A Tool for Enhanced Causal Discovery and Understanding: System Dynamics Modeling – Causal Miner (SDM-CMiner)

Ignacio J. Martinez-Moyano, Ph.D.
Argonne National Laboratory
imartinez@anl.gov

Nicholson Collier, Ph.D.
Argonne National Laboratory
ncollier@anl.gov

The University of Chicago
martinez-moyano@uchicago.edu

SDM-CMiner

The System Dynamics Modeling – Causal Miner (SDM-CMiner) is a prototype software tool designed for **exploring causal relationships** in system dynamics models **using data mining** processes and techniques.

SDM-CMiner

Using a causal model or map, a set of discovery rules, and a corpus of documents, the **SDM-CMiner tool attempts to uncover evidence** (or the lack thereof) **for the causal relationships** present in the model or map that can be found in the corpus of documents provided.

Process

1. **Extracts meta-data**
   Title, author, journal title, and date from the PDF.

2. **Sentences are identified and annotated** with a sentence annotation.

3. Each **sentence is tokenized** into its constituent words.

4. Each **word is tagged** with a part-of-speech annotation.

5. Each **word that conveys a positive or negative sentiment is annotated** as such.

Process

6. A **regular expression matcher is run** on the document text to identify the different variables in the system dynamics causal-loop diagram, annotating the matched text with the variable to which the text corresponds and the sentence in which it occurs.

7. The document's meta-data and the matched variable annotations are **written to a database**.

Regular Expression

- **A regular expression is a pattern** that describes the text to be matched.

For example, the simple regular expression "**colou?r**" will match "**color**" or "**colour**"

The "**?**" identifies the preceding "**u**" as an optional part of the match.

For more on regular expressions, see http://www.regular-expressions.info/.

Process--Continuation

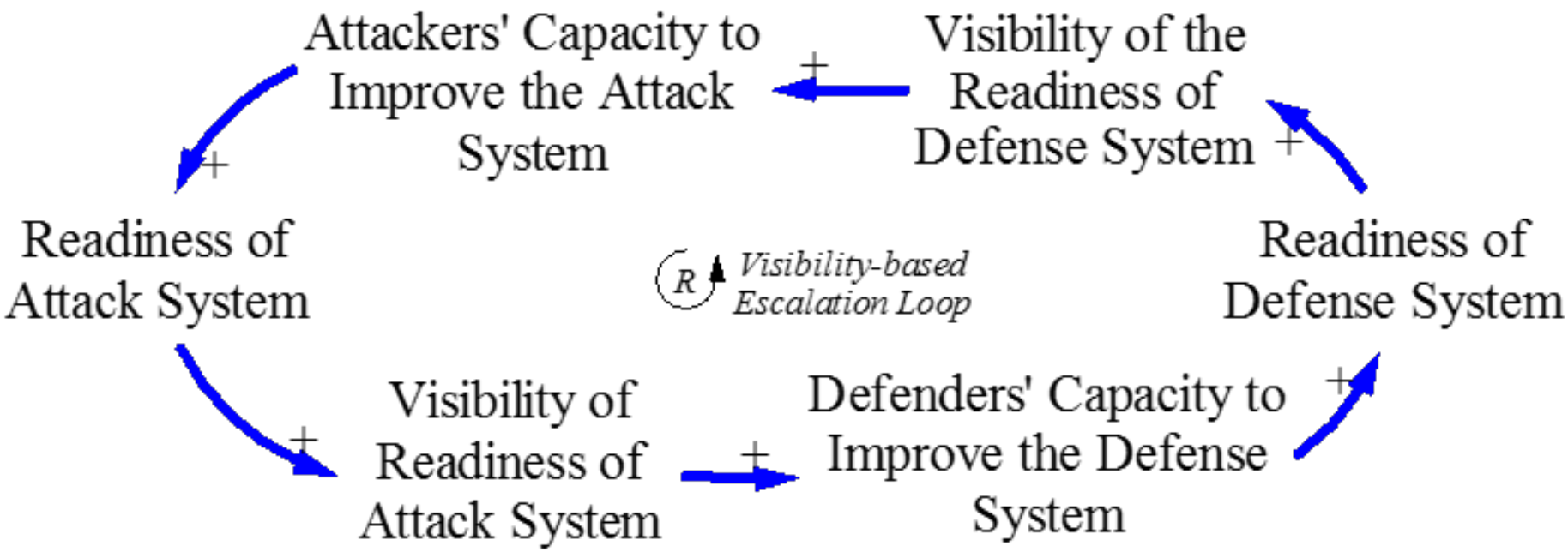
- The heart of this process is the **regular expression matching** that attempts to **identify document text that represents a corresponding variable** in the causal-loop diagram.
- The modeler provides regular expressions for each variable of interest, and the text-mining engine searches for these variables using regular expression matching and records their locations.
- *The text-mining engine does tag parts of speech and performs sentiment analysis*, although t**hese are not used** to identify variables in this version.

