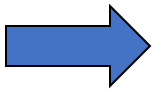
MR,QPLS,PEM,NN

MR: Easily found out miss input data.
Meaning parameter is clear.
Reverse design.

QPLS: Improve MR. But descriptors are tend to burst out.
(Two, three, more term interactions)
Effective selection of descriptors are needed.

PEM: Prediction ability is very high. No over-training.
Extrapolation ability.

NN: After check miss input data by MR, Check data type carefully (mountain type or something).
Descriptors are enough and minimum.
Then finally build by NN.

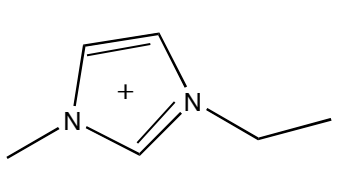


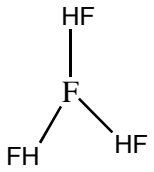
Need very careful usage.

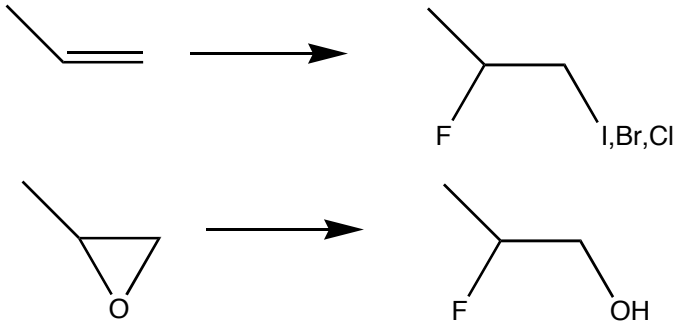
Ionic Liquids as Reaction Solvent

Journal of Fluorine Chemistry127 (2006) 29-35

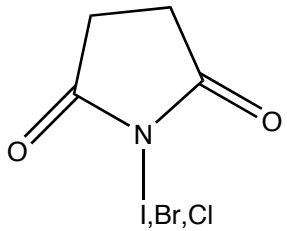
Ionic Liquid　EMI-F(HF)$_{2.3}$





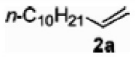
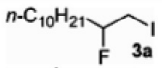
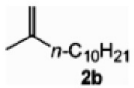
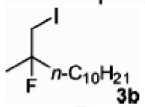
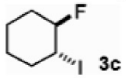
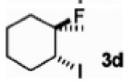
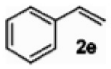
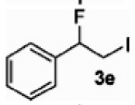
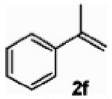
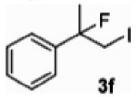
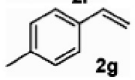
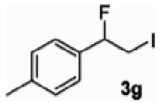


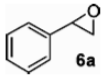
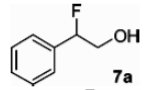
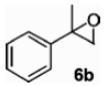
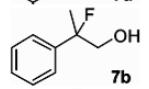
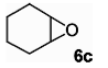
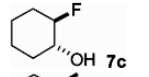
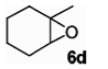
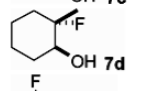
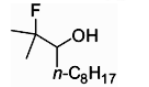
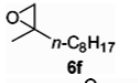
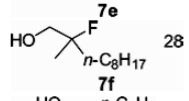
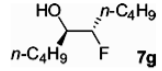
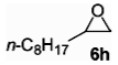
NIS:N-iodosuccinimide
NBS:Bromo
NCS:Chloro



CH2Cl2　Solvet

Yield of reaction

Entry	Alkene	Time (h)	Product	Yield (%)
1	 2a	3	 3a	90
2	 2b	1	 3b	91
3	 2c	1	 3c	95
4	 2d	1	 3d	98
5	 2e	1	 3e	70
6	 2f	1	 3f	88
7	 2g	1	 3g	78

Entry	Epoxide	Time (h)	Product	Yield (%)
1	 6a	16	 7a	81
2	 6b	16	 7b	74
3	 6c	16	 7c	55
4	 6d	6	 7d	65
5	 6e	24	 7e	37[a]
6	 6f	24	 7f 28	28[b]
7	 6g	48	 7g	7[c]
8	 6h	120	–	0[d]

