U.S. ARMY RESEARCH, DEVELOPMENT AND ENGINEERING COMMAND

Some Conclusions on Applying Statistical Design of Experiments to Numerical Weather Prediction (Paper TJ17.4)

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
Judah L. Cleveland

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.

Goal

- Provide warfighter a robust forecast ability to run with minimal intervention by “man-in-the-loop” right out of box

Objectives

- Reduce number of simulation runs required to efficiently explore a simulation output space.
- Quantify how parameterizations influence the atmospheric simulation to produce a forecast.
- Incorporate other factors, e.g., observation nudging weight or nesting ratios.

Challenges

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

Experiment Design

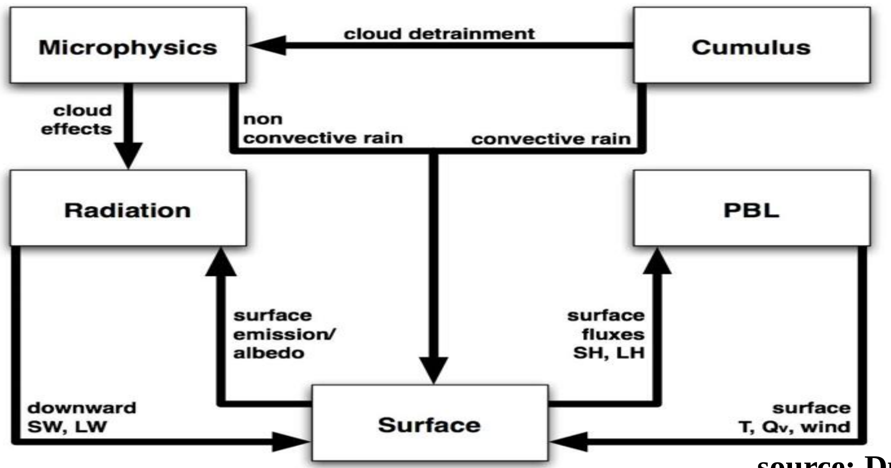
Theory

- Mathematically, a forecast is a mapping from a set of input conditions to some future set of conditions.
- We view the solver core as set f that represents all the supported physical parameterizations and configuration data.
- Mathematically, a forecast is a mapping from a set of input conditions to some future set of conditions:

$$f_i : x \rightarrow y$$

- x is initialization data and observational data used for data assimilation, and y is the model output distributed in space.
- Note: A *run* applies *treatment* $f_i \in f_i$ to x to produce y; the set of $f_i, i = 1, \dots, n$ is termed a n row *design matrix*.
- More details can be found in Smith et al (2018a, b).

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).

Parametrizations

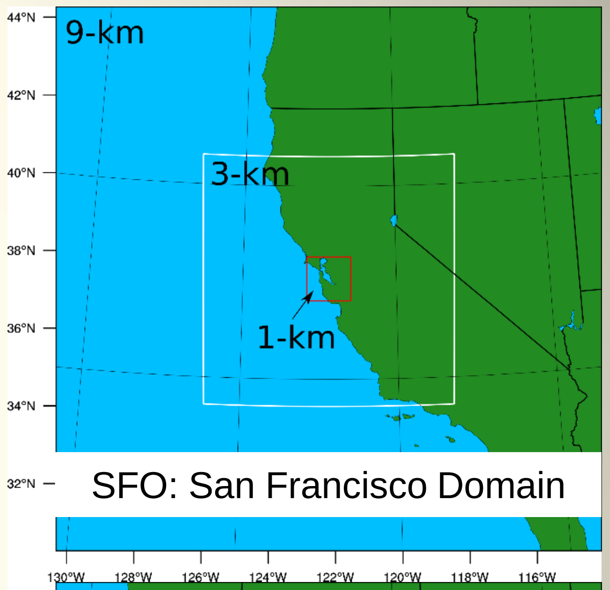
- Covered on the next slide.

