Statistical Design of Experiments in Numerical Weather Prediction: Emerging Results

Jeffrey A. Smith, Ph.D.
U.S. Army Research Laboratory
RDRL-CIE-M
WSMR, NM 88002-5501
jeffrey.a.smith1.civ@mail.mil

Richard S. Penc, Ph.D.
John W. Raby

UNCLASSIFIED / APPROVED FOR PUBLIC RELEASE

Problem Space

Background

- Statistical design of experiments (DoE) is supported by a body of literature extending back over 80 years, ranging from the pioneering work of Fisher (1935), to Box et al. (1978); Box and Draper (1987), and Montgomery (2013).
- Recent work expanded the use of DoE to high dimension computer codes, e.g., McKay et al. (1979); Sacks et al. (1989a); Sacks et al. (1989b); Santner et al. (2003); and Kleijnen (2015).
- DoE techniques have been successfully applied to computer simulations ranging from high dimensional force-on-force simulations (Sanchez et al. 2012) to computational fluid dynamics codes to study pollutant dispersion (Rahimi et al. 2014), and conduct design optimization (Berci et al. 2014; Zhu et al. 2015).
- Computational issues such as run time, high dimensional input spaces, and the modeling resolutions required to support Army tactical operations limit the effective number of samples we can make of a given NWP code.

Objectives

- Reduce the number of simulation runs required to efficiently explore a simulation output space.
- Quantify how parameterizations influence the atmospheric simulation to produce a forecast.
- Incorporate modeling parameters such as observation nudging weight or nesting ratios into an experimental design.

Challenges

- Incorporating and accounting for the variability of large scale (synoptic) weather features in experimental designs.
- Addressing the range of potential factors which can be purely numerical to those which may be categorical or ordinal as well.
- Creating experimental designs that allow us to extract the maximum information from a given, limited set of model runs.

Experiment Design

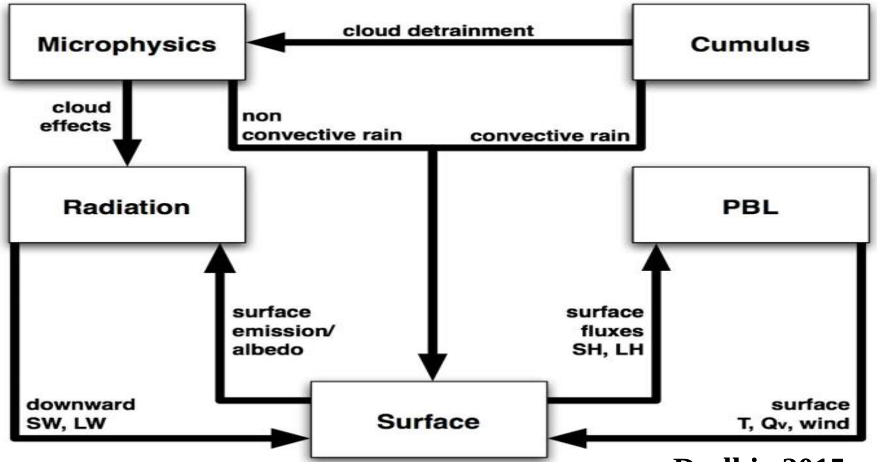
Theory

- Mathematically, a forecast is a mapping from a set of input conditions to some future set of conditions:

$$f : x \rightarrow y$$

- x is not only the initialization and observational data, but the model configuration data as well.
- y is the model output distributed in space.
- f represents the interaction of the solver core with the physical parameterizations.

Direct Interaction of Parameterizations



source: Dudhia 2015

Method

- Hold all inputs at nominal values save parameterizations:
  - Treat PBL/Surface Layer as a single factor (1x).
  - Cumulus, Microphysics, Shortwave and Long Wave Radiation schemes each as a factor (4x).
- Create a design that adequately explores the output space with a “few,” well chosen runs.

