Integration of Systems Engineering into Weather-Climate Model Optimization

Jiayu Zhou, Climate Mission, Office of Science and Technology Integration
David DeWitt, Climate Prediction Center, National Centers for Environmental Prediction

National Weather Service, National Oceanic and Atmospheric Administration, U.S. Department of Commerce

Introduction

In the meteorological model development, empirical tuning is often applied to improve performance. As tremendous efforts being made to press simulations closer to nature, climate-weather models are getting increasingly sophisticated (more physical, chemical and biological processes, higher horizontal and vertical resolutions, and complex interactions with added degrees of freedom). Because of resource constraints, it is impractical to conduct all tests to find the optimum configuration, causing progresses retarded helplessly. The urgency for model optimization has become another prominent issue in systems engineering since the Earth System Modeling Framework (ESMF) project launched to build a flexible software infrastructure to increase portability, interoperability, and code reuse.

The Orthogonal Array Test (OAT), a systems engineering approach of fractional factorial design, is widely used in industrial and agricultural production and proven to be effective to deal with multiple factors, levels and interactions with reliability and sensitivity analysis. It has been very successful in system configuration, parameter level selection and tolerance design *etc.* (SSTES 1975, Taguchi 1984)

In this introductory presentation for the meteorological community, the basic principles of OAT design are illustrated, followed by statistical analysis to determine dominant factors, significant interactions and percent contribution by individual component. Its ensemble capability to evaluate inherent variations and noises is also demonstrated. Finally, flexible designs to meet special application needs are briefly explored.

Fractional Factorial Design

- Factorial experiments - Experiments take on all possible combinations of levels across all factors.

  Difficulty: Prohibitively large combinations
- Fractional Factorial Design selects a limited number of experiments which produce the most information.

Orthogonal Array Testing

The philosophy of the OAT method is to design the product quality inspection into the production process, not to make it after the product being made. The OAT estimates the effects of control factors on the response mean and variation, making products robust that are insensitive to external environment.

I. Orthogonal array properties
- All levels are used in each column, in which each level occurs an equal number of times.
- Columns are mutually orthogonal.

II. Outstanding features
- Minimum number of experiments are conducted to optimize the model configuration
- Non-quantitative factors are enable
- Ensemble of repetition measures inherent variations

III. Analysis outcomes
- Dominant factors and significant interactions
- The percent contribution of each factor/interaction to the performance result.
- Expected outcome at the optimal configuration

OAT Methodology Illustration

1. **Experiment design**

A hypothetical case: *Model configuration optimization*

Factors: A. Physics 1 (*e.g.* cloud); B. Physics 2 (*e.g.* boundary layer);
C. Resolution; D. Initialization

Levels: A1 – param/pkg I; A2 – param/pkg II; B1 – param/pkg I; B2 – param/pkg II
C1 – higher; C2 – lower; D1 – scheme I; D2 – scheme II

Result measurement (y_i): Specific performance score

Orthogonal array $L_8(2^7)$ and factors assignment

Factor / Exp.	1 A	2 B	3 AXB	4 C	5	6	7 D	Perf. Score
1	1	1	1	1	1	1	1	y_1
2	1	1	1	2	2	2	2	y_2
3	1	2	2	1	1	2	2	y_3
4	1	2	2	2	2	1	1	y_4
5	2	1	2	1	2	1	2	y_5
6	2	1	2	2	1	2	1	y_6
7	2	2	1	1	2	2	1	y_7
8	2	2	1	2	1	1	2	y_8
I	I_1	I_2	I_3	I_4	I_5	I_6	I_7	Total $T=\sum_{i=1}^{8} y_i$
II	II_1	II_2	II_3	II_4	II_5	II_6	II_7	
I - II	I_1-II_1	I_2-II_2	I_3-II_3	I_4-II_4	I_5-II_5	I_6-II_6	I_7-II_7	
$(I - II)^2$	$(I_1 – II_1)^2$	$(I_2 – II_2)^2$	$(I_3 – II_3)^2$	$(I_4 – II_4)^2$	$(I_5 – II_5)^2$	$(I_6 – II_6)^2$	$(I_7 – II_7)^2$	
$\hat{w}$	$(I_1 – II_1)/8$	$(I_2 – II_2)/8$	$(I_3 – II_3)/8$	$(I_4 – II_4)/8$	$(I_5 – II_5)/8$	$(I_6 – II_6)/8$	$(I_7 – II_7)/8$	
S	$(I_1 – II_1)^2/8$	$(I_2 – II_2)^2/8$	$(I_3 – II_3)^2/8$	$(I_4 – II_4)^2/8$	$(I_5 – II_5)^2/8$	$(I_6 – II_6)^2/8$	$(I_7 – II_7)^2/8$	

$L_n(E^f)$ n – number of experiments, $n = (E - 1) \times f + 1$
f – maximum number of factors that can be accommodated
E – number of levels
I_i, II_i – summation of level 1 and level 2 results in column i, respectively
A x B Interaction effects of A and B
Columns 5 and 6 represent uncertainties.