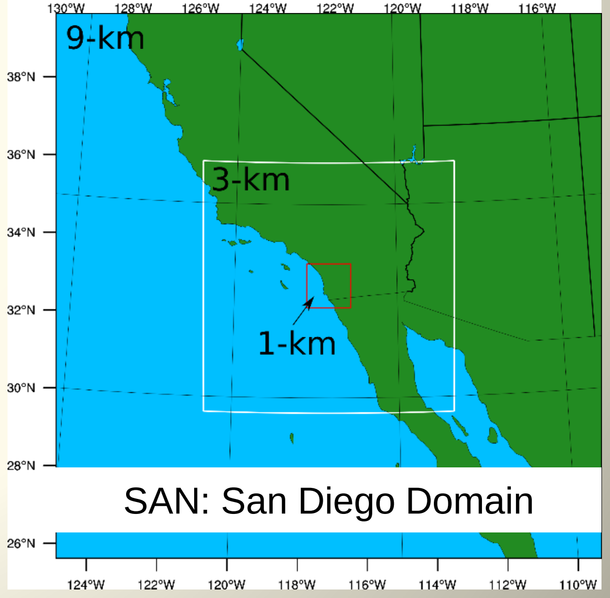
Data Assimilation

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

Forecast

- 18-h forecast (18-12 UTC)





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[4]
- 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 (*Not considered in this talk*).
3: Every run with the FLG long wave radiation scheme failed, but not every failed run used the FLG scheme.
4: Short wave FLG radiation scheme was not considered for this analysis.

Cases[1,2]

Case	Dates (2012)	San Francisco (SFO) Domain	San Diego (SAN) 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 an upper level 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. The dates and case numbers are used as proxies for the synoptic state of the atmosphere.
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.

Analysis Sketch

Basic Approach

- Use matched pair (MPR) data, post processed from WRF output data by MET Point-Stat.