Model and Domains

Version: WRF 3.8.1 (Skamarock et al., 2008)

Initialization

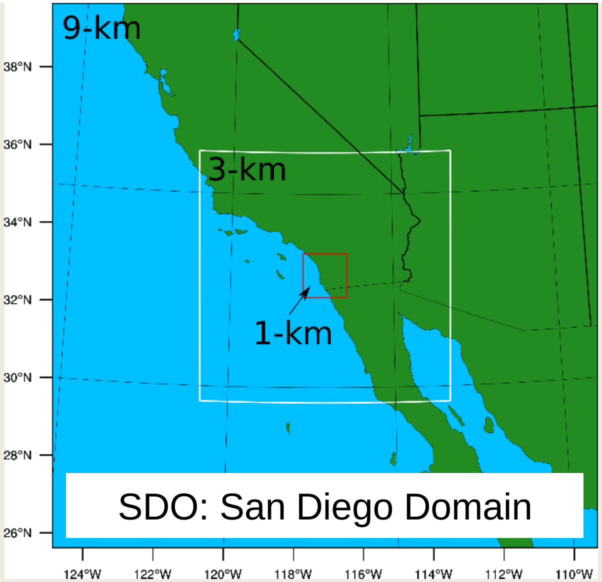
- Initial and boundary conditions from 0.5-degree GFS with observations analyzed onto initial conditions.
- 1/12 degree (~9 km) RTG SST.
- 1 km NOHRSC SNODAS snow where available (GFS snow elsewhere).

Data Assimilation

- 6-h pre-forecast with observation nudging (12-18 UTC) Observation nudging uses TAMDAR aircraft data and various MADIS datasets [standard surface observations, mesonet surface observations, maritime surface observations, profiler data, rawinsondes, and ACARS (aircraft) data].
- 18-h forecast (18-12 UTC).

Parametrization: Covered on a subsequent slide.

SFO: San Francisco Domain



SDO: San Diego Domain

U.S. ARMY RDECOM® ARL

Cases[1, 2]

Case	Dates (2012)	San Francisco (SFO) Domain	San Diego (SDO) Domain
1	Feb. 07–08	An upper level trough with associated frontal system moved onshore which led to widespread precipitation in the region that diminished mid-period.	Surface front / upper level trough moved onshore, which led to widespread precipitation in the region.
2	Feb. 09–10	Quiescent weather dominated the region with an upper level ridge remaining centered over central California	Quiescent weather was in place with a upper level ridge ridge centered over central California at 12 UTC.
3	Feb. 16–17	An upper level ridge located over northern California in combination with a surface high pressure area centered over the Rocky Mountains east of the domain produced quiescent weather in the region.	An upper-level low located near the California/Arizona border with Mexico at 12 UTC brought precipitation to that portion of the domain. This pattern moved south and east over the course of the day.
4	Mar. 01–02	A weak shortwave upper level trough with associated cold front resulted in considerable cloudiness and light precipitation over the region until after mid-period when conditions stabilized following frontal passage.	A weak shortwave trough resulted in precipitation in northern California at the beginning of the period that spread to Nevada, then moved southward and decreased in coverage.
5	Mar. 05–06	Weak surface pressure gradients at the surface and broad zonal flow aloft slowly gave way to stronger synoptic forcing in advance of a cold front that approached the region near the end of the period bringing increased cloudiness, but very limited precipitation.	Widespread high-level cloudiness due to weak upper-level low pressure but very limited precipitation.

1: Synoptic conditions for the case study days considered.
2: All case studies are 24 hours in length, running from 12 UTC to 12 UTC on the days listed with forecasts made on the hour.

