Journal of Education and Practice
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol.16, No.12, 2025

Comparing measure of systems thinking in a post-secondary classroom

Rebecca Jordan (Corresponding author), Steven Gray, and Amanda Sorensen
Department of Community Sustainability Michigan State University
480 Wilson Road, Room 151, East Lansing, MI, 48823 USA
E-mail: jordanre@msu.edu

Cindy Frantz, Paul Brehm, Rumi Shammin, John Petersen, Alita Boyse-Peacor and Johanna Jauernig
Oberlin College
Oberlin, OH, 44074

The research is financed by the US National Science Foundation, IUSE

Abstract

This study investigates how different methods of assessing systems thinking—surveys, case studies, and student-generated models—relate to one another in a lower-division biology course for non-majors. We found that while some measures align, others capture distinct cognitive skills. Our findings suggest that model creation and model interpretation may require different instructional supports and that deeper scaffolding is needed to help students engage with complex socio-ecological systems.

Keywords: Systems thinking, perspective taking, assessment, measurement

DOI: 10.7176/JEP/16-12-12

Publication date: November 30th 2025

1. Introduction:

The need for students to understand complex, socioecological systems has been highlighted across several disciplines including the life sciences (e.g., Momsen et al. 2023). Complex socioecological systems are those in which humans, society, and the environment interact (Palmer 2022). Such systems are characterized by dynamic entities with often unseen elements and interactions that are hard to predict. Pulling from prior literature (i.e., Cabera et al 2023; Arnold et al. 2015; Richmond and Peterson 2001), systems thinking can be broadly described as the ability to reason about systems in ways that lead to problem solving. Therefore, it is essential to teach systems thinking in the classroom if we hope for students to be able engage productively in finding both solutions and resolutions to pervasive, planetary-wide environmental problems and issues.

As Monroe et al. (2015) argue, simply increasing understanding about systems, often during short term classroom lectures, is not likely to promote skill in systems thinking. Monroe et al, therefore, build on previous literature and integrate models and modeling including simulation and associated explanation (e.g., Perkins and Grotzer 2005; Liu and Hmelo-Silver 2009; Reiss and Mischo 2010, etc.) into the undergraduate classroom to promote comprehensive skill in systems thinking. Over the past ten years, members of the author team have been designing and working with an online modeling tool to help students not only learn about socioecological systems but also to think about these systems (Gray et al. 2019; Sorensen et al. 2024).

This manuscript focuses on several ways in which systems thinking was measured during an undergraduate non-majors biology course, with the intent to guide future instruction and assessment. To contextualize the conceptualization and implementation of systems thinking in this manuscript, we use the following dimensions that we argue (i.e., Gray et al. 2019) are essential to develop: (1) system structure, (2) system function (and subsequently, behavior), (3) identification of leverage points for change (and subsequently, feedback and emergence), and (4) trade-off analysis. Despite instructional efforts, numerous studies have shown that students often struggle with these dimensions (Sweeney & Sterman, 2000; Rehmat et al., 2020; Jordan et al., 2023a).

Given these learner struggles, we focus on the use of a modeling tool, MentalModeler (explained below) with targeted instruction focused on socioecological issues. Our goal is to use multiple measures of systems thinking to determine how these assessment tools compare to each other and what tools are likely appropriate assessments for learning goals.

2. Methods:

Data were collected over a single semester (16 weeks) of a lower division biology and society type course offered to undergraduate non-science majors to meet their science requirement at a large midwestern public university. These data represent a subset of a larger study in progress, which focuses not only on systems thinking, but also on developing tools that allow for rapid model assessment in large classes. Of 112 students who provided consent, 63 identified as male, 49 as female, and 0 did not report gender identity. Below we provide the assignment structure, the modeling task and the survey instrument.

Assignment Structure

Students completed three modeling assignments:

1. **Biology-focused** (e.g., climate change and reproduction)
2. **Social science-focused** (e.g., Fetal Alcohol Syndrome)
3. **Student-selected issue** ("Choose Your Own Adventure")

Only the third assignment and a systems thinking survey were analyzed in this study.

Modeling Task

Students read two contrasting narratives and created two MentalModeler maps—one for each perspective. Models included causal arrows with direction and weight. Concepts were coded as drivers, receivers, or ordinary variables. For the modeling piece, students were asked to make a "mental model" for each perspective using MentalModeler (i.e., an online decision-support conceptual mapping tool) (Gray et al. 2013) which is Fuzzy Cognitive Mapping (FCM) based modeling language. These models contain causal arrows that indicate both a direction, which is positive or negative, and degree of influence indicated by weight represented by line thickness (see mentalmodeler.org). All students completed these assignments in the same order and roughly the same time across classes. (See www.perspective-taking.org for the case studies which can also be made available on request and are under instructional materials copyright).