- Augment the data using the design matrix; every data point in the MPR file is tied to a specific run configuration.
- Employ generalized least squares regression analysis to produce a meta model describing the bias as a function of the blocking factors (domain and synoptic state) and the parameterization:

$$y_{ijkl} = D_i + C_j + T_k + \varepsilon_{ijkl}$$

where y_{ijkl} is the bias at a particular station, D_i the effect due to the Domain location, C_j the effect due to Case Day, T_k represents the particular treatment (a combination of parameterization schemes), and ε_{ijkl} represents a residual error.

Why Use This Approach?

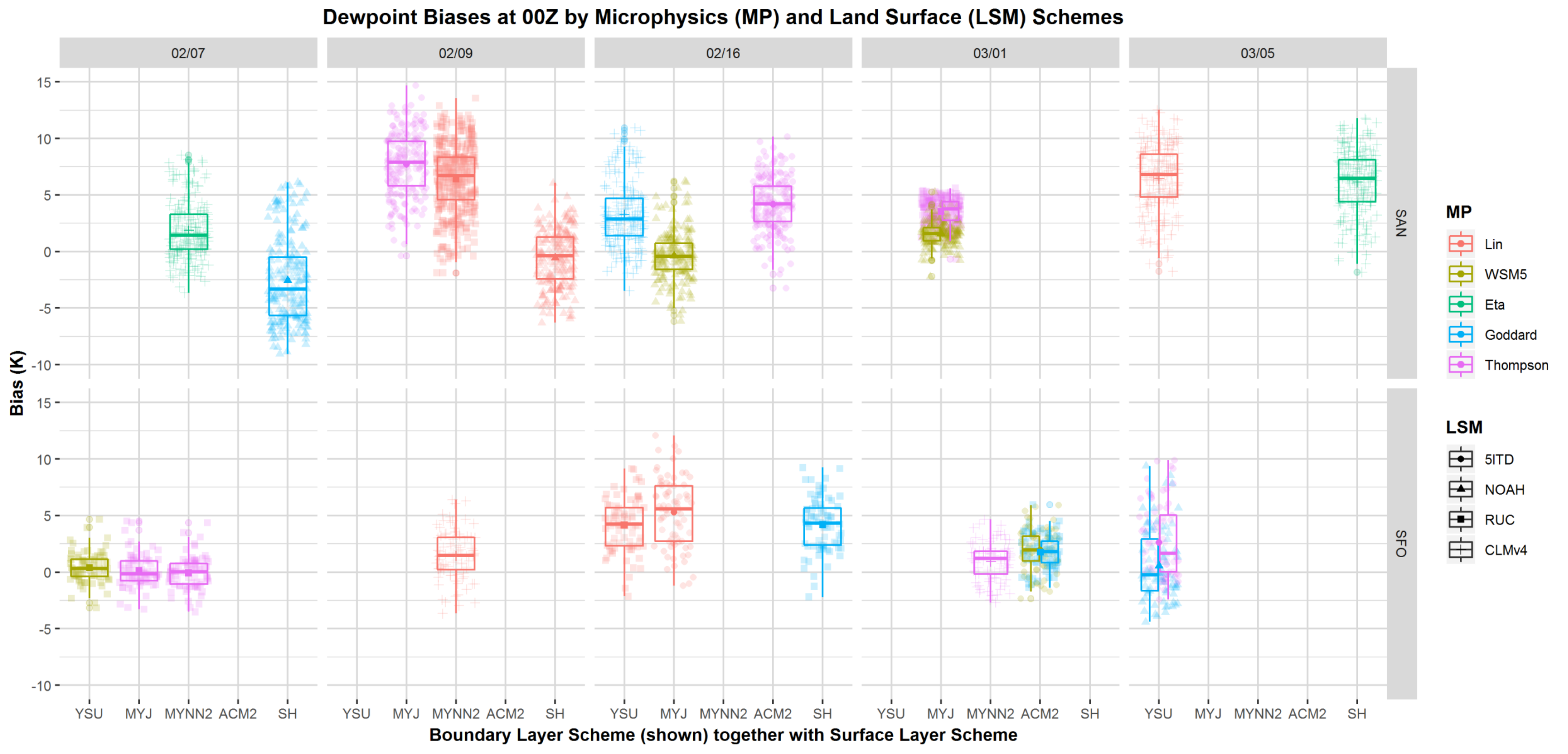
- A generalized linear model allows us to assign weights to each of the considered factors.
- Those weights that are statistically significant for a specific process, e.g., microphysics, indicate that that process is a significant contributor to the mean bias error.
- Thus, we have some macroscopic look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