Parameterization Space[1]

Planetary Bound. Layer, Surface (PBL, SL)
- 1, 1: YSU with revised MM5
- 2, 2: MYJ with ETA
- 5, 1: MYNN2 with revised MM5
- 7, 7: ACM2 with Pleim-Xu
- 11, 1: Shin-Hong with revised MM5

Cumulus (CU)[2]
- 1: KainFritsch (KF)
- 2: Betts-Miller-Janjic (BMJ)
- 6: Tiedke
- 16: New Tiedke
- 93: Grell-Devenyi

Microphysics (Micro)
- 2: Lin (Purdue)
- 4: WSM5
- 5: ETA (Ferrier)
- 7: Goddard
- 8: Thompson

Short Wave (RaSW)
- 1: Dudhia
- 2: Goddard
- 4: RRTMG
- 7: FLG
- 99: GFDL

Long Wave (RaLW)
- 1: RRTM
- 4: RRTMG
- 5: New Goddard
- 7: FLG[3]
- 99: GFDL

Land Surface Model (LSM)
- 1: 5 layer Thermal Diffusion
- 2: NOAH
- 3: RUC operational
- 5: CLMv4

1: For specific references for the various physics schemes please refer to Skamarock et al. (2008).
2: Cumulus scheme applied to the outer domain only.
3: Every run with the FLG long wave radiation scheme failed, but not every failed run used the FLG scheme We are investigating replacement schemes for these failed points in order to come closer to our desired 40 runs.

Initial Design

Blocking

- Allows us to control for “nuisance” factors, or sources of variation that are present but which we cannot control.

Balance

- Ensures that no one set of runs or factor combinations dominate the output space.

Orthogonality

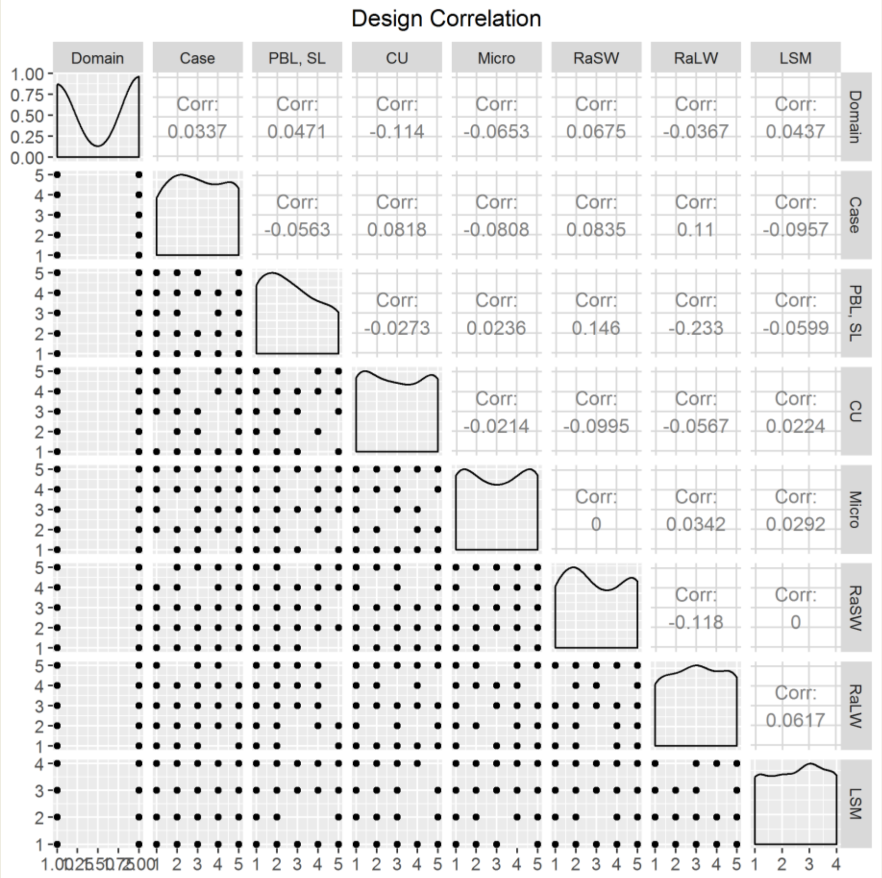
- Allows us to potentially infer factor effects.