Survey Instrument

Students completed a systems thinking survey with three sections:

- **A. Overall Systems Thinking**: STS (2018) and New Paradigm measures (Randall & Stroink, 2018)
- **B. Specific Skills**: Feedback loops, ripple effects, stock and flow, and connectedness (all designed by authors)
- **C. Case-Based Tasks**: Covid-19 (Kirk et al. 2024), landuse (designed by authors), and Abeesee (Kirk et al. 2018) scenarios

Model Sophistication Rubric

Model sophistication was scored using a rubric (Table 1) based on:

- Shape and complexity
- Logical coherence of connections

Inter-rater reliability was calculated (Cohen's $\kappa = 0.82$), indicating strong agreement.

Data analysis:

Simple correlation (Pearson's r) analysis, linear regression, and visual inspection of the data were used to compare system measures.

3. Results:

We received 112 complete answers to systems questions. Figure 1 shows the distributions of the systems questions and Figure 2 shows distributions from the Covid-19 Case, land use case, and the Abeesee Case. These cases were separated to allow us to discuss those separately. Correlations are reported in Table 3

Survey and Model Correlations

We analyzed correlations between survey scores and model sophistication. The appendix includes Pearson's r, p-values, and confidence intervals for all comparisons. Table 2 includes only the significant comparisons.

Significant correlations were found most between the model scoring (again see Table 1) and some of the other system measures. Not surprisingly, those measures were feedback and connectedness as well as the broad systems thinking scales. One broad systems measure (i.e., STS) was correlated with ripple effects (which characterize the notion of tweaking one system component and having those affects move downstream to affect other components in not necessarily linear ways.

Alternatively, a multiple regression analysis explored predictors of model sophistication. None of the dependent variables (i.e., other system measures) show statistically significant relationships with drawn model sophistication at the conventional $p < 0.05$ level. Most variables have very low R^2 values, indicating that student created models explains very little of the variance seen with the different system measures. Again, we focused on drawn models given the highest level of significant correlation with the other system measures.

4. Discussion:

In summary, we found different systems thinking assessments to be related to each other and to certain aspects of model-based learning. While the overall systems thinking instruments shared similar findings, one tool, STS, likely are related to ripple (or downstream type) effects. When thinking about creating models, students are more likely to think about feedback and connections. Both of which did not scale to model reading (to which we suggest might be a different skill than model building) and finally, the models/modeling tasks did not scale with the case studies; suggesting that we are getting at different constructs. We surmise that the cases represent a level of complexity not viewed by students when they had the option of drawing opposing sides of an issue of choice. Also, deeper instruction/scaffolding is likely warranted for students to represent these complex ideas in their models (i.e., these terms were only introduced in one lecture).

Certainly, others have found similar outcomes when looking at student models. Dauer et al. (2013) also found that student models tended to be simplistic at the initial time point, become more complicated after revision at a subsequent time point, and then simplified with time and tended to use more simplistic language. In our study, students modeled new cases through the duration of the course. In addition, Jordan et al. (2014) found that students earlier tended to mirror instructor language with little understanding of the parts. The language subsequently tended to be more colloquial but also was emblematic of greater understanding. The latter matches that of the more sophisticated (albeit rarer) models generated by students in our study and could lead to deeper consideration when thinking about socioenvironmental narratives and trade-offs.

Others have shown that the more students know about a construct, the better they can evaluate or read models (e.g., Nielsen and Nielsen 2021) and unsurprisingly certain representations are likely to evoke deeper understanding than others (e.g., Dauer et al. 2024 report students tended to do better reading schematics than box and arrow diagrams for certain biological constructs). Further the latter authors discuss noticing as an important skill in taking in a complex system representation. This, however, requires some sort of cue for students to attend to. Jordan et al. (2014) found that students were able to transfer system level concepts from pre- to post- when given cues called conceptual representations. The latter is a thinking tool, like a pneumonic, that can help students rely on a heuristic to remember which questions to ask themselves. For example, a student may remember to think about a system outcome or phenonmena by remembering the "P" in the PMC conceptual representation chosen by the authors of the latter study. This means that by remembering to say PMC aloud, students are reminded to ask themselves what the phenomenon is being modeled. Other approaches that use conceptual representations have focused on the overall task (e.g., Approach to Modeling (AtM); Bennet et al. 2020) or specific elements (e.g., focus on causal elements aka, Causal Mechanistic Reasoning (CMR) in Franovic et al. 2023). Alternatively, Krell et al. (2023) discuss engaging students more in the why along with the how of modeling.

