frac{r_s}{r_a})}$	$\left(\frac{\Delta(R_n - G) + \rho_a C_p \frac{(e_s - e_a)}{r_a}}{\Delta + \gamma(1 + \frac{r_s}{r_a})} \right)_i$

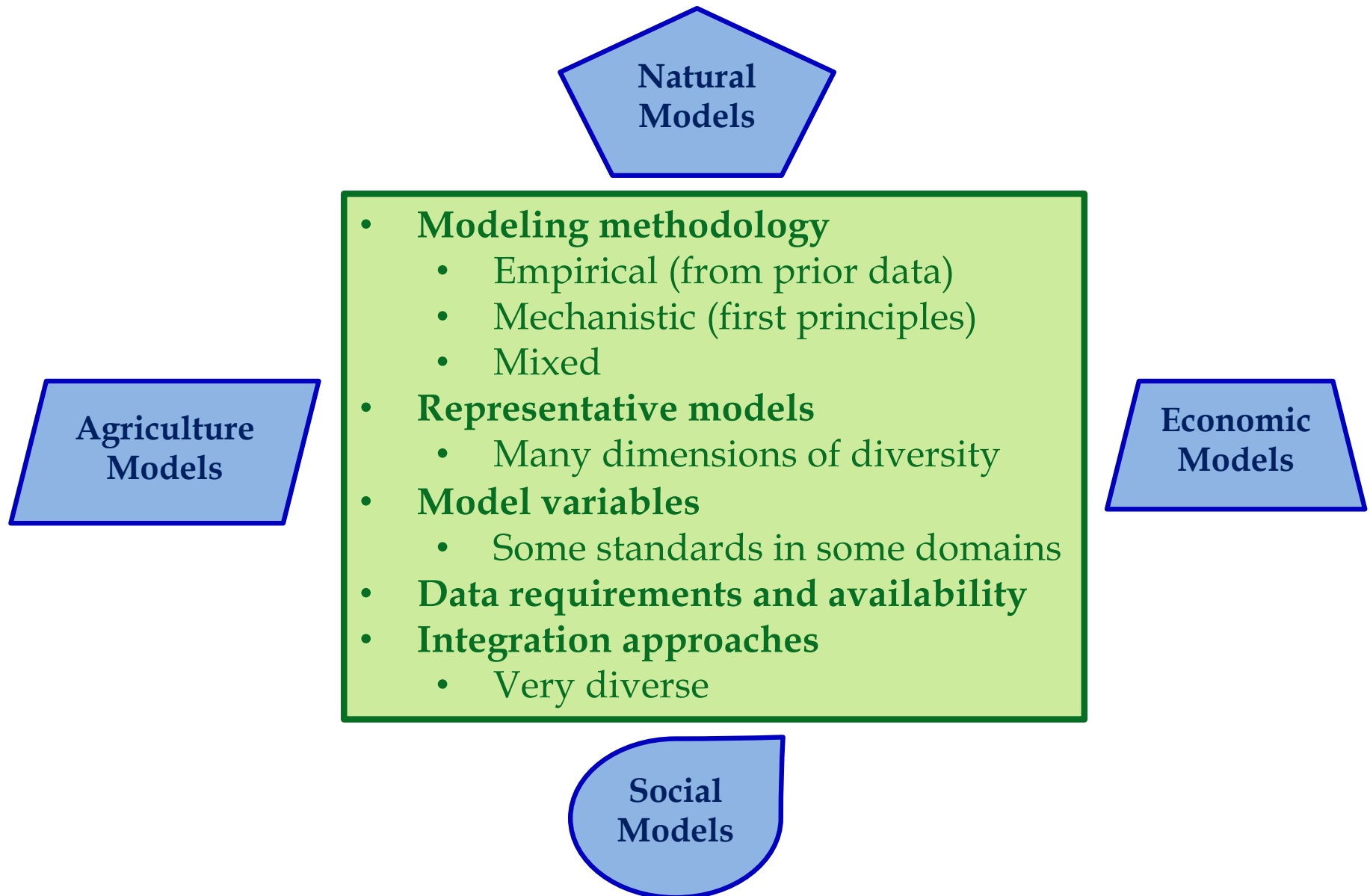
A Representative Agriculture Model



A Representative Social Model



Modeling in Different Disciplines: Diversity of Approaches



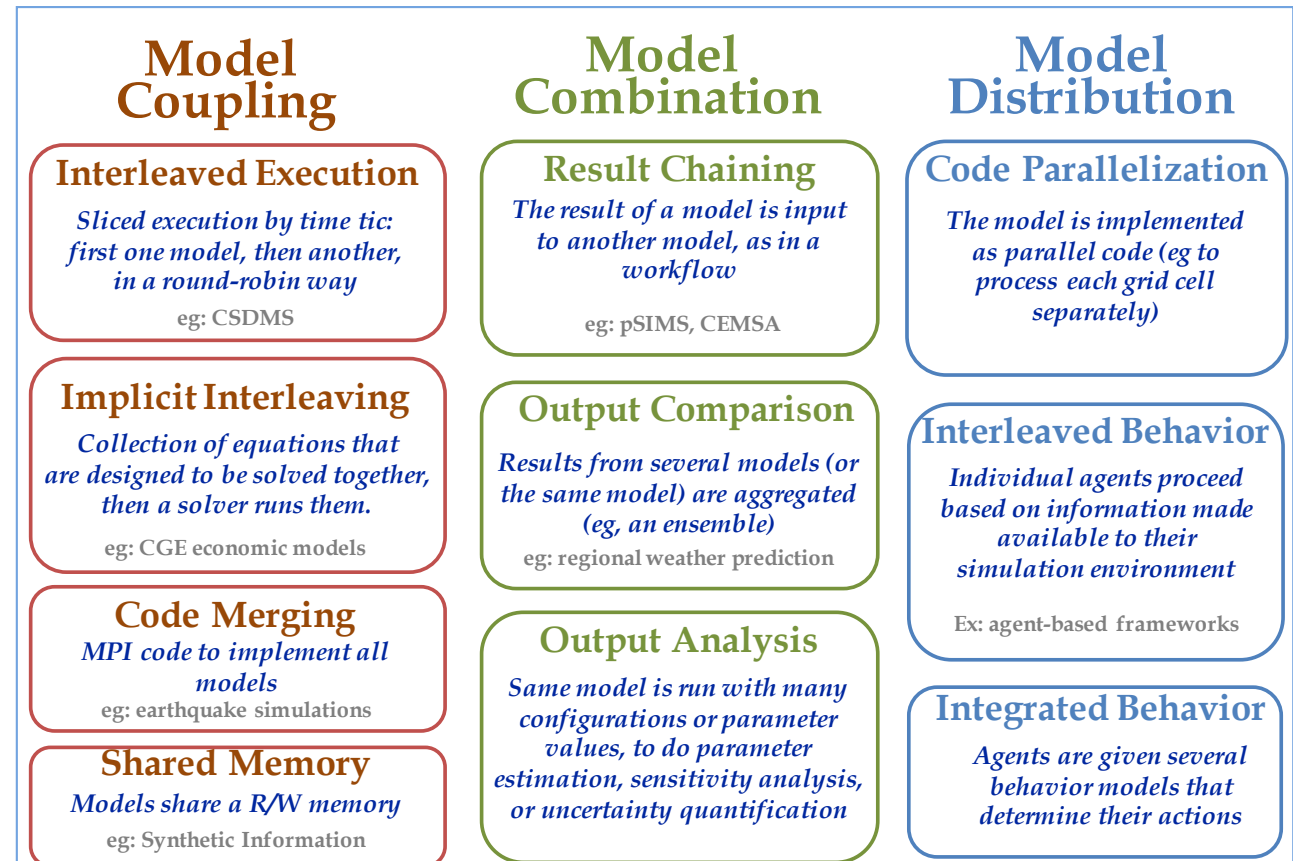
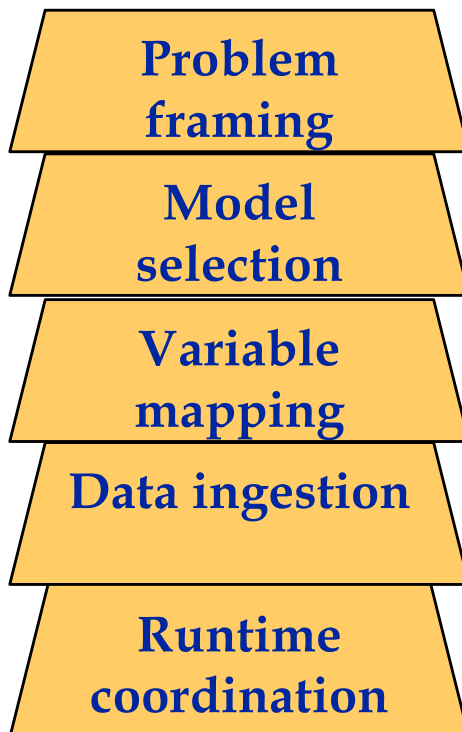
Integrated Modeling: Bridging Across Disciplines

Diversity across disciplines:

- Modeling approaches
- Representative models
- Model variables
- Integration frameworks



Mediation at many levels



How Can We Facilitate Model Integration?

