

AI for Understanding Complex Systems: Modeling Interventions

Yolanda Gil

Information Sciences Institute
and Department of Computer Science
Viterbi School of Engineering
University of Southern California



 Q52353442

May 11, 2022



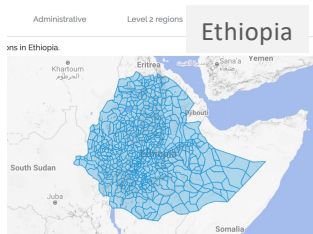
Understanding Complex Natural-Human Systems

[Gil et al iEMSs 2018; Gil et al TIIS 2021]

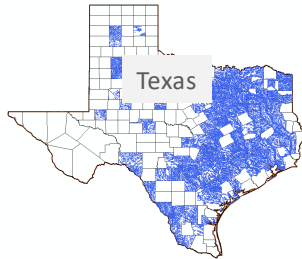


Collaboration with D. Garijo, D. Khider, C. Knoblock, E. Deelman, & R. Ferreira (USC/ISI), V. Kumar (UM), S. Peckham (CU), C. Duffy & A. Kemanian (PSU), K. Cobourn (VT), S. Pierce (UT) and many others

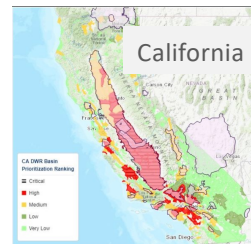
food insecurity



flooding

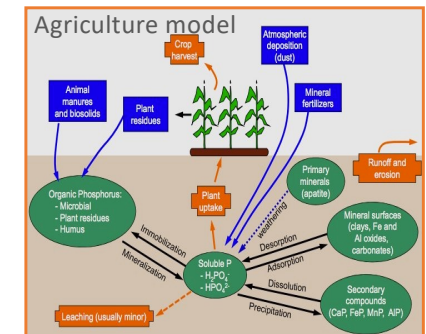
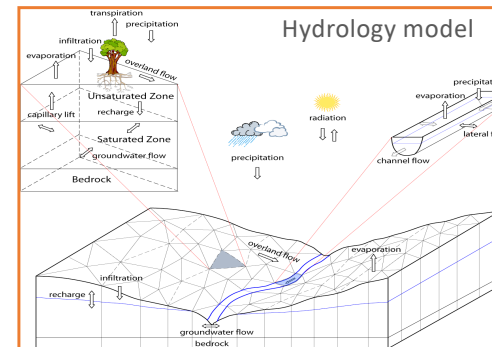
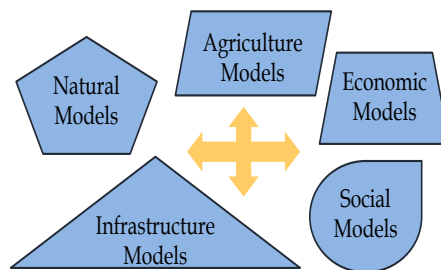
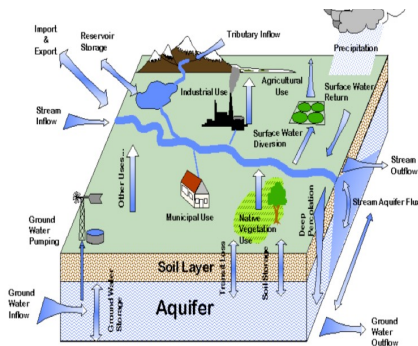


droughts



Integrated Modeling

- Increased demand
- Takes months or years
- Currently a craft



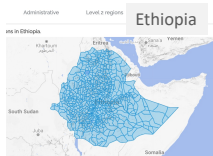
<http://water.state.co.us/SurfaceWater/SWRights/WaterDiagrams/Pages/default.aspx>

AI in MINT: From Modeling to Decisions

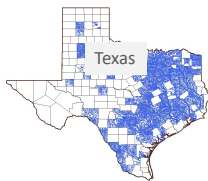
AI for Integrated Modeling

- Reduce effort from years to weeks
 - Automated reasoning
 - Machine learning
 - Intervention-driven user guidance

