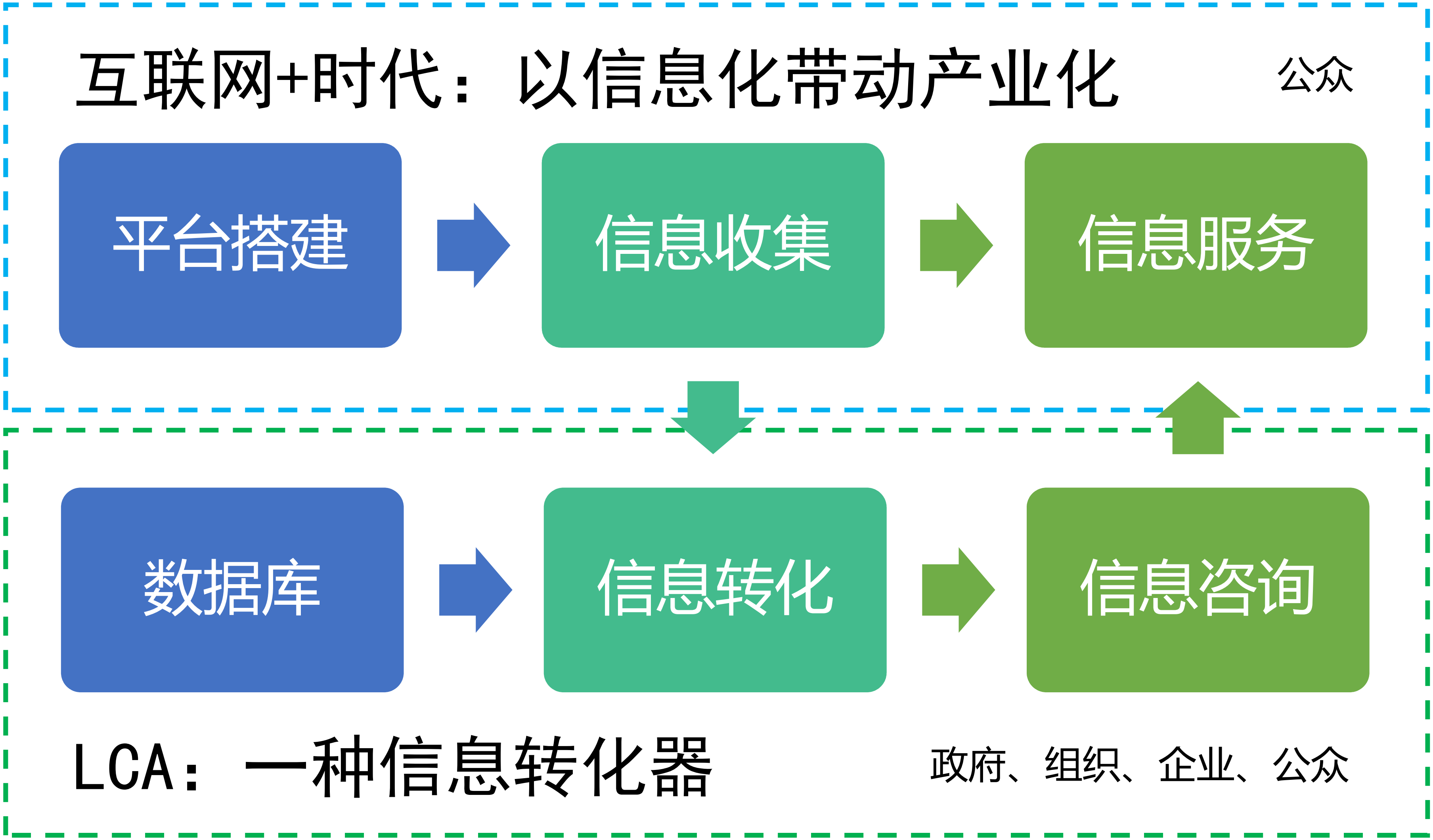
NANJING UNIVERSITY

LCI构建中的数据处理及其不确定性

盛虎 副研究员

南京大学环境学院

背景



互联网+时代：以信息化带动产业化

公众

平台搭建

信息收集

信息服务

数据库

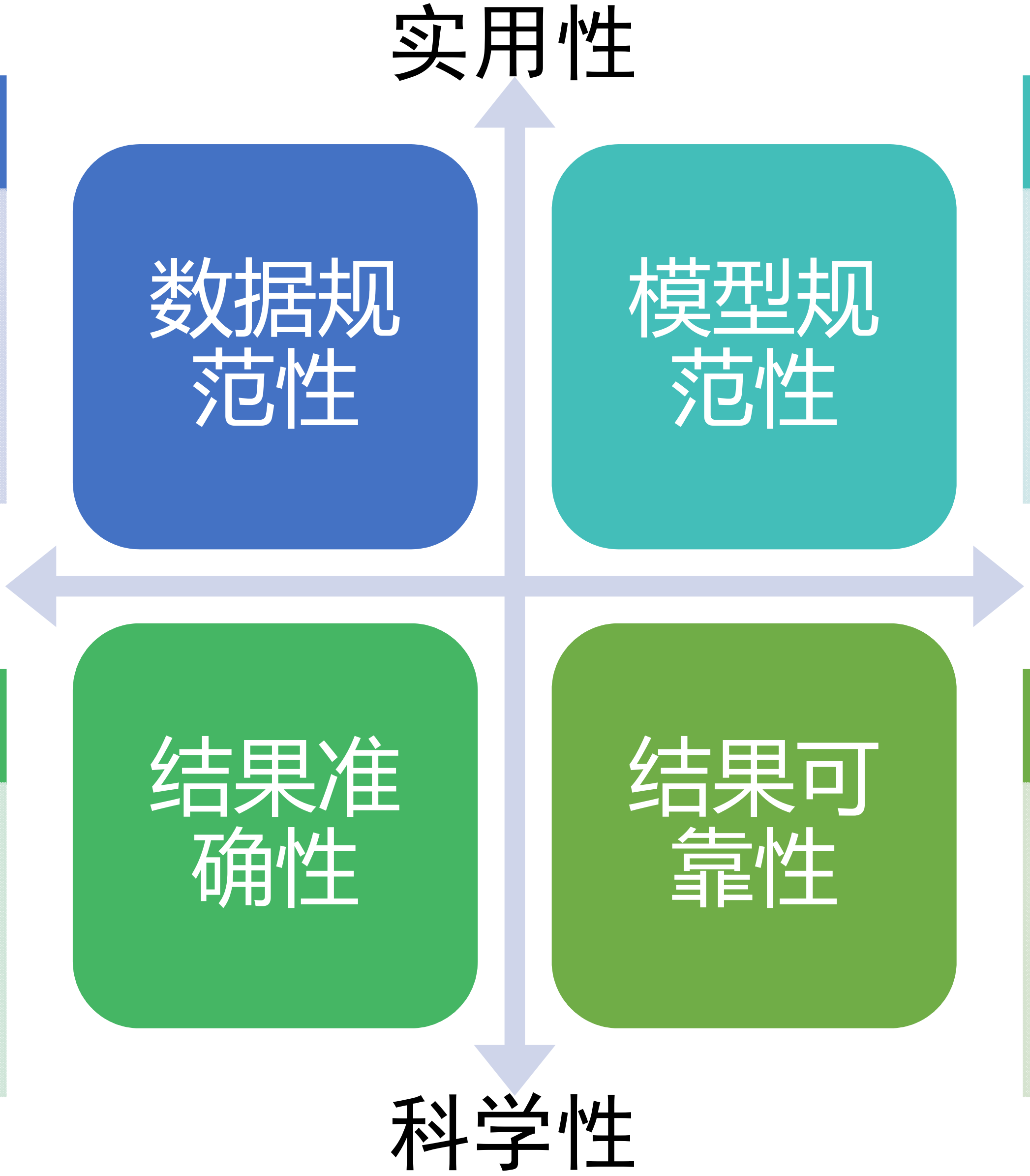
信息转化

信息咨询

LCA：一种信息转化器

政府、组织、企业、公众

需要解决的问题



实用性

数据规范性

模型规范性

结果准确性

结果可靠性

科学性

信息对接

- 标准数据库
- 数据接口
- 数据模型

大规模计算

- 标准模型库
- 高分辨率
- 分布式计算

不确定性评估

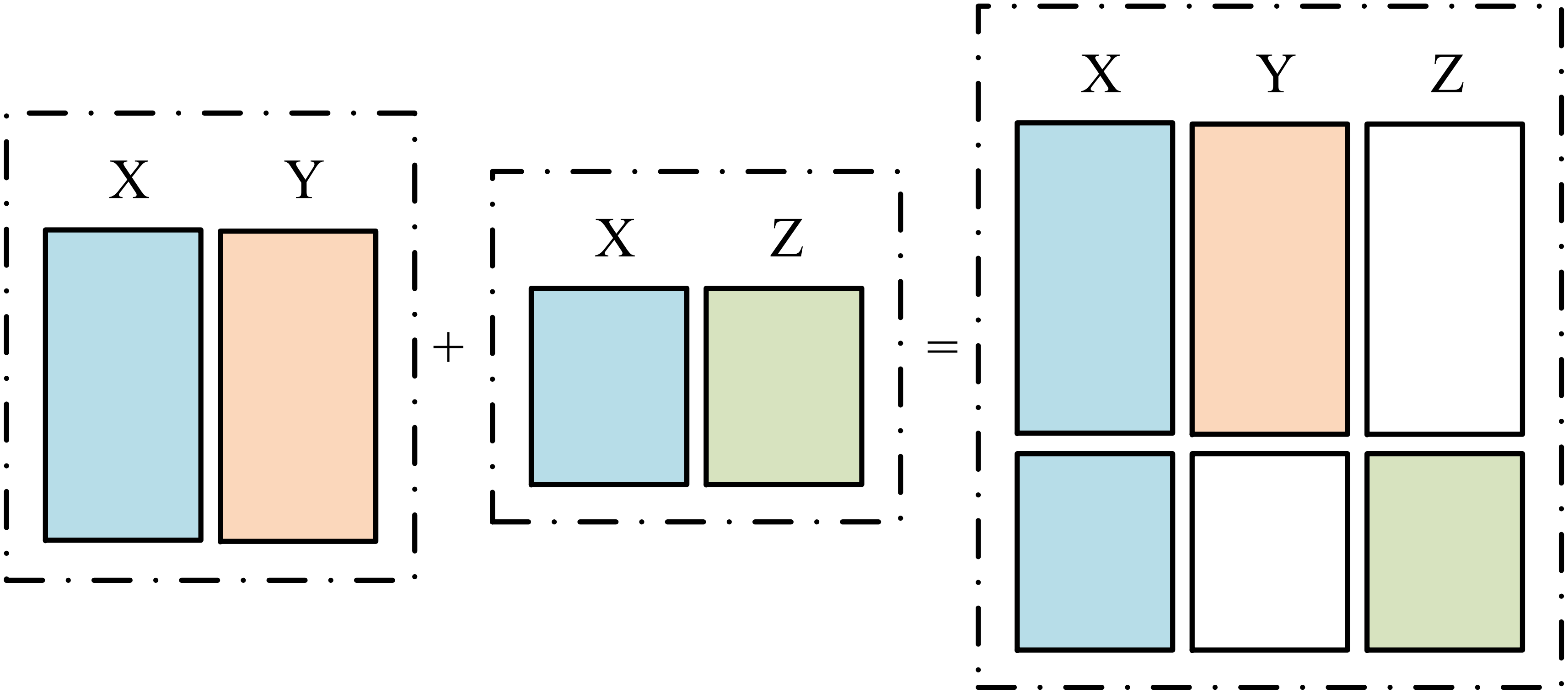
- 数据评估
- 模型评估
- 结果评估

结果验证

- 数据代表性
- 时间变异
- 空间变异

数据规范性

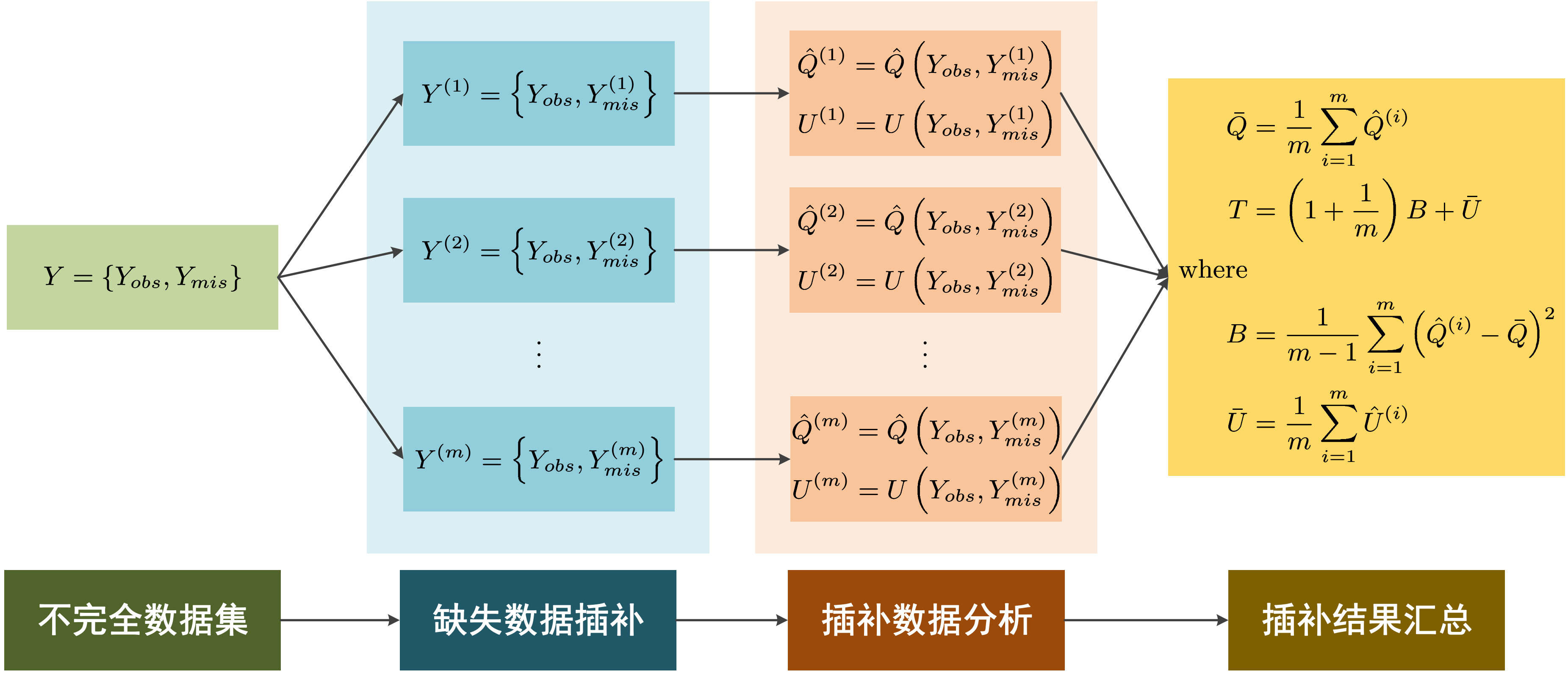
多源数据融合——求并集



数据源：A　　数据源：B　　数据融合：A∪B

出现数据缺失的问题

缺失数据的多重插补



$$Y = \{Y_{obs}, Y_{mis}\}$$

$$Y^{(1)} = \left\{Y_{obs}, Y_{mis}^{(1)}\right\}$$

$$Y^{(2)} = \left\{Y_{obs}, Y_{mis}^{(2)}\right\}$$

$$\vdots$$

$$Y^{(m)} = \left\{Y_{obs}, Y_{mis}^{(m)}\right\}$$

$$\hat{Q}^{(1)} = \hat{Q}\left(Y_{obs}, Y_{mis}^{(1)}\right)$$

$$U^{(1)} = U\left(Y_{obs}, Y_{mis}^{(1)}\right)$$

$$\hat{Q}^{(2)} = \hat{Q}\left(Y_{obs}, Y_{mis}^{(2)}\right)$$

$$U^{(2)} = U\left(Y_{obs}, Y_{mis}^{(2)}\right)$$

$$\vdots$$

$$\hat{Q}^{(m)} = \hat{Q}\left(Y_{obs}, Y_{mis}^{(m)}\right)$$

$$U^{(m)} = U\left(Y_{obs}, Y_{mis}^{(m)}\right)$$

$$\bar{Q} = \frac{1}{m}\sum_{i=1}^{m}\hat{Q}^{(i)}$$

$$T = \left(1 + \frac{1}{m}\right)B + \bar{U}$$

where