Process--Continuation

In this prototype, we are interested in the **sentence location of each variable with respect to all of the other variables**, such that sentence proximity is an indicator of a possible causal relationship.

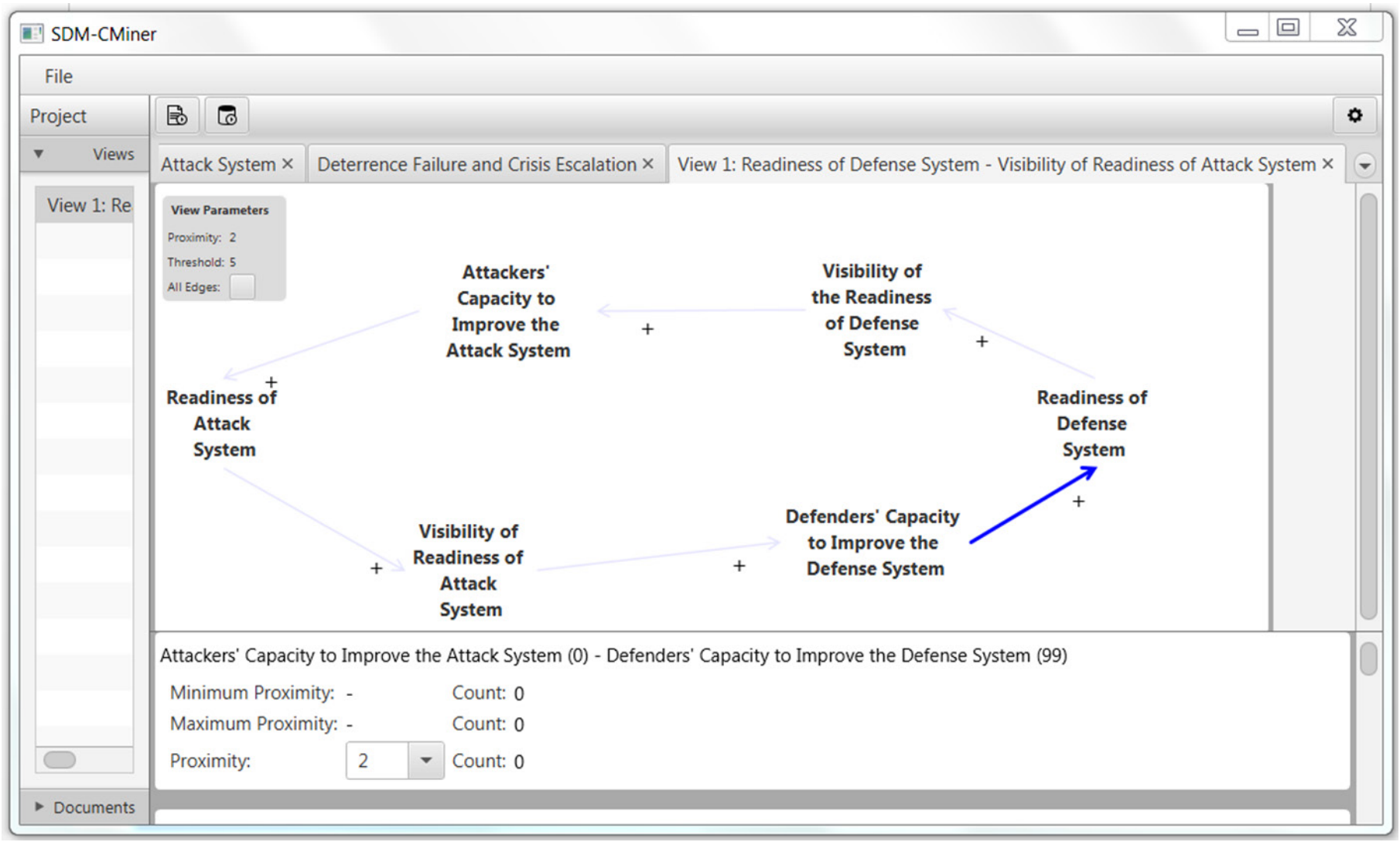
Proximity, however, **does not indicate either the direction of causality or a positive or negative polarity** in contrast to the actual causal-loop diagram, which typically indicates both.