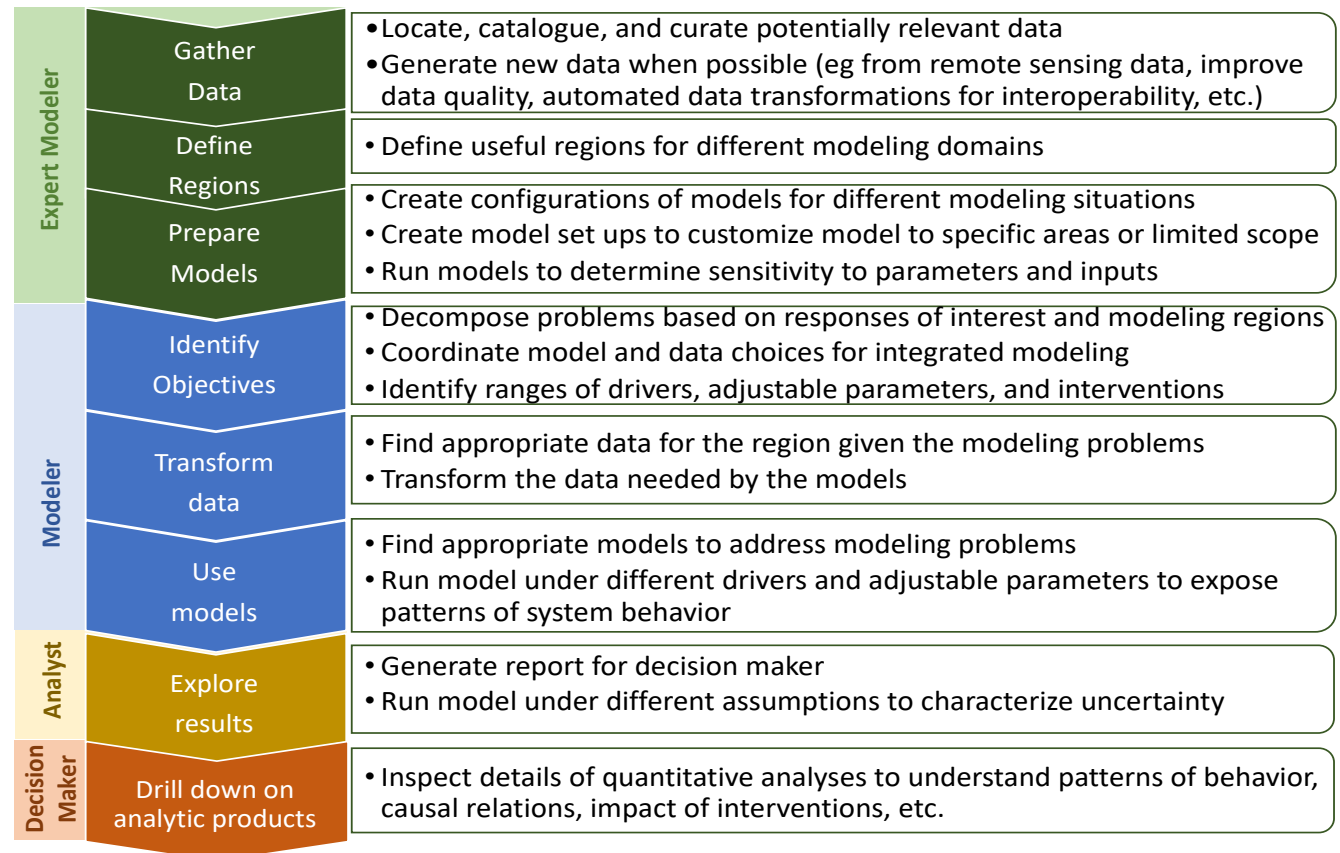
food insecurity



flooding



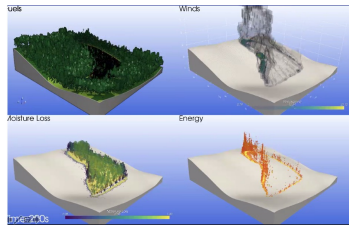
droughts



Proactive Management of Wildfires: Controlled Fires

Collaboration with M. Pazzani, B. Dilkina (USC) with I. Altintas (UCSD) & US Forest Svc & Los Angeles Fire Dept

Goals: Scale preventive fires from 1M to 10M acres by using AI to guide planners of controlled fires to use science-grade data and advanced fire models. Today plans are done based on experience, this is slow and does not scale.