Problem framing	Very diverse scope <i>Manual definition</i>
Model selection	Very diverse approaches <i>Manual selection</i>
Variable mapping	Very diverse variables <i>Manual mapping</i>
Data ingestion	Very diverse data needs <i>Disconnected tools</i>

Outline

- The need for integrated modeling in geosciences
- Diversity of models across disciplines
- MINT: knowledge-powered data science for integrated modeling
- Intelligent systems for geosciences

Knowledge-Powered Data Science for Integrated Modeling

Problem framing	Structured frameworks for scenario scoping
Model selection	Semantic descriptions of models and assumptions
Variable mapping	Ontologies of variables and relations
Data ingestion	Knowledge-guided information extraction and integration

MINT: Model INTegration [Gil et al iEMSs'18]

<http://mint-project.info>

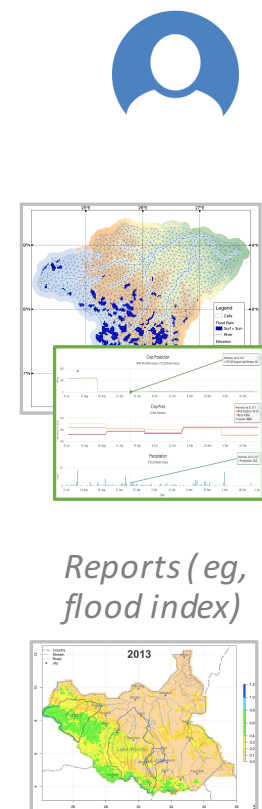
Collaboration with Daniel Garijo, Deborah Khider, Craig Knoblock, Ewa Deelman, Rafael Ferreira (USC/ISI), Vipin Kumar (UM), Scott Peckham (CU), Chris Duffy & Armen Kemanian (PSU), Kelly Cobourn (VT), Suzanne Pierce (UT)

User creates scenarios and guides the model integration process

Causal Analysis Graph (CAG) contains variables of interest

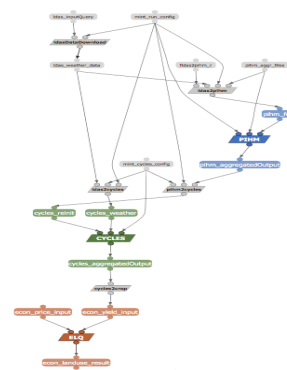
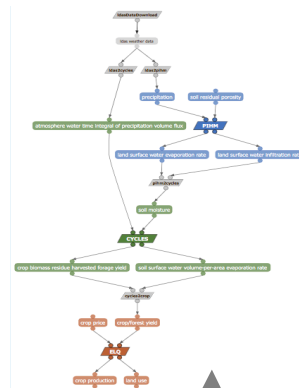
Model Graph has selected quantitative models for CAG variables

Workflow identifies data sources & includes data transformations

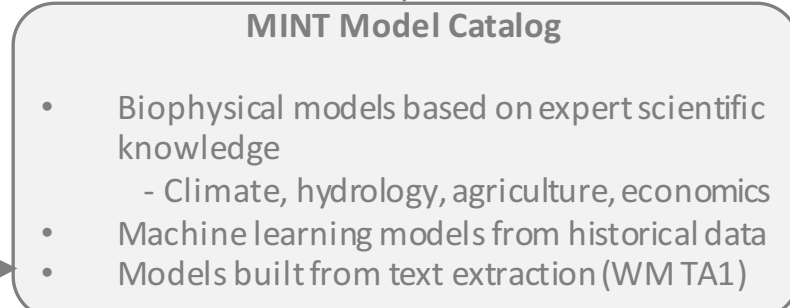
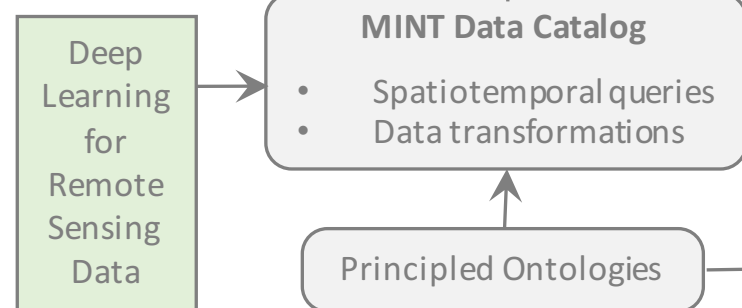
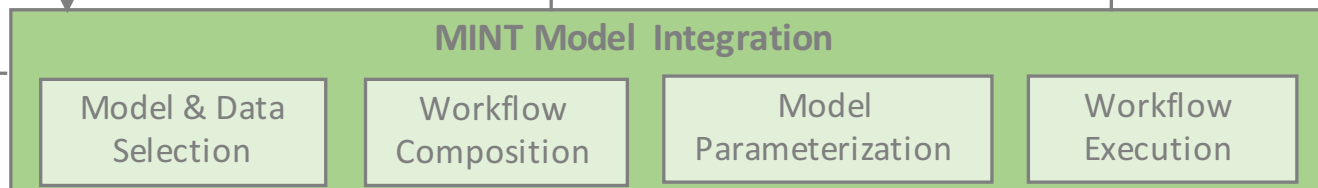


Reports (eg, flood index)

Uncertainty analysis (WM TA3)



Results and Provenance



Knowledge-Powered Data Science for Integrated Modeling in MINT

**Problem
framing**

Structured frameworks
for scenario scoping

**Model
selection**

Semantic descriptions
of models and
assumptions

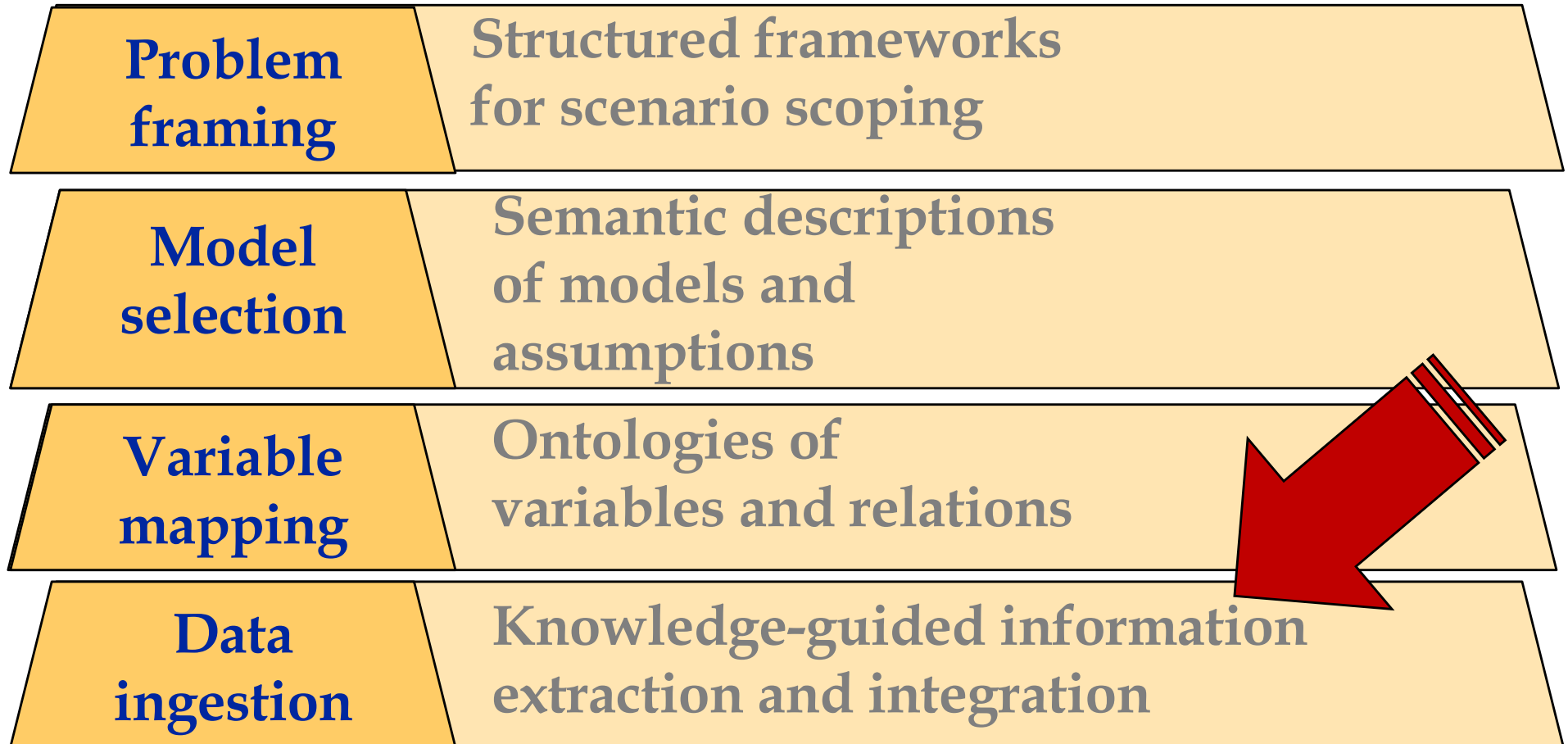
**Variable
mapping**

Ontologies of
variables and relations

**Data
ingestion**

Knowledge-guided information
extraction and integration

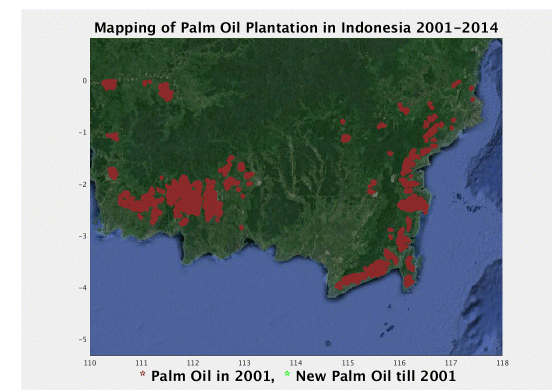
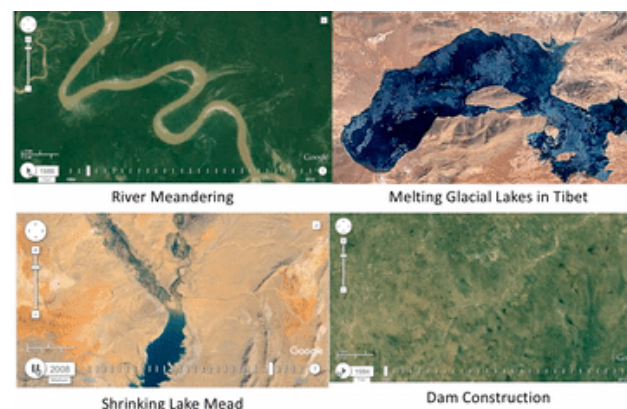
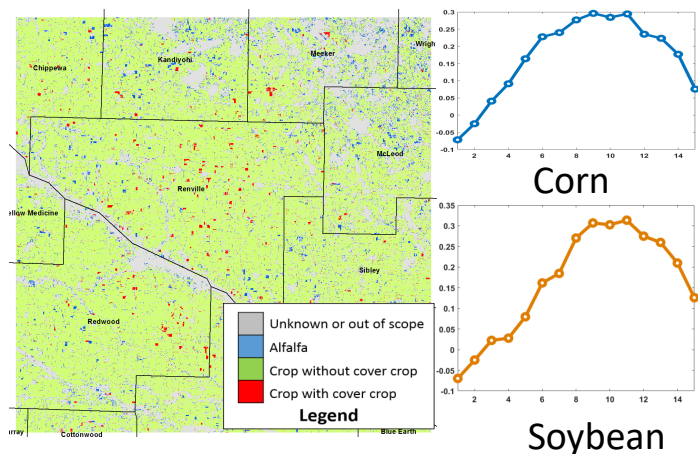
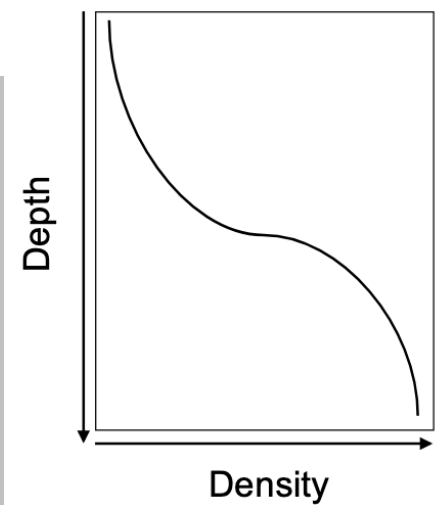
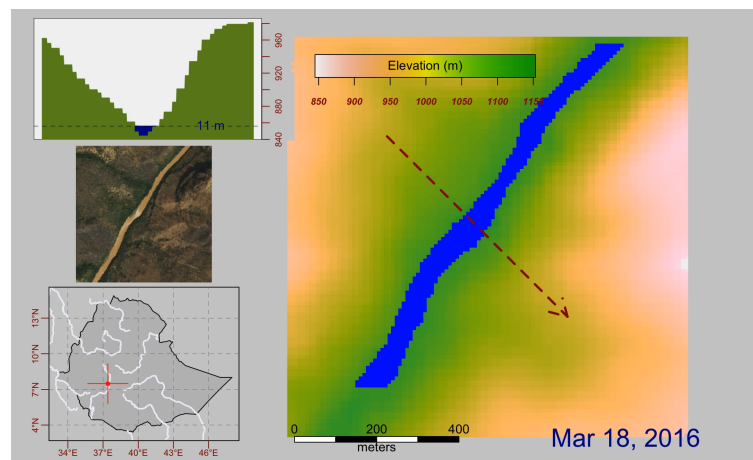
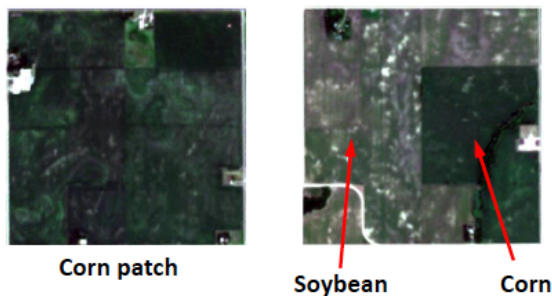
Knowledge-Powered Data Science for Integrated Modeling in MINT