Approach

- Apply the Mixed Integer Linear Programming approach developed at the Naval Post Graduate School, coupled with blocking to create the design.

Follows

- Clinical Paper at the Conference on Applied Statistics in Defense (2015).
- Poster paper at last years Annual AMS meeting.



Sample of the Design Matrix

First 10 runs of the initial 40 run design.

Run[1]	Domain	Case	PBL, SL	CU[2]	Micro	RaSW	RaLW	LSM	Success
1.1	SFO	7-Feb	7,7	16	5	2	7	1	No
2.1	SFO	9-Feb	5,1	16	2	1	5	5	Yes
3.1	SDO	7-Feb	11,1	1	7	2	1	2	Yes
4	SFO	7-Feb	5,1	1	8	4	5	3	Yes
5.1	SDO	5-Mar	11,1	16	5	2	1	5	Yes
6	SDO	9-Feb	11,1	93	2	99	4	2	Yes
7.1	SFO	16-Feb	11,1	1	7	99	4	3	Yes
8	SFO	5-Mar	1,1	93	8	1	1	1	Yes
9	SFO	16-Feb	1,1	2	2	99	1	3	Yes
10.1	SFO	9-Feb	1,1	16	4	4	7	5	No

1: Runs with a decimal digit were re-runs with either a change in the cumulus or land surface model package. All such changes were done as a uniform swap for the initial design point and done to address some noted execution failures.
2: Outer domain only.

Analysis Sketch

Basic Approach

- The data we plan to use for analysis will be post processed WRF output data generated by MET Point Stat. This will be matched pair (MPR) data.
- We will augment the data using the design matrix we developed. This will tie every data point in the MPR file with a specific run configuration used to generate those points.
- After using exploratory tools such as scatter and box plots conditioned on the various parameterizations, we will conduct a generalized least squares regression analysis to produce a meta model describing the mean bias as a function of the blocking factors and the parameterization:

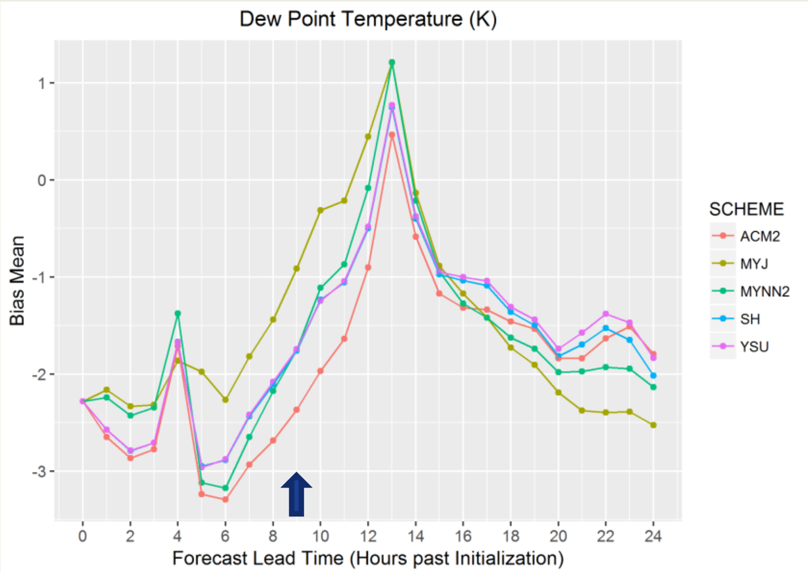
$$\mu_i = \beta^T x_i$$

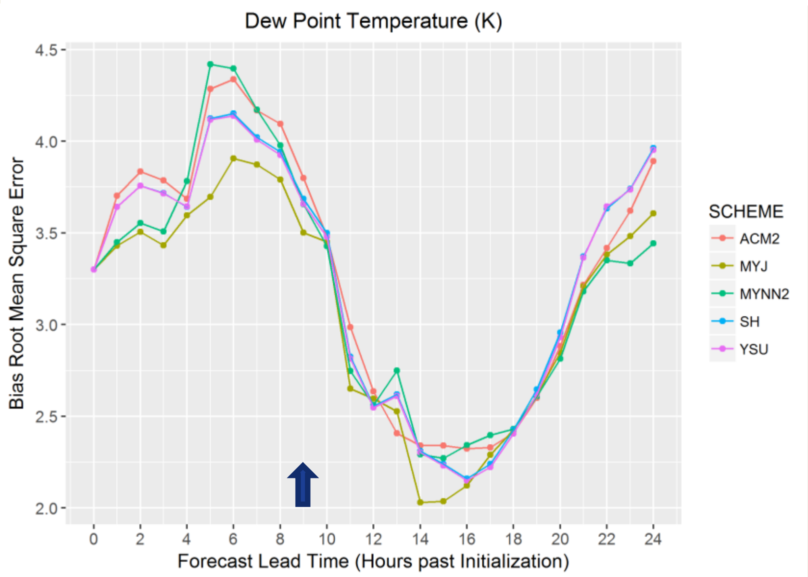
where x_i represents the design matrix elements for the $i'th$ run augmented with 1 to account for the constant term.

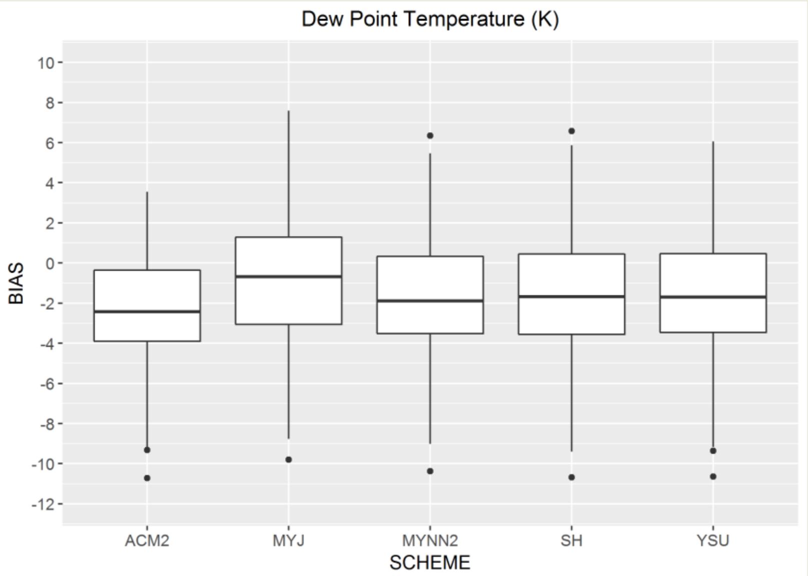
Why?

- A generalized linear model allows us to assign weights to each of the considered factors.
- Those weights that are statistically significant for a specific process, e.g., microphysics, indicate that that process is a significant contributor to the mean bias error.
- Thus, we have some macroscopic look into how the model is driving the error over the domain.

Z2 Dew Points (189 Points) at 21Z







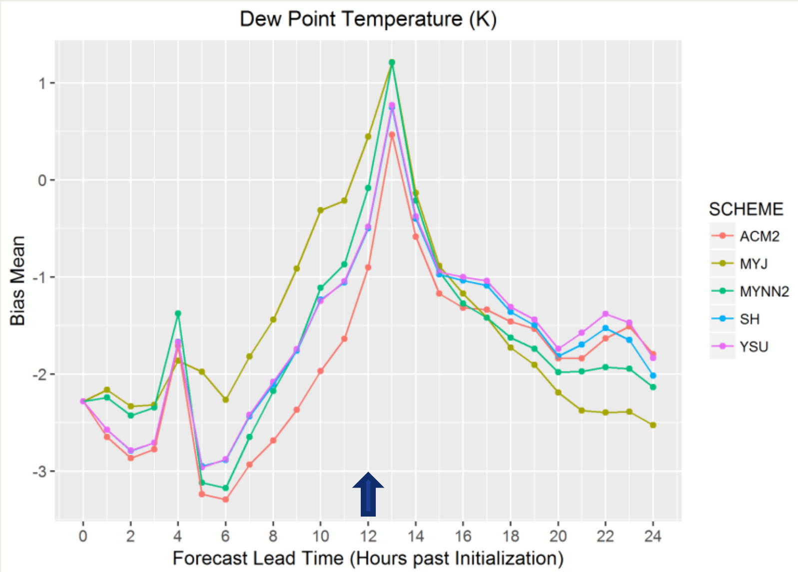
Tukey's Honest Significant Difference (0.99)