Dewpoint Biases for both Domains at 21Z and 00Z

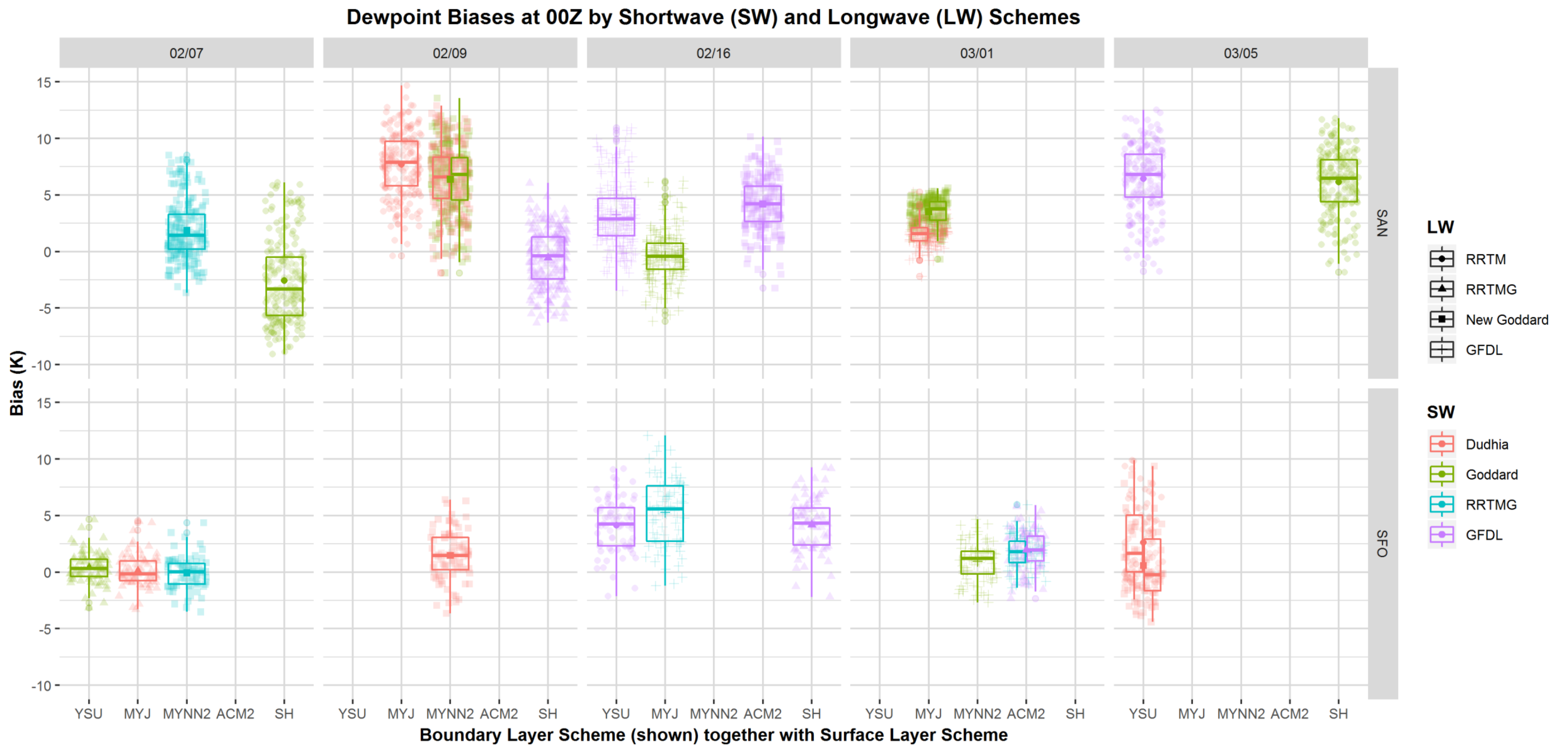
		Parameterization Scheme					Statistics (21Z)					Statistics (00Z)				
Domain	Case	Boundary Layer	Microphysics	Land Surface	Shortwave	Longwave	Points	Min	Mean	Median	Max	Points	Min	Mean	Median	Max
SAN	02/07	MYNN2	Eta	CLMv4	RRTMG	New Goddard	200	-6.25	0.67	-0.50	9.66	204	-3.66	1.88	1.43	8.53
SAN	02/07	SH	Goddard	NOAH	Goddard	RRTM	200	-12.28	-4.67	-6.42	6.21	204	-9.07	-2.57	-3.33	6.12
SAN	02/09	MYJ	Thompson	5lTD	Dudhia	RRTM	189	0.11	6.59	6.48	14.00	200	-0.38	7.71	7.88	14.67
SAN	02/09	MYNN2	Lin	RUC	Dudhia	New Goddard	189	-3.14	4.59	4.61	12.56	200	-1.89	6.44	6.60	12.91
SAN	02/09	MYNN2	Lin	RUC	Goddard	New Goddard	189	-2.61	4.71	4.93	12.09	200	-1.89	6.38	6.82	13.56
SAN	02/09	SH	Lin	NOAH	GFDL	RRTMG	189	-10.50	-1.60	-1.49	6.53	200	-6.29	-0.58	-0.37	6.05
SAN	02/16	YSU	Goddard	CLMv4	GFDL	GFDL	204	-3.13	2.23	1.99	8.82	198	-3.47	3.27	2.89	10.93
SAN	02/16	MYJ	WSM5	NOAH	Goddard	GFDL	204	-7.22	-1.14	-0.88	3.88	198	-6.17	-0.36	-0.42	6.23
SAN	02/16	ACM2	Thompson	5lTD	GFDL	New Goddard	204	-4.24	2.94	2.80	10.06	198	-3.23	4.19	4.22	10.16
SAN	03/01	MYJ	WSM5	NOAH	Dudhia	GFDL	205	-3.18	1.18	1.31	3.76	194	-2.21	1.61	1.57	5.27
SAN	03/01	MYJ	Thompson	RUC	Goddard	New Goddard	205	-0.53	3.76	3.89	6.36	194	-0.68	3.53	3.77	5.60