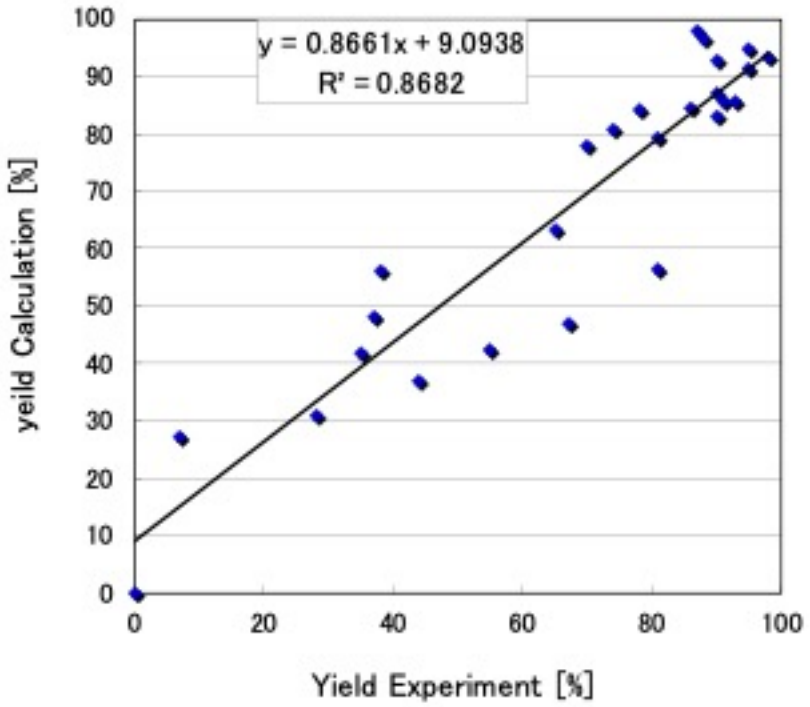
Analysis with Chemo-informatics

name	yield	time	eq	NIS	NBS	NCS	LUMO	Chr1-sui	Chr2-Firu
C10-3	90	3	2	1	0	0	1.156	-0.1663	-0.1379
isoC10	91	1	2	1	0	0	1.097	-0.171	-0.1146
cHex-1	95	1	2	1	0	0	1.17	-0.1488	-0.1488
MeC8EPO	28	24	0.1	0	0	0	2.254	-0.0073	0.0212
DiBuEPO	7	48	0.1	0	0	0	2.338	-0.0001	-0.0001
C8EPO	0	120	0.1	0	0	0	2.289	-0.0086	0.0005

Attach to Olefin

Attach to Epoxide

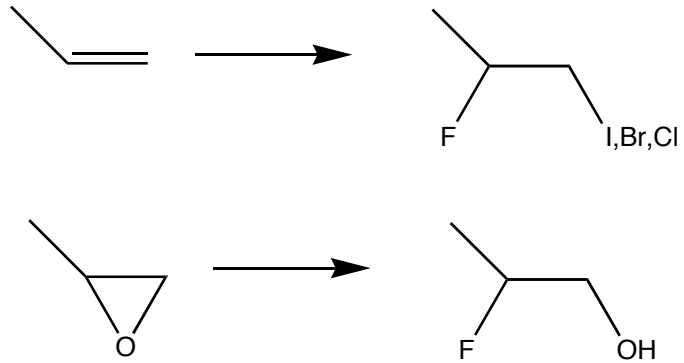
Fluorination of olefin and epoxide



name	yield	time	eq	NIS	NBS	NCS	LUMO	Chr1-sui	Chr2-Firu
C10-3	90	3	2	1	0	0	1.156	-0.1663	-0.1379
isoC10	91	1	2	1	0	0	1.097	-0.171	-0.1146
cHex-1	95	1	2	1	0	0	1.17	-0.1488	-0.1488
MeC8EPO	28	24	0.1	0	0	0	2.254	-0.0073	0.0212
DiBuEPO	7	48	0.1	0	0	0	2.338	-0.0001	-0.0001
C8EPO	0	120	0.1	0	0	0	2.289	-0.0086	0.0005

