

# ANALYSIS OF THE BIOMODELS DATABASE

{ JAMES HOLLINS \* & COLIN GILLESPIE }

## BIOMODELS DATABASE

Biomodels Database is an online resource for storing and serving quantitative models of biomedical interest.



The database was created in 2005.

The focus of the project is the curated branch of the database. Following the release on 11<sup>th</sup> August, there are 424 models in this branch, which have all been described in peer reviewed scientific literature.

## R

R is a language and environment for statistical computing and graphics, providing a wide variety of statistical



techniques. R is extensible by installing packages. An example is the package `rsbml` used to parse and extract information from SBML files. However, only 356 of the 424 curated models could be parsed in R using this package.

## SBML

SBML is a modelling standard used for exchanging models between different software tools. An example of SBML code is shown below:



```
<listOfSpecies>
<species metaid="_230475" id="C" name="Cyclin" compartment="cell"
  initialConcentration="0.01" substanceUnits="substance" sboTerm="SBO:0000252"/>
<species metaid="_230495" id="M" name="CDC-2 Kinase" compartment="cell"
  initialConcentration="0.01" substanceUnits="substance" sboTerm="SBO:0000252"/>
<species metaid="_230515" id="X" name="Cyclin Protease" compartment="cell"
  initialConcentration="0.01" substanceUnits="substance" sboTerm="SBO:0000297"/>
</listOfSpecies>
```

SBML represents the models as a list of chemical transformations, since every biological process in a cell can be described as a series of reactions. SBML is easy for computers to generate and parse but difficult for humans to read and write. Hence, R was used since R is easier to work with than SBML.

## TERMS

Suppose we have a chemical species,  $X$ . We model the rate of change of  $X$  using the following ODE:

$$\frac{dX(t)}{dt} = -k_1 X(t) + k_2.$$

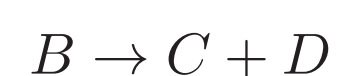
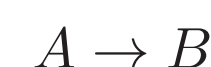
Where the amount of  $X$  is altered by the following processes:



- The entity  $X$  is a chemical species, for example an ion or a biological entity such as a protein binding site.
- The process altering the amount of  $X$  are reactions.
- $k_1$  and  $k_2$  are the reaction parameters. Parameters are the numbers used in the description of the rate laws of reactions.

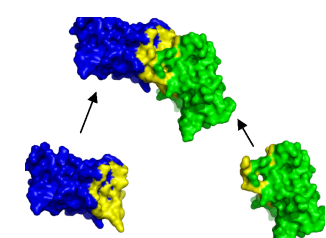
## CONNECTIONS

In the project, a species has a 'connection' for each reaction in which it is present. Consider the following reactions:



$A$ ,  $B$  and  $C$  each appear in two reactions and so have 2 connections.  $D$  and  $E$  each appear in just 1 reaction and so have just 1 connection.

The average number of connections is 5.72 (to 2 decimal places).



## SBO TERMS

Systems Biology Ontology (SBO) terms are used to provide additional information about model constituents.