AI approach:

- Knowledge Technologies to select best fire model for a region's characteristics (winds, vegetation, etc.)
- Physics-Informed Machine Learning for fire modeling
- Constraint Optimization to plan fires with target percentage vegetation burns, air quality ratings, and risk to buildings

Model catalog

Model Explorer

The WIFIRE Model Commons contains useful descriptions of science-grade models available to simulate controlled and wildfires. Each model can have separate configurations, each representing a unique functionality of that model (particular choices of regions, inputs, processes, etc). Each configuration can have separate setups that provide different default values for files and parameters. In the search bar below you can search models in several ways, which you can choose on the right. You can search by model name, description, type (e.g. fire), keywords (simulation), areas (e.g. California), variables (e.g. wind), index or intervention.

Search models

Search on
Name, description and keyword

Model Recommender

QUIC-Fire Rough Terrain High Wind High Complexity

Input Datasets for this model for the selected region and time period:

- Input **topo.dat** has data "CA-DEM-Placeholder", with the following characteristics:
 - Mean Terrain Elevation = 787.87 m
 - Mean Terrain Slope = 20.10 %
- Input **wind.csv** has data "Gusts-Georgia-Slide-Road-CA", with the following characteristics:
 - Wind Complexity = 1.00
 - Mean Wind Speed at 10m = 2.02 m/s

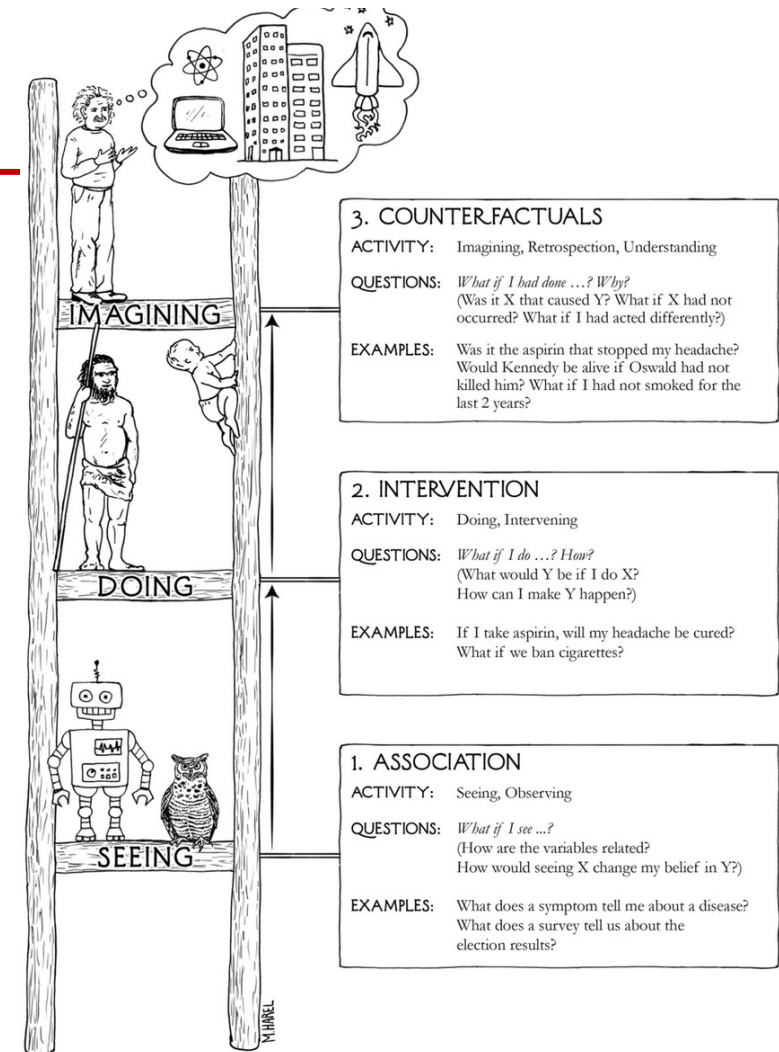
This Model is **RECOMMENDED** as there are datasets available for this region that satisfy model recommendations:

- ✓ Wind Complexity = 1
- ✓ Mean Terrain Slope >= 20%
- ✓ Mean Wind Speed at 10m >= 1 m/s

