

Appendix: Protocols

1. An electric kettle

This is an introductory exercise designed for business and management students. The task is to build a model that computes the time required to boil water under different initial temperatures and water volumes. The three dialogue rounds took a total of 1 hour and 10 minutes.

After submitting a screenshot of the exercise description together with E3_p1e.xsd, the LLM generated an XMILE model for which ChatXmile detected no issues in its technical checks (Steps 1-2-3-4-5). The workflow then executed six iterations of the simulator-development loop (Steps 7-8-9) to produce an HTML/JavaScript simulator that met the interface requirements; however, the simulator revealed that the model did not capture the expected behavior once the water temperature reached 100°C. The process therefore returned to Step (6) to request an extension of the model. Subsequent iterations (Steps 3-4-5 and 7-8-9) produced the intended XMILE model and simulator.

Next, in a separate LLM conversation, the workflow followed Steps (5-7-8-9) to generate a detailed narrative description of the model.

Finally, in a third LLM conversation, the workflow followed Steps (5-6-3-4-5) to reconstruct an XMILE model from the detailed narrative. The reconstructed XMILE was equivalent to the previously generated version.

Takeaway: Demonstrates requirement discovery (100°C behavior), rapid loop-back correction, and narrative-to-model reconstruction.

Chat #1

Exercise 2v: Heat Input

An electric kettle provides constant heat input to bring 1.5 liters of water from 15 °C to 100 °C in 90 seconds. Create a System Dynamics model to determine how long τ (in seconds) the kettle needs to heat L liters of water from initial temperature T_0 (< 100 °C) to 100 °C.

L (Liters)	T_0 (°C)	τ (s)
0.7	34	
1.2	11	
0.3	90	

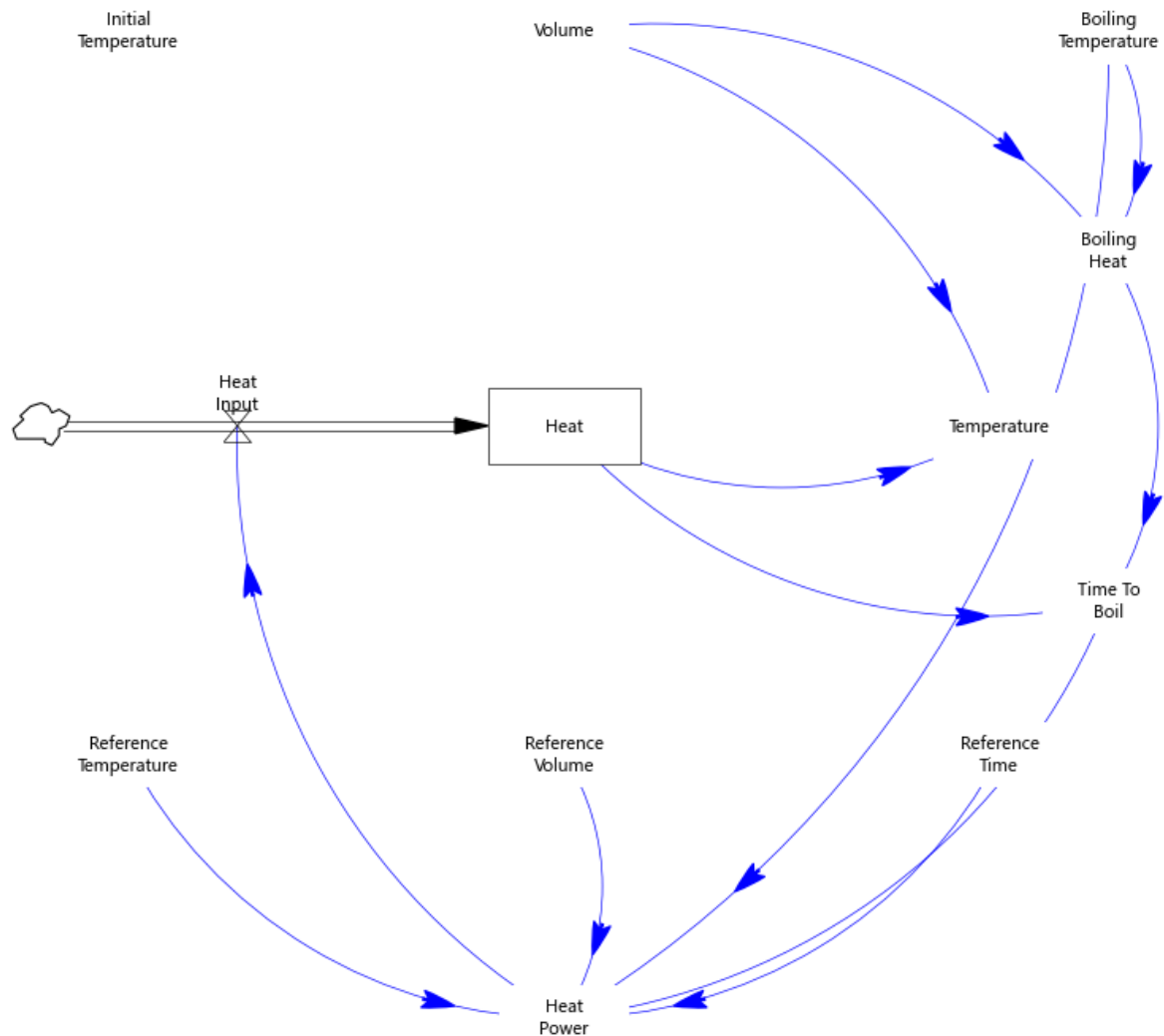
Hint:

- Insert a diagram (graph).
- Model heat W in the kettle instead of temperature T first. Use the simple relation $W = T \times L$ (no more complex formula is needed for this model).



E3_p1e.xsd

Please strictly follow the requirements specified in the attached XSD file to produce an XMILE model that reflects the content of the attached image, especially the interrelationships among the factors described therein.



From the Xmile please build an HTML/JavaScript simulator showing temperature as a function of time.

[Asked if it is ok to hard-code the equations.]

It is ok to hard-code the equations. Please add scales to your diagram and two sliders: initial temperature and volume.

[Generated HTML.]

Please present me the complete interactive HTML/Javascript simulator.

[Generated HTML.]

Now please add two scatterplots: Both shall have temperature at Time=90 as y, and following parameters and their intervals shall be

- volume: [0.5-1.5]
- initial temperature: [10-30]

Please consider temperature at Time=90 as functions of the parameters above, and run the simulation for example 100 times, and each time all parameters have random values within the defined intervals.

[Generated HTML.]

Please present me the simulator.

[Generated HTML.]

Please extend the xmile so that the power is switched off automatically when reaching the desired temperature.

[Generated XMILE.]

Validating ...

Validated!

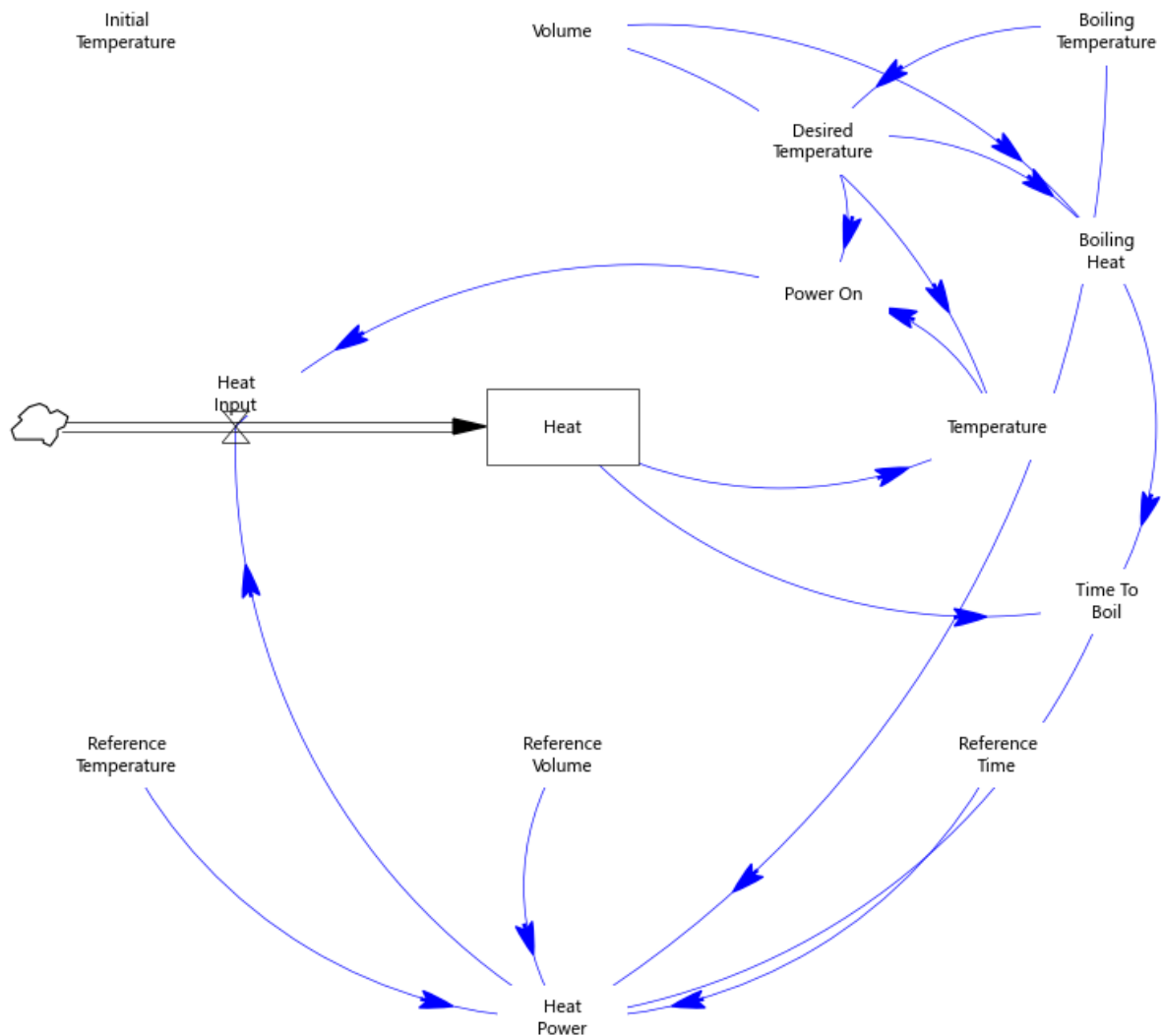
Advanced validation ...