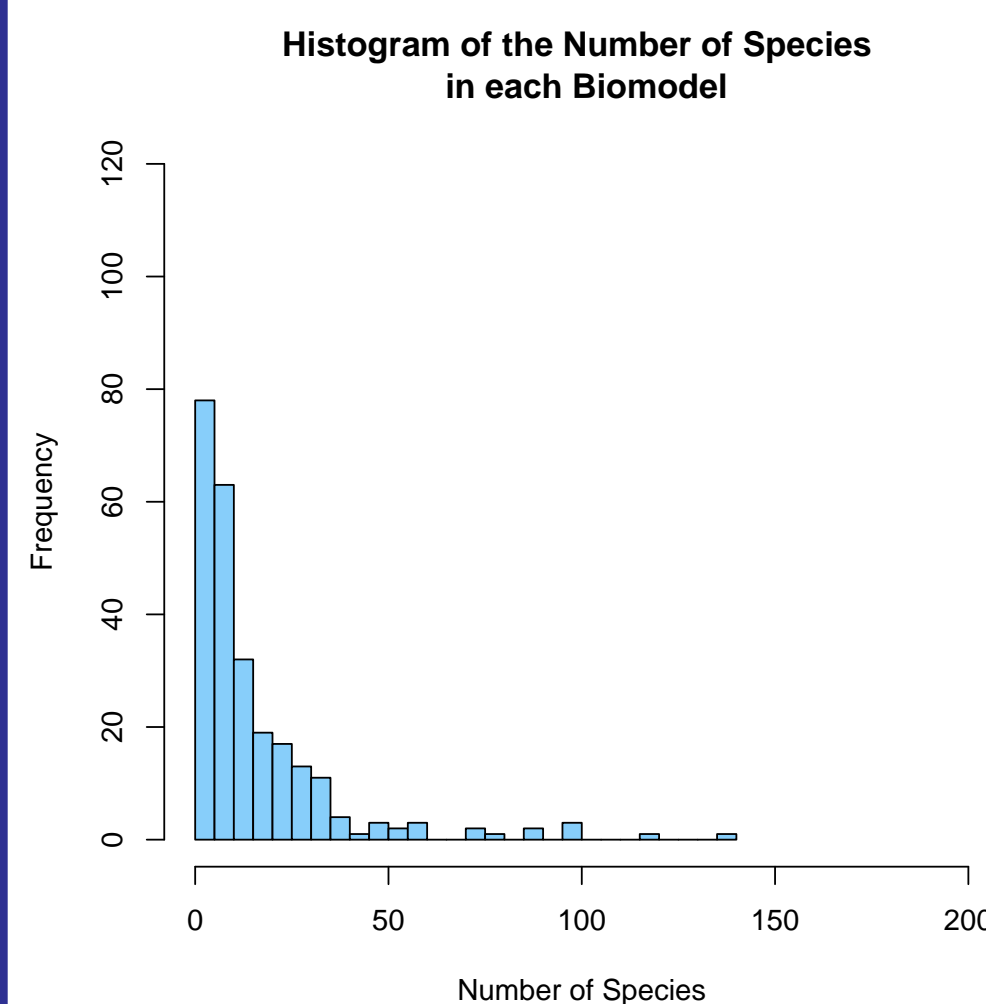
The possibility of using SBO terms to track which models certain species appeared in was explored. However, it was found that species do not necessarily have unique SBO terms

**Table 1:** Species SBO terms from one model

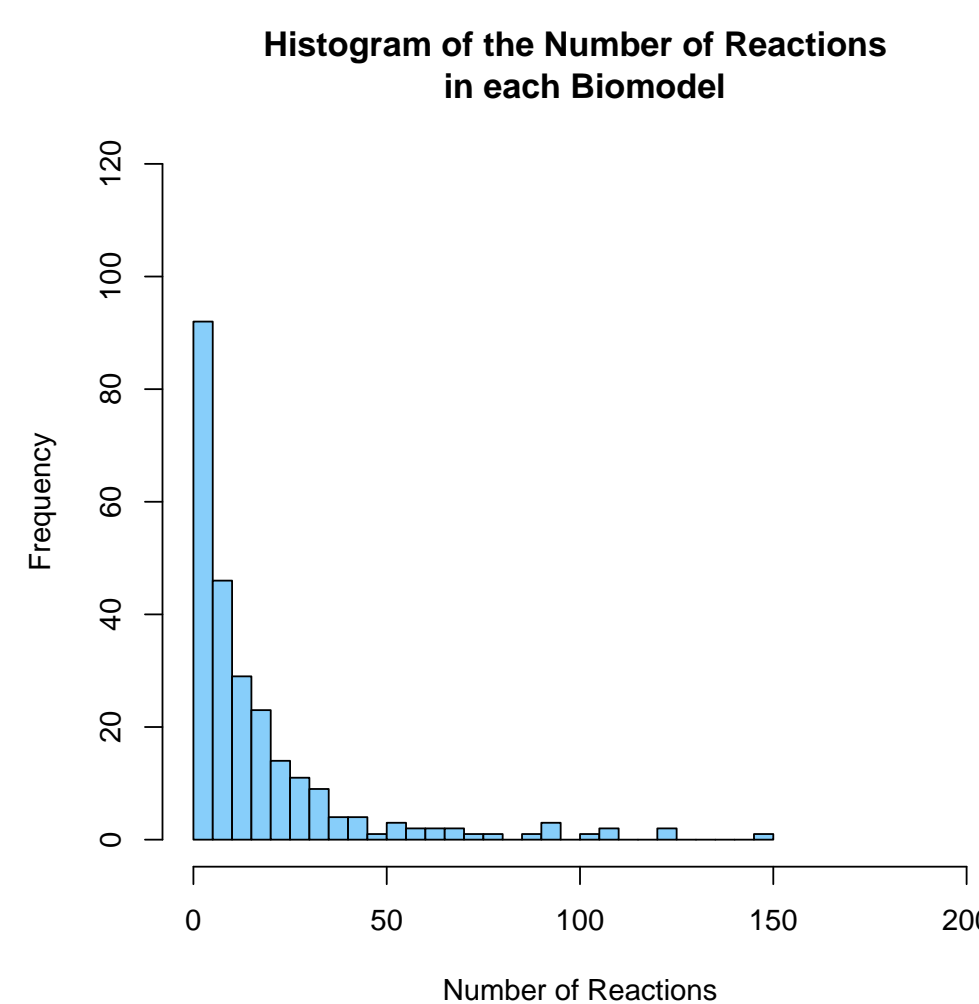
sboTerm	Model ID	Freq.
SBO:0000297	EPSP_Edelstein	8
SBO:0000420	EPSP_Edelstein	4