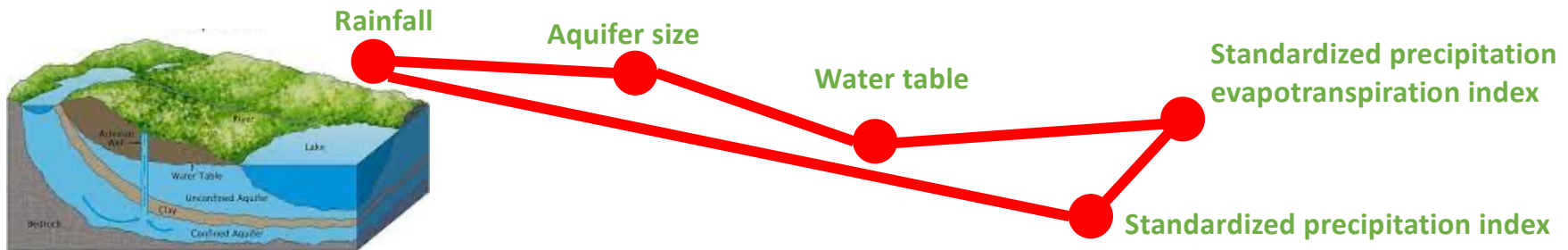
RUNNING...

<https://wifire.models.mint.isi.edu>

Interventions and the Causal Ladder [Pearl 2018]



Characterizing a Complex Dynamic System



- A **variable** is a measurable quantity of a complex dynamic system
 - It may be observed or estimated
- An **indicator** is a variable that is identified as playing a special role, namely to help characterize a complex property of a system being modeled that is meaningful to an end user (eg a decision maker)
 - A proxy indicator is an indirect way to measure variables of interest.
- An **index** is a combination of 2 or more indicators designed to give a useful aggregate value to assess a situation. They are also useful for comparison purposes

Modeling and Realizations (I)

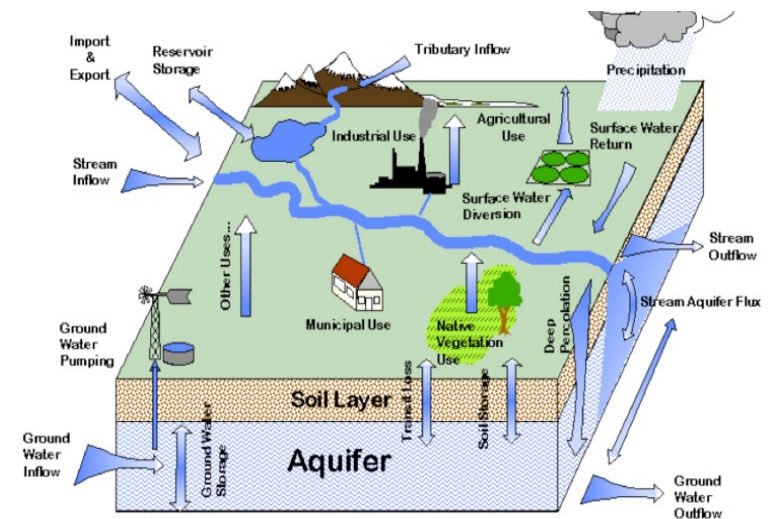
- A simulation **model** (or integrated model) is used to emulate processes in a dynamic system under different initial conditions in order to understand its behavior
- A **realization** is a single run of an integrated model, where given a fixed input situation and parameter settings (i.e., given input datasets and parameters), the integrated model generates a prediction (i.e., output datasets)
 - A realization is a workflow execution
 - Non-deterministic models can lead to several realizations, each corresponding to a different seed
- A **realization specification** consists of an input situation (input data and input parameter settings) together with the model(s) form which can then be submitted for execution in order to create a realization. A realization specification is an execution-ready workflow
- An **incongruent realization** is a realization with a complete execution that is considered erroneous because the predictions are not consistent with observations, physical laws, or known system behaviors

Modeling and Realizations (II)

- An **ensemble** is a collection of realizations that have a common theme
 - **Parameter ensemble**: a parameter value varies
 - **Data ensemble**: input data is obtained from alternative data sources or hypothetical situations (initial conditions)
 - **Model ensemble**: a modeling step is implemented with alternative models
- A reference realization is an execution of an integrated model that is compared against one or several others
 - This is often referred to as a **baseline**