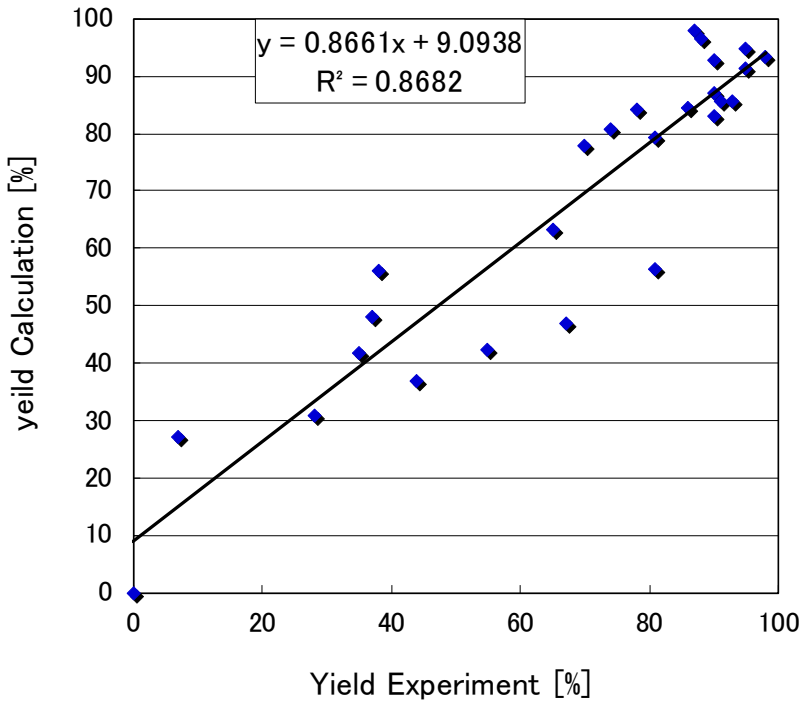
Attach to Olefin

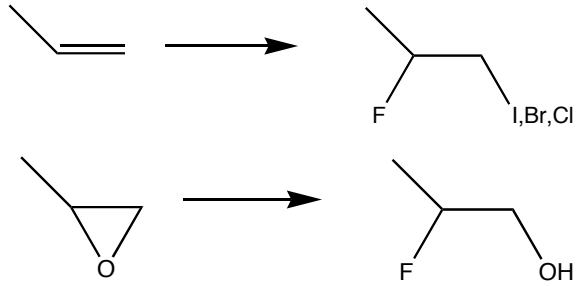
Attach to Epoxide



−179.253+100.19483*(
POWER((time*0.00000109+1),−1080.137)*
POWER((eq*1245.307+1),0.156)*
POWER((NIS*0.394+NBS*0.399+NCS*0.256+1),1.396)*
POWER((LUMO*8.190+1),0.00796)*
POWER((charge*−15.988+1),−0.458)*
POWER((charge(Crowd)*−10.285+1),−0.165))

Fluorination of olefin and epoxide





Although the charge on carbon of olefin and epoxy are very different, they can be treated in a unified manner.

Calculation of LUMO and charge with MOPAC.
We can also provide information on the yield of raw materials not found in the literature!

The reaction time has little effect on the yield rate.

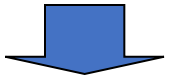
NCS is inferior in reaction.

−179.253+100.19483*(
POWER((Time*0.00000109+1),−1080.137)*
POWER((equivalent *1245.307+1),0.156)*
POWER((NIS*0.394+NBS*0.399+NCS*0.256+1),1.396)*
POWER((LUMO*8.190+1),0.00796)*
POWER((Charge Vacant*−15.988+1),−0.458)*
POWER((Charge crowded*−10.285+1),−0.165))

As a common sense chemist.

name	yield	time	eq	NIS	NBS	NCS	LUMO	Chr1-sui	Chr2-Firu
C10-3	90	3	2	1	0	0	1.156	-0.1663	-0.1379
isoC10	91	1	2	1	0	0	1.097	-0.171	-0.1146
cHex-1	95	1	2	1	0	0	1.17	-0.1488	-0.1488
MeC8EPO	28	24	0.1	0	0	0	2.254	-0.0073	0.0212
DiBuEPO	7	48	0.1	0	0	0	2.338	-0.0001	-0.0001
C8EPO	0	120	0.1	0	0	0	2.289	-0.0086	0.0005

0% yield after 120 hours of reaction



If we include this Exp. in the analysis,
we get the insane answer that
the yield does not depend on
reaction time.