To help student modeling practice to deal with the complexity of systems, there likely needs to be support for the various thinking tasks involved in each step of the modeling process. With this, are there ways that these thinking heuristics can be advanced such that students can ask themselves what are critical leverage points and when those are toggled what might trade-offs be? The mentalmodeler software can complete this function but our data suggests students are unlikely to use these tools unless prompted, though retention with model building at the very least is likely to be longer than didactic instruction alone (e.g., Dauer and Long 2015).

We can use ours and other data to suggest future directions which might result in greater model sophistication and enable student discussion of leverage points and trade-offs. Next steps for our research will include providing prompts for students to engage in scenario building. Akin to the think-aloud strategy used with teachers by Karga and Ceyan (2024), educators can focus on the dynamic systems thinking skills associated with modeling. With this, students will have the opportunity to meaningfully build scenarios, which is a major desire of the students with whom we work (e.g., Jordan et al. 2023b). By meaningful, we argue, that students will use terms about which they are familiar as opposed to copying the words used by instructors or others without knowing their meaning. The latter results in students manipulating inaccurate conceptions with no anchor to accuracy. The former of which can also be as simple as having students practice first building scenarios that allow students to formulate ideas using curated information. This information can be used first to gain practice and then build the complexity that can help learners make decisions that do not have right or wrong answers. Doing so would mean providing a safe space for learners to evaluate which leverage points to tweak and which outcomes to maximize based on values, data, and in some cases large amounts of uncertainty. Selection of such outcomes, we argue, is at the heart of socioenvironmental systems thinking and is likely the type of skill needed for the next century of wicked problems.

References

Arnold, R. D., & Wade, J. P. (2015). A definition of systems thinking: A systems approach. Procedia Computer Science, 44, 669–678. https://doi.org/10.1016/j.procs.2015.03.050

Bennett, S., Gotwals, A. W., & Long, T. M. (2020). Assessing students' approaches to modelling in undergraduate biology. International Journal of Science Education, 42(10), 1697–1714. https://doi.org/10.1080/09500693.2020.1765046

Cabrera, D., & Cabrera, L. (2023). What is systems thinking? In M. J. Bishop, E. Boling, J. Elen, & V. Svihla (Eds.), Learning, design, and technology: An international compendium of theory, research, practice, and policy (pp. 1495–1522). Springer International Publishing. https://doi.org/10.1007/978-3-319-17727-4_145

Dauer, J., Behrendt, M. G., Elliott, M., Gettings, B., Long, T., & Clark, C. (2024). Individual variation in undergraduate student metacognitive monitoring and error detection during biology model evaluation. Frontiers in Education, 9, 1–12. https://doi.org/10.3389/feduc.2024.00001

Dauer, J., & Long, T. M. (2015). Long-term conceptual retrieval by college biology majors following model-based instruction. Journal of Research in Science Teaching, 52(8), 1188–1206. https://doi.org/10.1002/tea.21237

Dauer, J. T., Momsen, J. L., Speth, E. B., Makohon-Moore, S. C., & Long, T. M. (2013). Analyzing change in students' gene-to-evolution models in college-level introductory biology. Journal of Research in Science Teaching, 50(6), 639–659. https://doi.org/10.1002/tea.21094

Franovic, C. G. C., Noyes, K., Stoltzfus, J. R., Schwarz, C. V., Long, T. M., & Cooper, M. M. (2023). Undergraduate chemistry and biology students' use of causal mechanistic reasoning to explain and predict preferential protein–ligand binding activity. Journal of Chemical Education, 100(5), 1716–1727. https://doi.org/10.1021/acs.jchemed.2c01010

Gray, S., Gray, L., Cox, S., & Henly-Shepard, S. (2013). Mental modeler: A fuzzy-logic cognitive mapping modeling tool for adaptive environmental management. In Proceedings of the 46th International Conference on Complex Systems (pp. 963–973).