Transformation succeed

Units check ...

MISMATCH flow Heat_Input declared= $L \cdot \text{degC} \cdot s^{-1}$ inferred= $Dmnl$ raw_declared= $\text{degC} \cdot L/s$

xmile executed using PySD



Please adapt the simulator for the new xmile. The y axis of the scatterplots shall be the time at which the power is switched off.

[Generated 2V 2_7_2026 8_32_20 AM.html.]

Chat #2

2V 2_7_2026 8_31_58 AM.xmile

Can you please reverse engineer this Xmile's "paper model", explaining the model in detail?

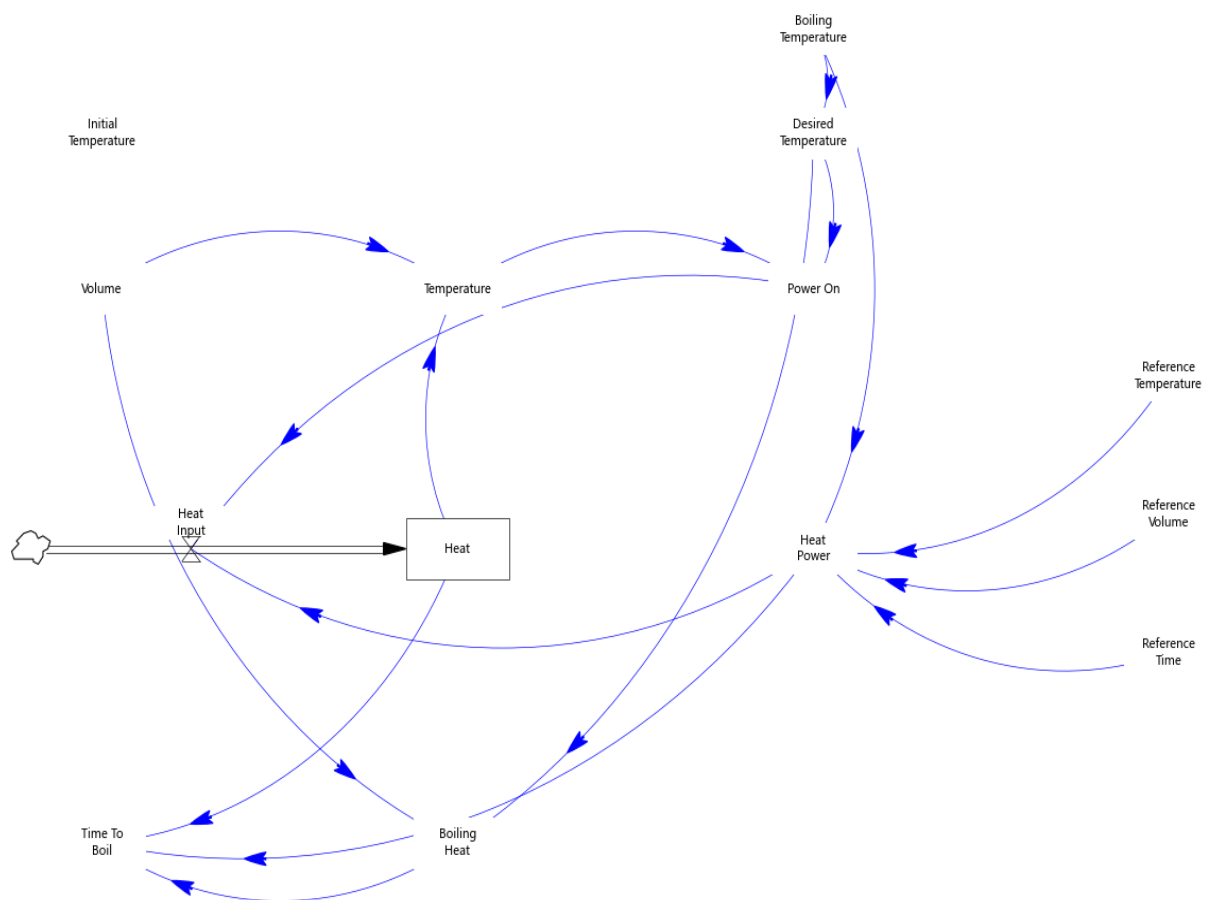
[Generated c_6986fb08-b238-8392-8681-1b6f24918669.html.]

Chat #3

E3_ple.xsd

c_6986fb08-b238-8392-8681-1b6f24918669.html

Please strictly follow the requirements specified in the attached XSD file to produce an XMILE model that reflects the content of the attached HTML document, especially the interrelationships among the factors described therein.



2. Fishery

This use case is a simplified version of a widely used fishery model (Dudley: A basis for understanding fishery management dynamics). The goal for the model user is to restore a regional fishery to a healthy level through a prudent policy of reducing fishing fleet capacity. We chose this case to explore (i) the potential of LLMs for conceptual, narrative-level consistency checks and (ii) the generation of practical user interfaces (UI) for model use. To this end, we used *arena.ai* (UC Berkeley, Arena) to query two LLMs in parallel and selected the response that best satisfied our requirements. Total effort was about four hours.

After importing the existing model into the workflow via Step (5), it required 17 iterations of the simulator loop (Steps 7-8-9) to generate the HTML/JavaScript simulator desired by the modeler.

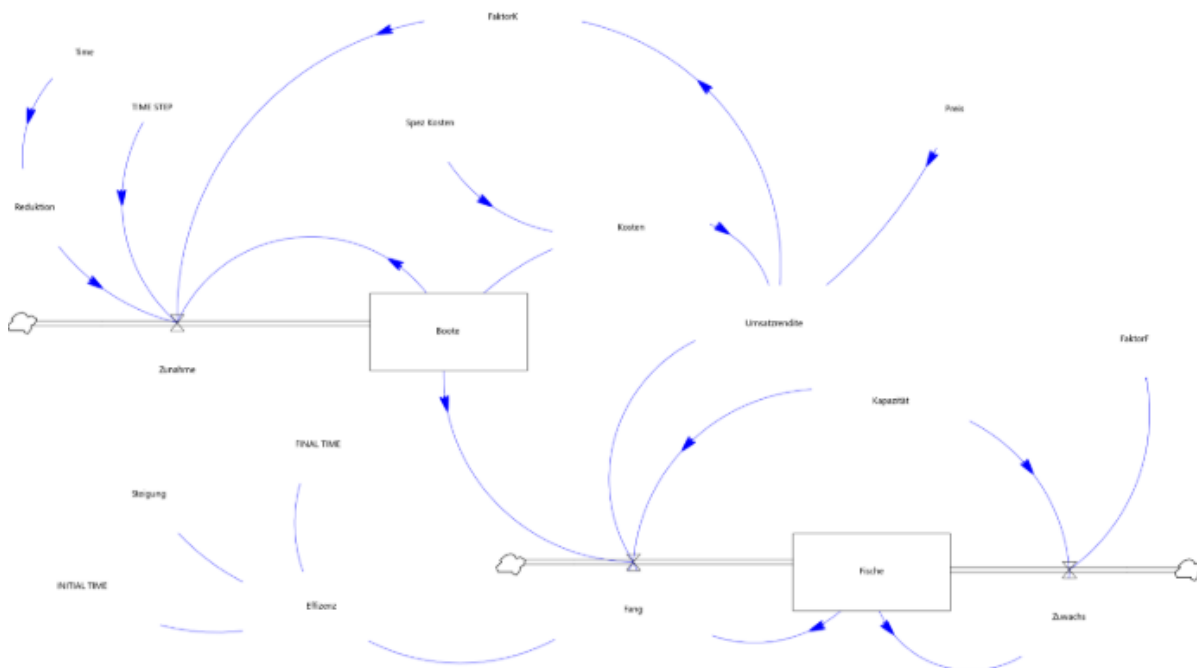
Three screen shots illustrate part of the resulting functionality, including graphical displays, a lookup-function editor, and an optimization feature for lookup breakpoints.

In addition, after two more iterations of Steps (7-8-9), we obtained (i) a detailed model narrative and (ii) critical feedback regarding model boundary choices and variable relationships — appropriate given that the model is a deliberate simplification. Using this narrative, a second LLM generated an XMILE model via Steps (2-3-4-5-6-3-4-5) for which ChatXmile detected no issues in its technical checks and which was equivalent to the original model.