Similar problems were found in each model. Therefore, it is not possible to use SBO terms to track where species appear.

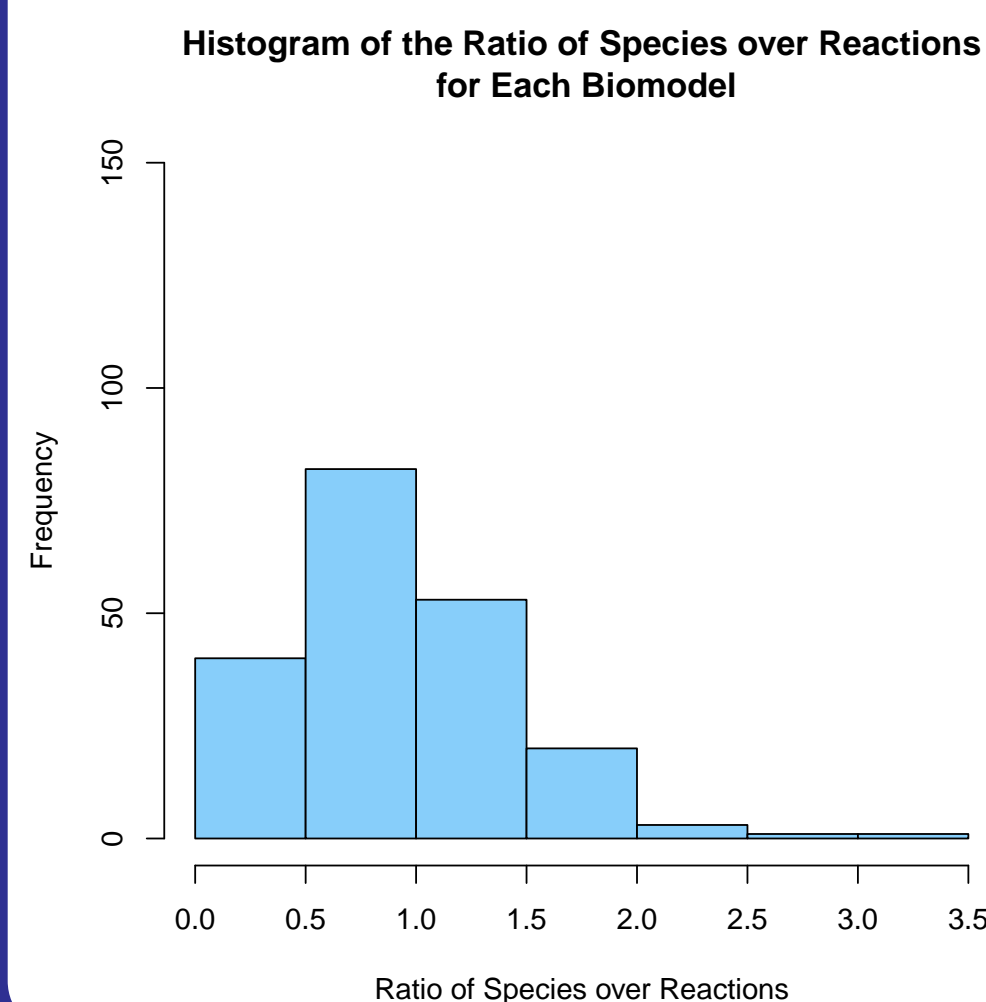
## SPECIES AND REACTIONS



As shown in the graph above, the majority of models have 10 or less species, suggesting that the models tend to have small numbers of species.