TGDS: Theory-Guided Data Science

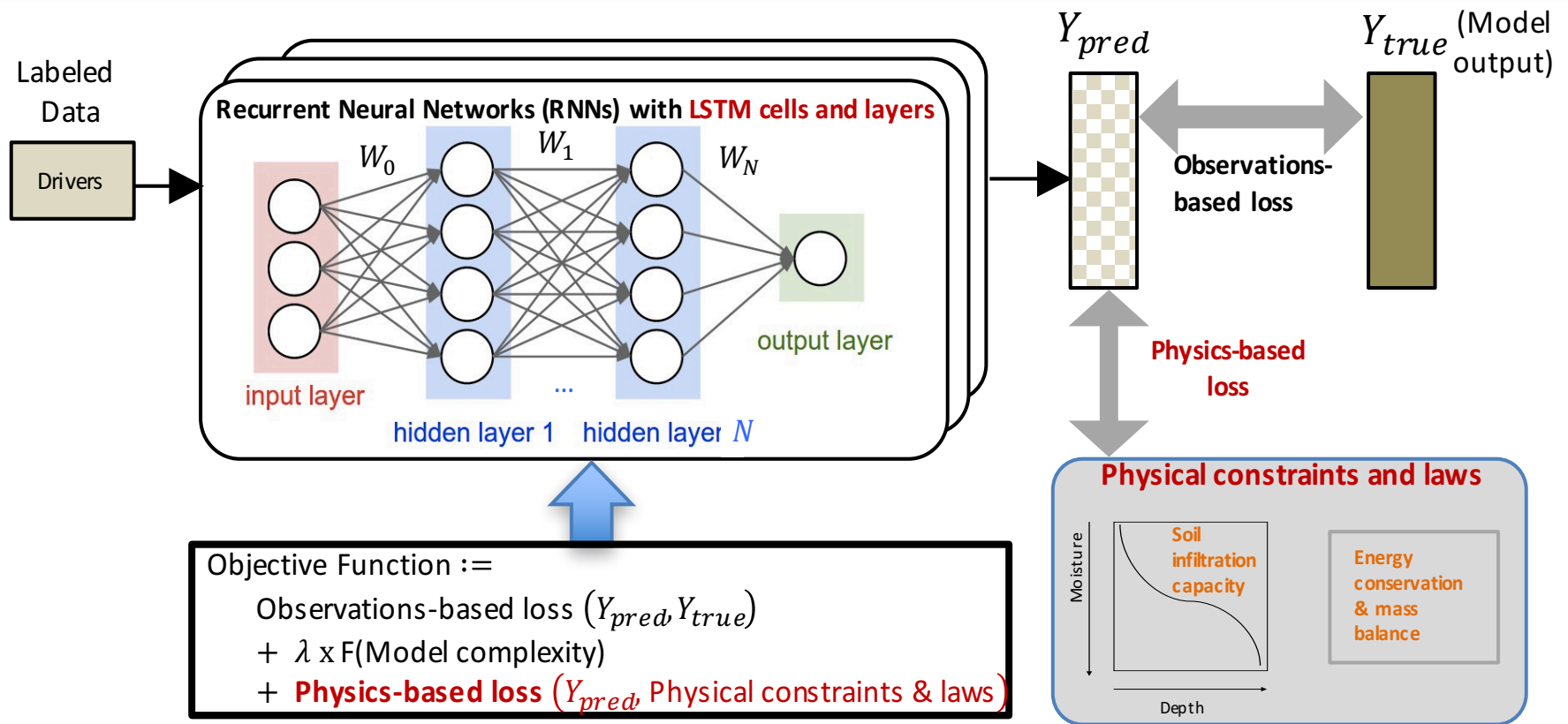
[Kumar et al 2017]

- Machine learning to generate physically consistent models
- Physics laws guide deep learning

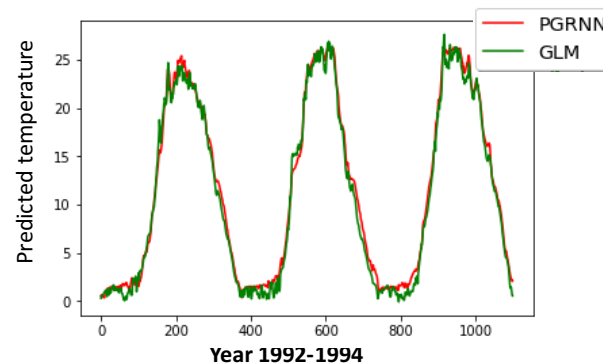


Physics-Guided Neural Networks (PGNNs)

[Karpatne et al 2017]



Karpatne et al., "Physics-guided neural networks (PGNN): An Application in Lake Temperature Modeling," arXiv: 1710.11431, 2017.



Automated Data Transformations

[Knoblock et al 2018]

wfp_food_prices_ethiopia

	A	B	C	D	E	F	G
1	date	cmname	unit	category	price	currency	country
2	#date	#item+name	#item+unit	#item+type	#value	#currency	#country+name
3	7/15/05	Sorghum - Wh	100 KG	cereals and t	238	ETB	Ethiopia
4	8/15/05	Sorghum - Wh	100 KG	cereals and t	250	ETB	Ethiopia
5	9/15/05	Sorghum - Wh	100 KG	cereals and t	248	ETB	Ethiopia
6	10/15/05	Sorghum - Wh	100 KG	cereals and t	233	ETB	Ethiopia
7	11/15/05	Sorghum - Wh	100 KG	cereals and t	252	ETB	Ethiopia

wfp_food_prices
D-REPR yml file

```
version: "1"
resources:
  default:
    type: csv
    delimiter: \t
variables:
  date: [2.., 0]
  crop: [2.., 1]
alignments:
  - type: dimension
    value: date:0 <=> crop:0
semantic_model:
  data_nodes:
    date: qb:Observation:1--sdmx-
    crop: qb:Observation:1--dcat-
    grain_yield: qb:Observation:1
```