$$B = \frac{1}{m-1}\sum_{i=1}^{m}\left(\hat{Q}^{(i)} - \bar{Q}\right)^2$$

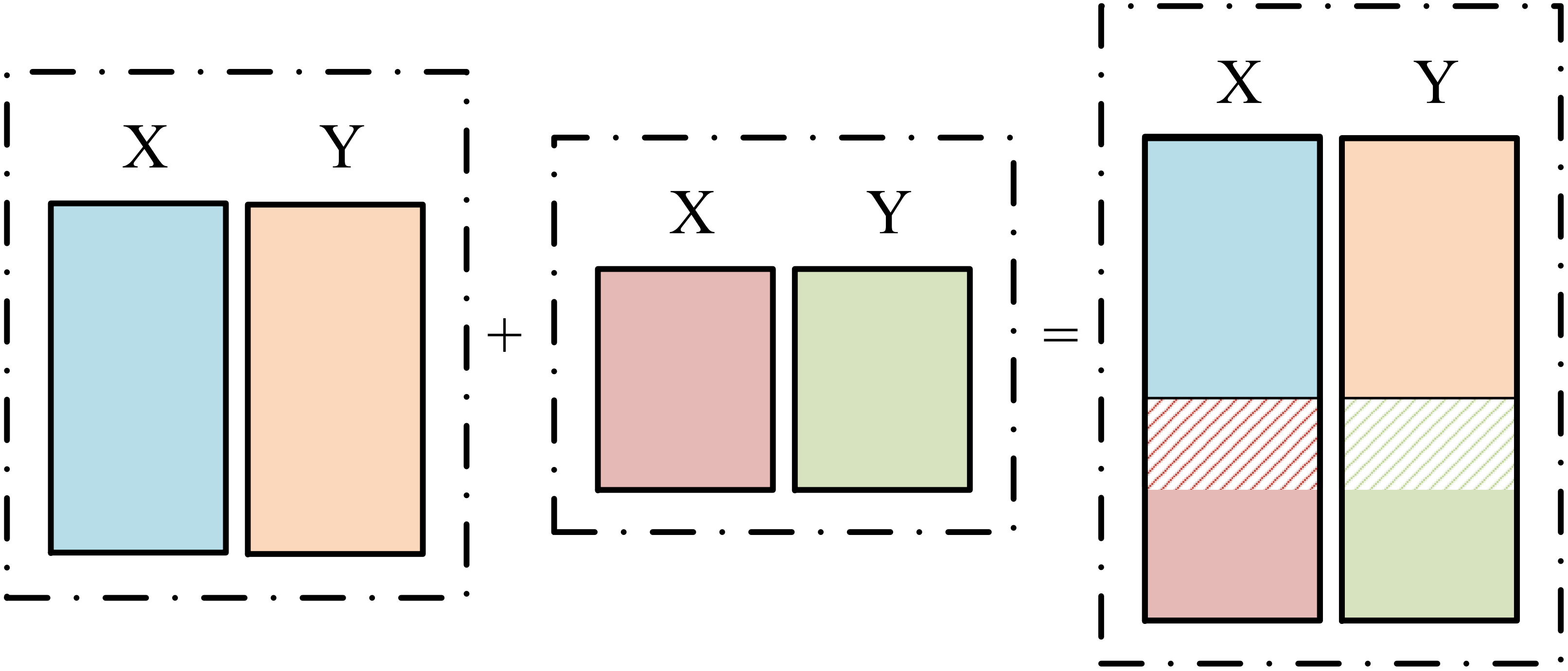
$$\bar{U} = \frac{1}{m}\sum_{i=1}^{m}\hat{U}^{(i)}$$

不完全数据集 → 缺失数据插补 → 插补数据分析 → 插补结果汇总

结果：得到了一个带有不确定信息的数据集

盛虎，郭怀成，数据缺失下流域模拟方法研究，科学出版社，2015

多源数据融合——求交集

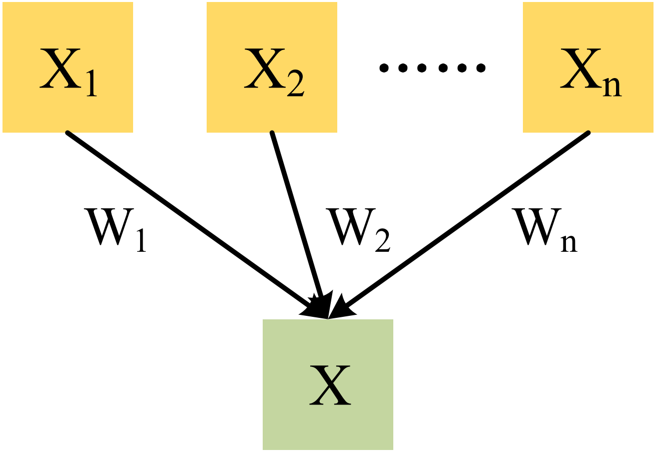


数据源：A　　数据源：B　　数据融合：A∩B

➡ 出现数据不一致的问题

不一致数据的处理

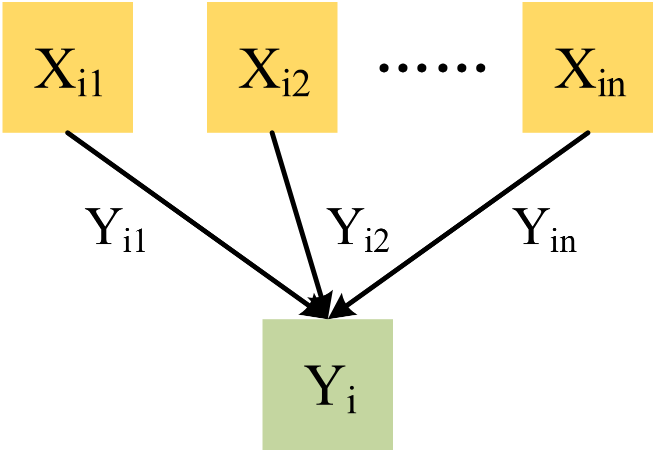
数值不一致



加权平均

$$X=\sum_{i=1}^{n} W_i X_i$$

时空尺度不一致



尺度转化

$$Y_{ij}=W(X_{ij})Y_i$$

结果：得到了一个带有不确定信息的数据集

标准化数据库

NAME	NOTE	UNIT	TIME	DATA	CV	DIST	P1	P2	P3
DRP	原矿产量	kt	Y2000	20	0.1	unif	16.536	23.464	
PRP	原矿元素含量	1	Y2000	0.5	0.1	triangle	0.378	0.622	0.500
PRR	矿石产率	1	Y2000	0.8	0.2	beta	4.200	1.050	0.000
DFP	化肥产量	kt	Y2000	15	0.1	unif	12.402	17.598	
PFP	化肥元素含量	1	Y2000	0.4	0.2	triangle	0.204	0.596	0.400
DCP	日化产量	kt	Y2000	5	0.1	unif	4.134	5.866	
PCP	日化元素含量	1	Y2000	0.3	0.3	triangle	0.080	0.520	0.300
PCR	化工废物回用率	1	Y2000	0.5	0.2	beta	12.000	12.000	0.000
DMP	作物产量	kt	Y2000	1000	0.1	unif	826.795	1173.205	
PMP	作物元素含量	1	Y2000	0.005	0.2	triangle	0.003	0.007	0.005
PHR	饲料占比	1	Y2000	0.4	0	none	0.400		
PNR	食品占比	1	Y2000	0.45	0	none	0.450		
DAP	动物产量	kt	Y2000	800	0.1	unif	661.436	938.564	
PAP	动物元素含量	1	Y2000	0.002	0.2	triangle	0.001	0.003	0.002
PAR	肉产率	1	Y2000	0.75	0.2	beta	5.500	1.833	0.000