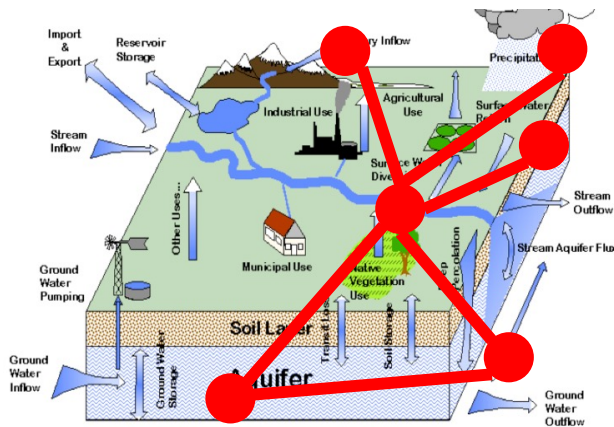
Interventions (I)

- An **intervention** is a possible action that could change the behavior of a system, typically carried out by humans to change an existing situation with the goal of eliminating/mitigating future undesirable outcomes
 - e.g., increasing the amount of fertilizer subsidies with the intention to increase crop production
 - e.g. carry out controlled fires in region R to burn 80% of the vegetation every two years

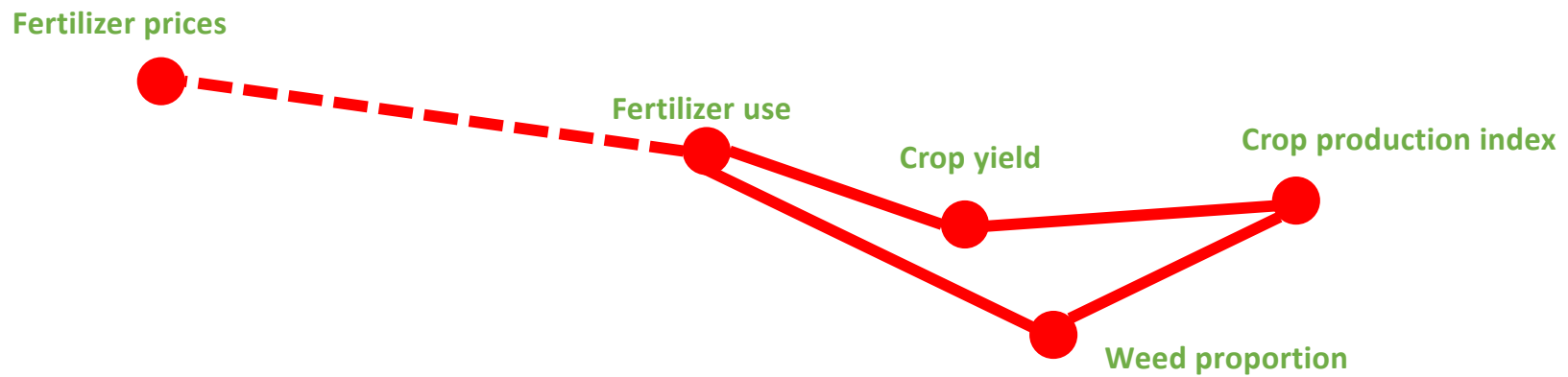


Interventions (II)

- An intervention itself is an action(s) that affects the system, but it is often not part of the model and may be represented in terms of the system variable(s) that the intervention can affect
 - e.g., reducing fertilizer prices to increase the use of fertilizers in agriculture



Example Intervention: Reduce Fertilizer Prices



Example

Intervention	Reduce fertilizer prices
Intervention target	Fertilizer use
Intervention result indicator	Crop yield
Intervention result index	Crop production index
Intervention risk indicator	Weed proportion
Intervention specification	Reduce fertilizer prices by 40% during Jan-June
Intervention intended outcome	Augment crop production index by 30%
Intervention baseline	Market-driven fertilizer prices

Specifying Interventions

- An **intervention target** (or manipulated variable) is a modeling input variable that represents measurable direct results of the actions taken in the intervention
- An **intervention result indicator** is a modeling variable that the user identifies as useful to quantify the effects of an intervention
- An **intervention risk indicator** is a modeling variable that the user identifies as useful to quantify the indirect or unintended (side) effects of an intervention
- An **intervention formulation** is a choice of intervention targets, result indicators, and risk indicators. An intervention formulation can be a theme for a workflow ensemble
- An **intervention specification** consists of the input data and parameters needed to specify the values of the intervention targets both in their initial state and throughout the modeling time period (If applicable)
 - May include qualifiers (eg magnitude, temporal, spatial)

Intervention	Reduce fertilizer prices
Intervention target	Fertilizer use
Intervention result indicator	Crop yield
Intervention result index	Crop production index
Intervention risk indicator	Weed proportion
Intervention specification	Reduce fertilizer prices by 40% during Jan-June
Intervention intended outcome	Augment crop production index by 30%
Intervention baseline	Market-driven fertilizer prices