Gray, S., Sterling, E. J., Aminpour, P., Goralnik, L., Singer, A., Wei, C., Akabas, S., Jordan, R. C., Giabbanelli, P. J., Hodbod, J., … (2019). Assessing (social-ecological) systems thinking by evaluating cognitive maps. Sustainability, 11(20), 5753. https://doi.org/10.3390/su11205753

Grohs, J. R., Kirk, G. R., Soledad, M. M., & Knight, D. B. (2018). Assessing systems thinking: A tool to measure complex reasoning through ill-structured problems. Thinking Skills and Creativity, 28, 110–130. https://doi.org/10.1016/j.tsc.2018.03.003

Jacobson, M. J., & Wilensky, U. (2006). Complex systems in education: Scientific and educational importance and implications for the learning sciences. Journal of the Learning Sciences, 15(1), 11–34. https://doi.org/10.1207/s15327809jls1501_4

Jordan, R. C., Gray, S., Boyse-Peacor, A., Sorensen, A. E., Frantz, C. M., Jauernig, J., Brehm, P., Shammin, M. R., & Petersen, J. (2023a). Promoting systems thinking through perspective taking when using an online modeling tool. Frontiers in Education, 8, 1–15. https://doi.org/10.3389/feduc.2023.1134567

Jordan, R. C., Sorensen, A. E., & Gray, S. A. (2023b). What undergraduate students know and what they want to learn about in climate change education. PLOS Sustainability and Transformation, 2(4), e0000055. https://doi.org/10.1371/journal.pstr.0000055

Jordan, R. C., Sorensen, A. E., & Hmelo-Silver, C. (2014). A conceptual representation to support ecological systems learning. Natural Sciences Education, 43(1), 141–146. https://doi.org/10.4195/nse2014.43.1.141

Karga, B., & Ceyhan, G. D. (2024). Investigating middle school science teachers' stock-flow, causal-loop, and dynamic thinking skills with scenario-based questions. International Journal of Science Education, 1–20. https://doi.org/10.1080/09500693.2024.000000

Kirk, E. T., Sadler, T. D., Ke, L., & Zangori, L. (2024). Design considerations for a multiple-choice assessment of socio-scientific systems thinking. International Journal of Designs for Learning, 15(1), 27–37. https://doi.org/10.14434/ijdl.v15i1.36492

Krell, M., Upmeier zu Belzen, A., & Krüger, D. (2012). Students' understanding of the purpose of models in different biological contexts. International Journal of Biology Education, 2(1), 1–34.

Momsen, J., Speth, E. B., Wyse, S., & Long, T. (2023). Using systems and systems thinking to unify biology education. CBE—Life Sciences Education, 21(2), 1–15. https://doi.org/10.1187/cbe.22-01-0015

Monroe, M., Plate, R. R., & Colley, L. (2015). Assessing an introduction to systems thinking. Journal of Natural Resources and Life Sciences Education, 44(1), 11–19. https://doi.org/10.4195/nse2015.44.1.11

Nielsen, S., & Nielsen, J. (2021). A competence-oriented approach to models and modelling in lower secondary science education: Practices and rationales among Danish teachers. Research in Science Education, 51, 685–709. https://doi.org/10.1007/s11165-019-09883-3

Palmer, M. (2022). What is a socio-environmental system? National Socio-Environmental Synthesis Center Resource: Explainer. https://www.sesync.org/resources/what-socio-environmental-system

Perkins, D. N., & Grotzer, T. A. (2005). Dimensions of causal understanding: The role of complex causal models in students' understanding of science. Cognition and Instruction, 23(3), 312–351. https://doi.org/10.1207/s1532690xci2303_1

Randle, J., & Stroink, M. (2018). The development and initial validation of the paradigm of systems thinking. Systems Research and Behavioral Science, 35(6), 645–657. https://doi.org/10.1002/sres.2510

Rehmat, A. P., Hmelo-Silver, C. E., Housh, K., Liu, L., & Cisterna, D. (2020). Assessing systems thinking across disciplines: A learning progression. In M. Gresalfi & I. S. Horn (Eds.), The interdisciplinarity of the learning sciences, 14th international conference of the learning sciences (ICLS) 2020, Volume 2 (pp. 899–900). International Society of the Learning Sciences.

Riess, W., & Mischo, C. (2009). Promoting systems thinking through biology lessons. International Journal of Science Education, 32(6), 705–725. https://doi.org/10.1080/09500690902769946

Richmond, B., & Peterson, S. (2001). An introduction to systems thinking. High Performance Systems, Incorporated.