基本数据

数据的不确定性

原则：①直观；②可累积；③含不确定性信息

数据接口

Long-type

Start	**End**	**Flow**
X1	X1	a11
X1	X2	a12
X1	X3	a13
……	……	……
X2	X1	a21
X2	X2	a22
X2	X3	a23
……	……	……
Y1	X1	b11
Y1	X2	b12
Y1	X3	b13
……	……	……

标准化数据库

⟷

数据转换

Wide-type

	X1	X2	X3	……
X1	a11	a12	a13	……
X2	a21	a22	a23	……
X3	a31	a32	a33	……
……	……	……	……	……

技术矩阵（经济矩阵）

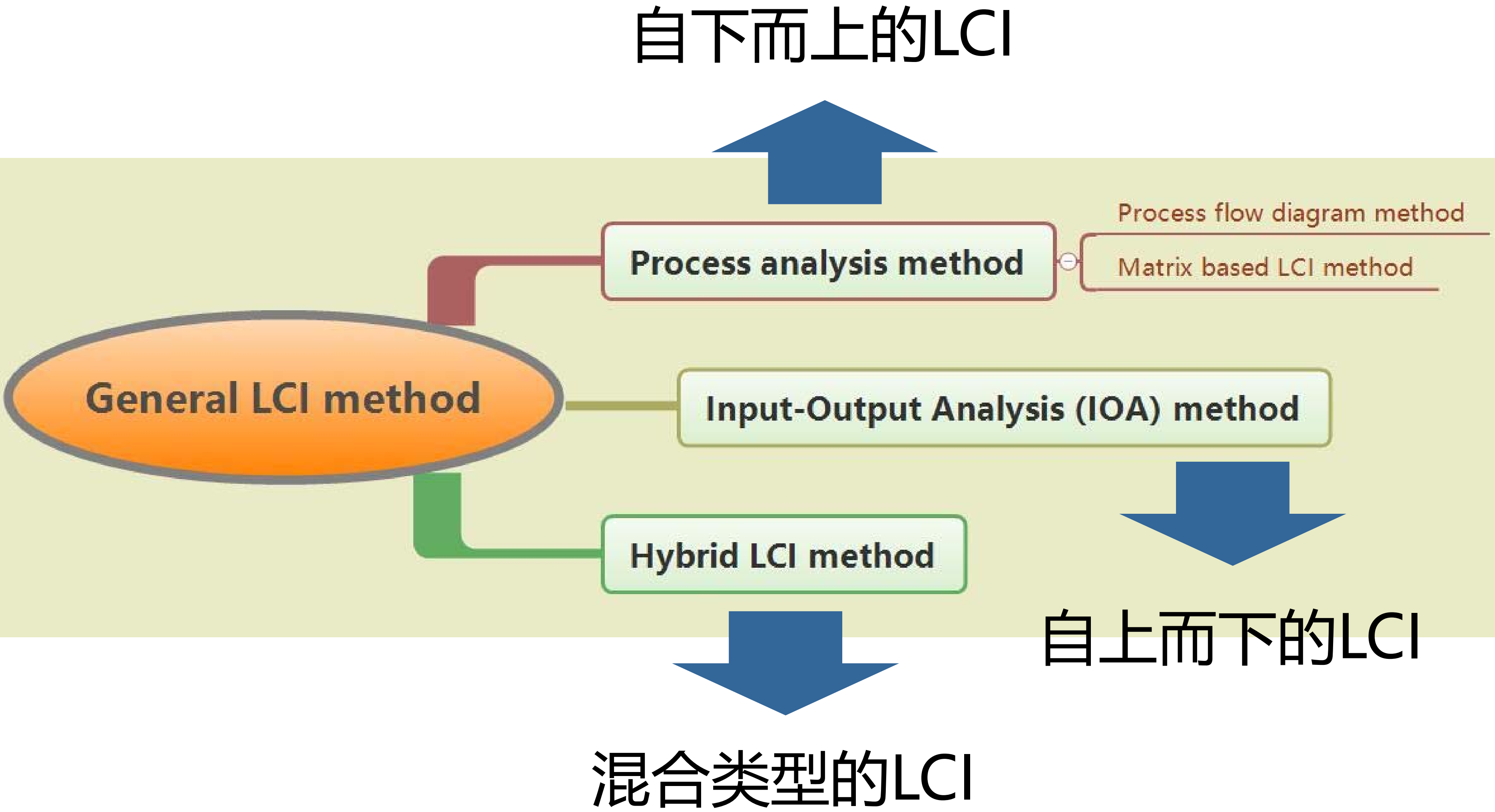
$$g = BA^{-1}f$$

	X1	X2	X3	……
Y1	b11	b12	b13	……
Y2	b21	b22	b23	……
……	……	……	……	……

干扰矩阵（环境矩阵）

模型规范性

LCI分析方法



自下而上的LCI

General LCI method

Process analysis method

Process flow diagram method

Matrix based LCI method

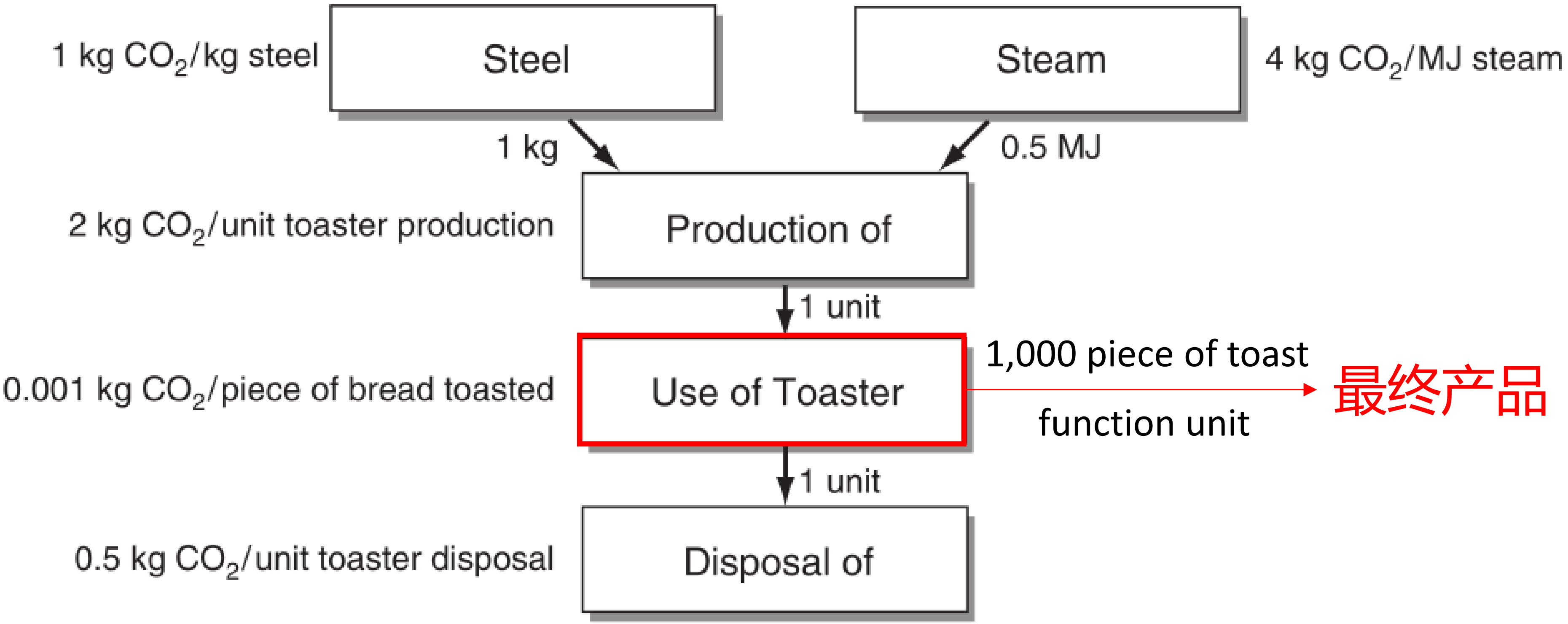
Input-Output Analysis (IOA) method

Hybrid LCI method

自上而下的LCI

混合类型的LCI

过程流图法



1 kg CO_2/kg steel

Steel

Steam

4 kg CO_2/MJ steam

1 kg

0.5 MJ

2 kg CO_2/unit toaster production

Production of

1 unit

0.001 kg CO_2/piece of bread toasted

Use of Toaster

1,000 piece of toast

function unit

最终产品

1 unit

0.5 kg CO_2/unit toaster disposal

Disposal of

Suppose that the toaster under study produces 1,000 pieces of toast during its life time.

	Steel Production	Steam Production	Toaster production	Bread toasting	Toaster disposal	LCI
CO_2	1	0.5*4	2	0.001*1000	0.5	6.5

矩阵算法

Process Commodity	Steel Production	Steam Production	Toaster production	Bread toasting	Toaster disposal	Reference flow
Steel	1	0	-1	0	0	0
Steam	0	1	-0.5	0	0	0
Toaster	0	0	1	-1	0	0
Bread toasted	0	0	0	1000	0	1000
Toaster disposed	0	0	0	1	-1	0
CO_2	1	4	2	1	0.5	?

$$As = f$$

LCI的矩阵化

$$g = Bs = BA^{-1}f$$

	Steel Production	Steam Production	Toaster production	Bread toasting	Toaster disposal	Reference flow
Steel	A					f
Steam						
Toaster						
Bread toasted						
Toaster disposed						
CO_2	B					g

投入产出法

$A = (a_{ij})_{n\times n}$ 消耗系数矩阵 $a_{ij} = \dfrac{\mathrm{input}_i}{\mathrm{output}_j}$ 参考流

	柴油	硬煤资源	动力煤	公路运输量	水电	火电	电网电力	最终产品
柴油	0	0	0	37.63	0	0	0	0
硬煤资源	0	0	1.09	0	0	0	0	0
动力煤	0	0	0	0	0	0.5	0	0
公路运输量	0	0	0	0	0	0.1	0	0
水电	0	0	0	0	0	0	0.161	0
火电	0	0	0	0	0	0	0.914	0
电网电力	0	0	0	0	0	0	0	1
SO_2	0	0	3	0.09	0	5.68	0	
NO_X	0	0	0	1.6	0	2.87	0	
CO_2	0	0	330	118	0	891	0	
CH_4	0	0	5.12	0.004	0.286	0	0	

$B = (b_{ij})_{m\times n}$ 排放系数矩阵 $b_{ij} = \dfrac{\mathrm{emission}_i}{\mathrm{output}_j}$