Read

Filter

UnitTrans

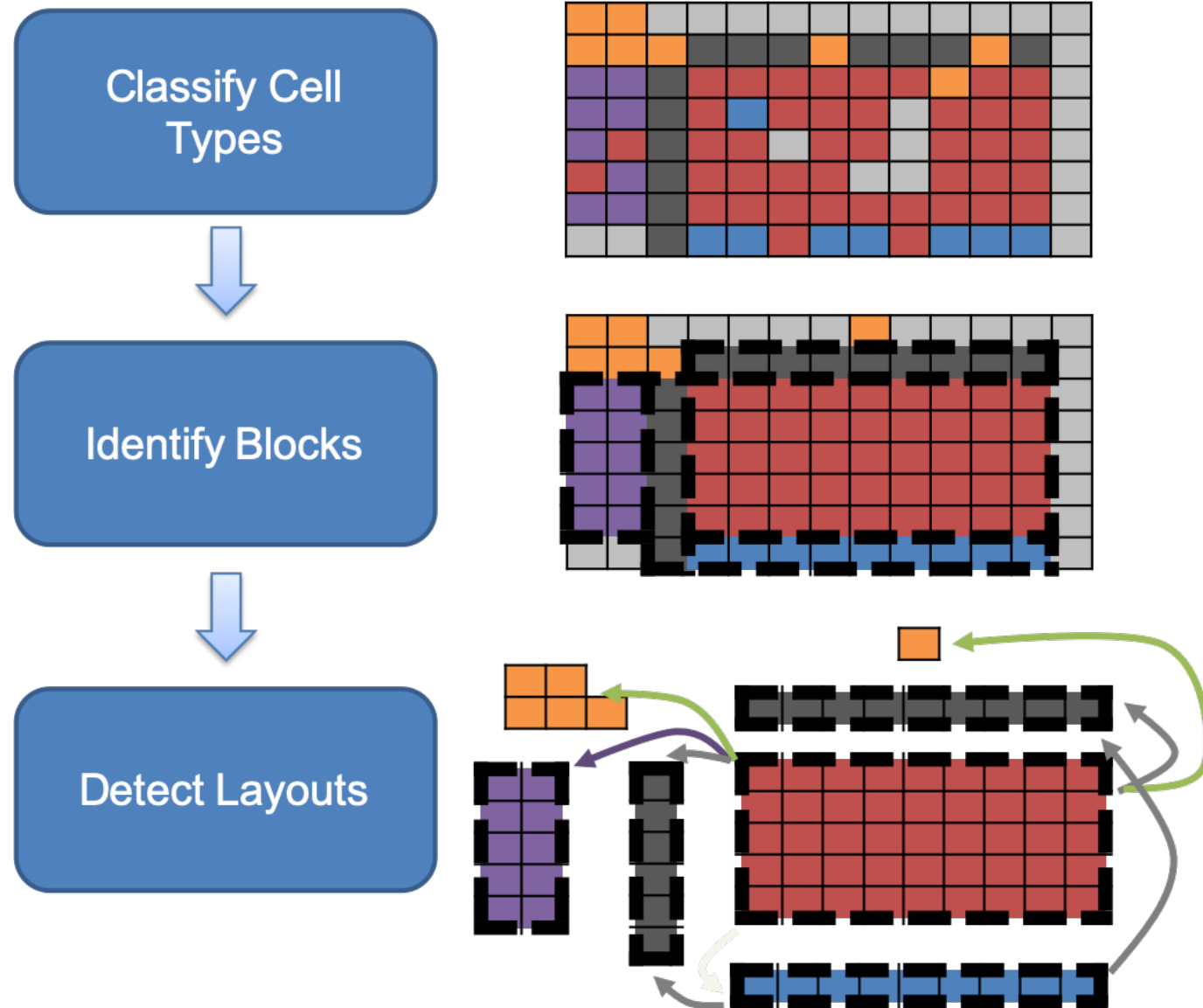
Write

wfp_food_prices_ethiopia.modified

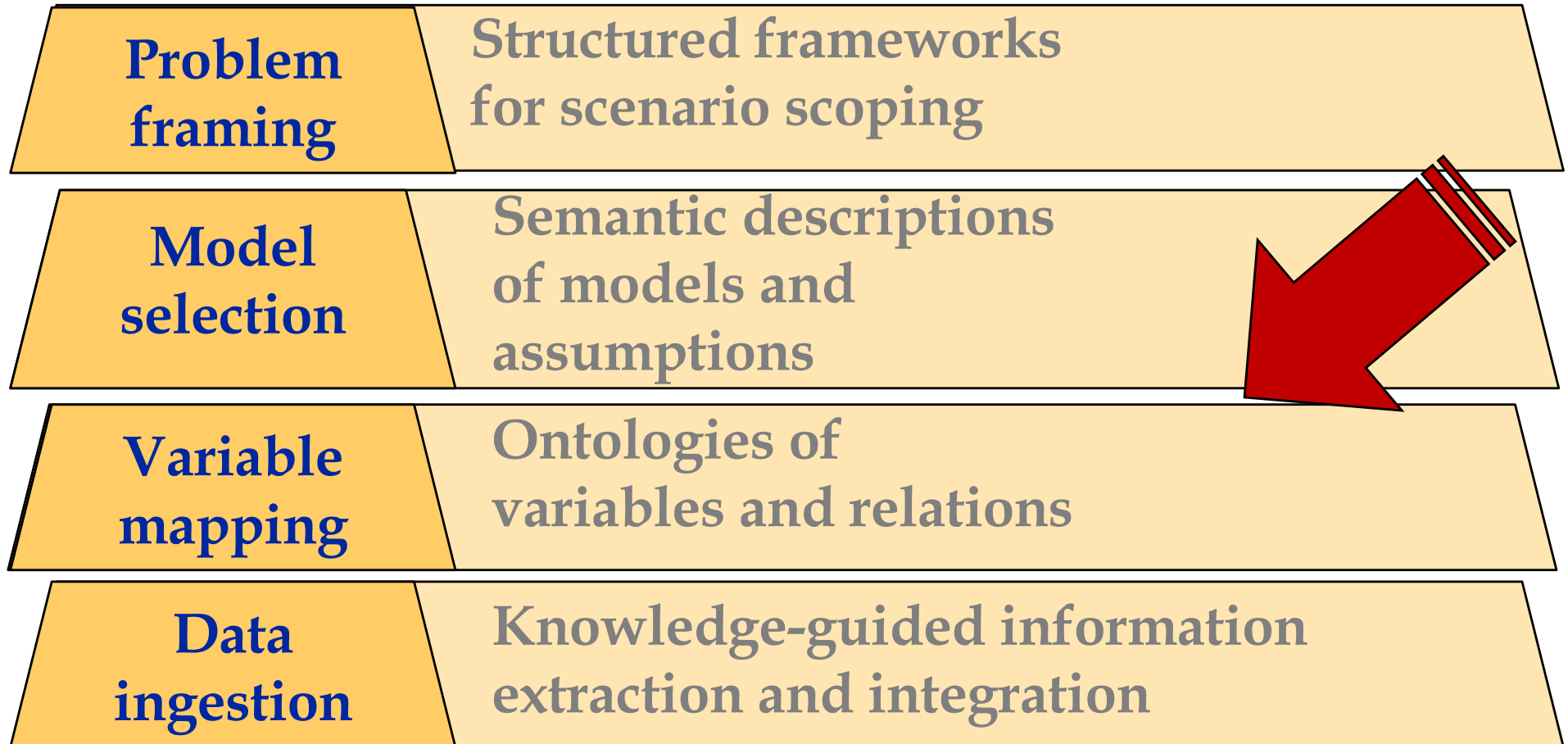
	A	B	C	D	E
1	sdmx-dimen	sdmx-attribute:refArea	sdmx-at	dcat-dimension:thing	dcat:measure_1_value
2	7/15/05	Ethiopia - Addis Ababa	\$/kg	Sorghum - Wholesale	0.082134514
3	8/15/05	Ethiopia - Addis Ababa	\$/kg	Sorghum - Wholesale	0.08627575
4	9/15/05	Ethiopia - Addis Ababa	\$/kg	Sorghum - Wholesale	0.085585544
5	10/15/05	Ethiopia - Addis Ababa	\$/kg	Sorghum - Wholesale	0.080408999
6	11/15/05	Ethiopia - Addis Ababa	\$/kg	Sorghum - Wholesale	0.086965956
7	12/15/05	Ethiopia - Addis Ababa	\$/kg	Sorghum - Wholesale	0.087311059

Table Understanding

[Vu et al WWW'19]



Knowledge-Powered Data Science for Integrated Modeling in MINT



Ontologies for Modeling Variables [Peckham & Stoica 2019]

Object: Soil

Processes (related to verbs):

Aeration, Bioturbation, Creep, Drainage, Erosion, Fertilization, Formation, Gelifluction, Movement, Slumping, Solifluction, Tilling, Weathering

Process Quantities:

creep_speed
erosion_rate
fertilization_time
tilling_depth

State Quantities:

age
alkalinity
brooks-corey_c_parameter
clay_volume_fraction
depth (to bedrock)
hydraulic_conductivity
organic_matter_volume_fraction
porosity
pressure_head
sand_volume_fraction
saturated_hydraulic_conductivity
water_volume_fraction

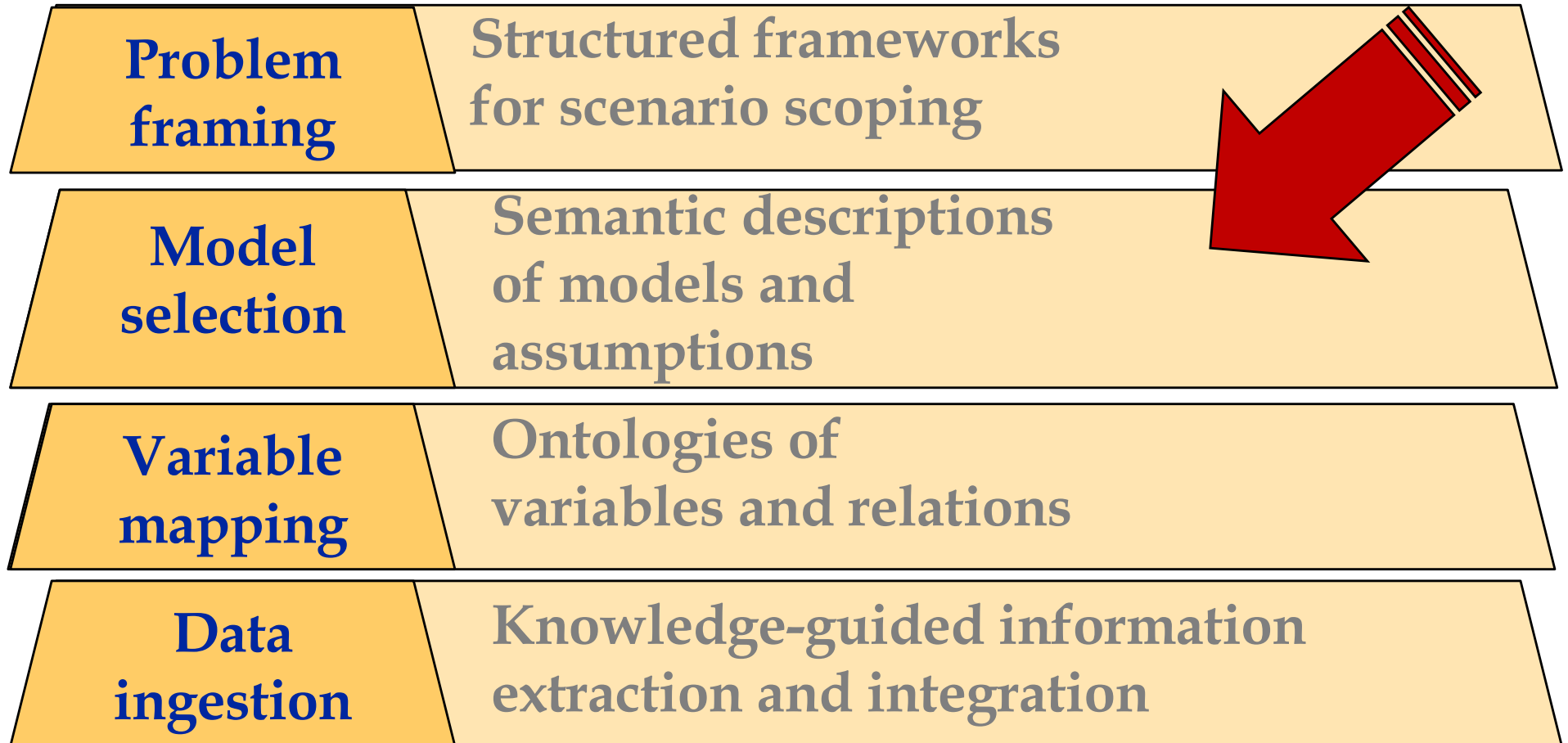
Properties:

color
fertility

- **Property** names are nominalizations of **adjectives**.
- **Quantity** is a subclass of **Property**. (Numerical, often has units.)
- A pairing of an **object** with a **quantity** is a **variable**.