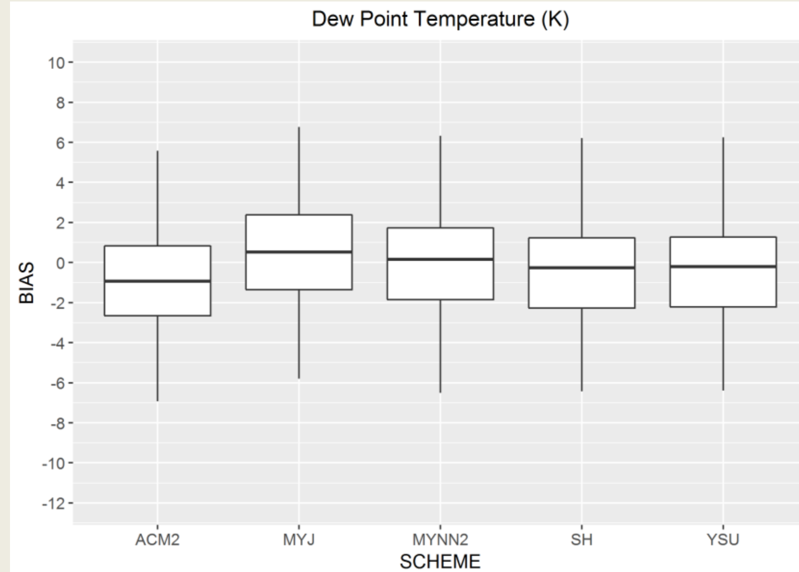
comparison	estimate	conf. low	conf.high	adj.p.value
MYJ-ACM2	1.45616	0.37689	2.53544	0.00012
MYNN2-ACM2	0.61841	-0.46087	1.69768	0.33418
SH-ACM2	0.60677	-0.47250	1.68604	0.35391
YSU-ACM2	0.62322	-0.45605	1.70250	0.32618
MYNN2-MYJ	-0.83776	-1.91703	0.24152	0.08416
SH-MYJ	-0.84940	-1.92867	0.22988	0.07703
YSU-MYJ	-0.83294	-1.91221	0.24633	0.08726
SH-MYNN2	-0.01164	-1.09091	1.06763	1.00000
YSU-MYNN2	0.00482	-1.07446	1.08409	1.00000
YSU-SH	0.01646	-1.06282	1.09573	1.00000