LCI计算公式

$$g = B(I - A)^{-1} f$$

$$f = (f_i)_{n\times 1}$$

LCI的一般化表达

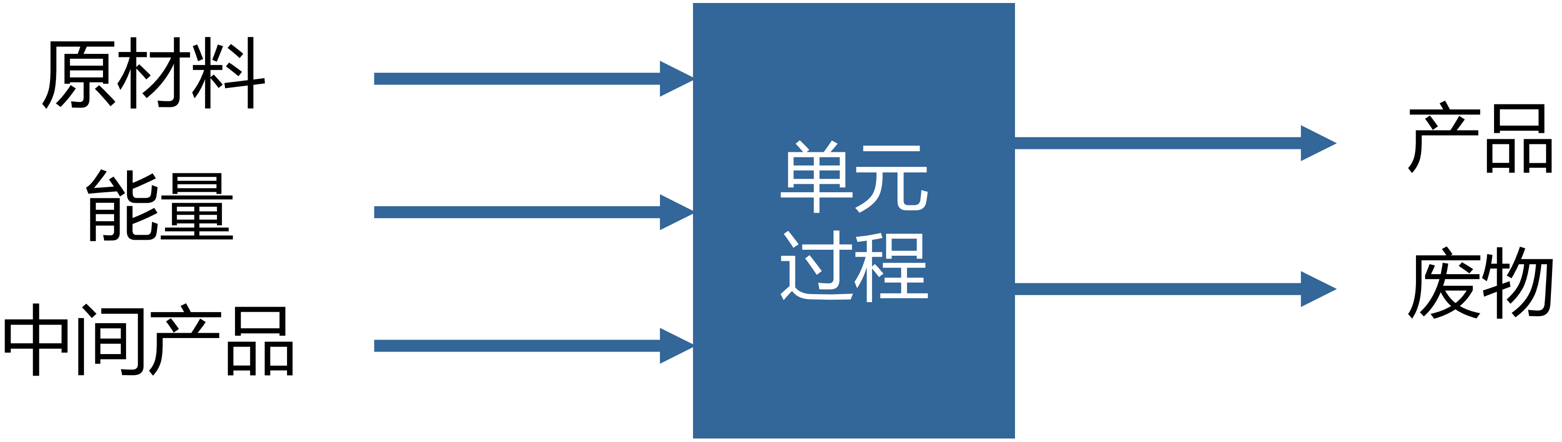
Process inventory → **Linear assumption**

$$\left[\begin{array}{c} -\mathrm{Input} \\ \mathrm{Output} \\ \hline \mathrm{Emission} \end{array}\right] \begin{array}{l} \left.\vphantom{\begin{array}{c} a \\ a \end{array}}\right\} \text{Technology (economic) flow} \\ \left.\vphantom{a}\right\} \text{Intervention (environmental) flow} \end{array}$$

PI_1	→	PI_2	→	…	→	PI_n
SF_1		SF_2				SF_n

$$\mathrm{LCI}_i = \sum_{j=1}^{n} \mathrm{SF}_j \cdot \mathrm{PI}_{ij} \quad \left\{ \begin{array}{c} \text{When } i \text{ is intermediate flow} \\ \mathrm{LCI}_i = 0 \end{array} \right.$$

LCI建模


原材料

能量

中间产品

单元过程

产品

废物

- 独立型计算：

  FLOW[节点i,节点j]=活动水平*转换系数
- 依附型计算：

  FLOW[节点i,节点j]=FLOW[节点k,节点i]*分配系数
- 平衡型计算：

  FLOW[节点i,节点j]=BALANCE(节点i|节点j)

标准化模型库

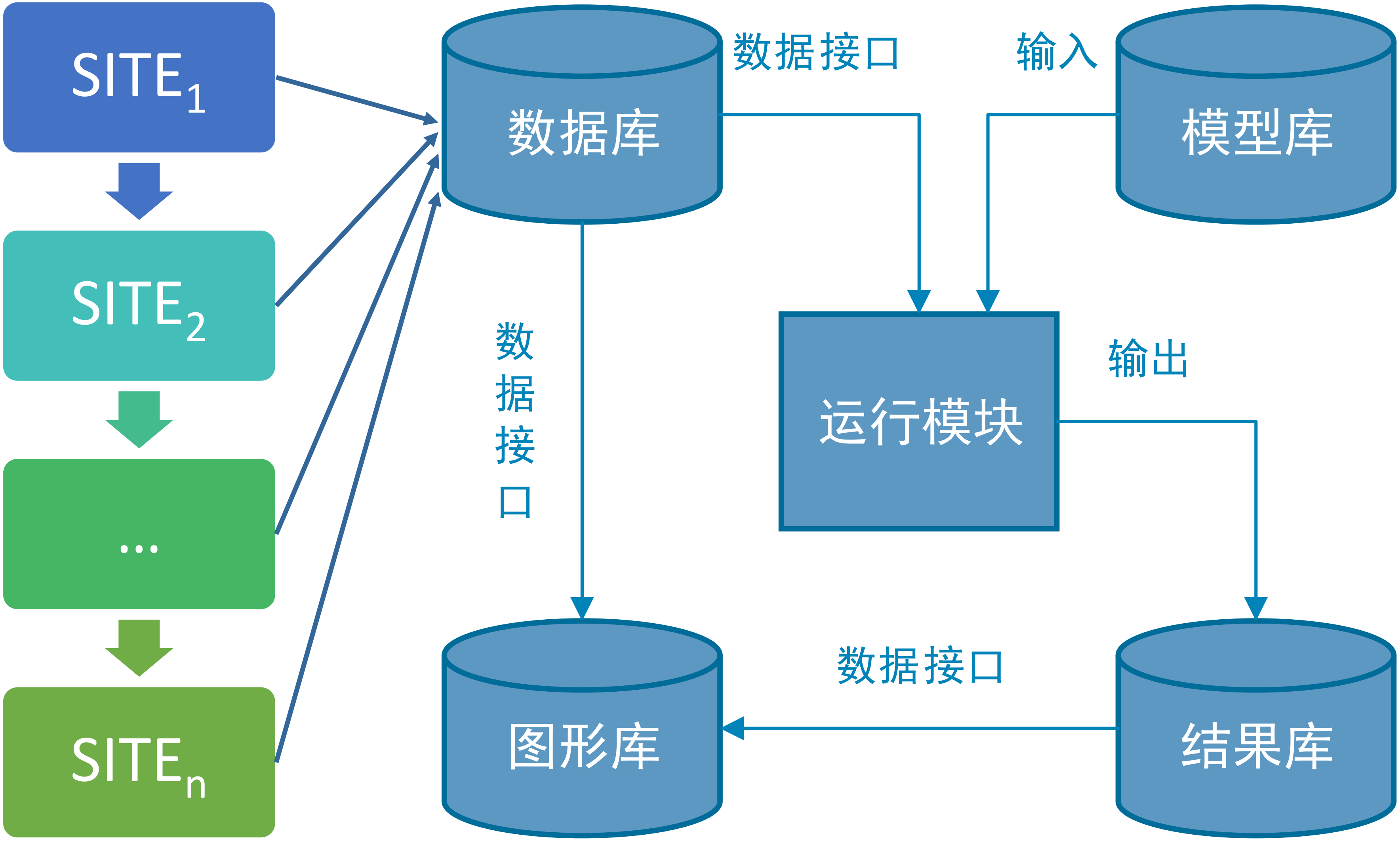
NAME	START	END	FUN
P rock production			
PF_01	R	Ch	DRP*PRP*PRR
PF_02	R	W	DRP*PRP-PF_01
P chemical production			
PF_03	Ch	Cr	DFP*PFP
PF_04	Ch	H	DCP*PCP
PF_05	W	Ch	PF_06*PCR
PF_06	Ch	W	(PF_01-PF_03-PF_04)/(1-PCR)
Crop production			
PF_07	Cr	A	DMP*PMP*PHR
PF_08	Cr	H	DMP*PMP*PNR
PF_09	Cr	W	PF_03-PF_07-PF_08
Animal production			
PF_10	A	H	DAP*PAP*PAR
PF_11	A	W	PF_07-PF_10
Human consumption			
PF_12	H	W	PF_04+PF_08+PF_10

第二步：定义流

第一步：模型结构

第三步：模型表达

大规模计算



SITE$_1$

SITE$_2$

…

SITE$_n$

数据库

数据接口

输入

模型库

数据接口

运行模块

输出

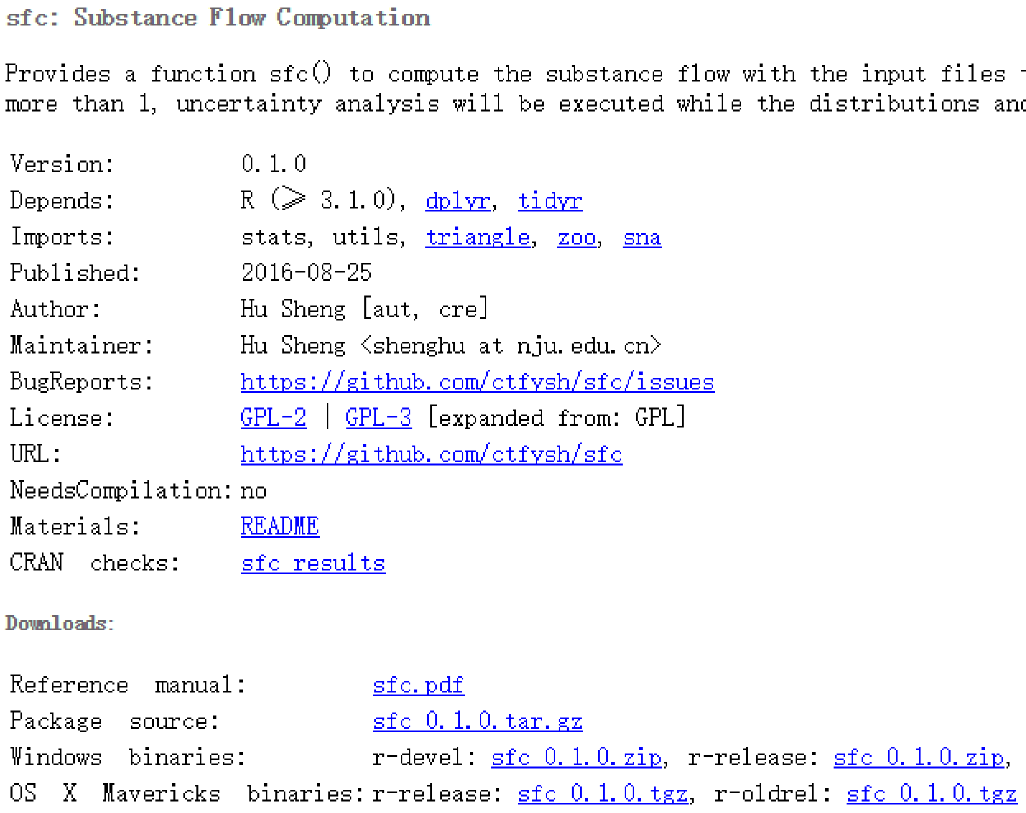
图形库

数据接口

结果库

sfc软件包及可视化界面

网址： https://cran.r-project.org/web/packages/sfc/index.html



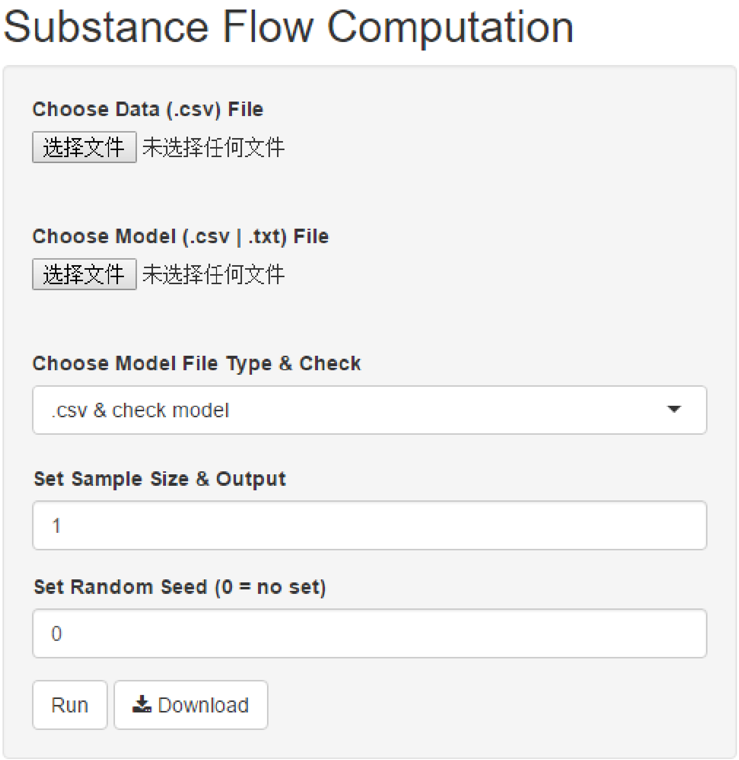
sfc: Substance Flow Computation

Provides a function sfc() to compute the substance flow with the input files — "data" and "model". If sample.size is set more than 1, uncertainty analysis will be executed while the distributions and parameters are supplied in the file "data".