SAN	03/05	YSU	Lin	CLMv4	GFDL	RRTM	189	-2.32	7.07	7.18	15.43	195	-1.77	6.43	6.80	12.51
SAN	03/05	SH	Eta	CLMv4	Goddard	RRTM	189	-2.06	7.19	7.42	14.06	195	-1.82	6.13	6.48	11.77
SFO	02/07	YSU	WSM5	RUC	Goddard	RRTMG	79	-4.61	-0.08	-0.01	4.35	78	-3.16	0.40	0.30	4.65
SFO	02/07	MYJ	Thompson	RUC	Dudhia	RRTMG	79	-4.33	0.22	0.08	5.01	78	-3.29	0.12	-0.15	4.50
SFO	02/07	MYNN2	Thompson	RUC	RRTMG	New Goddard	79	-5.03	-0.50	-0.57	3.96	78	-3.52	-0.08	0.03	4.37
SFO	02/09	MYNN2	Lin	CLMv4	Dudhia	New Goddard	80	-3.35	1.10	0.95	6.79	85	-3.66	1.46	1.47	6.39
SFO	02/16	YSU	Lin	RUC	GFDL	RRTM	78	-2.62	2.96	3.04	7.89	77	-2.15	4.16	4.24	9.14
SFO	02/16	MYJ	Lin	5lTD	RRTMG	GFDL	78	-1.41	5.42	5.39	11.15	77	-1.21	5.27	5.59	12.06
SFO	02/16	SH	Goddard	RUC	GFDL	RRTMG	78	-2.58	2.97	3.02	7.79	77	-2.19	4.19	4.34	9.24
SFO	03/01	MYNN2	Thompson	CLMv4	Goddard	GFDL	85	-2.49	0.56	0.48	5.28	82	-2.68	0.93	1.19	4.64
SFO	03/01	ACM2	WSM5	5lTD	GFDL	RRTMG	85	-1.05	1.60	1.60	5.97	82	-2.36	1.95	1.95	5.90
SFO	03/01	ACM2	Goddard	RUC	RRTMG	GFDL	85	-1.57	0.58	0.55	4.34	82	-1.40	1.77	1.80	5.94
SFO	03/05	YSU	Goddard	NOAH	Dudhia	New Goddard	81	-5.79	-0.79	-1.35	8.47	77	-4.41	0.57	-0.25	9.37
SFO	03/05	YSU	Thompson	5lTD	Dudhia	RRTM	81	-4.60	0.59	-0.27	9.89	77	-2.41	2.64	1.64	9.87