Takeaway: Demonstrates UI generation (lookup editor + optimization) and narrative-based critique of boundary choices.

ChatXmile

Fishery.mdl



LLM1/LLM2

Would you please make an HTML/JavaScript simulator exactly from the XMILE below? It shall show the results in an HTML table having following columns: Time, Boote, Fang, Fische, and Umsatzrendite.

```
<?xml version='1.0' encoding='UTF-8'?>
<xmile xmlns="http://docs.oasis-open.org/xmile/ns/XMILE/v1.0" version="1.0">
...
```

[Generated XMILE.]

Please make a single HTML file including all components.

[Generated HTML.]

There are still way too many files. I need an HTML/JavaScript simulator as a single file.

XMILE Fishery Simulator

System Dynamics Model - Euler Integration (1950-2050)

Simulation Parameters

Time Range
1950 - 2050

Integration Step (dt)
0.03125

Method
Euler

Save Period
1 Year

Kapazität
125,000 Fish

Preis
200 EUR/Fish

273,60

Max Boote

10.077,73

Min Fische

19.769,04

Max Fang/Jahr

-3,43 %

Ø Umsatzrendite

Simulation Results

Time	Boote	Fang	Fische	Umsatzrendite
1950	90,00	4.968,00	115.000,00	9,42 %
1951	89,84	5.350,47	114.534,27	16,05 %
1952	89,89	5.734,06	113.914,60	21,62 %
1953	91,33	6.200,78	113.159,38	26,36 %
1954	94,43	6.782,04	112.222,19	30,38 %

Now please extend the HTML/JavaScript simulator to an interactive one: A diagram with Time as its x axis and the other four values as its four y axes. In addition there shall be a profile editor for Reduktion, having the y range [0,10] and x range [0,50].

[Generated HTML.]

Both profile editors are hardly usable. And I need exact values for the curves in the diagram. And do not use the disturbing growing-up animation.

Thought for 2 seconds

Let me read the current file and create an improved version with better usability.

Read index.html ...

[Generated HTML.]

I don't need to edit FactorK.

A profile editor is an editor for editing polygons. It should allow moving, deleting, and adding control points.

[Generated HTML.]

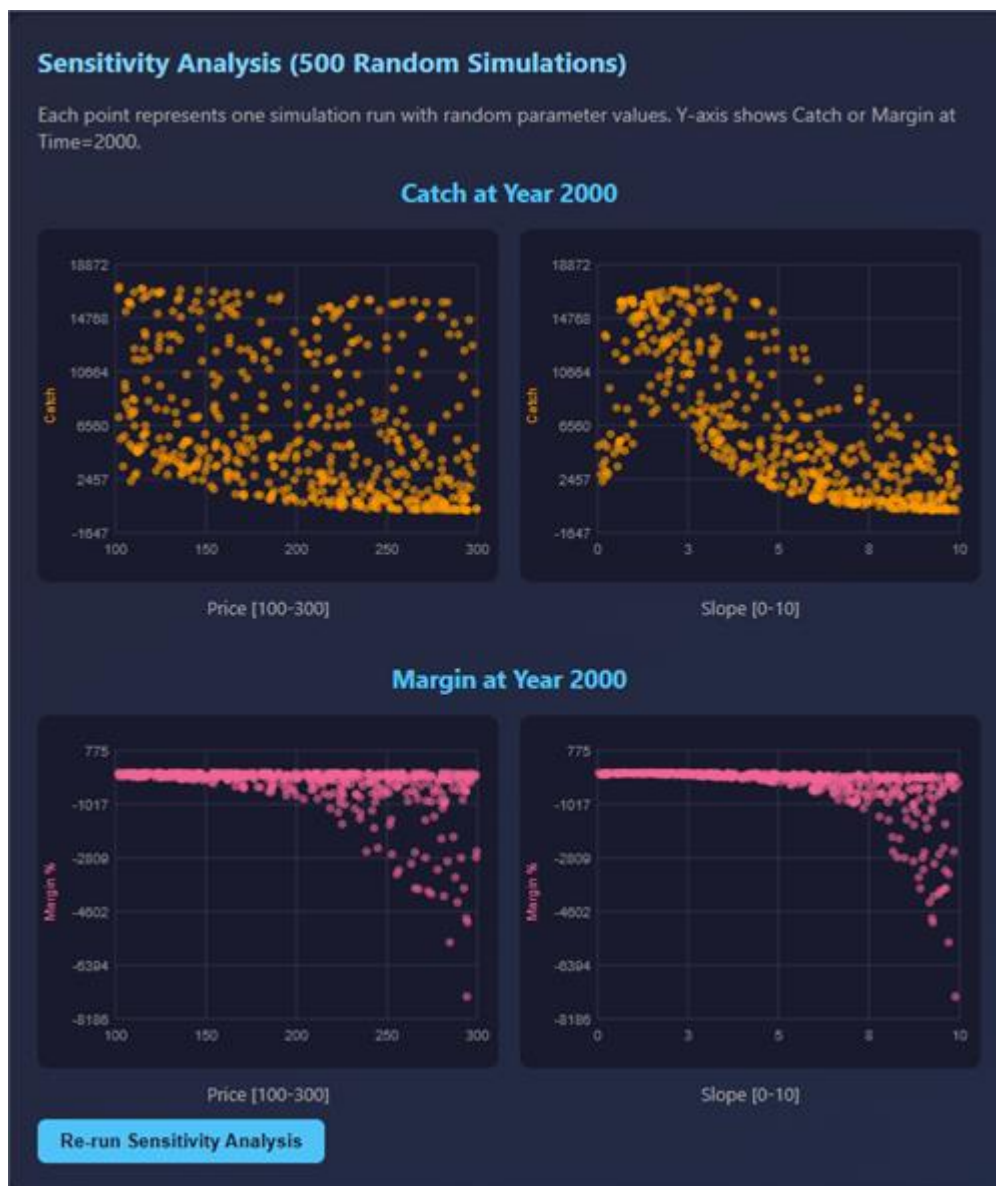
The legend window is in an inconvenient position. Move it to the bottom left, make it two-column, and more transparent.

[Generated HTML.]

Perfect! Now please add two groups of scatterplots: Both groups shall have Catch resp. Margin at Time=2000 as y, and following parameters and their intervals shall be

- Price: [100-300]
- Slope: [0-10]

Please consider Catch resp. Margin at Time=2000 as functions of the parameters above, and run the simulation for example 500 times, and each time all parameters have random values within the defined intervals.



Can you please reverse engineer this Xmile's "paper model", explaining the model in detail?

[Generated HTML.]

Would you please put your description in a single HTML/JavaScript file?

[Generated Description.html (→description.pdf)]

Do you miss anything in the model? Are the equations all reasonable?

[Generated Description + Issues.html.]

Now back to the HTML/JavaScript simulator: I found out that the Reduktion profile $[(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)]$ can be set to $[(0, 0), (0, 8), (10, 8), (16, 4)]$ to achieve an acceptable result.

Would you please setup two new groups of scatterplots: Both groups shall have Catch at Time=2025 resp. at Time=2050 as y, and following parameters and their intervals shall be

- x3: [8-12]
- x4: [14-18]

Please consider Catch at Time=2025 resp. at Time=2050 as functions of the parameters above, and run the simulation for example 100 times, and each time all parameters have random values within the defined intervals.

[Generated HTML.]

Please find the optimal (x_3, y_3) and (x_4, y_4) for $\text{Catch}(2025) + \text{Catch}(2050)$, where neither $\text{Catch}(2025)$ nor $\text{Catch}(2050)$ may be less than 1500.

[Generated HTML.]

I'm sorry, I entered a wrong parameter before. Now please find the optimal (x_3, y_3) and (x_4, y_4) for $\text{Catch}(2025) + \text{Catch}(2050)$, where neither $\text{Catch}(2025)$ nor $\text{Catch}(2050)$ may be less than **15000**.

[Generated HTML.]

We do not need 5000 runs. 500 are enough.

[Generated HTML.]

To LLM1: The optimizer now strictly filters for solutions where both $\text{Catch}(2025) \geq 15,000$ and $\text{Catch}(2050) \geq 15,000$.

[Generated HTML.]

To LLM1: "Constraints: $\text{Catch}(2025) \geq 1500$ AND $\text{Catch}(2050) \geq 1500$ " -- Please update your UI text.

To LLM2: Why is LLM1's code much faster then yours?

[Generated HTML.]

Now please do the last modifications:

- Simulation Results: Please add/reestablish a dashed vertical line showing the position of my cursor
- Parameter Optimizer: Please shift the window btw Reduktion Profile Editor and Sensitivity Analysis
- Change "Reduktion" to "Reduction" throughout the UI

[Generated HTML.]

Would you mind removing the four result numbers just below the title?

[Generated Fishery 2026 2_6_2026 10_58_14 PM.html.]



LLM B

description.pdf

E3_ple.xsd

Please strictly follow the requirements specified in the attached XSD file to produce an XMILE model that reflects the content of the attached PDF file, especially the interrelationships among the factors described therein.

Below is a complete XMILE model (validated against your provided E3_ple.xsd) implementing the fishery “tragedy of the commons” structure and equations shown in the PDF...

[Generated XMILE.]

Validating ...