Escalation in Adversarial Processes



Huth, P., & Russett, B. (1988). Deterrence failure and crisis escalation. International Studies Quarterly, 32(March), 29-45.

SDM-CMiner


SDM-CMiner
File
Project
Views
View 1: Re
Documents
Attack System
Deterrence Failure and Crisis Escalation
View 1: Readiness of Defense System - Visibility of Readiness of Attack System
View Parameters
Proximity: 2
Threshold: 5
All Edges:
Attackers' Capacity to Improve the Attack System
Visibility of the Readiness of Defense System
Readiness of Attack System
Readiness of Defense System
Visibility of Readiness of Attack System
Defenders' Capacity to Improve the Defense System
+
Attackers' Capacity to Improve the Attack System (0) - Defenders' Capacity to Improve the Defense System (99)
Minimum Proximity: - Count: 0
Maximum Proximity: - Count: 0
Proximity: 2 Count: 0

Regular Expression

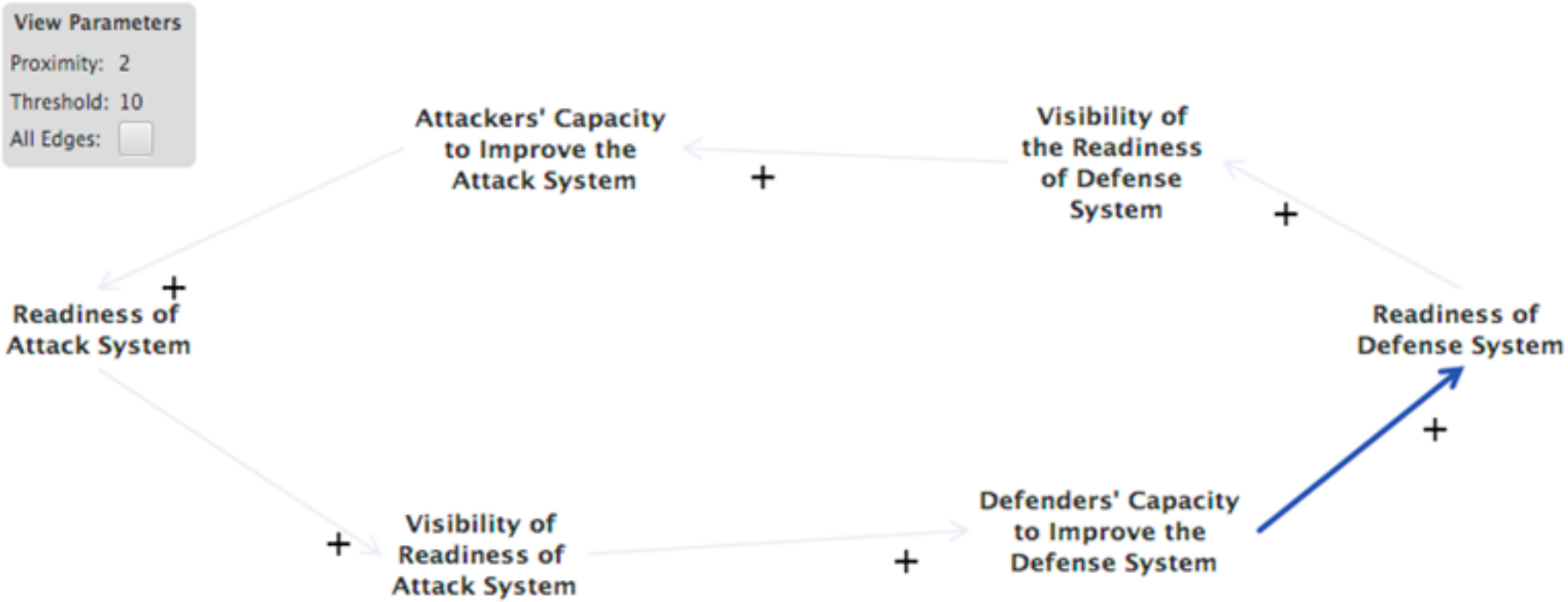
<Defenders' Capacity to Improve the Defense System>
deterrence