Version:	0.1.0
Depends:	R (≥ 3.1.0), dplyr, tidyr
Imports:	stats, utils, triangle, zoo, sna
Published:	2016-08-25
Author:	Hu Sheng [aut, cre]
Maintainer:	Hu Sheng <shenghu at nju.edu.cn>
BugReports:	https://github.com/ctfysh/sfc/issues
License:	GPL-2 | GPL-3 [expanded from: GPL]
URL:	https://github.com/ctfysh/sfc
NeedsCompilation:	no
Materials:	README
CRAN checks:	sfc results

Downloads:

Reference manual:	sfc.pdf
Package source:	sfc_0.1.0.tar.gz
Windows binaries:	r-devel: sfc_0.1.0.zip, r-release: sfc_0.1.0.zip,
OS X Mavericks binaries:	r-release: sfc_0.1.0.tgz, r-oldrel: sfc_0.1.0.tgz



Substance Flow Computation

Choose Data (.csv) File

选择文件 未选择任何文件

Choose Model (.csv | .txt) File

选择文件 未选择任何文件

Choose Model File Type & Check

.csv & check model

Set Sample Size & Output

1

Set Random Seed (0 = no set)

0

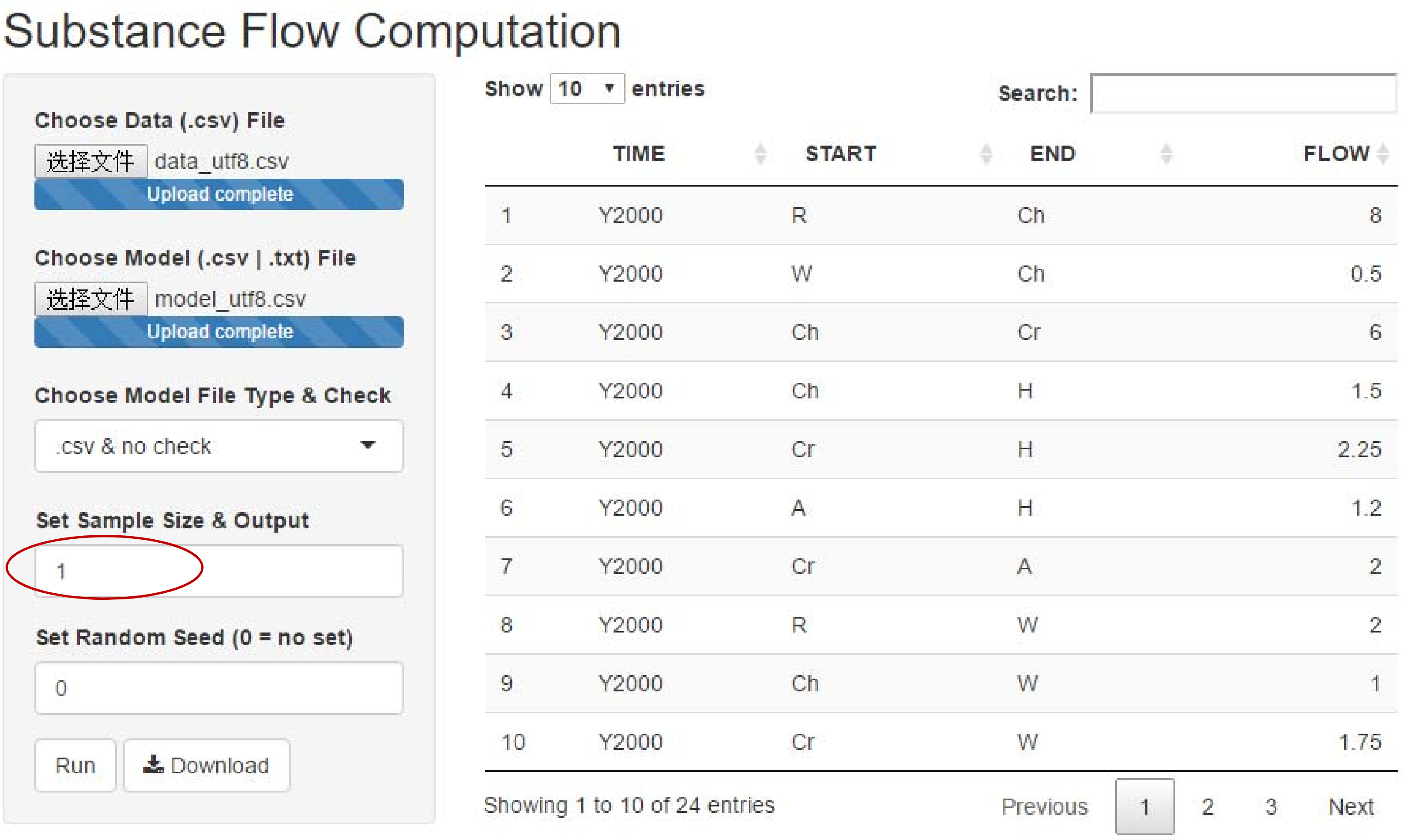
中文说明：

https://ctfysh.github.io/2016/08/27/sfc-package-manual.html

界面：

https://ctfysh.shinyapps.io/sample_app/

sfc进行流的计算



Substance Flow Computation

Choose Data (.csv) File

选择文件 data_utf8.csv

Upload complete

Choose Model (.csv | .txt) File

选择文件 model_utf8.csv

Upload complete

Choose Model File Type & Check

.csv & no check

Set Sample Size & Output

1

Set Random Seed (0 = no set)

0

Run | Download

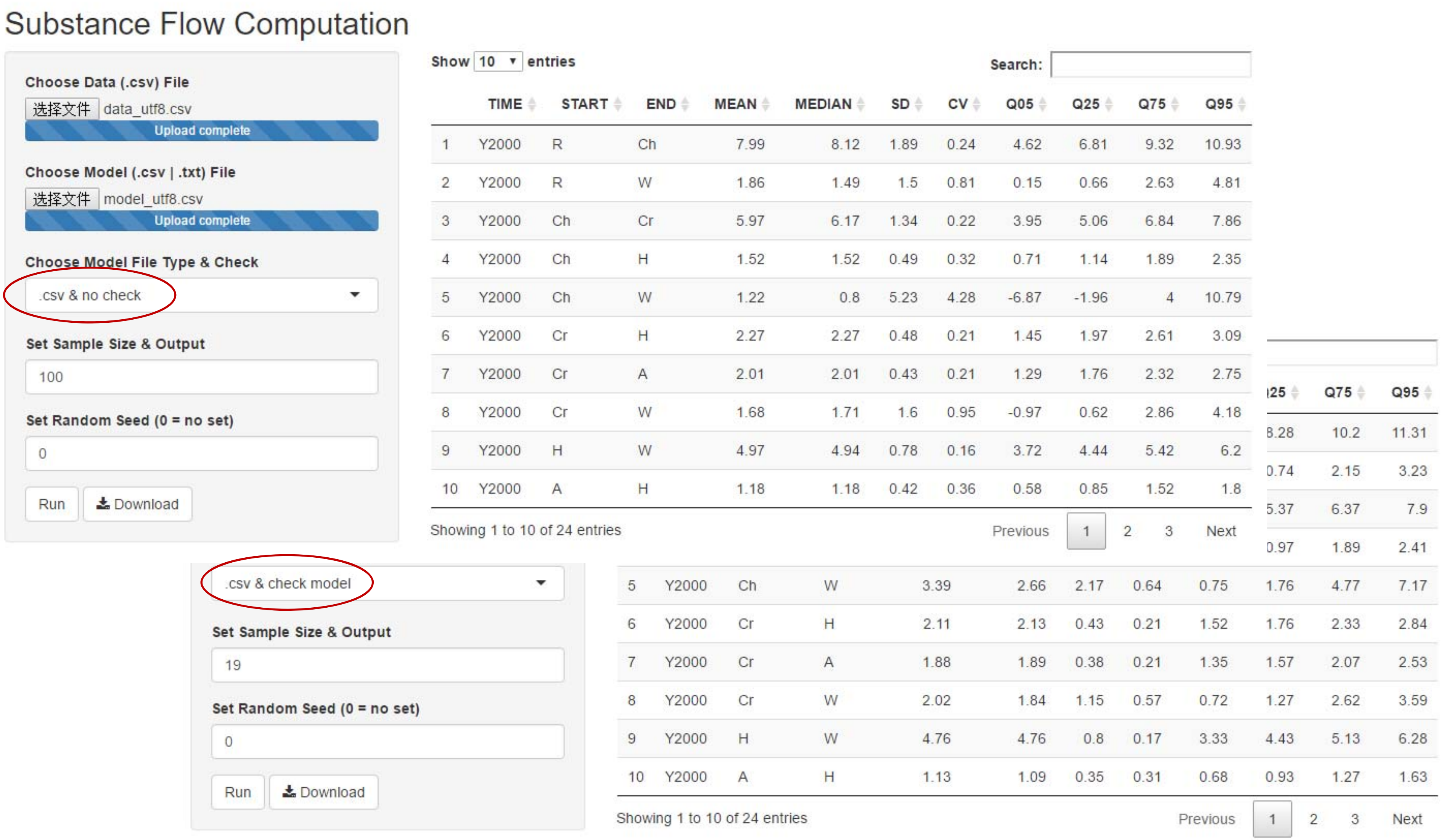
Show 10 entries | Search:

	TIME	START	END	FLOW
1	Y2000	R	Ch	8
2	Y2000	W	Ch	0.5
3	Y2000	Ch	Cr	6
4	Y2000	Ch	H	1.5
5	Y2000	Cr	H	2.25
6	Y2000	A	H	1.2
7	Y2000	Cr	A	2
8	Y2000	R	W	2
9	Y2000	Ch	W	1
10	Y2000	Cr	W	1.75

Showing 1 to 10 of 24 entries

Previous 1 2 3 Next

sfc进行不确定性分析



Substance Flow Computation

Choose Data (.csv) File

选择文件 data_utf8.csv

Upload complete

Choose Model (.csv | .txt) File

选择文件 model_utf8.csv

Upload complete

Choose Model File Type & Check

.csv & no check

Set Sample Size & Output

100

Set Random Seed (0 = no set)

0

Run | Download

Show 10 entries   Search:

	TIME	START	END	MEAN	MEDIAN	SD	CV	Q05	Q25	Q75	Q95
1	Y2000	R	Ch	7.99	8.12	1.89	0.24	4.62	6.81	9.32	10.93
2	Y2000	R	W	1.86	1.49	1.5	0.81	0.15	0.66	2.63	4.81
3	Y2000	Ch	Cr	5.97	6.17	1.34	0.22	3.95	5.06	6.84	7.86
4	Y2000	Ch	H	1.52	1.52	0.49	0.32	0.71	1.14	1.89	2.35
5	Y2000	Ch	W	1.22	0.8	5.23	4.28	-6.87	-1.96	4	10.79
6	Y2000	Cr	H	2.27	2.27	0.48	0.21	1.45	1.97	2.61	3.09
7	Y2000	Cr	A	2.01	2.01	0.43	0.21	1.29	1.76	2.32	2.75
8	Y2000	Cr	W	1.68	1.71	1.6	0.95	-0.97	0.62	2.86	4.18
9	Y2000	H	W	4.97	4.94	0.78	0.16	3.72	4.44	5.42	6.2
10	Y2000	A	H	1.18	1.18	0.42	0.36	0.58	0.85	1.52	1.8

Showing 1 to 10 of 24 entries   Previous 1 2 3 Next

.csv & check model

Set Sample Size & Output

19

Set Random Seed (0 = no set)