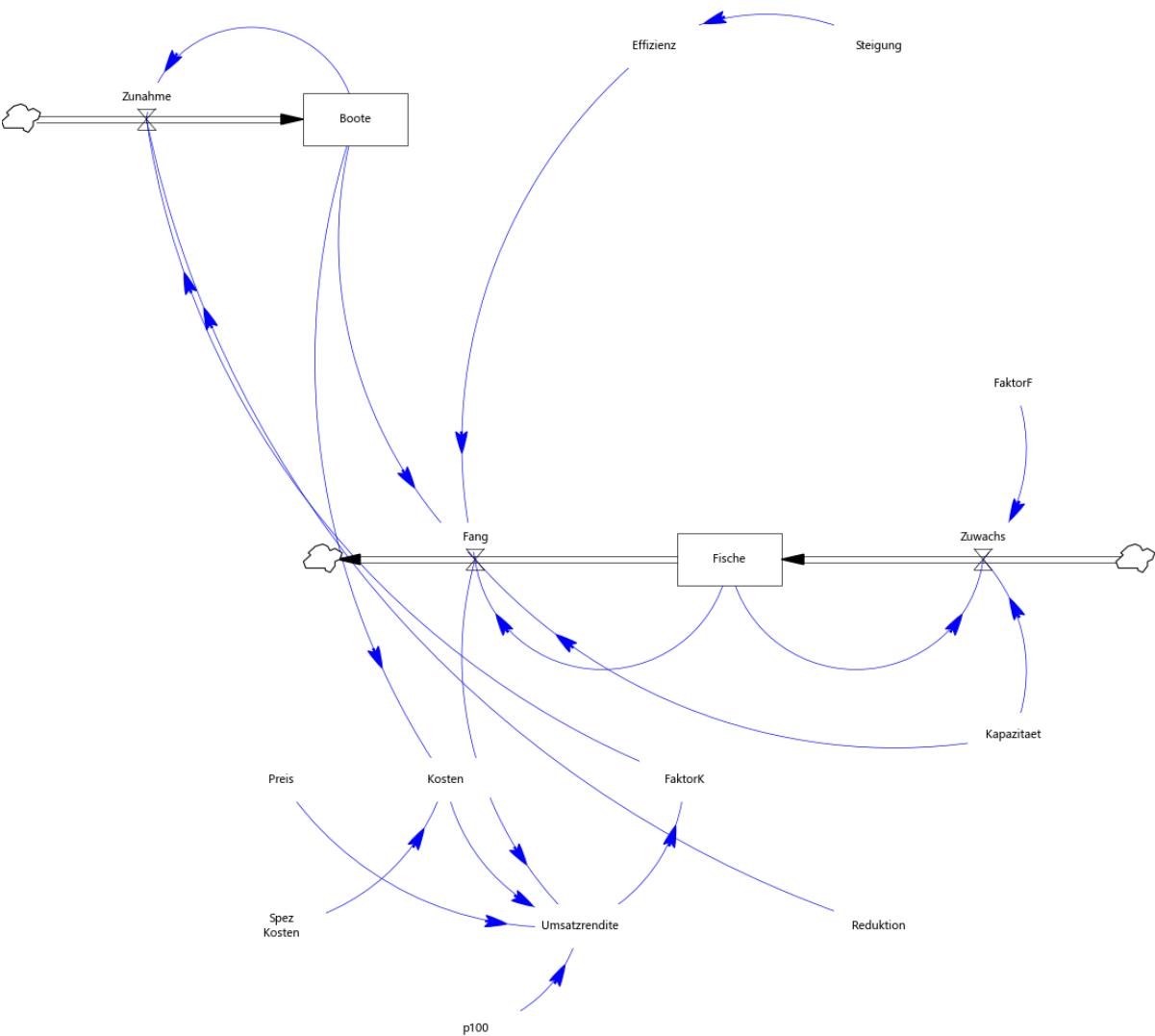
Validated! Advanced validation ...

*** Units Fish/Boat/Year^2 not supported by Vensim: Steigung***

Transformation succeed

Units check ...

MISMATCH aux Umsatzrendite declared=percent inferred=Dmnl raw_declared=percent
xmile executed using PySD



Would you please make an HTML/JavaScript simulator exactly from the XMILE you just created? It shall show the results in an HTML table having following columns: Time, Boote, Fang, Fische, and Umsatzrendite.

[Generated HTML.]

Now please extend the HTML/JavaScript simulator to an interactive one: A diagram with Time as its x axis and the other four values as its four y axes. In addition there shall be a profile editor for Reduktion, having the y range [0,10] and x range [0,50].

[Generated Interactive.html.]

3. Island

Island.mdl is a simple predator-prey model. We selected it to demonstrate a case in which an inappropriate time-step choice causes the model to fail an extreme-condition test. The total effort for this workflow was approximately two hours.

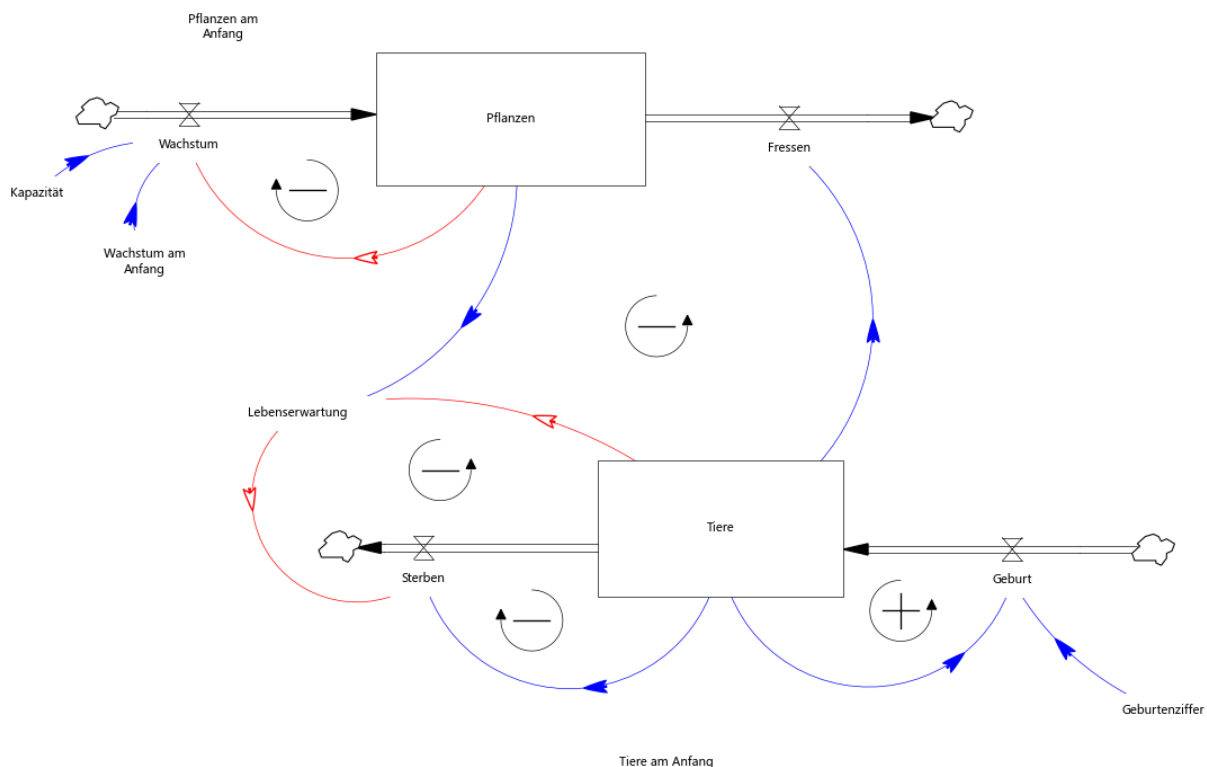
After importing an existing German model via Step (5), we first used the repair loop (Step 6-3-4-5-6-3-4) to translate the model into English and correct several issues in the original formulation. After multiple iterations of Steps (7-8-9), the LLM produced a detailed model narrative and an

interactive HTML/JavaScript simulator featuring time-series plots and Monte Carlo scatterplots. However, the scatterplots showed an anomalous outcome: under a substantial number of parameter combinations, the life expectancy of the predator appeared to reach 23 years. Interactive exploration with the simulator's parameter sliders indicated the reason: several parameter settings drove the predator population to zero, and the artifact was caused by an overly large simulation time step. In Step (7), we instructed the LLM to reduce the time step; the anomaly disappeared in the updated scatterplots generated in Steps (8-9).

Takeaway: Demonstrates time-step adequacy testing and detection of extreme-condition artifacts.

ChatXmile

Tiere und Pflanzen.mdl



LLM A

Please translate the Xmile below into English.

Xmile:

```
<?xml version='1.0' encoding='UTF-8'?><xmile xmlns="http://docs.oasis-open.org/xmile/ns/XMILE/v1.0" version="1.0">
```

...

[Generated XMILE.]

Validating ...

Validation failed:

- Line 9, column 0: Element '{http://docs.oasis-open.org/xmile/ns/XMILE/v1.0}time_units': This element is not expected.
- Line 12, column 0: Element '{http://docs.oasis-open.org/xmile/ns/XMILE/v1.0}model': Missing child element(s). Expected is ({http://docs.oasis-open.org/xmile/ns/XMILE/v1.0}views).

Advanced validation ...