<Readiness of Defense System>
\bbalance\W+(?:\w+\W+){0,6}?defender\b
\bdefender\W+(?:\w+\W+){0,6}?balance\b

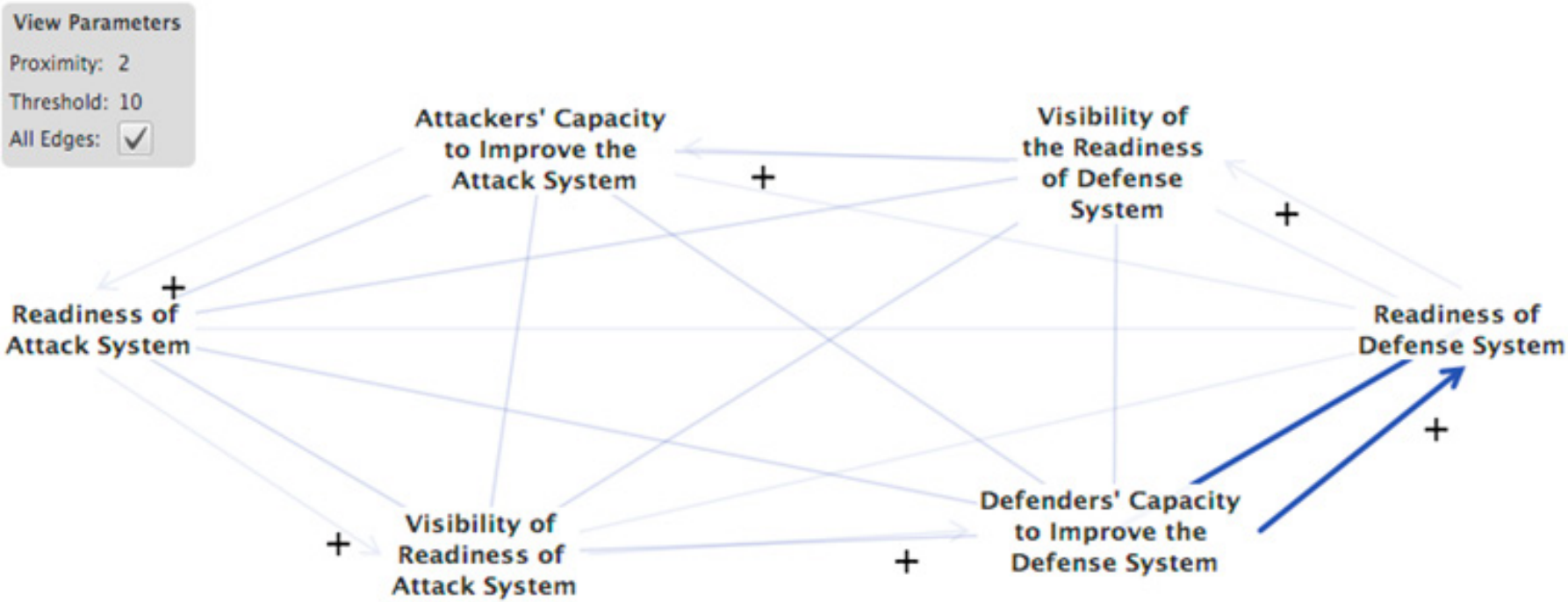
<Visibility of the Readiness of Defense System>
defender's reputation