Sorensen, A. E., Alred, A., Fontaine, J. J., & Dauer, J. M. (2024). A model-based instructional approach in a socio-ecological course-based undergraduate research experience (CURE). Ecosphere, 15(9), e70007. https://doi.org/10.1002/ecs2.70007

Society for Simulation in Healthcare. (2018). STS manual. https://case.edu/nursing/sites/default/files/2018-04/STS_Manual.pdf

Sweeney, L. B., & Sterman, J. D. (2000). Bathtub dynamics: Initial results of a systems thinking inventory. System Dynamics Review, 16(4), 249–286. https://doi.org/10.1002/sdr.198

Table 1: Model coding scheme (with associated figures below)

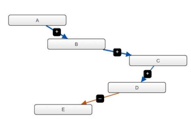
Model coding scheme:		
Part A: Model duplication:	Do the first and second models use the same components?	Yes or No
Part B: Model sophistication:**		
Model Shape	See figure below for linear, wagon wheel or wheels, star, or other.	Linear & Ball = 1 Star & Wagon Wheel(s) = 2 Other = 3
Model complexity	This is an assessment of how many lines cross in the model picture.	Simple (no crosses) = 1 Moderate (falls in btwn) =2 Highly (multiple crosses) =3
Model connections	This is an assessment of the extent to which the connections make sense.	Take 10 connections, do more than 5 make sense? (Yes or No)

** Model sophistication was measured as an average of the three variables detailed in Table 1. Briefly, the combination entailed the following code averages with 1 = simple complexity linear, No on sensical connections; 1.75 = simple complexity, star, No on sensical connections; 2 = moderate complexity, wheel(s), star, and No on sensical connections, 2.5 = highly complexity wheels, star, or other, and Yes on sensical connections, and 2.75 = high or moderate complexity, other, and Yes on sensical connections. (Note: all other combinations of the coding variable were not evident and why other values, e.g., 2.25, are not present).

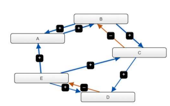
Final scores were reported under two variables (model duplication, and model sophistication with the latter being calculated in the correlation matrix).

Figure associated with Table 1:

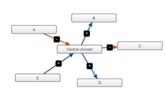
Linear



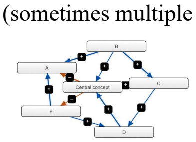
Ball



Star



Wagon wheel
(sometimes multiple are connected)



Other

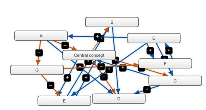
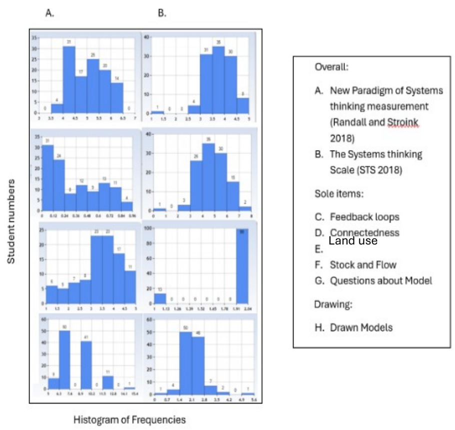


Table 2: Significant pairings for the correlation analysis.

Variable 1	Variable 2	Pearson's r	P-value	95% CI Lower	95%CI Upper
Model Score	STS	0.375	0.0032	0.134	0.574
Model Score	New Par	0.264	0.0416	0.011	0.485
Model Score	Feedbacks	0.296	0.0215	0.046	0.512
Model Score	Landuse	0.261	0.0437	0.008	0.483
STS	Ripple Effect	0.333	0.0094	0.086	0.541

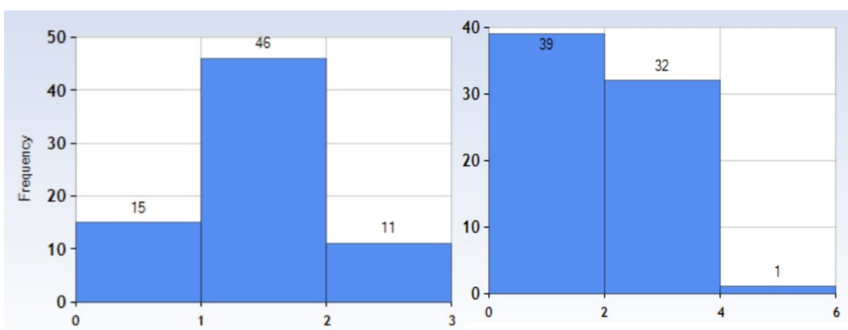
Note: full table can be found in appendix.