- Data summarized by domain, case and study hour.
- Points: The number of stations available for a specific combination of the Domain, Case and Parameterization Schemes.
- FLG shortwave radiation scheme removed from this data.

Dewpoint Analysis (16L)



Dewpoint Analysis (16L)



Dewpoint Model (16L)

Method

- Split the study data randomly into two halves.
- Build a model against one half, test against the other half.
- Compute the root mean square error (RMSE) based on model – prediction error in the test set.

Model

- lm(BIAS ~ DOM + CASE + BL + MP + LSM + SW + LW)

RMSE

- Depending on the particular model, the RMSE tended to be on the order of 3 Kelvin.

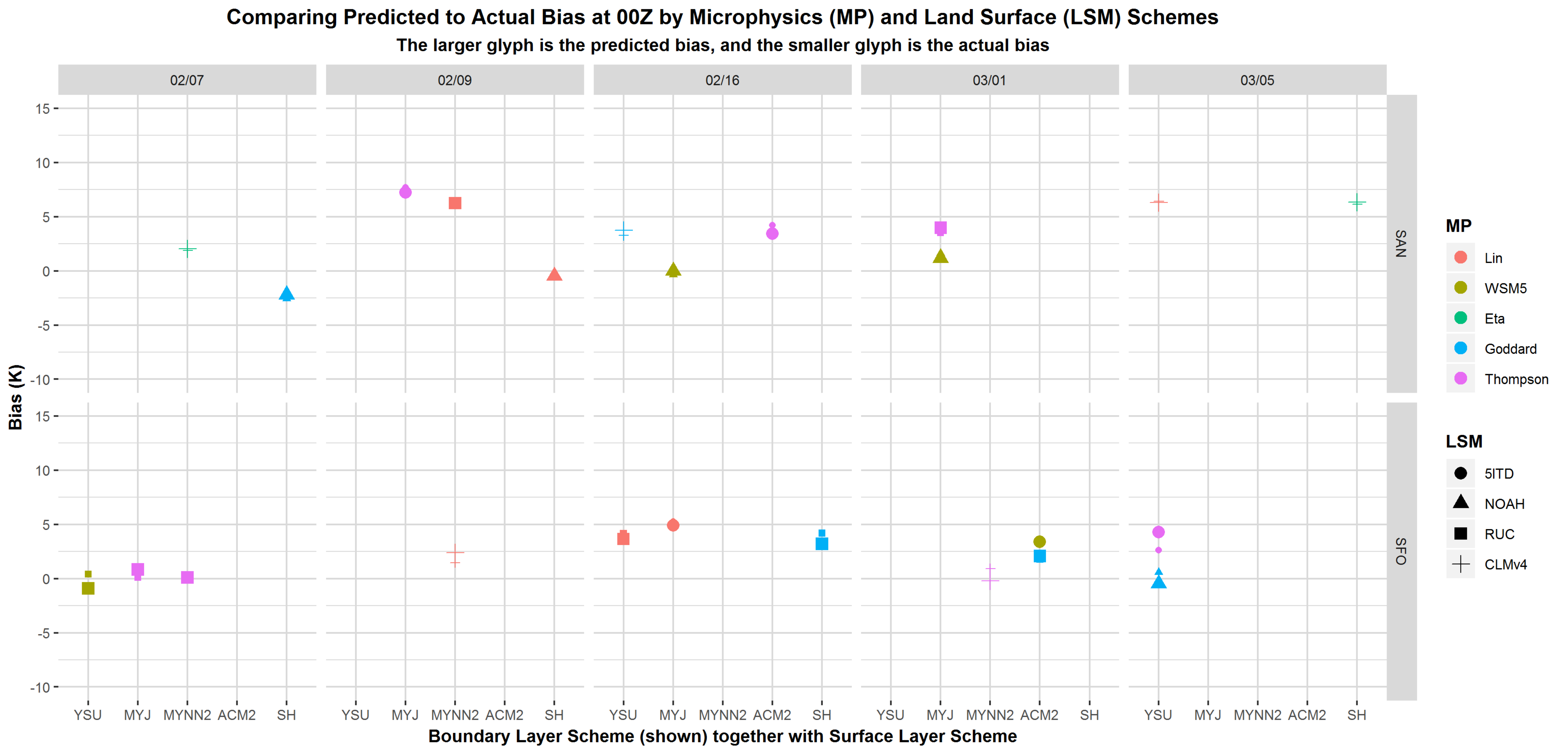
Observations

- The estimate value and its significance varied based on the set used to train the model (expected).
- Over many trials, some estimates were significant, and other times not.
- However, BL_ACM2 and MP_Goddard typically showed estimates that were not significant.