"balance of power with respect to the defender" as indicative of the defender's readiness

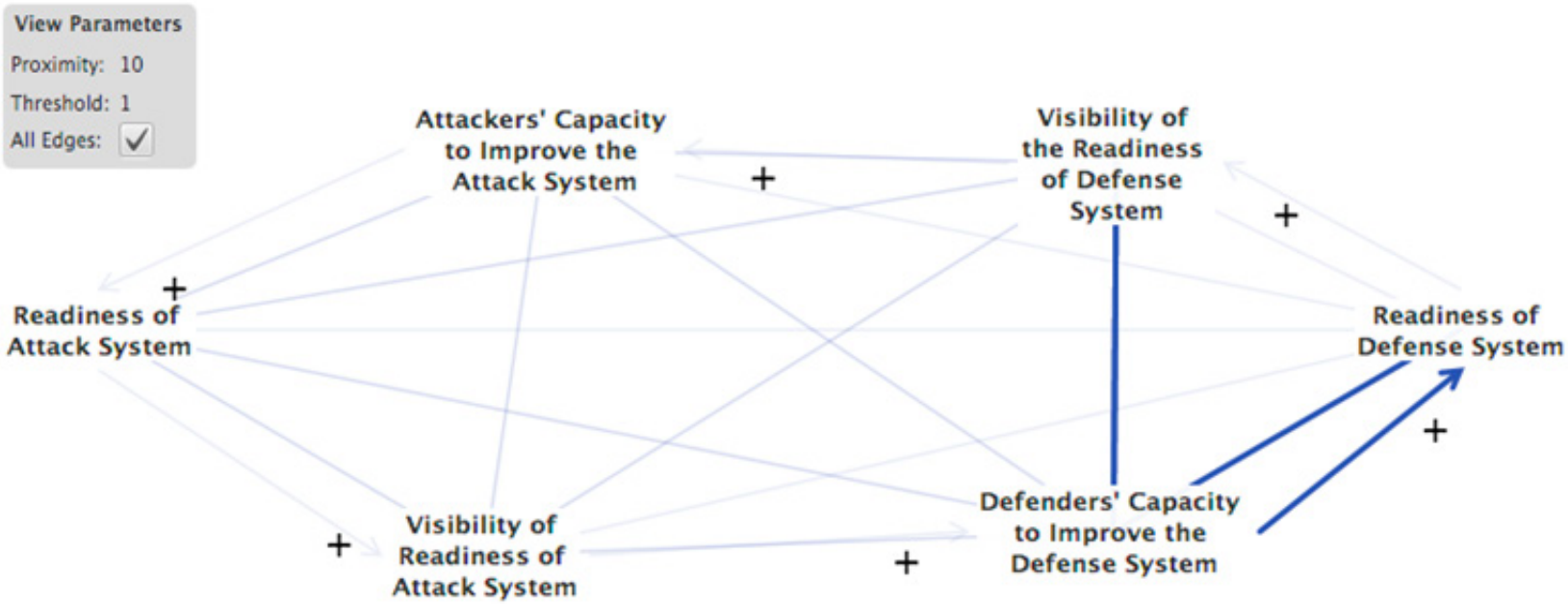
Causal Map Results



“All Edges” Results



“All Edges” Results



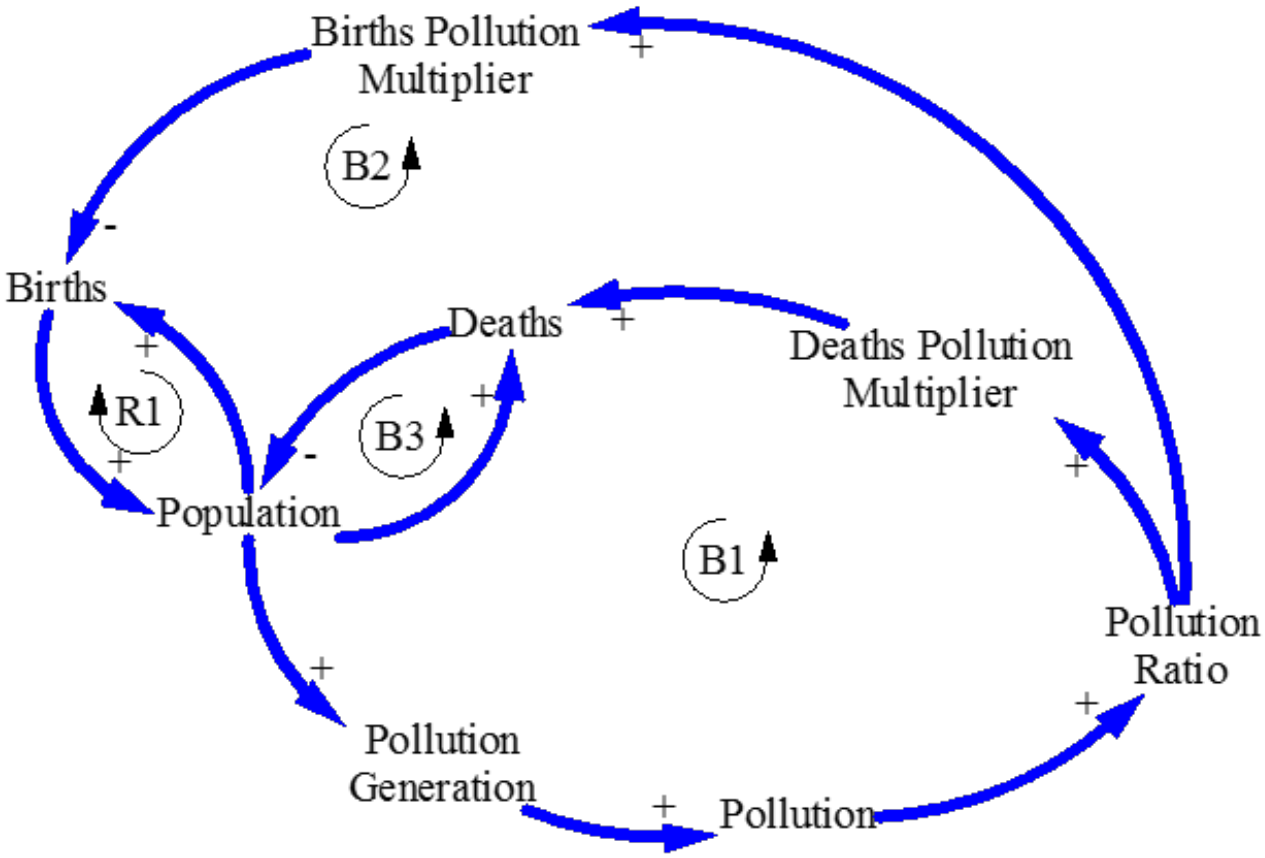
Proximity and Representative Sample of Text

Readiness of Defense System (12) – Visibility of the Readiness of Defense System (1)

Minimum Proximity:	42	Count: 1
Maximum Proximity:	167	Count: 1
Proximity:	80	Count: 4