- An **index** or **scale** numerically quantifies a property (eg, drought index, flood index)
- **Index** is a subclass of **Quantity**


Knowledge-Powered Data Science for Integrated Modeling in MINT



Knowledge-Rich Catalogs of Models

[Garijo et al eScience 2019]


1. Model invocation
2. Data formats
3. Model variables
4. Constraints
5. Adjustable parameters
6. Interventions



Model Catalog

Search models Search on Full text

1 version




Category: Hydrology
Type: Theory-

[The Soil & Water Assessment Tool \(SWAT\)](#)

The Soil & Water Assessment Tool (SWAT) is a small watershed to river basin-scale model used to simulate the quality and quantity of surface and ground water and predict the environmental impact of land use; land management practices and climate change (<https://swat.tamu.edu/> 2019)

Keywords: Soil, watershed, surface water, ground water, en... [More details](#)

2 versions




Category: Agriculture
Type: Theory-

[Cycles](#)

Cycles simulates the productivity and the water-carbon and nitrogen balance of soil-crop systems subject to climate conditions and a large array of management constraints

Keywords: agriculture, cycles, crop growth, weather, soil, cr... [More details](#)

4 versions



Category: Economy
Type: Theory-

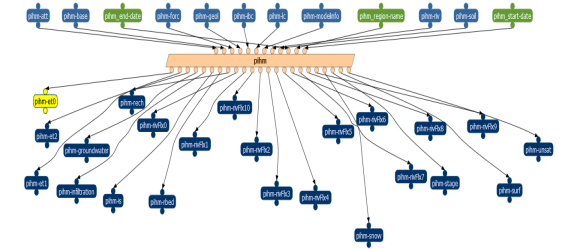
[Economic aggregate crop supply response model \(EACS\)](#)

The Aggregate crop supply response model (EACS) describes the aggregate crop supply response model for the country of South Sudan. This is a regional-scale aggregate model of agricultural supply for a specified set of crops (cassava; groundnuts; maize; sesame seed; and sorghum).

Keywords: economy, land use, crop production, fertilizer c... [More details](#)

1) Model Invocation

- Configuration: selected processes
- Setup: calibrated for specific areas



Creation date: 2016






Category: Agriculture

Model type: Theory Guided



Cycles simulates the productivity and the water, carbon and nitrogen balance of soil-crop systems subject to climate conditions and a large array of management constraints. Overall the model is set up to be daily. Some processes such as water balance are temporally nested (subdaily)

- **Authors:** Armen Kemanian
- **Publisher:** The Pennsylvania State University

- Select a configuration
-  Cycles configuration (v0.9.4) exposing weed fraction and fertilizer rate  
-
- Select a configuration setup
-  Cycles calibrated model (v0.9.4) for the Pongo region with planting dates. Weather file can be c  

Overview

Parameters and Files

Variables

Assumptions

Compatible Software

Technical Information

2) Data Formats



Creation date: 2016

Category: Agriculture

Model type: Theory Guided

Cycles

Cycles simulates the productivity and the water, carbon and nitrogen balance of soil-crop systems subject to climate conditions and a large array of management constraints. Overall the model is set up to be daily. Some processes such as water balance are temporally nested (subdaily)

- **Authors:** Armen Kemanian
- **Publisher:** The Pennsylvania State University

 Select a configuration
Cycles configuration (v0.9.4) exposing weed fraction and fertilizer rate  

 Select a configuration setup
Cycles calibrated model (v0.9.4) for the Pongo region with planting dates. Weather file can be c  

Overview

Parameters and Files

Variables

Assumptions

Compatible Software

Technical Information

Output files

Name	Description
<code>cycles_outputs</code>	Cycles season configuration file
<code>cycles_crop</code>	Cycles crop output file
<code>cycles_nitrogen</code>	Nitrogen file. Results in this file are for the sum of all layers in the soil profile, i
<code>cycles_water</code>	Cycles water file
<code>cycles_weatherOutput</code>	Cycles weather output file
<code>cycles_season</code>	The season.dat file provides information about each crop harvest.
<code>cycles_soilProfile</code>	Results in this file are for the sum of all layers in the soil profile, including surf
<code>cycles_som</code>	Cycles annual SOM file. This file provides annualized measurements of the ca ratio by soil layer. One column will be created for each layer in the soil profile heading indicates LAYER 1..x
<code>cycles_summary</code>	The summary file provides a summarized output of total C inputs over the du annual rates for N cycling processes.

3) Model Variables

■ Ontology of standard scientific names [Peckham iEMSs 2014; Peckham & Stoica 2019]

- Eg SSN: watershed_outlet_water__volume_outflow_rate is more precise than “streamflow” or “discharge”

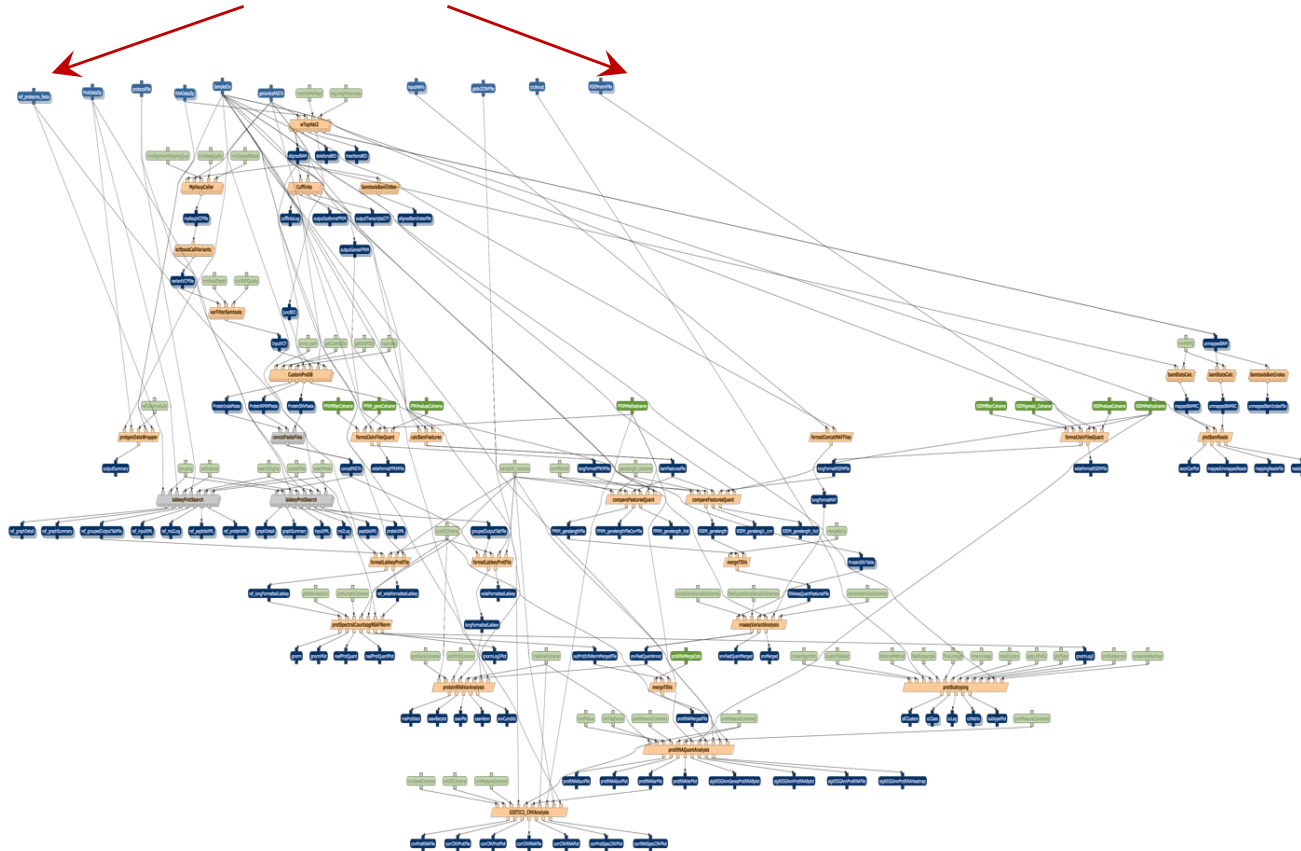
`cycles_soil` Cycles soil description file. Soil files typically have a suffix of .soil, but any naming convention can be used as long as it matches the soil file name

Label	Long Name	Description	Standard Name	Units
RV	rock volume	Rock volume expressed as volume over total volume	soil_rock__volume_fraction	m3 m-3
DZ	soil layer thickness	Soil layer thickness	soil_layer__thickness	m
SLOPE	slope of the field	Average slope of field of interest	land_surface__slope	m m-1
LAYER	soil layer number	Soil layer number from 1 to an integer that is user defined (or database defined)		
SOM	soil organic matter	Soil organic matter per unit of non-rock soil expressed as percentage over total mass	soil_matter~organic__mass_fraction	kg kg-1 x-100
SILT	silt percentage	Silt mass per unit of soil mass (no rocks) and expressed as a percentage	soil~no-rock_silt__mass_fraction	%
BD	bulk density	Soil mass dry and without rock divided by the sampled volume	soil~no-rock~dry__mass-per-volume_density	Mg m-3
CLAY	clay percentage	clay particle size fraction size fraction of each soil layer in %.	soil_clay_particle__volume_fraction	%

4) Constraints and Preconditions

- Daily rainfall data vs monthly rainfall data

Must use the same forecast dataset



5) Drivers and Adjustable Parameters

- Drivers are typically input data files (eg weather forecast)
- Adjustable parameters are those useful to explore what-if scenarios

Parameter	Description	Relevant for intervention	Value on setup ?
<code>start_year</code>	Year when the simulation started		2000 (default)
<code>end_year</code>	Year when the simulation ended		2017 (default)
<code>crop_name</code>	Name of the crop to run the simulation for. Accepted values are: Maize, Sorghum, Peanut		Maize (default)
<code>start_planting_day</code> The range is from 1 to 365	Day of the year for the start of the planting window	Planting Windows	100 (default)
<code>end_planting_day</code> The range is from 1 to 365	Day of the year for the end of the planting window	Planting Windows	149 (default)
<code>fertilizer_rate</code> The range is from 0 to 1250	Mass of nitrogen fertilizer added each year (kg/ha)		0 (default)
<code>weed_fraction</code> The range is from 0 to 1	Areal fraction of weed	Weed Control	0 (default)
<code>use_forcing</code>	Use forcing data from a hydrology model (when available)		FALSE