0

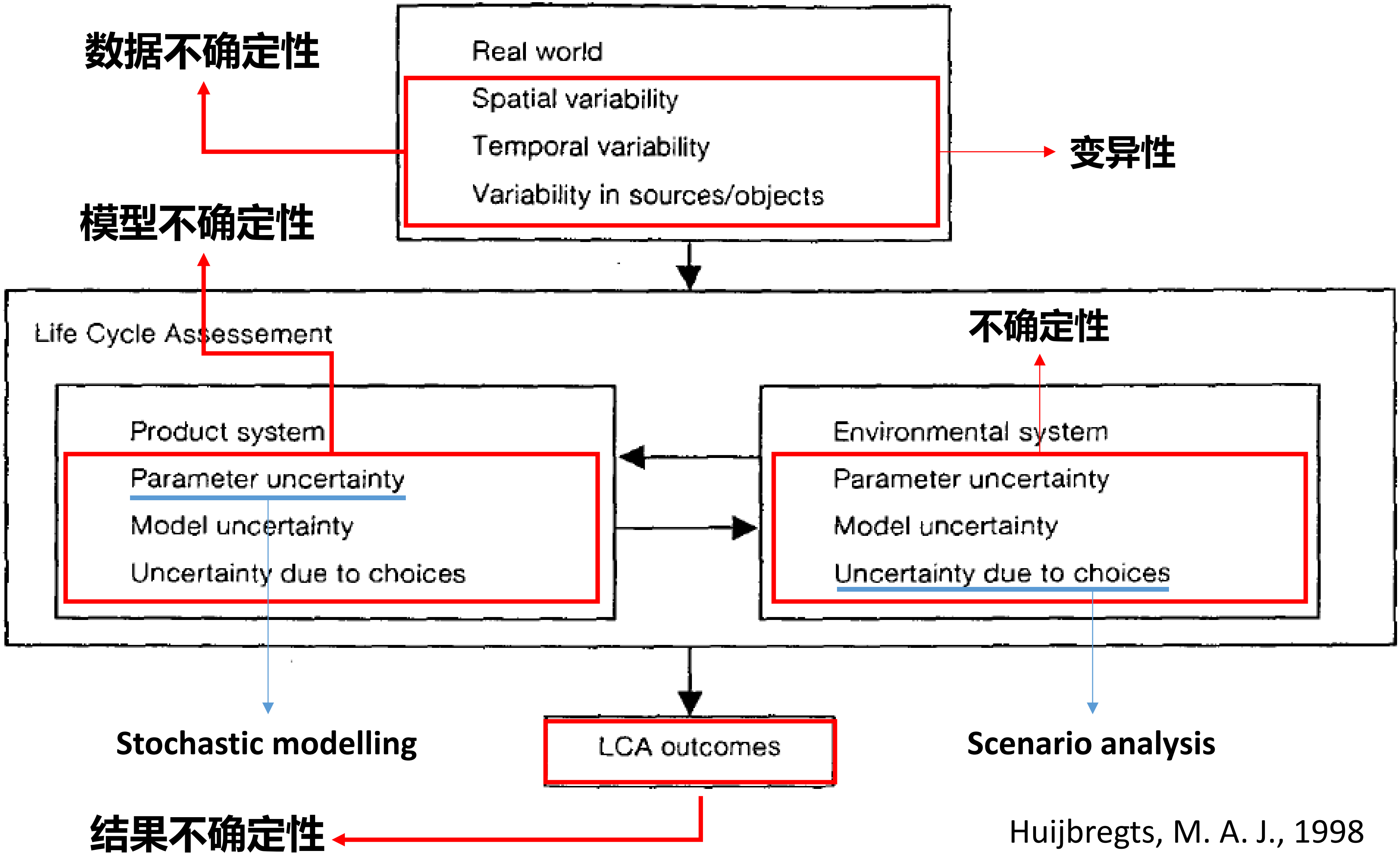
Run | Download

									25	Q75	Q95
									8.28	10.2	11.31
									0.74	2.15	3.23
									5.37	6.37	7.9
									0.97	1.89	2.41
5	Y2000	Ch	W	3.39	2.66	2.17	0.64	0.75	1.76	4.77	7.17
6	Y2000	Cr	H	2.11	2.13	0.43	0.21	1.52	1.76	2.33	2.84
7	Y2000	Cr	A	1.88	1.89	0.38	0.21	1.35	1.57	2.07	2.53
8	Y2000	Cr	W	2.02	1.84	1.15	0.57	0.72	1.27	2.62	3.59
9	Y2000	H	W	4.76	4.76	0.8	0.17	3.33	4.43	5.13	6.28
10	Y2000	A	H	1.13	1.09	0.35	0.31	0.68	0.93	1.27	1.63

Showing 1 to 10 of 24 entries   Previous 1 2 3 Next

结果准确性

不确定性类型



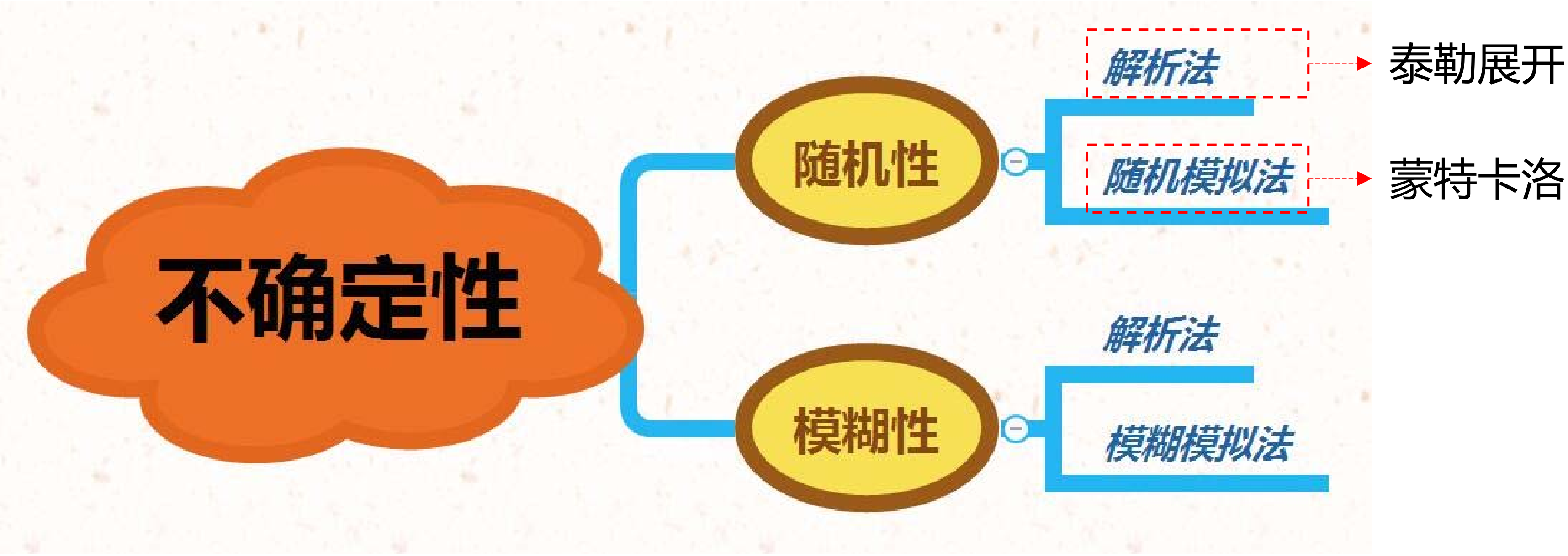
Huijbregts, M. A. J., 1998

不确定性分析

- 对象：模型 → 任何一个过程都可能导致不确定性

$$\underbrace{y_i}_{\text{模型输出}} = \underbrace{f_i}_{\text{模型结构}}(\underbrace{x_1, x_2, \cdots, x_m}_{\text{模型输入}} | \underbrace{\theta_1, \theta_2, \cdots, \theta_k}_{\text{模型参数}}), \quad i = 1, 2, \cdots, n$$

- 目的：**将输入的不确定传递给输出**



解析法 → 泰勒展开

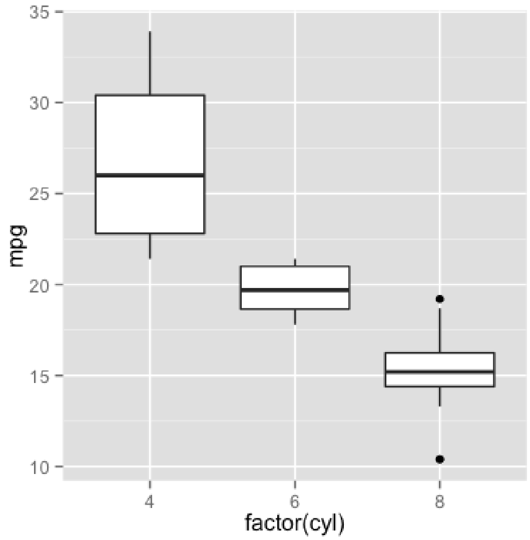
随机模拟法 → 蒙特卡洛

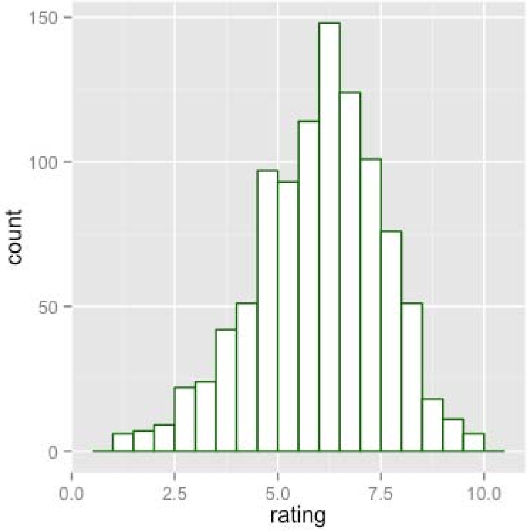
不确定性的表达

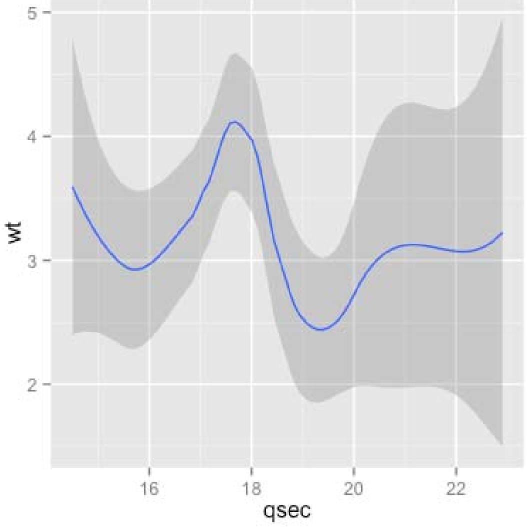
- 数值

均值： $$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$ 方差： $$S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

- 图表


箱线图


直方图


范围图

LCI不确定性解析法

Taylor展开：

$$\Delta z = \frac{\partial z}{\partial x}\Delta x + \frac{1}{2}\frac{\partial^2 z}{\partial x^2}\Delta x^2 + \frac{1}{6}\frac{\partial^3 z}{\partial x^3}\Delta x^3 + \cdots$$

忽略高阶项：

$$\Delta z = \frac{\partial z}{\partial x}\Delta x + \frac{\partial z}{\partial y}\Delta y$$

$$z = f(x, y)$$

求变量的方差：

$$\mathrm{var}(z) = (\frac{\partial z}{\partial x})^2 \mathrm{var}(x) + (\frac{\partial z}{\partial y})^2 \mathrm{var}(y) + 2\frac{\partial z}{\partial x}\frac{\partial z}{\partial x}\mathrm{cov}(x, y)$$

独立性假设：

$$\mathrm{var}(z) = (\frac{\partial z}{\partial x})^2 \mathrm{var}(x) + (\frac{\partial z}{\partial y})^2 \mathrm{var}(y)$$

$$\frac{\partial \ln(z)}{\partial \ln(x)} = \frac{x}{z}\frac{\partial z}{\partial x}$$

