USGS Notes for Extracting Entities for the Model Catalog:

Entities:

- Method: Machine learning or statistical model (NN, SVM, etc.)
- Task: e.g. Image classification, segmentation, etc.
- Application: e.g. tracking rainfall thresholds for landslides, etc.

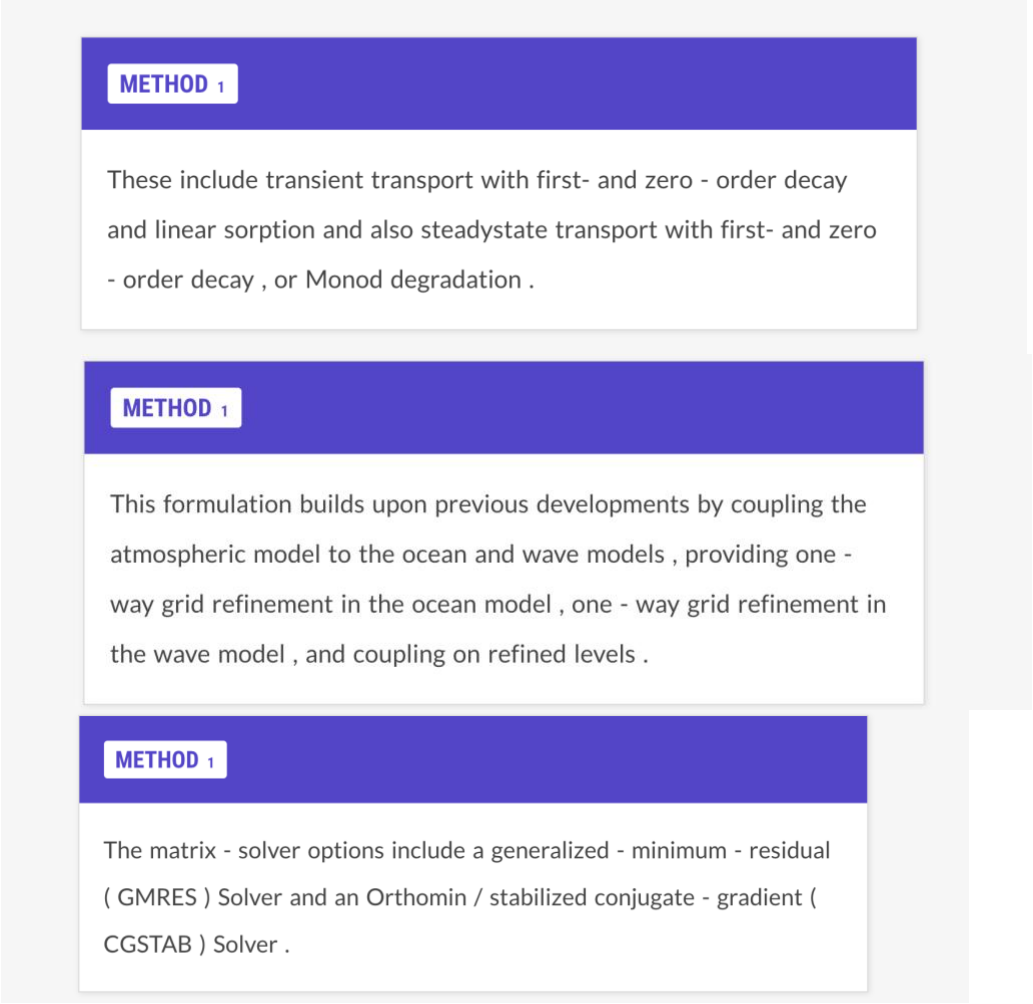
- Methods = GEO model or ML/Statistical Model (try separate and together)

Observations while manually scraping text:

- Tasks and applications may overlap. Difficulty distinguishing between the two for some articles.
- Some texts may not have a method.
- One article had a one-sentence section called "Introduction" as well as a short "Model Overview and System Requirements", but no Abstract section.
- Other sections to consider (though not consistent): "purpose and scope", "summary", "Model Overview and System Requirements"
- May reference models outside of text – Future models, other models …
- Mostly specific mathematical models for waterflow, watershed, water balance, landslides, etc. (not machine learning methods)

- GEO either has a lot of material to highlight or to try to weed out what are actual GEO models. ML has maybe 1-2 recognizable ML/Statistical Model in 20 texts (most are geological modeling). few examples for ML-method, so couldn't train well.
- Models built on prior models (are those models the method?)
- Accuracy correct? models this far label just about everything as an entity
- Need to tease out the geologic models.
- When trying GEO alone, I realized that I started to label any clear geology noun (volcano, stream, ocean, reservoir, etc.). These are not necessarily geologic models, but could be useful… the models are much harder to separate see GeoModel Notes and examples below
- GeoModel Notes:
    - digital elevation model, diffuse layer model,
    - 3D method of columns (landslides)
    - Vertical fluid flow (seepage flux)
    - "building a 3D domain"
    - Limit-equilibrium analysis, slope-stability analysis, slope stability
    - Aqueous geochemical calculations
    - (1D) transport calculations
    - Ion aqueous model, Pitzer aqueous model
    - Peng-Robinson equation (for gas solubility)
    - Transport simulations, geochemical simulations, heat transport equations,
    - Flow and transport: aquifer, pollutants, sediment, solute, heat, thermal energy,

    - Population (density, estimation, prediction, etc)
    - Water-balance model, water balance modeling,
    - Migration/Migration Cycle
- See examples below for help labeling:



build a tree:

Simulation/model of:
Uber -Geo-categories(A, B, C, …)

- Within each geo-category (geo-methods(A1, A2, A3), (B1), (C1, C2, C3, C4, C5) + ML/Statistical methods)
- This gets too complicated though because many methods are the same or could be built on other methods.

Volcano (volcanic ash, ??)
Earthquakes (subduction zone, ground motion prediction

Flooding/Waves
Migration
Reservoir water balance
Aqueous geochemical model
Precipitation/Slope Stability Analysis
Hydrodynamics, Vertical Fluid Flow
data estimation from isopleth map
solute transport, thermal energy transport, sediment transport
watersheds and redevelopment
water quality
evolution of coastal zone
climate change