6) Interventions

- Associated with specific input parameters

Parameters:

Parameter	Description	Relevant for intervention	Value on setup ?
<code>start_year</code>	Year when the simulation started		2000 (default)
<code>end_year</code>	Year when the simulation ended		2017 (default)
<code>crop_name</code>	Name of the crop to run the simulation for. Accepted values are Peanut		Maize (default)
<code>start_planting_day</code> The range is from 1 to 365	Day of the year for the start of the planting window	Planting Windows	100 (default)
<code>end_planting_day</code> The range is from 1 to 365	Day of the year for the end of the planting window	Planting Windows	149 (default)
<code>fertilizer_rate</code> The range is from 0 to 1250	Mass of nitrogen fertilizer added each year (kg/ha)		0 (default)
<code>weed_fraction</code> The range is from 0 to 1	Areal fraction of weed	Weed Control	0 (default)
<code>use_forcing</code>	Use forcing data from a hydrology model (when available)		FALSE

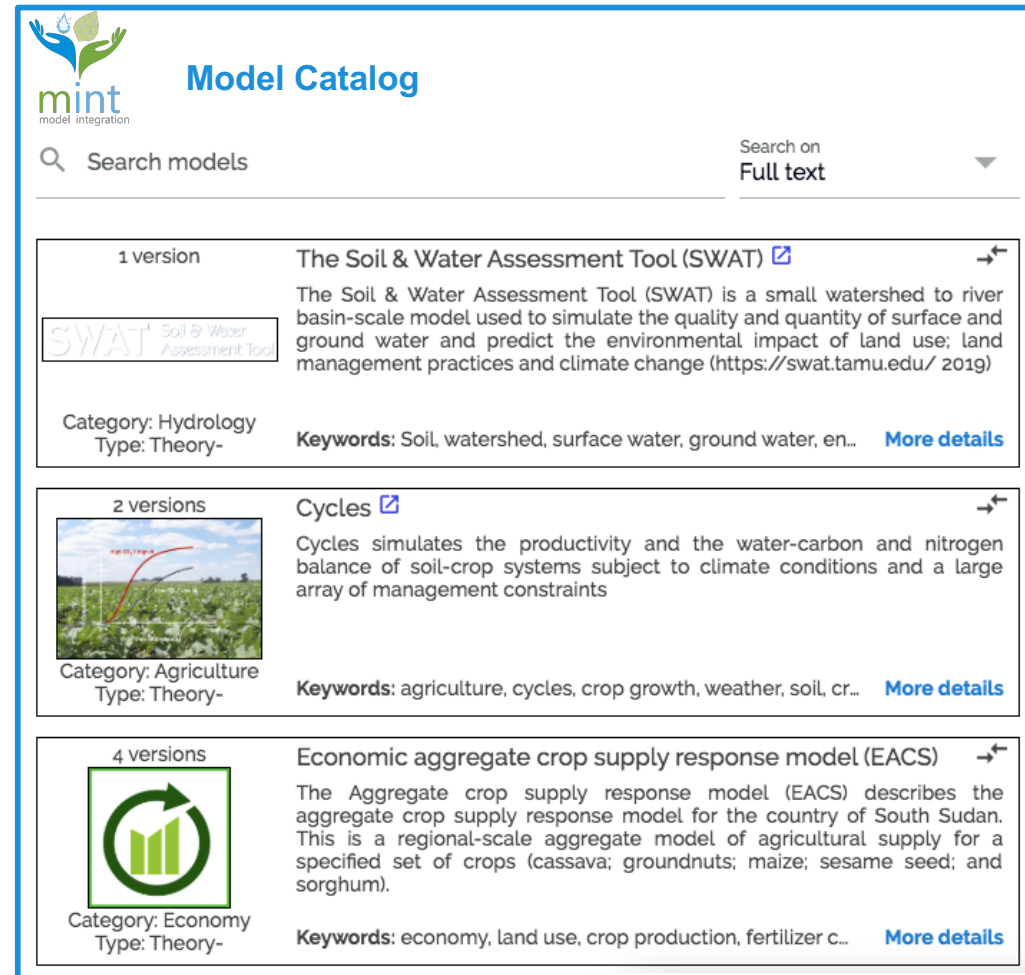
Interventions that force specific target planting windows can be expressed in this model as a start or end planting date

Knowledge-Rich Catalogs of Models

[Garijo et al eScience 2019]




1. Model invocation
2. Data formats
3. Model variables
4. Constraints
5. Adjustable parameters
6. Interventions

and much more!!

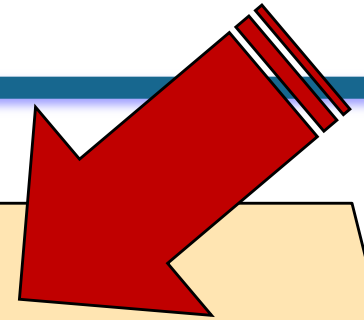


The screenshot displays the MINT Model Catalog interface. At the top, the MINT logo (model integration) is on the left, and the title "Model Catalog" is on the right. Below the logo is a search bar with the text "Search models" and a search icon. To the right of the search bar is a dropdown menu labeled "Search on Full text".

The catalog lists three models:

- 1 version**
The Soil & Water Assessment Tool (SWAT) [↗](#)
The Soil & Water Assessment Tool (SWAT) is a small watershed to river basin-scale model used to simulate the quality and quantity of surface and ground water and predict the environmental impact of land use; land management practices and climate change (<https://swat.tamu.edu/> 2019)

Category: Hydrology
Type: Theory-
Keywords: Soil, watershed, surface water, ground water, en... [More details](#)
- 2 versions**
Cycles [↗](#)
Cycles simulates the productivity and the water-carbon and nitrogen balance of soil-crop systems subject to climate conditions and a large array of management constraints

Category: Agriculture
Type: Theory-
Keywords: agriculture, cycles, crop growth, weather, soil, cr... [More details](#)
- 4 versions**
Economic aggregate crop supply response model (EACS) [↗](#)
The Aggregate crop supply response model (EACS) describes the aggregate crop supply response model for the country of South Sudan. This is a regional-scale aggregate model of agricultural supply for a specified set of crops (cassava; groundnuts; maize; sesame seed; and sorghum).

Category: Economy
Type: Theory-
Keywords: economy, land use, crop production, fertilizer c... [More details](#)

Knowledge-Powered Data Science for Integrated Modeling in MINT



**Problem
framing**

Structured frameworks
for scenario scoping

**Model
selection**

Semantic descriptions
of models and
assumptions

**Variable
mapping**

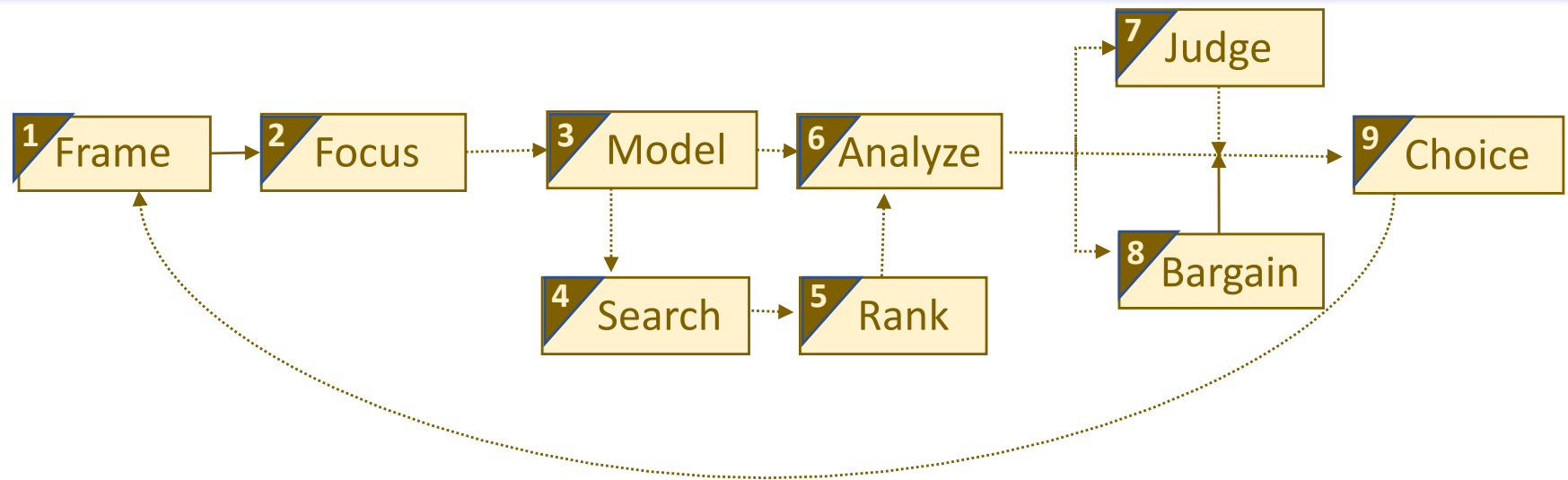
Ontologies of
variables and relations

**Data
ingestion**

Knowledge-guided information
extraction and integration

Decision Making Process

[Pierce et al 2019]



DECISION MAKING STEPS:

1. **Frame:** Recognize and identify key issues
2. **Focus:** Define the context, key variables, reference behaviors, and decision problem
3. **Model:** Construct/Run Analysis Pipeline – run base case
4. **Search:** Explore solution space, consider interventions
5. **Rank/Filter:** Select interesting solutions
6. **Analyze:** Create summary of solutions explored, assess uncertainty
7. **Judge:** Adjudicate possible actions
8. **Bargain:** Consider tradeoffs
9. **Choice:** Make recommendation, take action, converge on decisions

Iterate to refine/reframe/refocus modeling goals

Knowledge-Driven Problem Framing: Indicators and Adjustable Variables

► ETHIOPIA

Explore Areas

Prepare Models

Browse Datasets

Use Models

Prepare Reports

MESSAGES

LOGOUT MINT@ISI.EDU

< Problem statements

Choose an existing problem from the list below or click add to create a new problem statement.