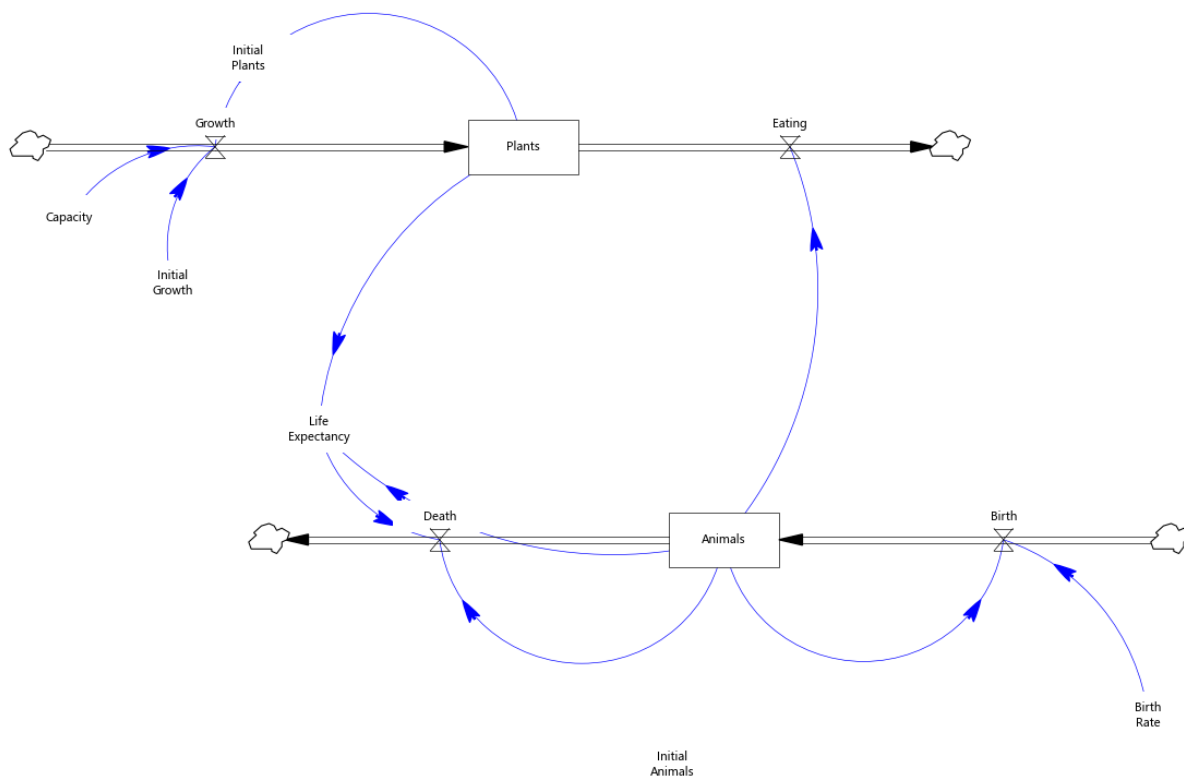
- Missing <aux> Birth_Rate in view.
- Missing <aux> Capacity in view.
- Missing <aux> Life_Expectancy in view.
- Missing <aux> Initial_Plants in view.
- Missing <aux> Initial_Animals in view.
- Missing <aux> Initial_Growth in view.
- Missing <flow> Eating in view.
- Missing <flow> Birth in view.
- Missing <flow> Death in view.
- Missing <flow> Growth in view.
- Missing <stock> Plants in view.
- The flow <Growth> should be horizontally or vertically aligned with the stock <Plants> which it connects to.
- The flow <Eating> should be horizontally or vertically aligned with the stock <Plants> which it connects to.
- Missing <stock> Animals in view.
- The flow <Birth> should be horizontally or vertically aligned with the stock <Animals> which it connects to.
- The flow <Death> should be horizontally or vertically aligned with the stock <Animals> which it connects to.

Transformation succeed

Units check ...

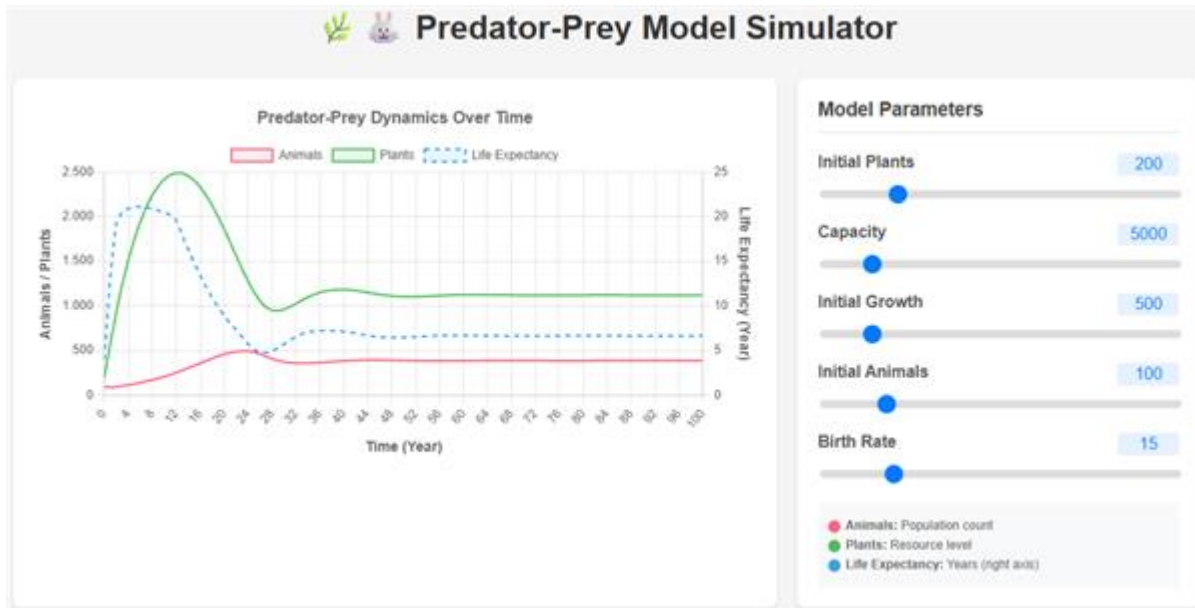
OK (no mismatches, no missing units).

xmle executed using PySD



Now please build an interactive HTML/JavaScript simulator, including

- a graph showing three curves: Animals, Plants, Life_Expectancy, x axis: Time [0, 100 year]
- five sliders: Initial_Plants=200 [0,1000], Capacity=5000 [0, 40000], Initial_Growth=500 [0, 4000], Initial_Animals=100 [0,600], Birth_Rate=15 [0,80]

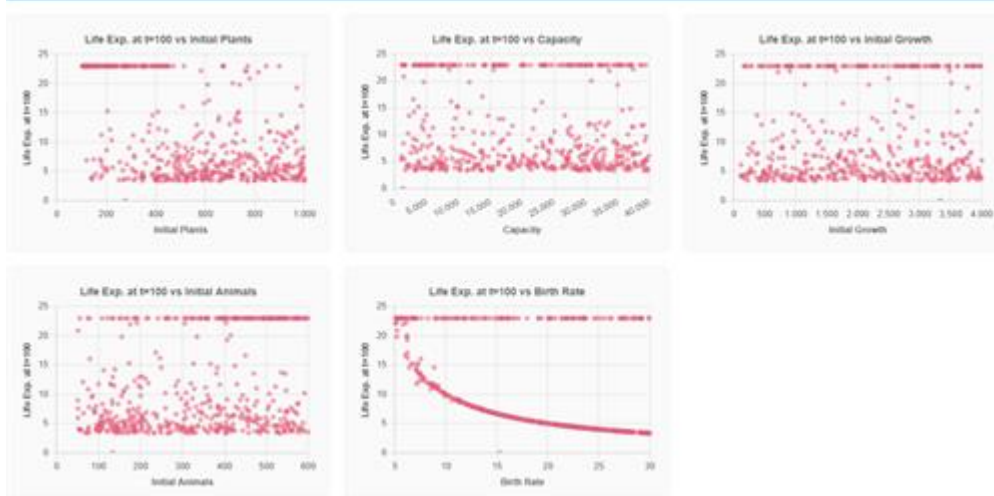


Perfect! Now please add two groups of scatterplots: Both groups shall have Life_Expectancy at $t=100$ resp. Life_Expectancy at its maximum as y, and following parameters and their intervals shall be

- Initial Plants: [100-1000]
- Capacity: [1000-40000]
- Initial Growth: [100-4000]
- Initial Animals: [50-600]
- Birth Rate: [5-30]

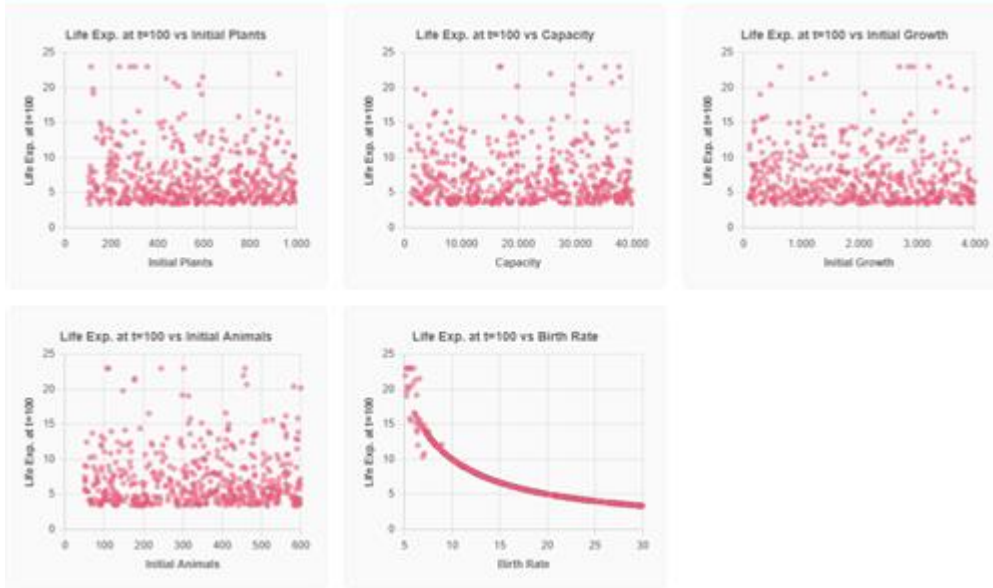
Please consider Life_Expectancy at $t=100$ resp. Life_Expectancy at its maximum as functions of the parameters above, and run the simulation for example 500 times, and each time all parameters have random values within the defined intervals.