Monte Carlo方法

Monte Carlo方法基本思路：

$$\boxed{\mathbb{E}(X)} \xrightarrow[X\sim f(x)]{\text{continuous}} \boxed{\int_{-\infty}^{+\infty} x f(x)\mathrm{d}x} \xrightarrow[X_i\sim f(x)]{\text{discrete}} \boxed{\frac{1}{n}\sum_{i=1}^{n} X_i}$$

随机变量的期望值

微积分

数值计算

求方差的数值解：

$$\mathrm{var}(Z) = \mathrm{var}(f(X,Y)) = \frac{1}{n-1}\sum_{i=1}^{n}[f(X_i,Y_i) - \bar{f}(X_i,Y_i)]^2$$

Monte Carlo方法是用一种随机方法来计算确定数值的过程

LCI数据不确定性

正态分布

$$\begin{cases} P_1 = \mu \\ P_2 = \mathrm{CV} \cdot \mu \end{cases}$$

均匀分布

$$\begin{cases} P_1 = (1 - \sqrt{3} \cdot \mathrm{CV})\mu \\ P_2 = (1 + \sqrt{3} \cdot \mathrm{CV})\mu \end{cases}$$

$$\mathrm{CV} = \frac{\sigma}{\mu}$$

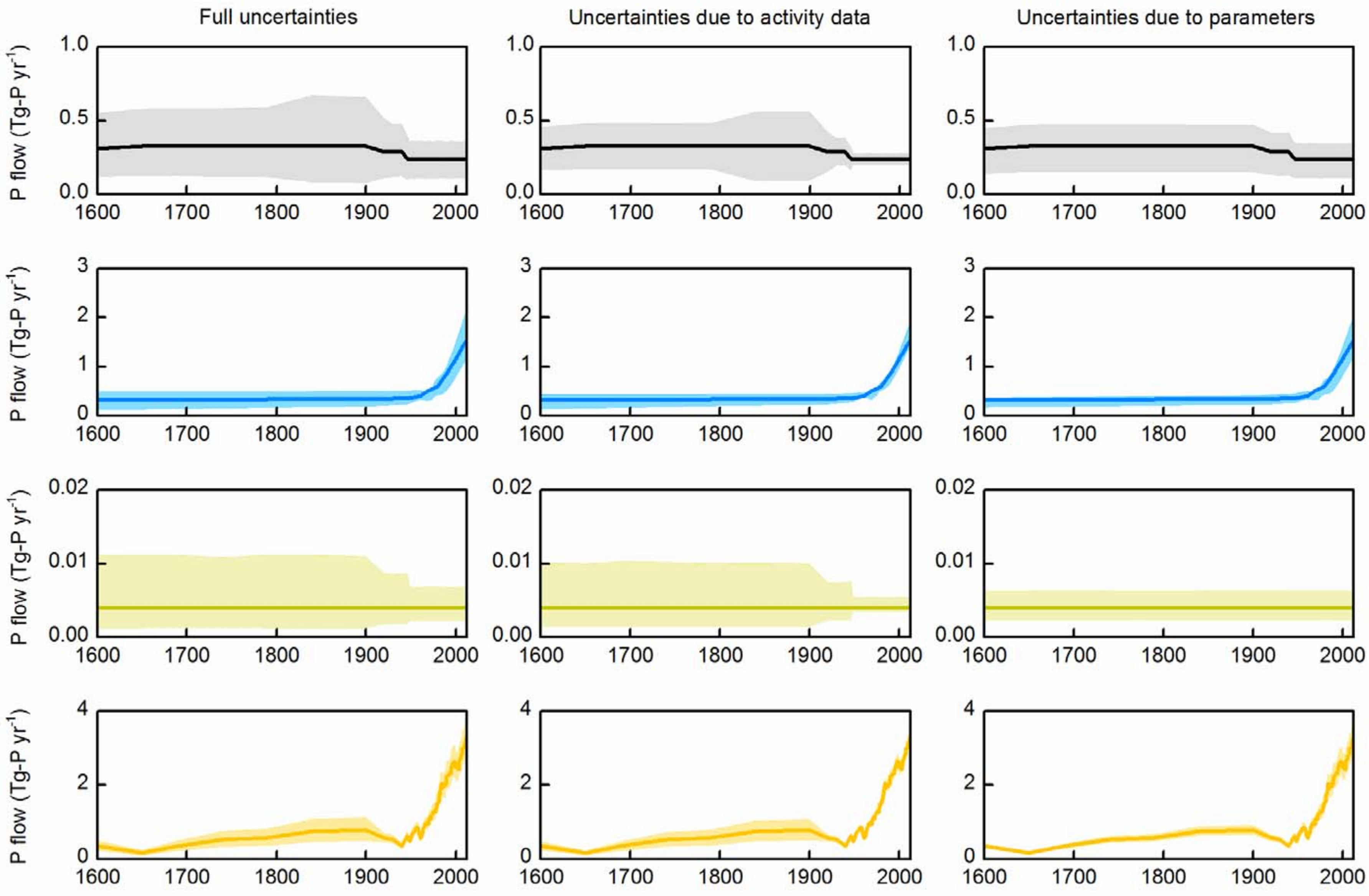
$$\begin{cases} P_1 = \mu \\ P_2 = \sqrt{\ln(\mathrm{CV}^2 + 1)} \end{cases}$$

对数正态分布

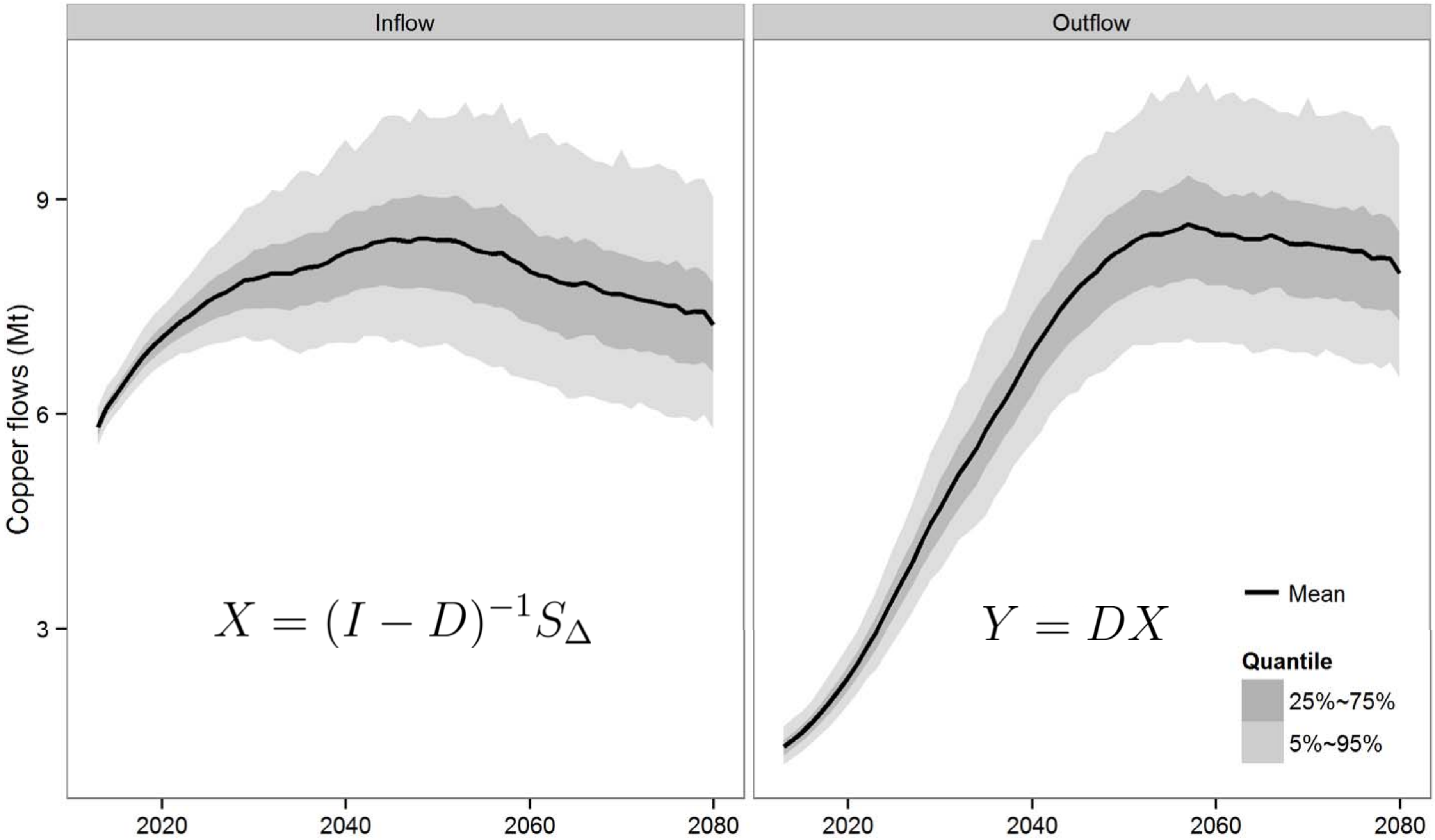
$$\begin{cases} P_1 = (1 - \sqrt{6} \cdot \mathrm{CV})\mu \\ P_2 = (1 + \sqrt{6} \cdot \mathrm{CV})\mu \end{cases}$$