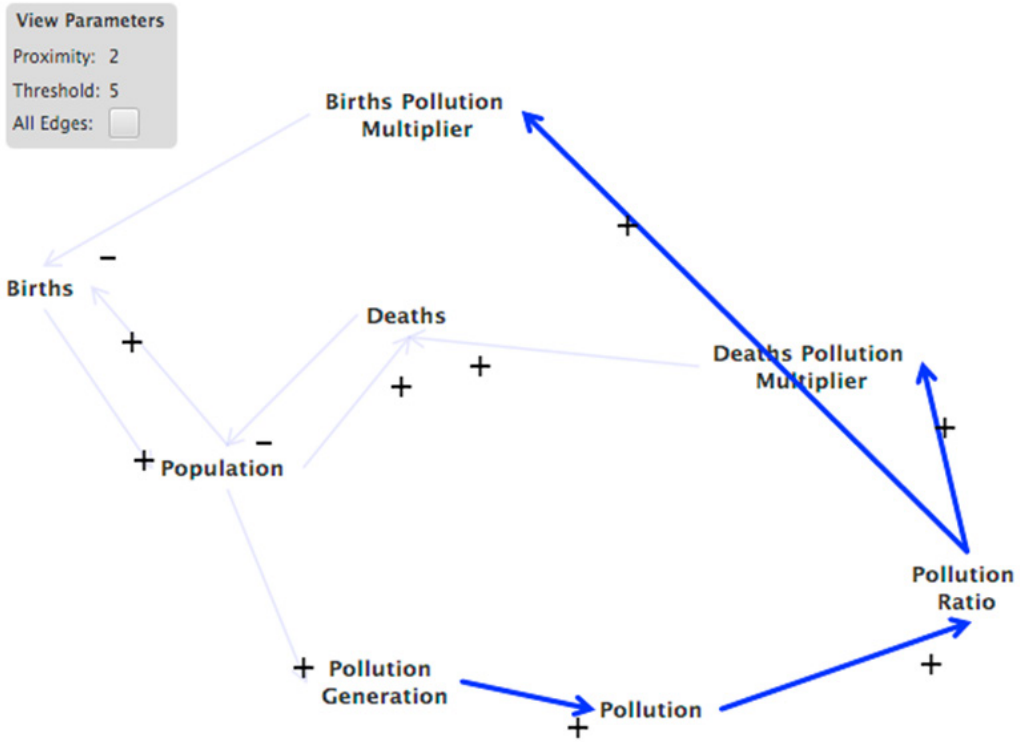
normative inhibitions associated with this disproportion made it absurd to consider nuclear use a real possibility in, for example, the Indonesian confrontation with Britain in 1964, Britain's conflicts with Guatemala over Belize in the 1970s, or the French-Libyan confrontation of 1983. Furthermore, in a majority of cases the nuclear power also possessed sufficiently strong conventional forces, including naval, to impose high costs on the potential attacker in the event of an armed conflict; where the intermediate balance did not favor the defender (e.g., the British-Indonesian confrontation), deterrence failed. Two cautions about this interpretation are necessary. First, the Quemoy-Matsu crisis may be an exception (see Huth, 1988: chapter 5). The immediate balance of forces probably favored the United States, though marginally; the United States not only possessed nuclear weapons but made overt nuclear threats, and deterrence worked. We do not know enough about Chinese decision making to know precisely why. Second, we are generalizing from a fairly small number of cases involving 12 Janice Gross Stein (1987) reports that in her case set, unlike in ours, deterrence often failed even when the military balance favored the defender. Most of her cases, however, involve deterrence of acts so much less provocative than those

Section of World Dynamics Model *(see Forrester, 1973, p 20-21)*



Clancy, Goodman, Sinclair, and Dockery (2002) and Nieuwenhuijsen et al. (2014).

Section of World Dynamics Model *(see Forrester, 1973, p 20-21)*



Unadjusted fertility rates declinedwith increasing air pollution levels (Fig. 1). After adjustment for covariates, we found reduced fertility rates with increases of all pollutants, but only for the PMcoarse fraction was this statistically significant (Table 3). Two pollutant models including the PMcoarse fraction and any of the other pollutants did not change risk estimate materially for the PMcoarse fraction (Table A2). Sensitivity analyses using the ratio of the number of children in age group of 0–4 and fertile women showed significantly reduced estimates found a statistically significant reduction of fertility rates with an increase in traffic related air pollution levels, particularly for the PMcoarse fraction. These results may therefore hint that air pollution could be associated with lower fertility rates in humans.

Thanks!

Ignacio J. Martinez-Moyano, Ph.D.

imartinez@anl.gov
martinez-moyano@uchicago.edu