Figure 1: Histograms of student responses and subsequent codes for each of the systems thinking and modeling assessments.



n (Kirk et al. 2024) and Abeesee case (Grohs et al. 2018). The first was rated on a scale of 1-3 And the second, 1-4. In both cases, the higher number was the more Sophisticated response. Students struggled more with the Abeesee Case.

Histogram of Frequency (number of students).



Our findings suggest that different systems thinking assessments capture distinct skills. Paradigm-based measures aligned with feedback and stock-flow reasoning, while STS measures were more loosely connected to ripple effects.
Model creation and model reading appear to be separate skills. The landscape case (model reading) did not correlate with model creation, supporting this distinction. Similarly, the Abeesee case did not correlate with modeling tasks, possibly due to differences in complexity and framing.

Appendix Table:

Variable 1	Variable 2	Pearson's r	P-value	95% CI Lower	95% CI Upper
SES Covid	SES Abeesee	-0.044	0.7374	-0.298	0.215
SES Covid	Model Score	-0.122	0.3548	-0.367	0.14
SES Covid	STS	-0.094	0.4739	-0.343	0.167
SES Covid	New Par	-0.149	0.2552	-0.391	0.112
SES Covid	Feedbacks	-0.208	0.1113	-0.441	0.052
SES Covid	Ripple Effect	-0.106	0.4183	-0.354	0.155
SES Covid	Stock & flow	-0.031	0.8151	-0.286	0.228
SES Covid	Landuse	-0.062	0.6359	-0.314	0.198
SES Covid	Connectedness	0.068	0.6031	-0.192	0.32
SES Abeesee	Model Score	0.114	0.3838	-0.147	0.361
SES Abeesee	STS	-0.204	0.1179	-0.438	0.056
SES Abeesee	New Par	-0.09	0.4918	-0.34	0.17
SES Abeesee	Feedbacks	-0.051	0.699	-0.304	0.209
SES Abeesee	Ripple Effect	-0.085	0.5173	-0.335	0.176
SES Abeesee	Sock & flow	0.038	0.7759	-0.222	0.292
SES Abeesee	Landuse	-0.056	0.672	-0.308	0.204
SES Abeesee	Connectedness	0.027	0.8389	-0.232	0.282
Model Score	STS	-0.085	0.517	-0.335	0.176
Model Score	New Par	0.078	0.5527	-0.182	0.329
Model Score	Feedbacks	0.087	0.51	-0.174	0.336
Model Score	Ripple Effect	-0.166	0.2045	-0.406	0.095
Model Score	Stock & flow	-0.051	0.7015	-0.304	0.209
Model Score	Landuse	-0.219	0.0933	-0.45	0.041
Model Score	Connectedness	-0.225	0.0844	-0.455	0.034
STS	New Par	0.375	0.0032	0.13	0.576
STS	Feedbacks	0.264	0.0416	0.007	0.488
STS	Ripple Effect	0.296	0.0215	0.043	0.514
STS	Stock & flow	0.253	0.0512	-0.004	0.479
STS	Landuse	0.144	0.2732	-0.118	0.386
STS	Connectedness	0.261	0.0437	0.005	0.486
New Par	Feedbacks	-0.014	0.9132	-0.27	0.243
New Par	Ripple Effect	0.055	0.6737	-0.204	0.308
New Par	Stock & flow	0.333	0.0094	0.083	0.543
New Par	Landuse	0.153	0.2427	-0.108	0.395
New Par	Connectedness	0.179	0.1704	-0.081	0.417
Feedbacks	Ripple Effect	0.191	0.1429	-0.069	0.427
Feedbacks	Stock & flow	-0.025	0.849	-0.28	0.233
Feedbacks	Landuse	0.051	0.7009	-0.209	0.304
Feedbacks	Connectedness	-0.008	0.9487	-0.265	0.249
Ripple Effect	Stock & flow	-0.015	0.91	-0.271	0.243
Ripple Effect	Landuse	0.001	0.992	-0.256	0.258
Ripple Effect	Connectedness	0.156	0.2325	-0.105	0.397
Stock &flow	Landuse	-0.082	0.5358	-0.332	0.179
Stock & flow	Connectedness	0.169	0.1974	-0.092	0.408
Landuse	Connectedness	0.072	0.5845	-0.188	0.323