Cost-Benefit Analysis for Interventions

- The **intervention cost** quantifies the cost of the action as well as the overall effects and risks that the intervention incurs
 - A cost-benefit analysis can be done through separate cost variables and cost models
- An **intervention realization** is an intervention specification, a possible realization of it, and the values of intervention result and risk indicators
 - An **intervention baseline**, corresponding to the "do nothing" alternative, may be used as a reference realization
- The intervention cost is often compared with the intervention baseline

Causality, Combining Interventions

- Causality is a fundamental approach for understanding interventions
 - Interventions are often combined together, which would result in potentially interacting impact pathways to amplify their effects (positive or negative)
- An **intervention impact pathway** is a graph that links intervention targets to result and risk indicators through causal relations.
- An **intervention impact network** is a graph with multiple interventions and their associated impact pathways which can potentially share the same variables
- Interventions can either be **tactical interventions** (the user reacts to a developing situation) or **strategic interventions** (these decisions are long-term and require planning).

Modeling Questions About Complex Systems

[Gil et al CogSci'22]

Association

Is X associated with Y?

Diagnostic

What was observed about X given the last few events?

Intervention

What would I observe if I change X?

Counterfactual

What outcome if event E had not occurred in the past?

Backcasting

What event G will have future outcome, what E,F before G?

Characterization

Do X, Y, Z conform to a pattern/function?

Prefactual

What outcome if event E occurs now?

Semifactual

Same outcome if event E had occurred instead of F?

Prediction

What future outcome if event E occurs now?

Postdiction

What event E caused F that caused current outcome?

Semantic Representations for Model Metadata [Garijo et al 2019; Garijo et al 2018]



Cycles

Cycles simulates the productivity and conditions and a large array of management as water balance are temporally nested

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

Creation date: 2016
Category: Agriculture
Model type: Theory Guided

Select a configuration
Cycles configuration (v0.9.4)

Select a configuration setup
Cycles calibrated model (v0.9.4)

Overview Parameters and Files Variables

Relevant for calculating index:

- [crop_seasonal_production_index](#)

Parameters:

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

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	m ³ 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	%

Framing Modeling Problems



Identify
Objectives

Use
Models

AI for reasoning about

- Driving variables
- Response variables
- Interventions

Forecasting: Investigate food security in South Sudan for August 2017

- What are expected crop yields for maize and sorghum during the summer of 2017?
- What flooding is expected in the Pongo Basin for the 2017 lean season?**
- Will flooding impact transportation in the Pongo Basin area during the 2017 lean season?
- What is the modeled yield for maize and sorghum in the Pongo Basin during the 2017 lean season?
- What is the effective modeled yield for maize and sorghum effective yields in the Pongo Basin during the 2017 lean season?
- Would fertilizer subsidies affect maize and sorghum effective yields in the Pongo Basin during the 2017 lean season?
- What flooding is expected in the Gel Aliab area for the 2017 lean season?
- Will flooding impact transportation in the Gel

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. The response variable indicates the kind of results that you're interested in. An optional driving variable indicates the kind of inputs that you want to use to drive the results.

Variables

Response Variables* +

- European Flooding Index

Driving Variables +

- Precipitation

Notes

Browse Datasets

Rivers Search Datasets

Found 2 results

South Sudan Rivers and Ports

Additional Metadata

Source: The Creative Commons Attribution

South Sudan Rivers and Ports

Additional Metadata

Source: The Creative Commons Attribution

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

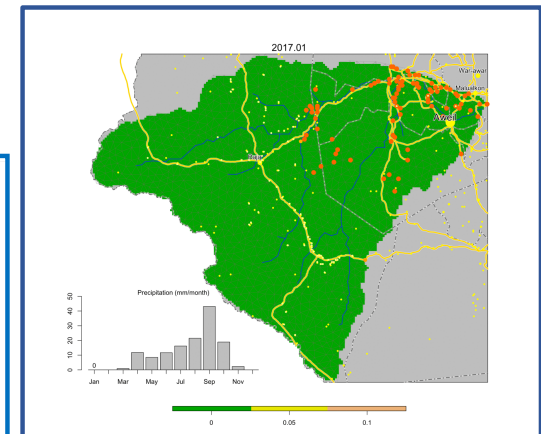
Model: PIHM v2.2 - Pongo Basin Calibration

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.