Crop elasticities for the Econ model
2018-01-01 to 2018-12-31

Forecasting potential crop production in Baro region
2018-01-01 to 2018-12-31

Forecast flooding in the Baro region
2018-01-01 to 2018-12-31

Crop production with crop elasticities
2018-01-01 to 2018-12-31

ADD

TASKS

Several modeling tasks can be created for a given problem statement. [Read more](#)

Potential Crop Production: Ethiopia: 2018-01-01 - 2018-12-31
Response of maize to fertilizer with no weeds

Potential Crop Production: Ethiopia: 2018-01-01 - 2018-12-31
Average conditions

Modeling threads

For a given task, you can investigate different initial conditions and create a new modeling thread for that task. [Read more](#)

Average conditions

Average conditions with a large number of runs

Indicators/Response of interest

Crop Production

Adjustable Variables

Fertilizer cost

Note on fertilizer cost: Interventions concerning fertilizer subsidies can be expressed in this model as a percentage of fertilizer prices

Models

Datasets

Setup

Runs

Results

Visualize

Forecasting and Interventions

Cycles
2 versions, 2 configs
Category: Agriculture
Type: Theory Guided

Keywords: Agriculture, crop yield, crop failure, weather, fertilizer, crop management

MODFLOW
1 version, 1 config
Category: Hydrology
Type: Theory Guided

Keywords: groundwater modeling, steady and nonsteady flow, areal recharge, evapotranspiration

Penn State Integrated Hydrology Model (PIHM)
2 versions, 2 configs
Category: Hydrology
Type: Theory Guided

Keywords: channel routing, surface overland flow, subsurface flow, interception, snow melt, evap.

Decision Support System for Agrotechnology Transfer (DSSAT)
1 version, 1 config
Category: Agriculture
Type: Theory Guided

Keywords: Decision Support System, Agrotechnology, crop growth simulation, weather, soil

Files:

Input files

Name	Description	Value on setup	Format
<u>cycles_weather</u>	Cycles weather file	-	weather
<u>cycles_soil</u>	Cycles soil file	pongo.soil	soil
<u>cycles_crops</u>	Cycles crops file	crops.crop	crop

Parameters:

Name	Description	Value on setup
<u>crop_name</u>	Name of the crop to run the simulation for	Maize (default)
<u>start_planting_day</u>	Day of the year when the planting started The range is from 1 to 365	100 (default)
<u>end_planting_day</u>		
<u>fertilizer_n</u>		
<u>weed_fraction</u>		

Variables:

cycles_weather Cycles weather file

Label	Long Name	Description	Standard Name
RHn	relative humidity minimum	relative humidity minimum	atmosphere_air_water-vapor__min_of_relative_humidity
Solar	solar radiation of the day	Shortwave incoming radiation on	land_surface-horizontal_radiation-incoming-shortwave

Processes and Configurations:

Selected configuration: Cycles configuration (v0.9.4) exposing weed fraction and fertilizer rate

Cycles configuration (version 0.9.4) exposing additional parameters such as weeds fraction and fertilizer rate
Authors: Rafael Silva

- Time interval: 1 day
- Grid details:
 - Type: PointBasedGrid
 - Dimensions: 0D
 - Spatial resolution: Point
- Processes: Respiration, Nitrogen mineralization and immobilization, Biomass growth, Humification, Transpiration, Nitrogen uptake, Precipitation, Solar radiation, Management, Nitrogen transport
- Download: [cycles-0.9.4-alpha.zip?raw=true](#)

Selected configuration setup: Cycles calibrated model (v0.9.4) for the Pongo region with planting dates. Weather file can be chosen

Cycles calibrated model (v0.9.4) for the Pongo region. Planting dates can be selected as parameter values and the weather file can be chosen as an input
Authors: Rafael Silva

- Region: Pongo Basin (South Sudan)
- Time interval: 1 day

1 Identify variables of interest

2 Identify variables of interest

3 Compare models

4 Set up and run model

5 Adjust model to explore interventions, identify problem areas

6 Prepare modeling products for analyst

mint model integration

Forecasting: Investigate food security in South Sudan for August 2022

What flooding is expected in the Pongo Basin for the 2022 lean season?

Model: PHM v2.2 - Pongo Basin Calibration

Adjustment to precipitation: 0.10, +

Adjustment to temperature: 0.10, +

Model: PHM v2.2 - Pongo Basin Configuration

Subsidies for sorghum fertilizer will decrease sesame production

If sorghum prices fall, sesame and maize production increase

If sesame prices fall, groundnuts production will increase

MINT User Interaction

1 Identify results of interest

What flooding is expected in the Pongo Basin for the 2017 lean season ?

Variables Models Datasets Setup Runs Results Visualize

This step is for selecting driving and response variables for your analysis that you're interested in. An optional driving variable indicates the kind of

Variables

Response Variables* +

European Flooding Index

Driving Variables +

Precipitation

Notes

2 Identify relevant models

What flooding is expected in the Pongo Basin for the 2017 lean season ?

Variables Models Datasets Setup Runs Results Visualize

This step is for selecting models that are appropriate for the response variables that you selected earlier.

Models

The models below generate data that includes the response variables that you selected earlier: European Flooding Index. Other models that are available do not generate that kind of result.

Model	Category	Calibration Region	Relevant Output
<input checked="" type="checkbox"/> PIHM v2.2 - Pongo Basin Configuration	Hydrology	Pongo Basin (South Sudan)	Flooding Index
<input type="checkbox"/> PIHM v2.2 - Pongo Basin Calibration	Hydrology	Pongo Basin (South Sudan)	Flooding Index
<input checked="" type="checkbox"/> TopoFlow 3.5 - Pongo Basin Configuration	Hydrology	Pongo Basin (South Sudan)	Flooding Index

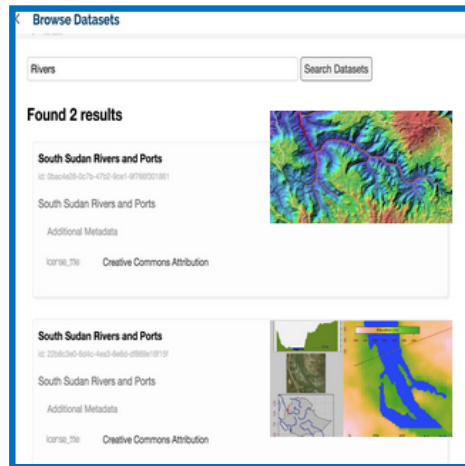
COMPARE SELECTED MODELS CANCEL SELECT & CONTINUE

Notes

Selected the models that have been manually configured for the region.

MINT User Interaction (Cont'd)

3 Find and transform datasets



5 Adjust model to explore interventions, identify problem areas

4 Set up and run models

What flooding is expected in the Pongo Basin for the 2017 lean season ?

Variables Models Datasets Setup Runs Results Visualize

This step is for setting up and running the models that you selected earlier.

Setup Models

Model: PIHM v2.2 - Pongo Basin Calibration

Setup the model by specifying values below. You can enter more than one value (comma separated) if you want several runs

Adjustable Variable	Values
Adjustment to precipitation This is a percentage. The range is from -20 to 20	0, 10, -10
Adjustment to temperature This is a percentage. The range is from -20 to 20	0, 10, -10

Model: PIHM v2.2 - Pongo Basin Configuration

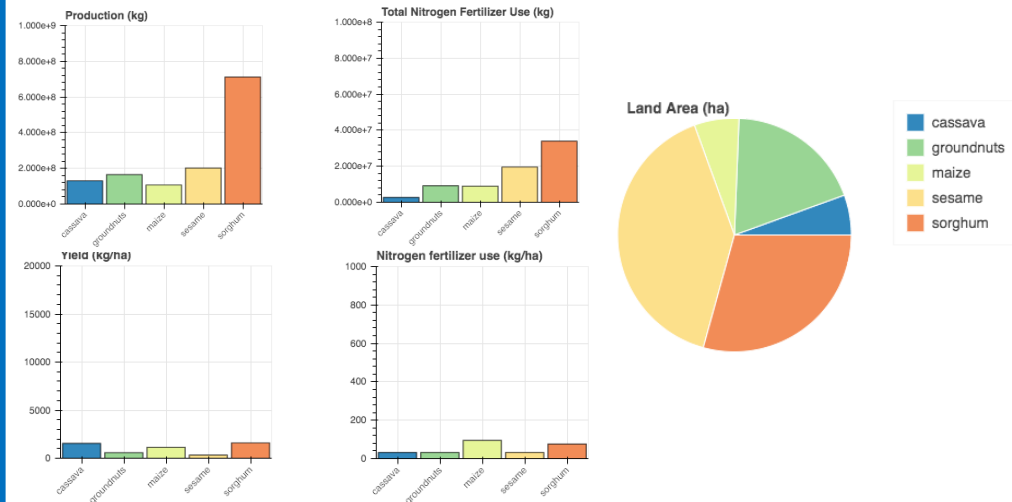
Setup the model by specifying values below. You can enter more than one value (comma separated) if you want several runs

Cassava:
Land Cost Adjustment: 0
Nitrogen Fertilizer Cost Adjustment: 0
Crop Price Adjustment: 0

Maize:
Land Cost Adjustment: 0
Nitrogen Fertilizer Cost Adjustment: 0
Crop Price Adjustment: 0