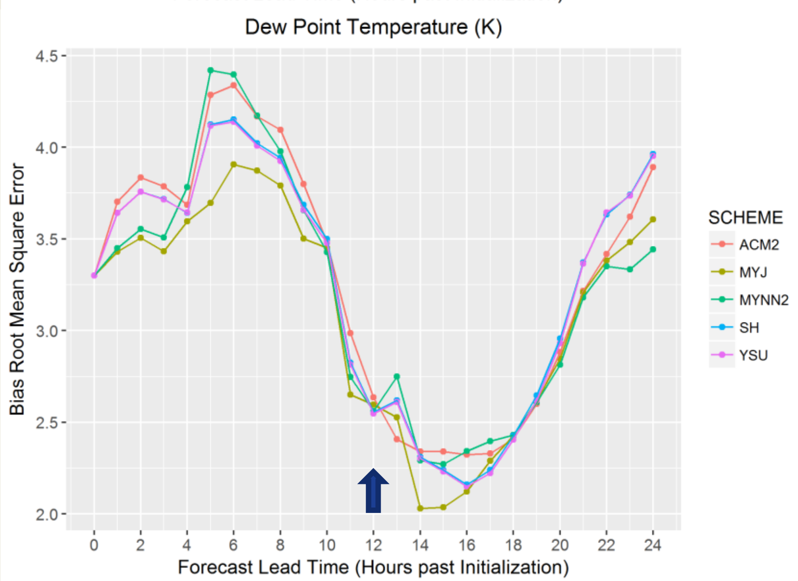
U.S. ARMY RDECOM

Z2 Dew Points (200 Points) at 24Z

ARL







Tukey's Honest Significant Difference (0.99)

comparison	estimate	conf. low	conf.high	adj.p.value
MYJ-ACM2	1.34775	0.52318	2.17231	0.00000
MYNN2-ACM2	0.81773	-0.00683	1.64229	0.01094
SH-ACM2	0.40435	-0.42021	1.22891	0.49744
YSU-ACM2	0.41977	-0.40479	1.24434	0.45849
MYNN2-MYJ	-0.53002	-1.35458	0.29454	0.22176
SH-MYJ	-0.94340	-1.76796	-0.11883	0.00186
YSU-MYJ	-0.92797	-1.75254	-0.10341	0.00234
SH-MYNN2	-0.41338	-1.23794	0.41118	0.47454
YSU-MYNN2	-0.39795	-1.22252	0.42661	0.51380
YSU-SH	0.01543	-0.80914	0.83999	1.00000

Observations, Summary and Next Steps

Observations

- Creating the design, managing the run configurations, and management of the data is a challenge; some of which presented bigger hurdles than we expected at the outset.
- Out of the original 40 runs, approximately 50% of the runs crashed at the outset. We adapted the design to account for "bad" configurations and recovered approximately 8 runs.
- With 28 runs, we are set up to consider the blocking factors (Domain and Case) plus 5 of the 7 remaining factors.
- Other work suggests that hardware and software considerations can be a nuisance factor in the execution of the experiment. We need to incorporate indicators for this if we split runs across various machines and environments.

Summary

- We do have some preliminary evidence that a design of experiments approach can lead to deeper analysis; however, we have not yet completed this analysis.

Next Steps

- Attempt to recover a few more runs. If successful, we can consider 6 (or more) factors.
- A number of the parameterization combinations we have explored have not been documented in the literature, especially those where our runs failed; document these.

Acknowledgements

The authors wish to thank:

- Dr. Brian Reen, ARL
- Mr. Bob Dumais, ARL
- Ms. Leelinda Dawson, ARL.

- Dr. Reen and Mr. Dumais contributed a wealth of experience with WRF as well as several support tools used in the production of data.

- Ms. Dawson wrote a number of python scripts to automate post processing of that data produced by WRF.