Common Sense for chemist

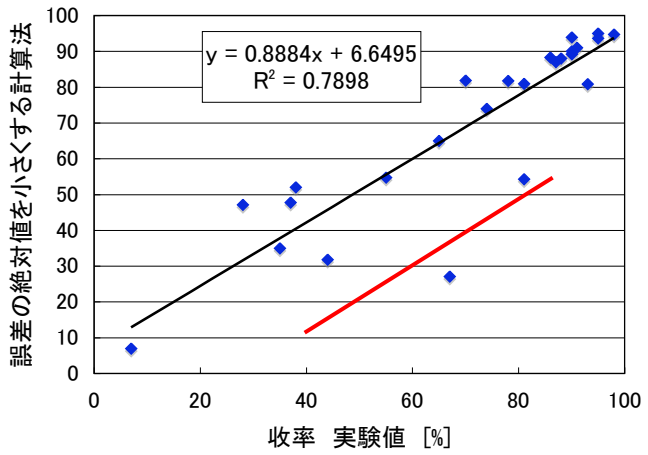
If 0% at 120hr, 0% at 1hr
(If it is a sequential reaction, it will be destroyed if the time is increased.)

There is no need for a yield of more than 100%.

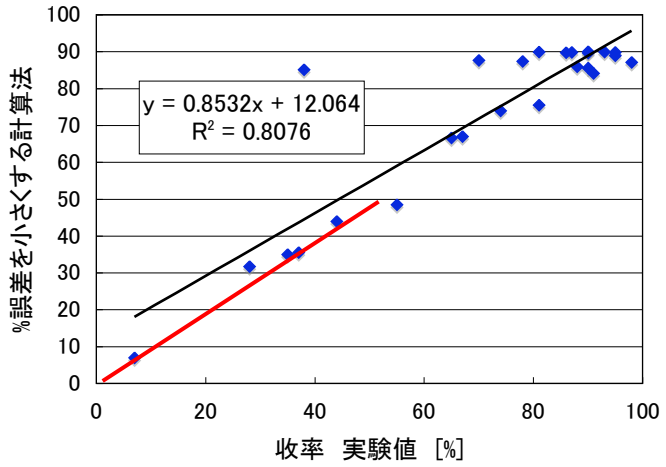
There are no byproducts below 0%.

You can't just look at the analysis results in R^2!
The chemo part of chemo-informatics is important

Interpreting the results by taking the error



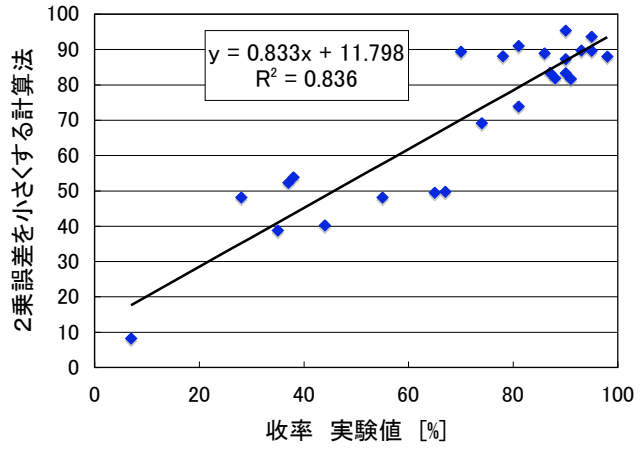
If we reduce the **absolute value of the error**, we can see what doesn't fit. 。



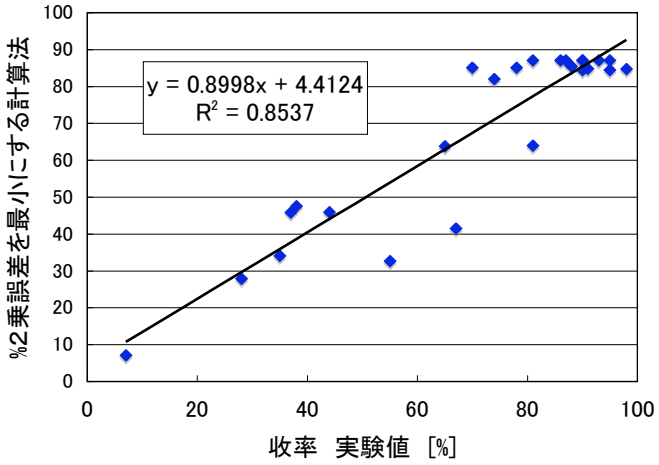
When the **absolute value of % error** is reduced, the closer to the origin, the better the fit.

If the analysis program is homemade...