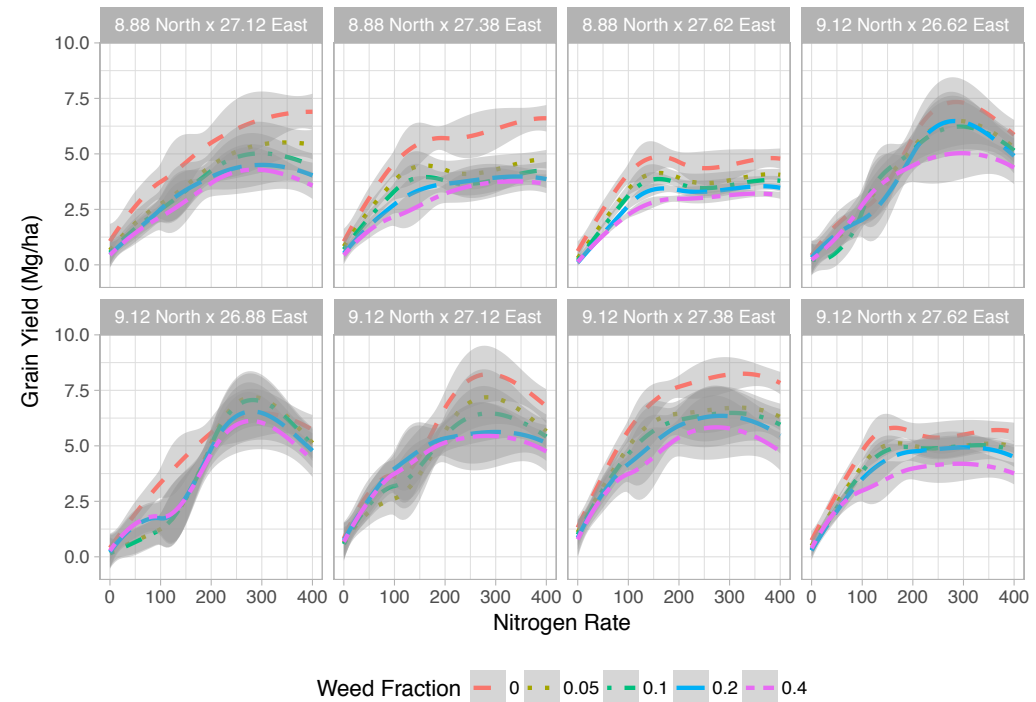
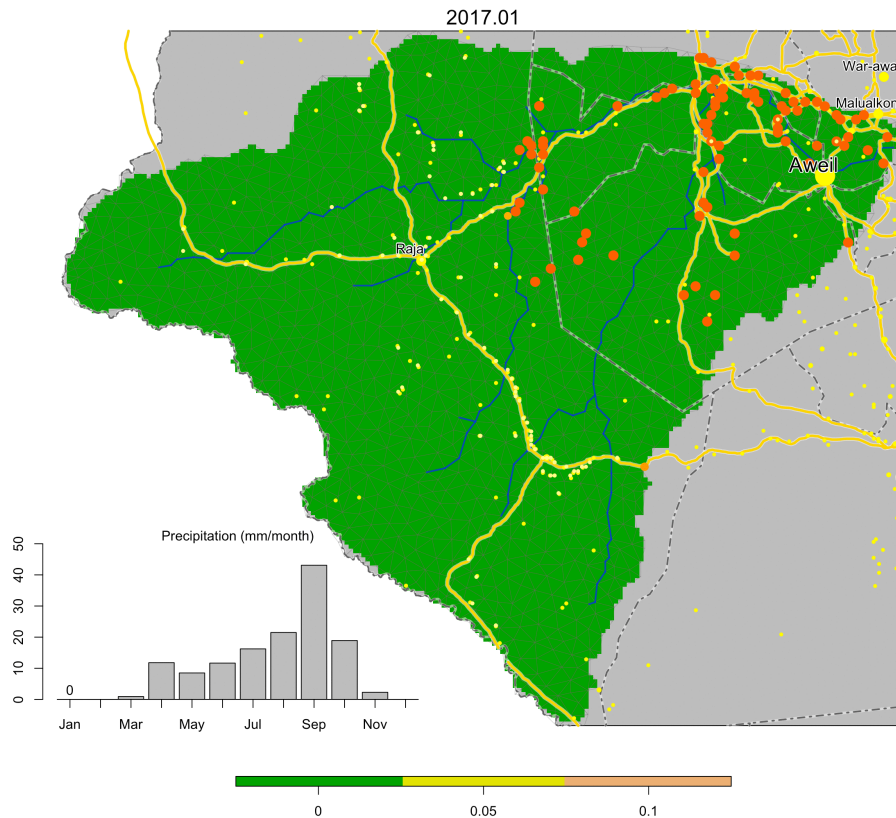
Groundnuts:
Land Cost Adjustment: 0
Nitrogen Fertilizer Cost Adjustment: 0
Crop Price Adjustment: 0

Sorghum:
Land Cost Adjustment: 0
Nitrogen Fertilizer Cost Adjustment: 0
Crop Price Adjustment: 0



Future Work in MINT

- Causal models
- Forecasting
- Interventions
- Uncertainty



Knowledge-Powered Data Science for Integrated Modeling in MINT

- **Knowledge-guided machine learning**
- **Knowledge-rich catalogs of models and data**
- **Knowledge-driven problem framing**

**Problem
framing**

Structured frameworks
for scenario scoping

**Model
selection**

Semantic descriptions
of models and
assumptions

**Variable
mapping**

Ontologies of
variables and relations

**Data
ingestion**

Knowledge-guided information
extraction and integration

Outline

- The need for integrated modeling in geosciences
- Diversity of models across disciplines
- MINT: knowledge-powered data science for integrated modeling
- Intelligent systems for geosciences

REVIEW ARTICLES

Intelligent Systems for Geosciences: An Essential Research Agenda

By Yolanda Gil, Suzanne A. Pierce, Hassan Babaie, Arindam Banerjee, Kirk Borne, Gary Bust, Michelle Cheatham, Imme Ebert-phoff, Carla Gomes, Mary Hill, John Horel, Leslie Hsu, Jim Kinter, Craig Knoblock, David Krum, Vipin Kumar, Pierre Lermusiaux, Yan Liu, Chris North, Victor Pankratius, Shanan Peters, Beth Plale, Allen Pope, Sai Ravela, Juan Restrepo, Aaron Ridley, Hanan Samet, Shashi Shekhar

Communications of the ACM, January 2019, Vol. 62 No. 1, Pages 76-84

10.1145/3192335

[Comments](#)

VIEW AS:



SHARE:



SIGN IN for Full Access

User Name

Password


» [Forgot Password?](#)

» [Create an ACM Web Account](#)



Many aspects of geosciences pose no systems research. Geoscience data is to be uncertain, intermittent, sparse scale. Geosciences processes and object spatiotemporal boundaries. The lack model evaluation, testing, and computing these challenges requires breakthrough transform intelligent systems, while geosciences in turn. Although there beneficial interactions between the geosciences communities,^{4,12} the potential research in intelligent systems for ge





Report of the 2015 NSF Workshop on Intelligent Systems for Geosciences

March 26-27, 2015
Arlington VA

<http://www.IS-GEO.org>

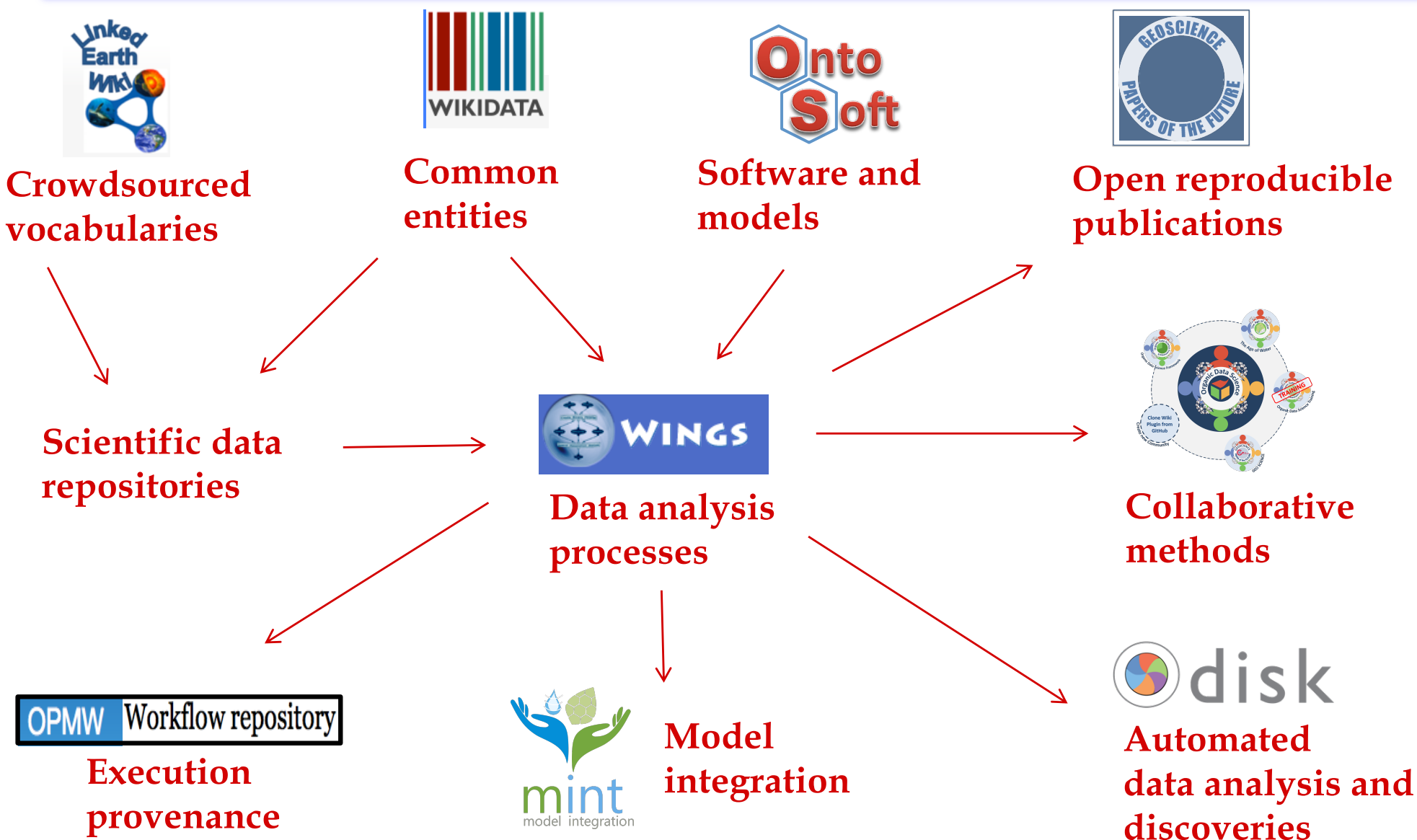
Co-Chairs:

Yolanda Gil (gil@isi.edu)

Suzanne Pierce (suzpierce@jsg.utexas.edu)



Open Knowledge Networks



The Future is Here:

Knowledge-Powered Data Science

- Focus shifted from data to **models**
 - Model characterization, reuse, and integration
- Need to **incorporate model-centered science knowledge** about phenomena and context
 - Knowledge about **physical, geological, chemical, biological, ecological, and anthropomorphic factors**
 - Knowledge about the user **goals and context**
- This would enable novel forms of reasoning, integrating, visualizing, managing, learning, and discovery with geosciences data

Continue the conversation at <http://www.IS-GEO.org>



A 20-Year Community Roadmap for Artificial Intelligence Research in the US

Yolanda Gil, USC, AAI President
Bart Selman, Cornell, AAI President-Elect

<https://arxiv.org/abs/1908.02624>



CCC

Computing Community Consortium
Catalyst



AAAI
Association for the Advancement
of Artificial Intelligence