对称三角分布

不确定性分析——中国磷流分析



不确定性分析——中国铜存量分析

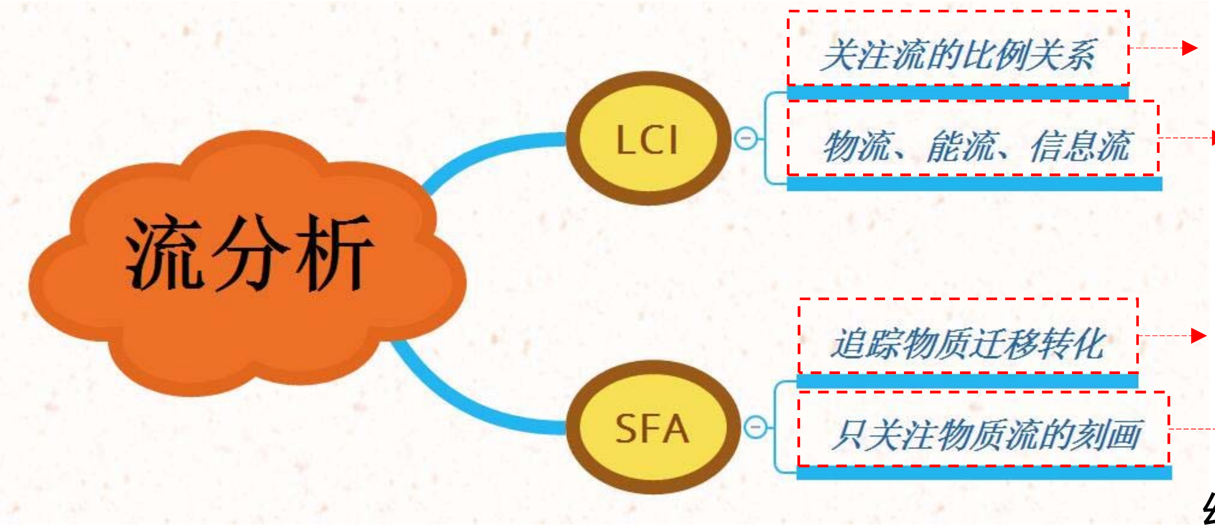


$$X = (I - D)^{-1} S_{\Delta}$$

$$Y = DX$$

结果可靠性

LCI与SFA 、排放清单的比较



得到单位功能单元的流量

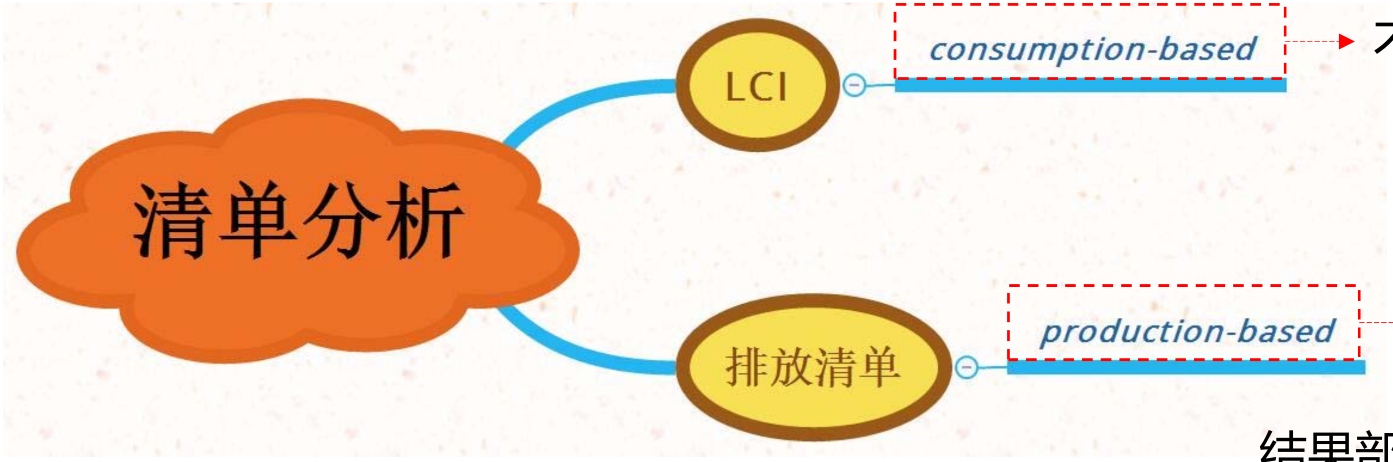
不统一量纲

结果无法直接验证

得到系统范围内的总流量

统一量纲

结果部分能被实测数据验证



不考虑排放点位的信息

结果无法直接验证

考虑排放点位信息

结果部分能被实测数据验证

从个案到整体

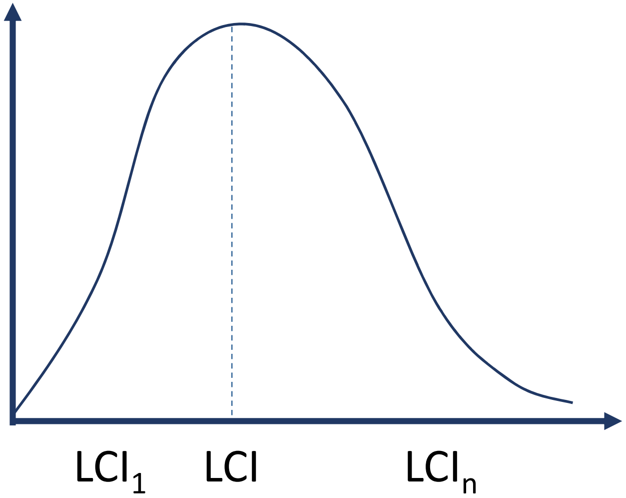
企业1LCI

企业2LCI

……

企业nLCI

行业LCI



从企业到行业：

- 行业总产能：

$$M=\sum_{i=1}^{n} M_i$$

单个企业的产能

- 行业LCI：

$$\mathrm{LCI}=\sum_{i=1}^{n}\frac{M_i}{M}\cdot \mathrm{LCI}_i$$

- 行业总的LCI：

单个企业的LCI

$$\mathrm{LCI}_T=M\cdot \mathrm{LCI}=\sum_{i=1}^{n} M_i\cdot \mathrm{LCI}_i$$

蛋糕的切法

条件期望：

$$\mathbb{E}\underbrace{(Z)}_{\text{LCI}} = \mathbb{E}[\mathbb{E}_X(Z\underbrace{|X)]}_{\text{企业}} = \mathbb{E}[\mathbb{E}_Y(Z\underbrace{|Y)]}_{\text{工艺}}$$

条件方差：

$$\mathrm{var}(Z) = \mathbb{E}[\mathrm{var}_X(Z|X)] + \mathrm{var}[\mathbb{E}_X(Z|X)]$$

- 行业总产能：$$M = \sum_{i=1}^{n} \underbrace{M_i}_{\text{单种工艺的产能}}$$
- 行业LCI：$$\mathrm{LCI} = \sum_{i=1}^{n} \underbrace{\frac{M_i}{M}}_{\text{单种工艺的占比}} \cdot \mathrm{LCI}_i$$
- 行业总的LCI：$$\mathrm{LCI}_T = M \cdot \mathrm{LCI} = \sum_{i=1}^{n} M_i \cdot \underbrace{\mathrm{LCI}_i}_{\text{单种工艺的LCI}}$$

空间化LCI

传统LCI：

$$\mathrm{LCI}_i = \sum_{j=1}^{n} \mathrm{SF}_j \cdot \mathrm{PI}_{ij}$$

→ 无空间信息

空间化LCI：

$$\mathrm{LCI}_i^{(k)} = \sum_{j=1}^{n} \mathrm{SF}_j \cdot \mathrm{PI}_{ij}^{(k)}$$

→ 空间信息

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

$$\left[\begin{array}{c|c} A_1 & A_2 \\ \hline A_3 & A_4 \end{array}\right] \qquad A = \textstyle\sum_{i,j} a_{ij},$$

$$(i,j) \in I(A_k), \quad A_k = \textstyle\sum_{i,j\in I(A_k)} a_{ij},$$

空间分配模型 →

$$B = B_1 + B_2 + B_3 + B_4$$

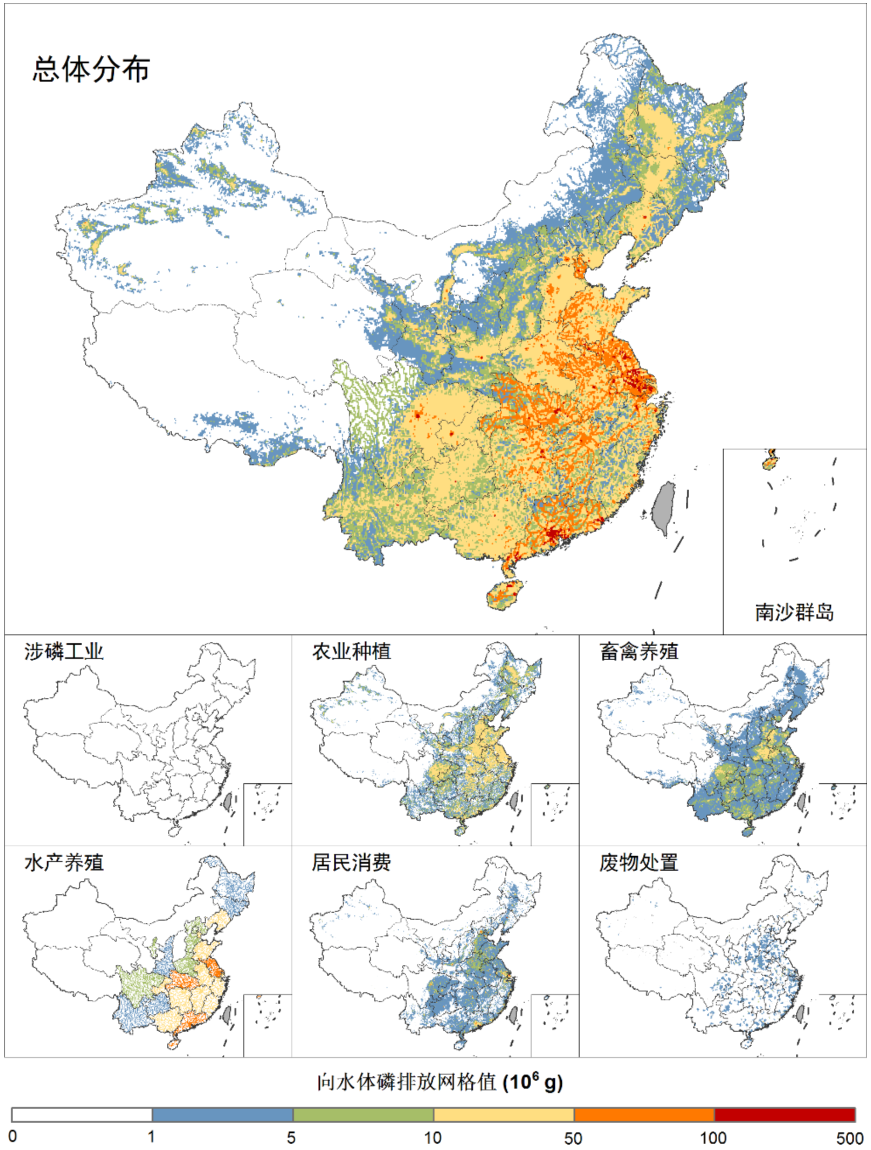
如果不知道B_1, B_2, B_3, B_4而只知道B

$$b_{ij} = B \cdot \frac{a_{ij}}{A}$$

如果知道B_1, B_2, B_3, B_4

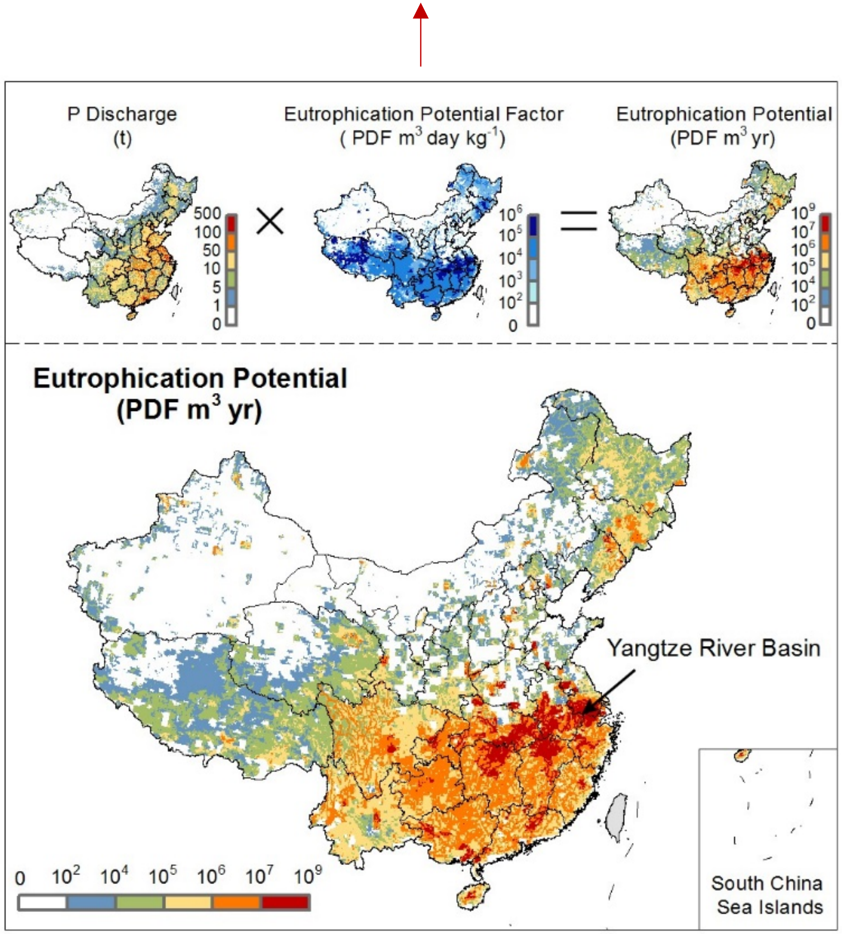
$$b_{ij} = B_k \cdot \frac{a_{ij}}{A_k} \qquad (i,j) \in I(A_k)$$

中国涉磷行业空间化LCI



→ 空间化的LCI

富营养化潜势评估



总结

1. 构建标准数据库，实现数据富集
2. 构建标准模型库，实现算法富集
3. 推行不确定性分析，评估结果准确性
4. 突破时空格局分析，促进结果验证

谢谢！

课题组网站：http://www.njumce.com/
微信公众号：Material_Cycles
个人邮箱：shenghu@nju.edu.cn
个人主页：https://ctfysh.github.io/