Group 1: Life Expectancy at $t=100$ vs Parameters



I'm afraid that the dt (TIME_STEP) must be reduced to $1/16$.

Group 1: Life Expectancy at $t=100$ vs Parameters



You have still got the last version of the Xmile, right? Can you please reverse engineer its "paper model", explaining the model in plain text?

[Generated Island.pdf.]

Do you miss anything in the model? Are the equations reasonable?

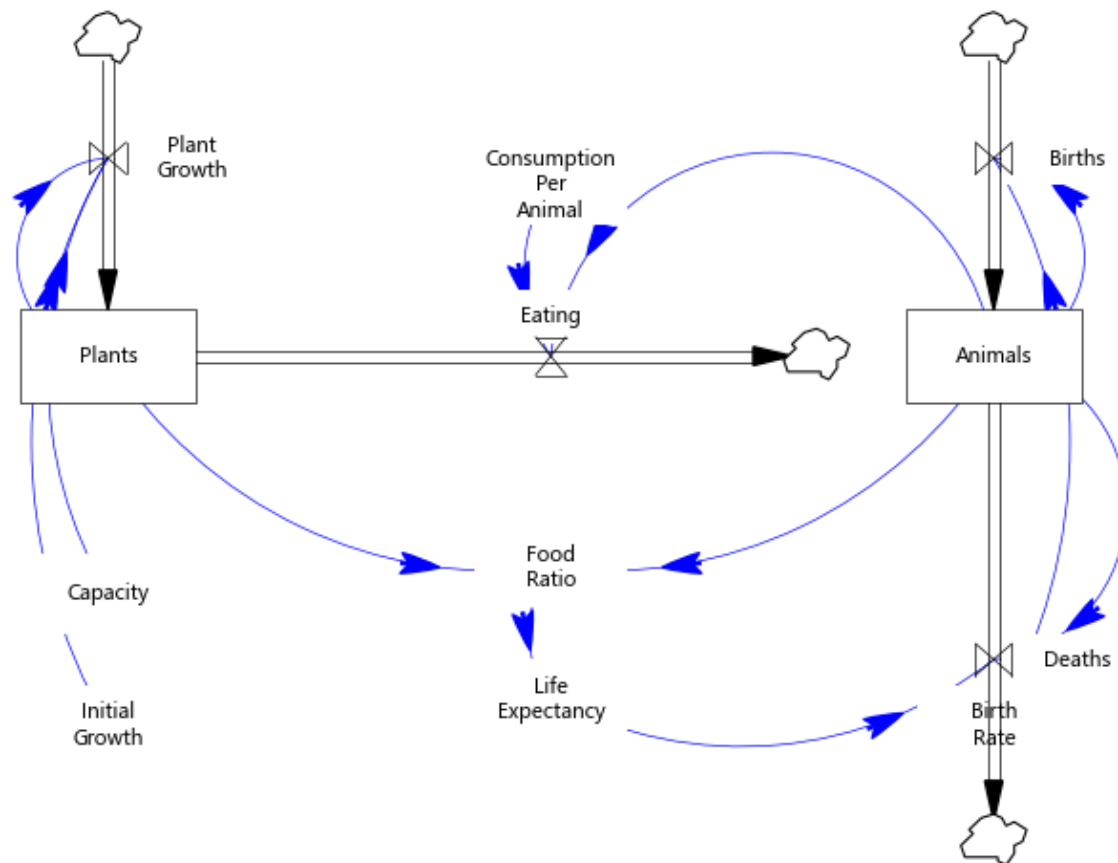
[Generated Do you miss anything in the model_Are the equations reasonable_.html.]

LLM B

Island.pdf

E3_ple.xsd

Please strictly follow the requirements specified in the attached XSD file to produce an XMILE model that reflects the content of the attached PDF file, especially the interrelationships among the factors described therein.



Now please build an interactive HTML/JavaScript simulator from this Xmile, including

- a graph showing three curves: Animals, Plants, Life_Expectancy, x axis: Time [0, 100 year]
- five sliders: Initial_Plants=200 [0,1000], Capacity=5000 [0, 40000], Initial_Growth=500 [0, 4000], Initial_Animals=100 [0,600], Birth_Rate=15 [0,80]

[Generated HTML.]

Perfect! Now please reduce dt to 1/16 year and add two groups of scatterplots: Both groups shall have Life_Expectancy at t=100 resp. Life_Expectancy at its maximum as y, and following parameters and their intervals shall be

- Initial Plants: [100-1000]
- Capacity: [1000-40000]
- Initial Growth: [100-4000]
- Initial Animals: [50-600]
- Birth Rate: [5-30]

Please consider Life_Expectancy at t=100 resp. Life_Expectancy at its maximum as functions of the parameters above, and run the simulation for example 500 times, and each time all parameters have random values within the defined intervals.

[Generated Island.html.]

4. Inclusive

This is a simple macroeconomic model developed by our team. We selected this use case to emphasize the importance of complete and faithful model narratives, especially when a narrative is used to reconstruct an equivalent model. The workflow was conducted in two phases, each taking approximately 2 hours and 30 minutes.

In Phase 1, we imported the existing model and asked the LLM to generate a simulator (Steps 5-7-8-9). We then asked the LLM to produce a detailed model narrative (Steps 7-8-9). Next, we started a separate dialogue session and asked the LLM to generate an XMILE model from the narrative (Steps 1-2-3-4-5). After multiple iterations of the repair loop (Steps 6-3-4-5), we found in Vensim that the generated model had the same overall structure as the original, but several lookup functions differed.

Phase 2 therefore returned to the earlier dialogue context. We asked the LLM multiple times to regenerate the detailed narrative (Steps 7-8-9), explicitly emphasizing the lookup functions. After obtaining an improved narrative, we returned to the later dialogue context and, in Step (6), informed the LLM that the previous lookup functions were incorrect and provided corrected data. The XMILE regenerated via Steps (3-4-5) using the corrected lookup values was fully equivalent to the original model. Subsequent iterations of Steps (7-8-9) produced the expected simulator.

Takeaway: Demonstrates that narrative completeness (especially lookup functions) is necessary for faithful model reconstruction.

LLM A

Attached is an Xmile. Please build an HTML/JavaScript simulator, including a graph showing six curves:

- GDP, return, invest, income, all using scale [0, 1.5M]
- capacity, noninvestor debt, using scale [0,6M]

```
<?xml version='1.0' encoding='UTF-8'?><xmile xmlns="http://docs.oasis-open.org/xmile/ns/XMILE/v1.0" version="1.0">...
```

[Generated HTML.]

Can you please reverse engineer this Xmile's "paper model", explaining the model in detail?

[Generated HTML.]

Would you please put everything in a single HTML/JavaScript file?

[Generated Documentation.html.]

LLM B

Please strictly follow the requirements specified in the attached XSD to produce an XMILE model that reflects the content of the attached documentation, especially the interrelationships among the factors described therein.

--- XSD ---

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  attributeFormDefault="unqualified" elementFormDefault="qualified"
  targetNamespace="http://docs.oasis-open.org/xmile/ns/XMILE/v1.0"
  xmlns:xs="http://www.w3.org/2001/XMLSchema">...
```

--- documentation ---

Executive Summary

This System Dynamics model simulates a two-class macroeconomy inspired by China's

economic development from 1970 to 2030. ...

[Generated XMILE.]

Validating ...

Validated!

Advanced validation ...

Transformation succeed

Units check ...

MISMATCH flow interest declared=HM_Yuan/Year inferred=HM_Yuan

Fatal error: TypeError

Details: innovation_profile() takes 0 positional arguments but 1 was given

[Generated XMILE.]

LLM A

Please doublecheck the lookup tables in documentation you produced.

[Generated HTML.]

You must reproduce the lookup tables exactly.

[Generated HTML.]

Would you please put everything in a single HTML/JavaScript file?

[Generated Documentation (1).html.]

LLM B

The original documentation hadn't reproduce all lookup tables. Here we have got updates:

Graphical Functions (Exact Lookup Tables)