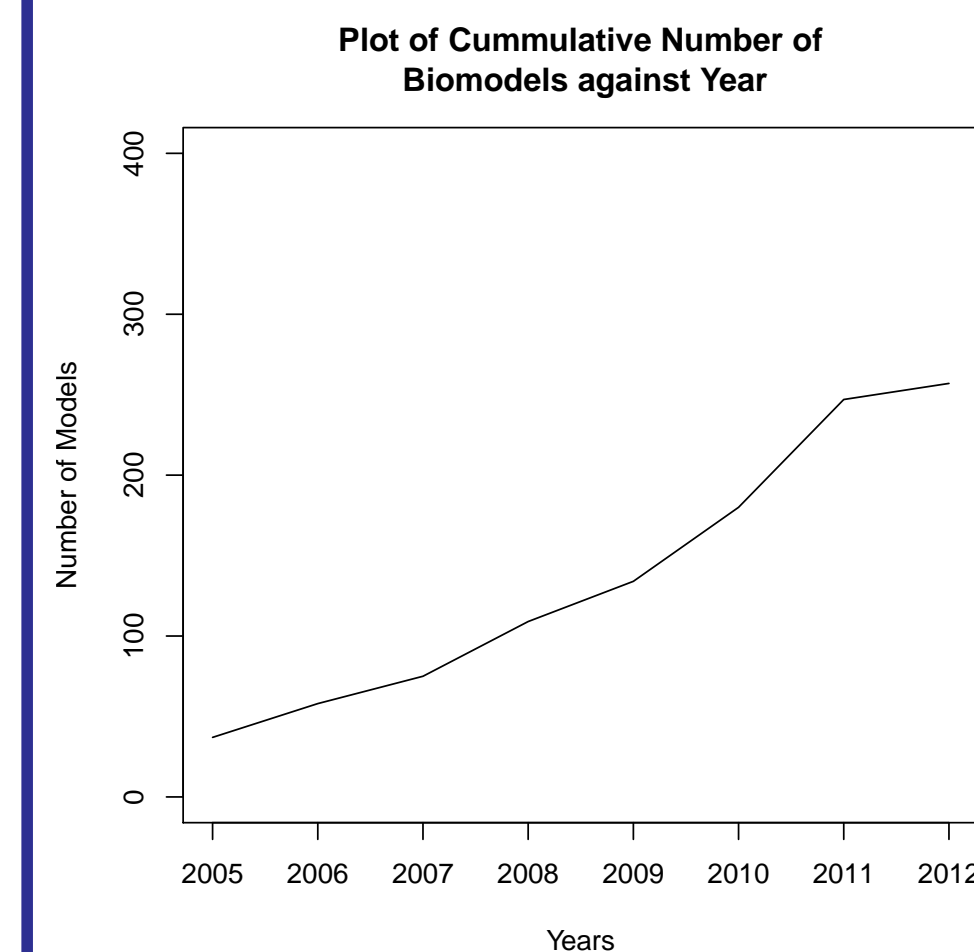
The species and reactions histograms have similar patterns, suggesting that the models tend to also have low numbers of reactions.



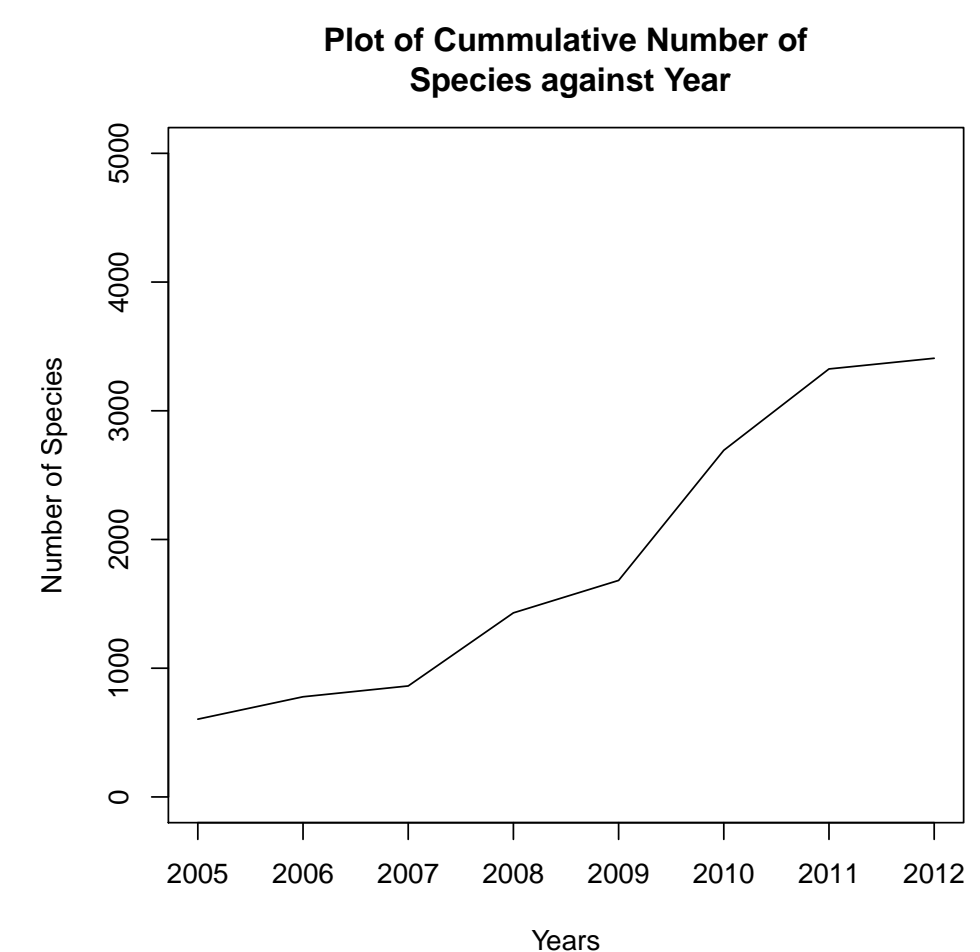
The most frequent range of values for the ratios of species to reactions is 0.5-1.0, with the majority of models having ratios less than 2.