References

1. Bathke, A., 2004: The ANOVA F test can still be used in some balanced designs with unequal variances and nonnormal data. *Journal of Statistical Planning and Inference*, **126,** 413-422.
2. Berci, M., V. V. Toropov, R. W. Hewson, and P. H. Gaskell, 2014: Multidisciplinary multifidelity optimisation of a flexible wing aerofoil with reference to a small UAV. *Structural and Multidisciplinary Optimization*, **50,** 683-699. doi: 10.1007/s00158-014-1066-2.
3. Box, G. E. P., and N. R. Draper, 1987: *Empirical model-building and response surfaces.* Wiley.
4. Box, G. E. P., W. G. Hunter, and J. S. Hunter, 1978: *Statistics for experimenters: An introduction to design, data analysis, and model building.* Wiley.
5. Dudhia, J., 2015: Overview of WRF Physics. *2015 Basic WRF Tutorial*, National Center for Atmospheric Research.
6. Fisher, R. A., 1935: *The design of experiments.* Oliver and Boyde.
7. Kleijnen, J. P. C., 2015: *Design and analysis of simulation experiments.* 2nd ed. Springer.
8. McKay, M. D., R. J. Beckman, and W. J. Conover, 1979: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, **21,** 239-245. doi: 10.2307/1268522.
9. Montgomery, D. C., 2013: *Design and analysis of experiments.* 8th ed. Wiley.
10. National Center for Atmospheric Research, 2016: Model Evaluation Tools Version 5.2 (METv5.2). User's Guide 5.2. [Available online at http://www.dtcenter.org/met/users/.]
11. R Core Team, 2017: R: A language and environment for statistical computing. [Available online at https://www.R-project.org/.]
12. Rahimi, A., T. Tavakoli, and S. Zahiri, 2014: Computational Fluid Dynamics (CFD) modeling of gaseous pollutants dispersion in low wind speed condition: Isfahan Refinery, a case study. *Petroleum Science and Technology*, **32,** 1318-1326. doi: 10.1080/10916466.2011.653701.
13. Sacks, J., S. B. Schiller, and W. J. Welch, 1989a: Designs for computer experiments. *Technometrics*, **31,** 41-47. doi: 10.2307/1270363.
14. Sacks, J., W. J. Welch, T. J. Mitchell, and H. P. Wynn, 1989b: Design and analysis of computer experiments (includes comments and rejoinder). *Statistical Science*, **4,** 409-435. doi: 10.1214/ss/1177012413.
15. Sanchez, S. M., T. W. Lucas, P. J. Sanchez, C. J. Nannini, and H. Wan, 2012: Designs for large-scale simulation experiments, with applications to defense and homeland security. *Design and Analysis of Experiments*, K. Hinkelmann, Ed., John Wiley & Sons, Inc., 413-441. doi: 10.1002/9781118147634.ch12.
16. Santner, T. J., B. J. Williams, and W. I. Notz, 2003: *The design and analysis of computer experiments.* Springer-Verlag.
17. Schloerke, B., and Coauthors, 2017: GGally: Extension to 'ggplot2'. R package version 1.3.2. [Available online at https://CRAN.R-project.org/package=GGally.]
18. Skamarock, W. C., and Coauthors, 2008: A description of the advanced research WRF version 3. NCAR Technical Note NCAR/TN-475+STR.
19. Smith, J. A., and R. S. Penc, 2016: A design of experiments approach to evaluating parameterization schemes for numerical weather prediction: Problem definition and proposed solution approach. *Joint Statistical Meetings Proceedings, Section on Statistics in Defense and National Security, Conference on Applied Statistics in Defense 2015***,** 4183-4192.
20. Smith, J. A., R. Penc, and J. W. Raby, 2017: Is there a role for statistical design of experiments in numerical weather prediction? *97th Annual AMS Meeting, Joint with 28th Conference on Weather Analysis and Forecasting / 24th Conference on Numerical Weather Prediction*, Poster 616.
21. Vieira, H., Jr., S. Sanchez, K. H. Kienitz, and M. C. N. Belderrain, 2011: Generating and improving orthogonal designs by using mixed integer programming. *European Journal of Operational Research*, **215,** 629-638. doi: 10.1016/j.ejor.2011.07.005.
22. Vieira, H., Jr., S. M. Sanchez, K. H. Kienitz, and M. C. N. Belderrain, 2013: Efficient, nearly orthogonal-and-balanced, mixed designs: an effective way to conduct trade-off analyses via simulation. *Journal of Simulation*, **7,** 264-275. doi: 10.1057/jos.2013.14.
23. Wickham, H, 2017.: tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1. [Available online at https://CRAN.R-project.org/package=tidyverse.]
24. Zhu, B., X. Wang, L. Tan, D. Zhou, Y. Zhao, and S. Cao, 2015: Optimization design of a reversible pump-turbine runner with high efficiency and stability. *Renewable Energy*, **81,** 366-376. doi: 10.1016/j.renene.2015.03.050.