MINT: User Interaction



1 User specifies question of interest

A screenshot of the MINT web application interface. The top navigation bar includes the MINT logo and a series of tabs: Variables, Models, Datasets, Setup, Runs, Results, and Visualize. The 'Variables' tab is currently selected. On the left, a sidebar titled 'Forecasting: Investigate food security in South Sudan for August 2017' contains a list of questions. The main content area displays the question 'What flooding is expected in the Pongo Basin for the 2017 lean season?' and provides instructions for selecting variables. Under 'Response Variables', 'European Flooding Index' is selected. Under 'Driving Variables', 'Precipitation' is selected. A 'Notes' field is at the bottom.

Forecasting: Investigate food security in South Sudan for August 2017

- What are expected crop yields for maize and sorghum during the summer of 2017?
- What flooding is expected in the Pongo Basin for the 2017 lean season?
- Will flooding impact transportation in the Pongo Basin area during the 2017 lean season?
- What is the modeled yield for maize and sorghum in the Pongo Basin during the 2017 lean season?
- What is the effective modeled yield for maize and sorghum in the Pongo Basin during the 2017 lean season?
- Would fertilizer subsidies affect maize and sorghum effective yields in the Pongo Basin during the 2017 lean season?
- What flooding is expected in the Gel Aliab area for the 2017 lean season?
- Will flooding impact transportation in the Gel

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

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

Variables

Response Variables* +

- European Flooding Index

Driving Variables +

- Precipitation

Notes

2 System recommends models

A screenshot of the MINT web application interface, showing the 'Models' step. The top navigation bar is the same as in the previous screenshot, but the 'Models' tab is now selected. The main content area displays the same question: 'What flooding is expected in the Pongo Basin for the 2017 lean season?'. It provides instructions for selecting models. A table lists three recommended models, all of which are selected. At the bottom, there are three buttons: 'COMPARE SELECTED MODELS', 'CANCEL', and 'SELECT & CONTINUE'. A 'Notes' field at the bottom contains the text: 'Selected the models that have been manually configured for the region.'

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

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
☒ PIHM v2.2 - Pongo Basin Configuration	Hydrology	Pongo Basin (South Sudan)	Flooding Index
☐ PIHM v2.2 - Pongo Basin Calibration	Hydrology	Pongo Basin (South Sudan)	Flooding Index
☒ 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.

Browse Datasets

Found 2 results

South Sudan Rivers and Ports

ID: 2008-06-12-12:12 (Date: 01/03/2018)

South Sudan Rivers and Ports

Additional Metadata

100% 75% Creative Commons Attribution

South Sudan Rivers and Ports

ID: 2008-06-12-12:12 (Date: 01/03/2018)

South Sudan Rivers and Ports

Additional Metadata

100% 75% Creative Commons Attribution

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

Groundnuts:

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

Sorghum:

Land Cost Adjustment: 0

Nitrogen Fertilizer Cost Adjustment: 0

Crop Price Adjustment: 0

Production (kg)

Crop	Production (kg)
cassava	~1,500,000
groundnuts	~2,500,000
maize	~1,500,000
sesame	~2,500,000
sorghum	~7,500,000

Total Nitrogen Fertilizer Use (kg)

Crop	Total Nitrogen Fertilizer Use (kg)
cassava	~500,000
groundnuts	~1,500,000
maize	~1,500,000
sesame	~2,000,000
sorghum	~3,000,000

Yield (kg/ha)

Crop	Yield (kg/ha)
cassava	~2,000
groundnuts	~500
maize	~1,000
sesame	~500
sorghum	~1,000

Nitrogen fertilizer use (kg/ha)

Crop	Nitrogen fertilizer use (kg/ha)
cassava	~50
groundnuts	~50
maize	~100
sesame	~50
sorghum	~100

Land Area (ha)

Legend:

- cassava
- groundnuts
- maize
- sesame
- sorghum

Future Work

- Incorporating representations of uncertainty
 - Many sources of uncertainty in modeling
 - Can be associated with all input values of any model
 - Must be propagated through the model and reflected in the model outputs
 - Many potential initial conditions must be considered
- Incorporating representations of sensitivity analysis
 - Sensitivity analysis evaluates the impact of the uncertainty contained in each input value on the outputs
 - Sensitivity analysis allows modelers to understand which variables' are important to specify and how accurately