2. **Analysis of variance (ANOV)**

Factor	S	df	S/df	F	Significance
A	S_1	E_1-1	S_1/df_A	$\frac{S_1/dfA}{S_e/dfe}$	Comparing with $F_\alpha(df_X, df_e)$: * at $\alpha = 0.05$significance level ** at $\alpha = 0.01$significance level Blank - Insignificant $ $S_e=S_5+S_6$, when AXB is significant. Otherwise, $Se=S_3+S_5+S_6$
B	S_2	E_2-1	S_2/df_B	$\frac{S_2/dfB}{S_e/dfe}$	
C	S_4	E_4-1	S_4/df_C	$\frac{S_4/dfC}{S_e/dfe}$	
D	S_7	E_7-1	S_7/df_D	$\frac{S_7/dfD}{S_e/dfe}$	
AXB	S_3	$(E_1-1)(E_2-1)$	S_3/df_{AXB}	$\frac{S_3/dfAxB}{S_e/dfe}$	
e	$ S_e	$n-1-(df_A+df_B+df_C+df_D+df_{AxB})$	S_e/df_e	*NA*	

e – reference of uncertainties; df – degree of freedom; $E_1= E_2= E_4= E_7 \equiv E$;

3. **Ensemble capability**

Embracing uncertainties

I. Perturbation of ICs/BCs.

II. Sensitiveness of factor parameters

Repeat each experiment *m* times with the uncertainties.

$y_i = \sum_{k=1}^{m} y_{i,k}/m$ (i=1, n)

The ANOV follows the same procedure described in previous section.

4. **Flexible applications**

The orthogonal array can be constructed to have as many schemes as possible with maximum number of factors with different levels for the smallest number of experimental runs, *e.g.* $L_8(2^7)$, $L_{16}(2^{15})$, $L_9(3^4)$, $L_{32}(4^9)$, $L_{25}(5^6)$ *etc.* (Bolboacă and Jäntschi 2007)

OAT design principals:
- Use an OAT array that has more rows than df required.
- Different factors/interactions can't be assigned to a same column.
- The interaction between two columns of $L_n(E^f)$ occupies E-1 columns indicated by the interaction table.

Meeting various application needs:
- ***Column merging*** : Assign factors having different levels in an OA simultaneously.
- ***Dummy levels***: Assign factors having less levels to OA of more levels
- ***Compounding factors***: Assign factors having more levels to OA of less levels
- ***Fractional addition***: Make additional tests with a few new levels for a factor found having some kind of trend to influence the performance result.
- ***Dividing zones***: Repeat costly experiments less times than inexpensive ones.

Prospects

Orthogonal Array Test technique selects a set of test cases from a universe of tests and makes testing efficient and effective, having advantages of multiformity, parallelity and synthetic comparability. The optimum configuration resulted from OAT is the best combination among not only the test conditions but also all conditions of possible combinations in a given case.

Beside promoting model improvement, OAT has a lot of potential for meteorological applications, such as assessing the dependence of satellite retrieved atmospheric profiles on physical and statistical parameters of the data assimilation system, and transforming model output into sensible climate/weather parameters, for example.

In practice, professional knowledge in understanding the fundamental processes inherent in the system of investigation is helpful to make experiments more efficient, *e.g.* knowing some interactions nonexistent could considerably decrease the level of effort.

References

Bolboacă, S. and L. Jäntschi, 2007: Design of experiments: useful orthogonal array for number of experiments from 4 to 16. *entropy*, **9**, 198-232.

SSTES, 1975: *Orthogonal Array Test Design – A Multifactor Test Method* (in Chinese). Shanghai People's Press, 229 pp.

Taguchi, G., 1984: Introduction to quality engineering-Taguchi methods. American Supplier Institute, Inc., Romulus, MI.