This suggests that in the majority of models, the species tend to appear in multiple reactions, since if every species in a model appeared in just one reaction, the ratio would be at least 2.

## MODELS AND SPECIES OVER TIME

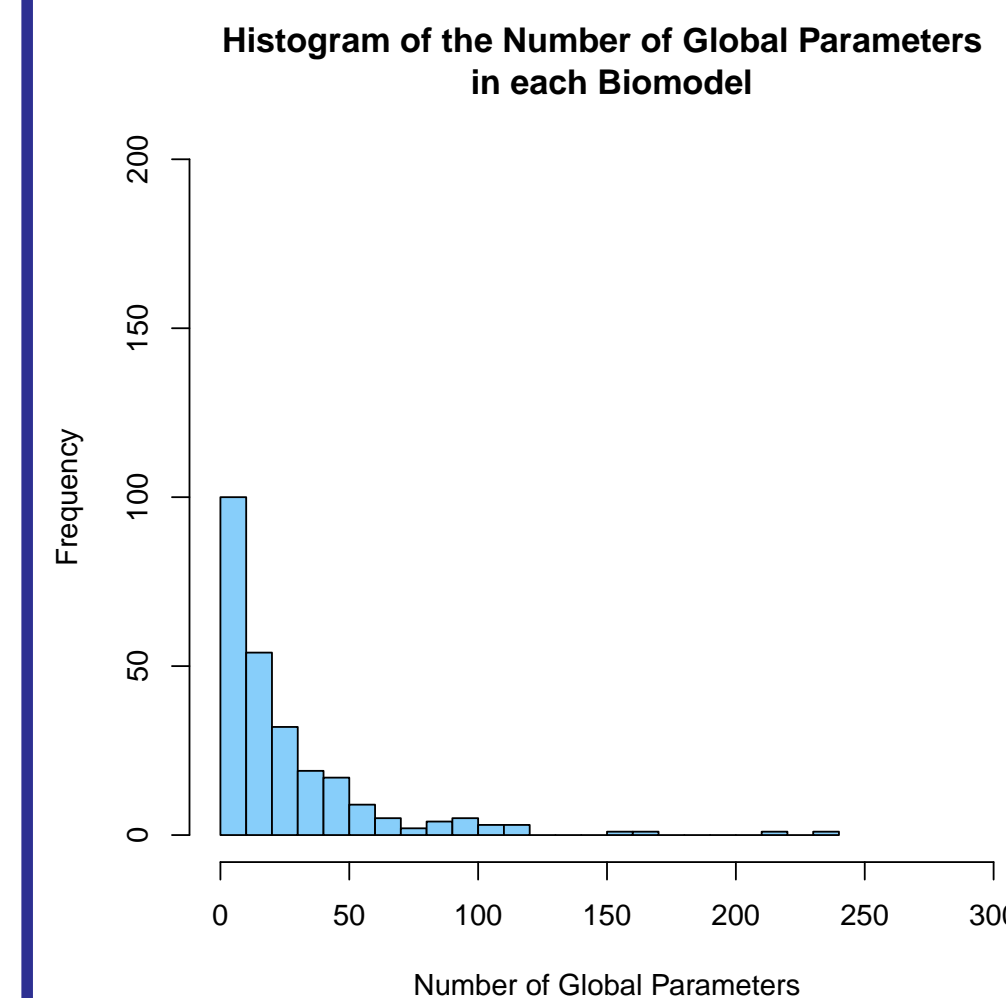


The increase in the number of curated models appears to be almost linear, suggesting that models are being added at a reasonably constant rate.