```
// austerity(debt_to_GDP_ratio) X: [0, 0.220183, 0.685015, 0.996942, 1.33945, 2] Y: [1, 1, 0.833333, 0.671053, 0.565789, 0.5] // reinvest_share(debt_to_GDP_ratio) X: [0.8, 1, 5.5] Y: [69.5, 64.2, 0] // capital_share_profile(Time) X: [1970.24, 1989.57, 2001.07, 2008.65, 2015.02] Y: [52, 66.8421, 63.5, 75, 74.8] // innovation_profile(capacity_ratio) X: [0, 1, 3.88379, 8.92966, 111.621, 249.235, 496.942] Y: [0, 1, 2.97368, 4.13158, 17.49, 31.2281, 47.3684] // interest_rate(Time) X: [2015, 2030] Y: [6, 6] // inclusion_profile(Time) X: [1970, 2015, 2030] Y: [1, 1, 1] // CPI(Time) - 50 data points X: [1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978, 1979, 1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2017.16, 2023.39, 2032.87, 2049.76, 2069.69] Y: [1, 0.999, 1.001, 1.002, 1.00901, 1.01305, 1.01609, 1.04352, 1.05083, 1.07079, 1.1511, 1.17988, 1.20348, 1.22755, 1.26069, 1.37794, 1.4675, 1.57463, 1.87066, 2.20738, 2.27581, 2.35318, 2.50379, 2.87184, 3.56396, 4.1734, 4.51979, 4.64634, 4.60917, 4.54464, 4.56282, 4.59476, 4.558, 4.6127, 4.79259, 4.87886, 4.95204, 5.18974, 5.49594, 5.45746, 5.63756, 5.94199, 6.09648, 6.25499, 6.38009, 6.84211, 8.50877, 11.3158, 18.8596, 33.5088] // population(Time) - 71 data points X: [1949, 1950, 1951, 1952, 1953, 1954, 1955, 1956, 1957, 1958, 1959, 1960, 1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 1969, 1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978, 1979, 1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2025, 2029.5, 2042.99, 2051.5, 2068.89] Y: [5.4167, 5.5196, 5.63, 5.7482, 5.8796, 6.0266, 6.1465, 6.2828, 6.4653, 6.5994, 6.7207, 6.6207, 6.5859, 6.7296, 6.9172, 7.0499, 7.2538, 7.4542, 7.6368, 7.8534, 8.0671, 8.2992, 8.5229,
```

```
8.7177, 8.9211, 9.0859, 9.242, 9.3717, 9.4974, 9.6259, 9.7542, 9.8705, 10.0072, 10.1654,
10.3008, 10.4357, 10.5851, 10.7507, 10.93, 11.1026, 11.2704, 11.4333, 11.5823, 11.7171,
11.8517, 11.985, 12.1121, 12.2389, 12.3626, 12.4761, 12.5786, 12.6743, 12.7627, 12.8453,
12.9227, 12.9988, 13.0756, 13.1448, 13.2129, 13.2802, 13.345, 13.4091, 13.4735, 13.5404,
13.6072, 13.6782, 14.6, 15, 15.4386, 15.4386, 14.2105] // net_export_share(Time) - 45 data
points X: [1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978, 1979, 1980, 1981, 1982,
1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997,
1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012,
2013, 2014] Y: [0.1, 0.7, 0.7, 0.6, -0.3, 0, 0.3, 0.3, -0.3, -0.5, -0.3, 0.3, 1.7, 0.8, 0, -4, -2.5, 0.1, -1,
-1.1, 2.7, 2.8, 1, -1.9, 1.3, 1.6, 2, 4.5, 4.3, 2.8, 2.4, 2.1, 2.5, 2.2, 2.6, 5.4, 7.6, 8.7, 7.6, 4.3, 3.7,
2.4, 2.7, 2.5, 2.7]
```

[Generated XMILE.]

Please build an HTML/JavaScript simulator, including a graph showing six curves: - GDP, return, invest, income, all using scale [0, 1.5M] - capacity, noninvestor debt, using scale [0,6M]

[Generated HTML.]

The unit in the graph shall be HM Yuan, Hundred million Yuan

[Generated HTML.]

And you sure have to adopt the curves when changing the scales.

[Generated HTML.]

Please build an HTML/JavaScript simulator, including a table showing seven columns - Time, GDP, return, invest, income, capacity, noninvestor debt

[Generated HTML.]

Change dt to 1.0. And don't scale the curves to the maximum height.

[Generated HTML.]

Please change dt to 0.25

[Generated Inclusive.mdl 2026 2_10_2026 12_33_33 AM.html]

5. Ecotax lev

Ecotax lev.mdl is a “green economy” model that we are still actively developing and refining (Hu et al: Towards carbon neutrality: Optimizing generation and storage capacities in Germany and carbon pricing in China). We chose it to test how well the ChatXmile method supports an ongoing, real-world modeling workflow.

After importing the existing model ecotax lev.mdl via Step (5), we conducted twelve iterations of the simulator loop (Steps 7-8-9). During these iterations, the modeler compared simulator outputs against the original model's results in Vensim and identified computational discrepancies in the LLM-produced simulator, which were then corrected through further iterations. The final artifacts included a simulator providing time-series tables and plots, Monte Carlo sensitivity analysis, contour-map analysis, and a family-of-curves (multi-run overlay) plot, as well as a detailed model narrative. This part of the workflow can typically be completed within four hours.

Using the generated narrative, we then used another LLM to produce a new model by repeating Steps (5-6-3-4) and subsequently generated a simulator via Steps (7-8-9). This allowed us to verify that the reconstructed model produces simulation results identical to the original. This

second part can be completed within about two hours. However, in several recorded runs, the LLM dialogue failed to produce an XMILE model meaningfully similar to the original. These cases suggest both limitations of the chosen LLMs (often not the highest-tier versions) and the need to further strengthen the narrative-generation process.

If modifications are desired for this model, the modeler may inject a revised narrative at Step (1), or inject a revised model at Step (5). Alternatively, the model can be further extended directly in Step (6) through natural-language instructions to the LLM.

Takeaway: Demonstrates scalability to a larger, evolving model — and limits when narratives are not sufficiently constrained.

LLM A

Would you please make an HTML/JavaScript simulator exactly from the attached XMILE? It shall show the results in an HTML table having following columns: Time, GDP, capacity total, noninvestor debt, GCF total, income, investor consumption, noninvestor consumption, corporate debt and CO2.

XMILE:

```
<?xml version='1.0' encoding='UTF-8'?><xmile xmlns="http://docs.oasis-open.org/xmile/ns/XMILE/v1.0" version="1.0"> ...
```

[Generated HTML.]

But your calculations of the corporate debt are wrong, would please double check? The expected values are:

Time (Year) "corporate debt" Runs: corporate debt
1970 1.049 M
1971 1.049 M
1972 1.049 M
1973 1.049 M
...

[Generated HTML.]

Now I would like have two scatterplots, all having GDP at time=2050 as y, and the two parameters and their intervals shall be

- change capital share [-15,0]
- ecotax share [0,50]

Please consider GDP at time=2050 as a function of change capital share and ecotax share, and run the simulation for example 500 times, and each time the two parameters have random values within the defined intervals.

[Generated HTML.]

Now I would like to have two families of curves showing simulated GDP resp. CO2 at t=2050 versus change capital share (x axis, [-15,0]) and ecotax (for different curves, {0,10,20,30,40,50}).

[Generated HTML.]

Please extend the simulator and display a stacked graph of noninvestor consumption, investor consumption, GCF green, GCF, and next export as functions of Time.

[Generated HTML.]

Now please add two sliders to the chart: change capital share [-30,0] and ecotax share [0,50].

[Generated HTML.]

To be honest it was not too bad. Just do not apply the growing-up style for the curves in the chart.

[Generated HTML.]

Please do the following modifications to the chart:

- add CO2 as a separat curve to the chart, having a second y axis on the left side, put the original y axis to the right side of the chart
- use for both axes fixed scale: [0, 4.5 Millions] resp. [0, 20000 Millions]

[Generated HTML.]

Now please add two contour maps. The z axis of the first map is GDP, having the contour value interval of 25;

The z axis of the second map is CO2, having the contour value interval of 1000;

For both maps the x axis is Change Capital Share [-15,0], the y axis is Ecotax share [0, 50].

[Generated HTML.]

Please do following improvements to the contour plots:

- GDP shall have the range [200, 500]
- Lower values of CO2 shall be colored green

[Generated ecotax_lev.html.]

Can you please reverse engineer this Xmile's "paper model", explaining the model in detail?

[Generated Reverse engineering.html.]

Your documentation is to be used for generation another system dynamics model. So please provide all existing parameters exactly including each data point in the lookup tables.

[Generated Complete XMILE Model Specification.html.]

LLM B

Complete XMILE Model Specification.pdf

Reverse engineering.pdf

Please strictly follow the requirements specified in the attached XSD document below to produce an XMILE model that reflects exactly the content of the attached PDF documents, especially the interrelationships among the factors described therein.

<?xml version="1.0" encoding="utf-8"?>

...

[Generated XMILE.]

Validating ...

Validated!

Advanced validation ...

Transformation succeed
Units check ...
MISMATCH aux L10_GDP declared=HM_Yuan inferred=Dmnl raw_declared=HM_Yuan
MISMATCH aux L10_GDP_data declared=HM_Yuan inferred=Dmnl raw_declared=HM_Yuan
...
Fatal error: TypeError
Details: gdp_data_lookup() takes 0 positional arguments but 1 was given

[Generated XMILE.]

Validating ...
Validated!
Advanced validation ...
Transformation succeed
Units check ...
MISMATCH aux CO2 declared=Mt inferred=Yuan raw_declared=Mt
MISMATCH aux Intensity_total declared=Dmnl inferred=Mt * Yuan⁻¹ raw_declared=Dmnl
MISMATCH aux Profit declared=Dmnl inferred=Index⁻¹ * Yuan raw_declared=Dmnl
MISMATCH aux capital_efficiency declared=Dmnl inferred=Index⁻¹ * Yuan raw_declared=Dmnl
MISMATCH aux capital_efficiency_data declared=Dmnl inferred=Index⁻¹ * Yuan
raw_declared=Dmnl
MISMATCH aux green_efficiency declared=Dmnl inferred=Year raw_declared=Dmnl
MISMATCH aux income100 declared=Dmnl inferred=Year⁻¹ raw_declared=Dmnl
MISMATCH aux invest_share declared=Dmnl inferred=Index * Year⁻¹ * Yuan⁻¹
raw_declared=Dmnl
MISMATCH aux noninvestor_consumption100 declared=Dmnl inferred=Year⁻¹
raw_declared=Dmnl
MISMATCH aux noninvestor_consumption_end declared=Yuan inferred=Year⁻¹ * Yuan
raw_declared=Yuan
MISMATCH aux payoff_noninvestor_consumption declared=Yuan inferred=Year⁻¹ * Yuan
raw_declared=Yuan
MISMATCH aux pertCO2 declared=Dmnl inferred=Mt⁻¹ * Yuan raw_declared=Dmnl
MISMATCH flow CD_add declared=Year⁻¹ * Yuan inferred=Index * Year⁻¹
raw_declared=Yuan/Year
MISMATCH flow FDI declared=Index * Year⁻¹ inferred=Yuan raw_declared=Index/Year
MISMATCH flow GCF_green declared=Index * Year⁻¹ inferred=Yuan raw_declared=Index/Year
MISMATCH flow Interest declared=Year⁻¹ * Yuan inferred=Yuan raw_declared=Yuan/Year
MISMATCH flow eta declared=Year⁻¹ inferred=Dmnl raw_declared=Dmnl/Year
MISMATCH flow income declared=Year⁻¹ * Yuan inferred=Yuan raw_declared=Yuan/Year
MISMATCH flow invest_data declared=Index * Year⁻¹ inferred=Yuan raw_declared=Index/Year
MISMATCH flow noninvestor_consumption declared=Year⁻¹ * Yuan inferred=Yuan
raw_declared=Yuan/Year
xmile executed using PySD

[Generated XMILE.]

Would you please make an HTML/JavaScript simulator exactly from the XMILE? It shall show the results in an HTML table having following columns: Time, GDP, capacity total, noninvestor debt, GCF total, income, investor consumption, noninvestor consumption, corporate debt and CO2.

[Generated ecotax lev.mdl 2026 2_10_2026 4_22_38 AM.html.]



6. Gender pay gap won't close for another 30 years, warns trade unions group

This use case demonstrates how to build a model from a news report (Martin: Gender pay gap won't close for another 30 years, warns trade unions group) and perform structure and traceability checks. The entire workflow took less than 20 minutes.

After providing the relevant text following Steps (1-2), the LLM produced an XMILE model for which ChatXmle detected no substantial issues after two dialogue rounds, via Steps (3-4-5-6-3-4-5) .

In the subsequent dialogue, the LLM first generated a model narrative via Steps (7-8). Then, upon further request, it produced an annotated version of the original text via Steps (9-7-8), marking sentence-by-sentence correspondences between model elements and the source article. This enabled traceable, concept-level verification of how the model was derived from the original narrative.

Takeaway: Demonstrates rapid text-to-model generation and traceable narrative-to-model correspondence.

LLM

Gender pay gap won't close for another 30 years, warns trade unions group

The gender pay gap is prominent even in industries where women make up the majority of employees, such as health and social care

Moves to close the gender pay gap will not be successful until 2056 if progress remains at its current rate, according to the Trades Union Congress (TUC).

The disparity between average wages for men and women is 12.8%, or £2,548 per year, according to TUC analysis of official pay data. It is widest in the finance and insurance industry, at 27.2%. In the leisure service sector it was only 1.5%.

The gender pay gap measures the difference in salaries paid to men and women in the same industries. Employers with more than 250 UK staff must report pay data.

The government said it was "tackling the root causes of the gender pay gap and backing women to succeed at work".

Even in industries where woman make up a majority of the workforce, such as education and health and social care, the gender pay gaps are 17% and 12.8%, respectively.

Pay disparity between genders meant that the average woman employee "effectively works for 47 days of the year for free," the TUC said.

"Women have effectively been working for free for the first month and a half of the year compared to men," TUC general secretary Paul Nowak said.

"With the cost of living still biting hard, women simply can't afford to keep losing out. They deserve their fair share".

Nowak pointed to the recent Employment Rights Act as a "an important step forward for pay parity" but said the government needed to increase access to paid parental leave so "mums and dads can better share care".

The gender pay gap is widest for workers aged 50-59, which the group of unions argues is partly due to the long-term effects of women pausing or reducing their career ambitions and pay to prioritise caring responsibilities.

The TUC, the national federation of trade unions, wants improved access to flexible working and better access to childcare in order to help close the gap.

Throughout the debate around increasing worker rights under the Employment Rights Act business groups argued that increased benefits and leave provisions would increase their costs and reduce their willingness to hire staff.

"The cost of doing business is already leading to firms cutting jobs. With major changes to employment laws coming down the line, the government must be extra careful not to add to those pressures," said Matthew Percival, the Future of Work and Skills Director at business group the CBI.

Employers will soon have to publish plans for how they plan to reduce the gender pay gap.

A government spokesperson said: "Combined with changes to flexible working, stronger protections for expectant and new mothers, and wider action to review parental leave and to expand childcare entitlements, we are tackling the root causes of the gender pay gap and backing women to succeed at work."

Please strictly follow the requirements specified in the XSD below to produce an XMILE model that reflects the content of the text above, especially the interrelationships among the factors described therein.

```
<?xml version="1.0" encoding="utf-8"?>
```

```
...
```

[Generated XMILE.]

Validating ...

Validated!

Advanced validation ...

*** Do not use if() or IF(), use IF THEN ELSE() instead ***

Not used: Cost_of_Living_Pressure

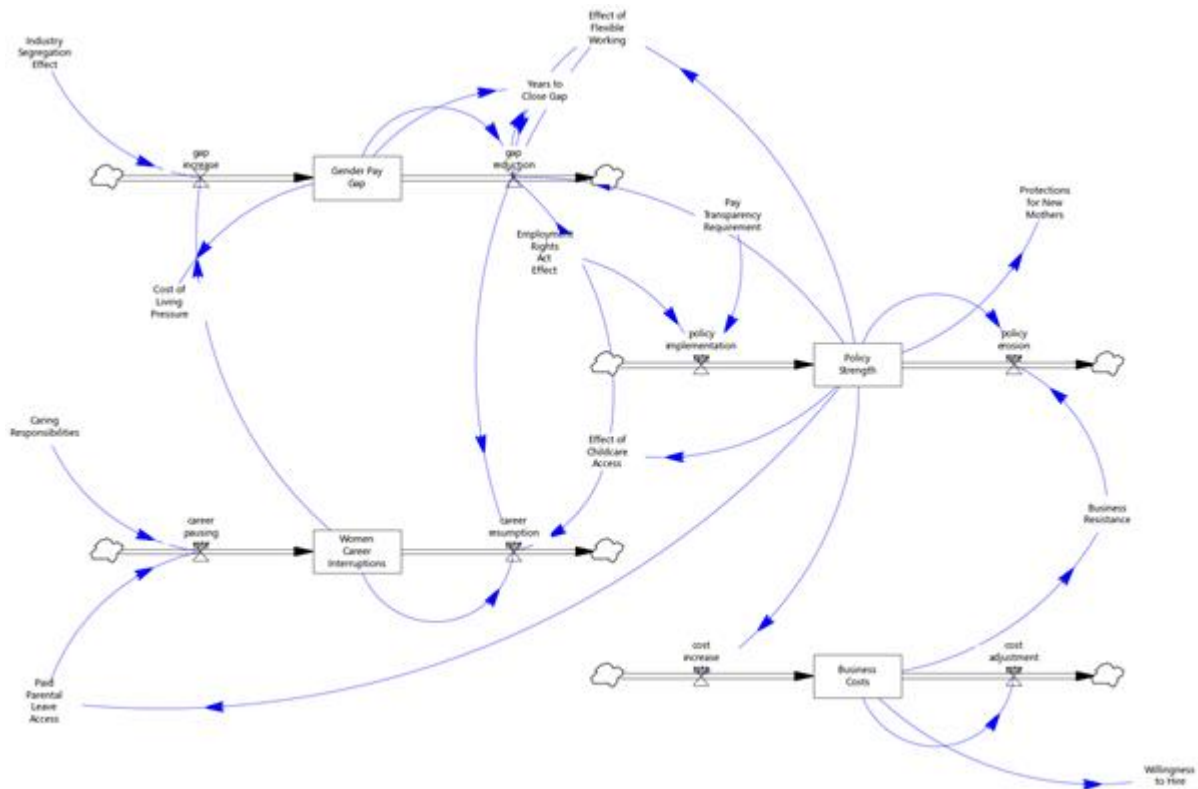
Transformation succeed

Units check ...

MISMATCH aux Career_Interruption_Effect declared=Dmnl inferred=percent

raw_declared=Dimensionless

[Generated XMILE.]



Would you please provide a detailed description of your model?

[Generated HTML.]

Would you please make a commented version of the original text marking all textual bases and showing for which stock, parameter, feedback loop and Policy levers they are used?

[Generated Gender Pay Gap 2_17_2026 1_33_15 AM.html]