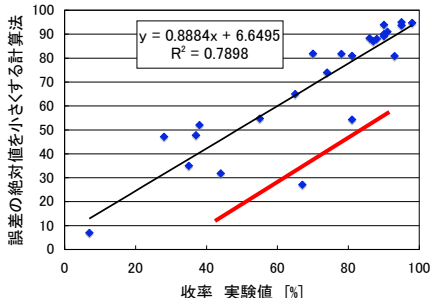
The smaller the **squared error**, the better the average fit.

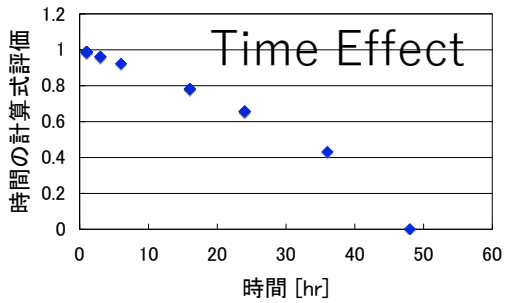


When the % squared error is reduced, it will fit on average from the origin.



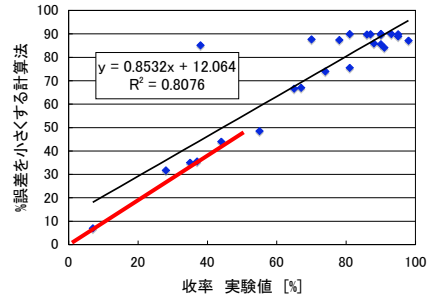
Which model tells the truth?

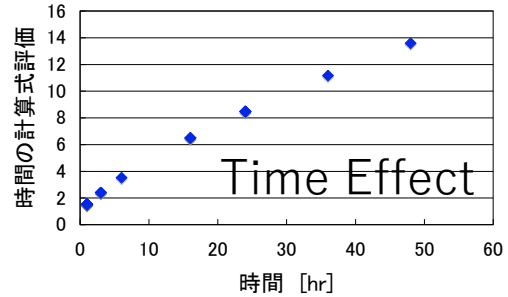




Minimum absolute error value

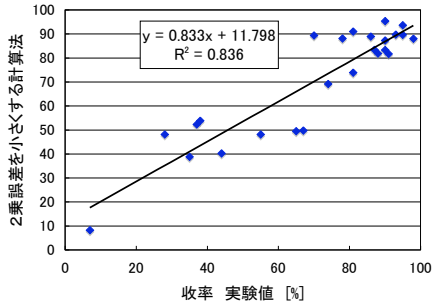
We consider it as a sequential reaction in which the product is broken down.
If you are trying to discuss something with a high yield rate in the upper right, this is what you want to do.

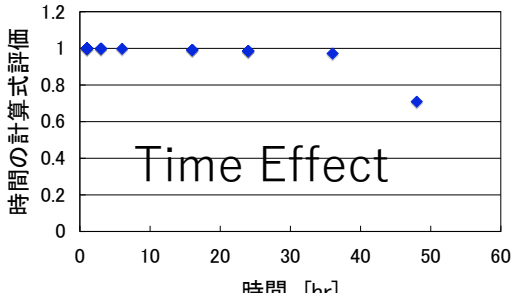




(% of absolute error) Minimum

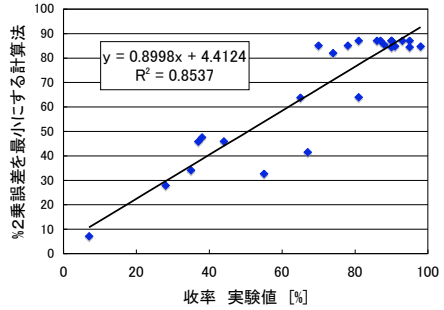
This model is the best for this case.
Easy to discuss what doesn't fit.

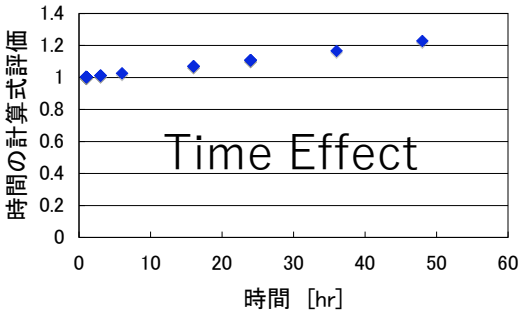




least-squares error

This is usually used for neural nets or whatever.
It's perfect for a paper.





(% of error) squared minimum

If you want to have an average discussion around the origin, this is the one.