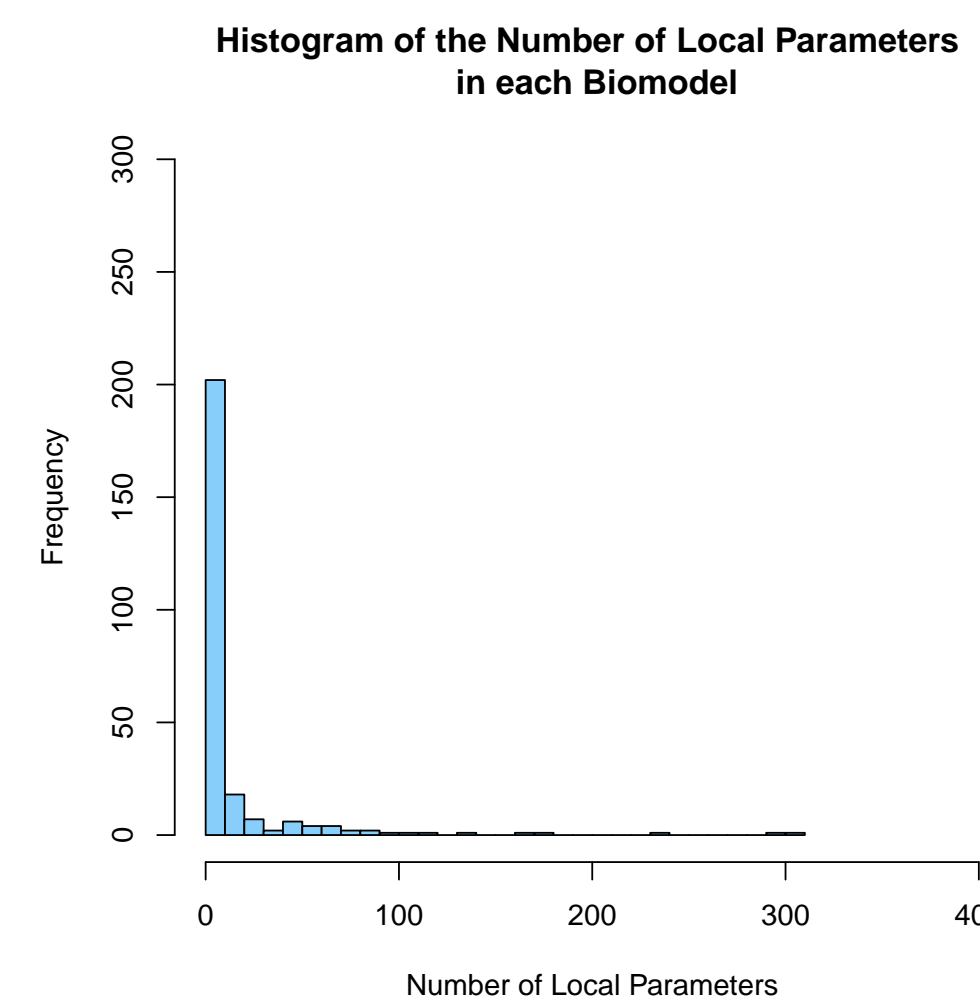


There appears to be a pattern that a large increase in the number of species in one year precedes a smaller increase in the next year.

## PARAMETERS



The majority of models have twenty or fewer global parameters. This suggests the models tend to have a low number of global parameters.



The majority of models have fewer than ten local parameters. This suggests the models tend to have a low number of local parameters.

## REFERENCES

- [1] Wolkenhauer, O., Wellstead, P., Cho, K.H. Essays in Biochemistry volume 45 Systems Biology
- [2] Wilkinson, D. Stochastic Modelling for Systems Biology
- [3] BioModels Database (<http://www.ebi.ac.uk/biomodels-main/>)
- [4] R Project (<http://www.r-project.org/>)