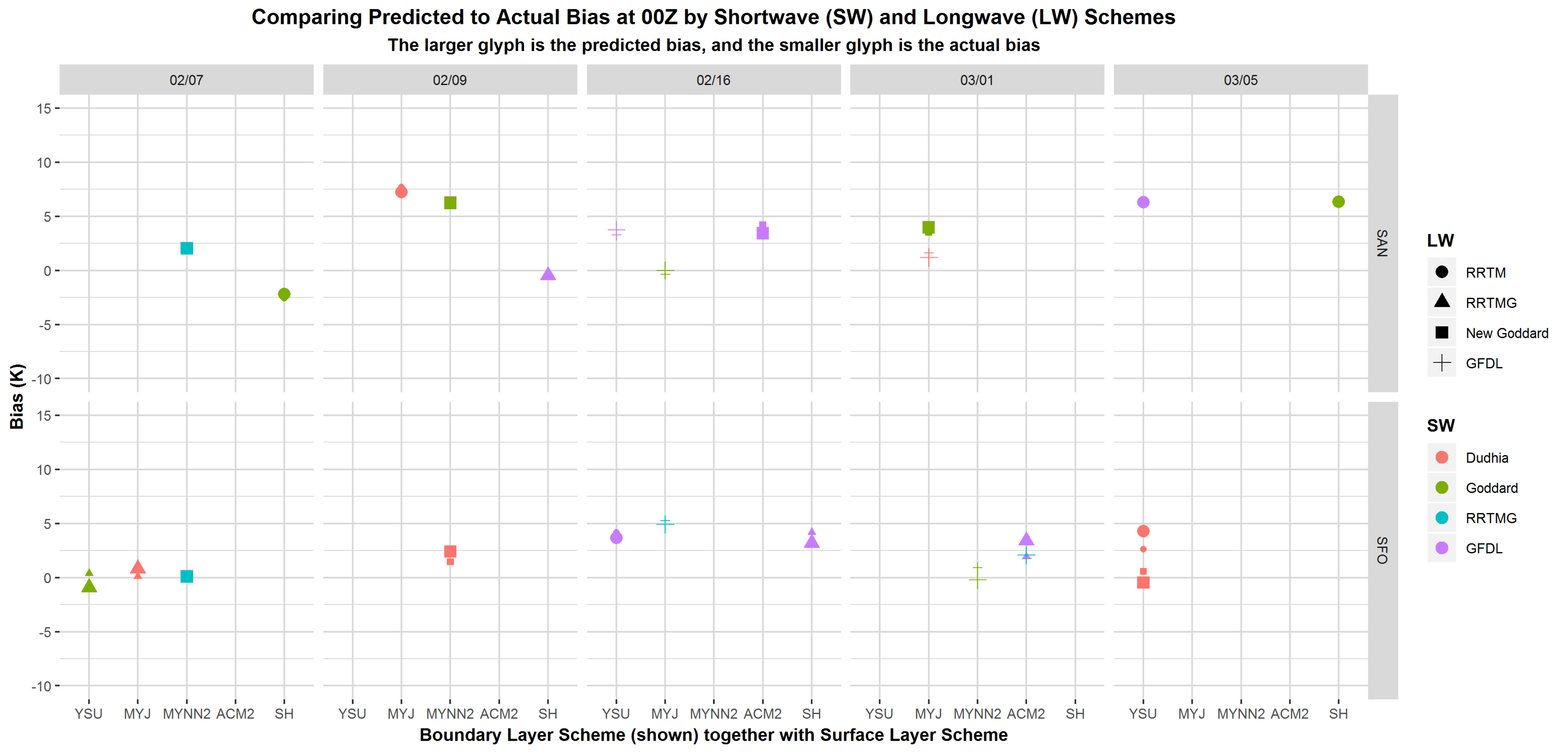
Table 1: Model at 00Z

term	estimate	std.error	statistic	p.value
(Intercept)	3.2255	0.6560	4.9172	0.0000
DOM_SFO	-0.7524	0.4906	-1.5337	0.1253
CASE_02/09	1.9336	0.4251	4.5488	0.0000
CASE_02/16	-1.3285	1.0809	-1.2290	0.2192
CASE_03/01	-0.2141	0.7890	-0.2714	0.7861
CASE_03/05	3.6871	0.4823	7.6454	0.0000
BL_MYJ	3.9715	0.6239	6.3656	0.0000
BL_MYNN2	2.1925	0.5284	4.1495	0.0000
BL_ACM2	-0.0472	0.4295	-0.1099	0.9125
BL_SH	1.2113	0.5013	2.4165	0.0158
MP_WSM5	0.4116	0.4892	0.8413	0.4003
MP_Eta	2.4274	0.7461	3.2533	0.0012
MP_Goddard	1.0650	0.3544	3.0048	0.0027
MP_Thompson	-1.8781	0.3047	-6.1640	0.0000
LSM_NOAH	-7.6303	0.5166	-14.7696	0.0000
LSM_RUC	-1.0054	0.3307	-3.0401	0.0024
LSM_CLMv4	-4.1309	0.6161	-6.7053	0.0000
SW_Goddard	-0.0644	0.2919	-0.2207	0.8254
SW_RRTMG	-1.6186	0.5156	-3.1391	0.0017
SW_GFDL	3.5143	0.9821	3.5784	0.0004
LW_RRTMG	-2.7333	0.5954	-4.5907	0.0000
LW_New Goddard	-0.0631	0.4884	-0.1292	0.8972
LW_GFDL	1.4158	0.8343	1.6969	0.0899

Predicted Mean vs Actual Mean Bias (16L)

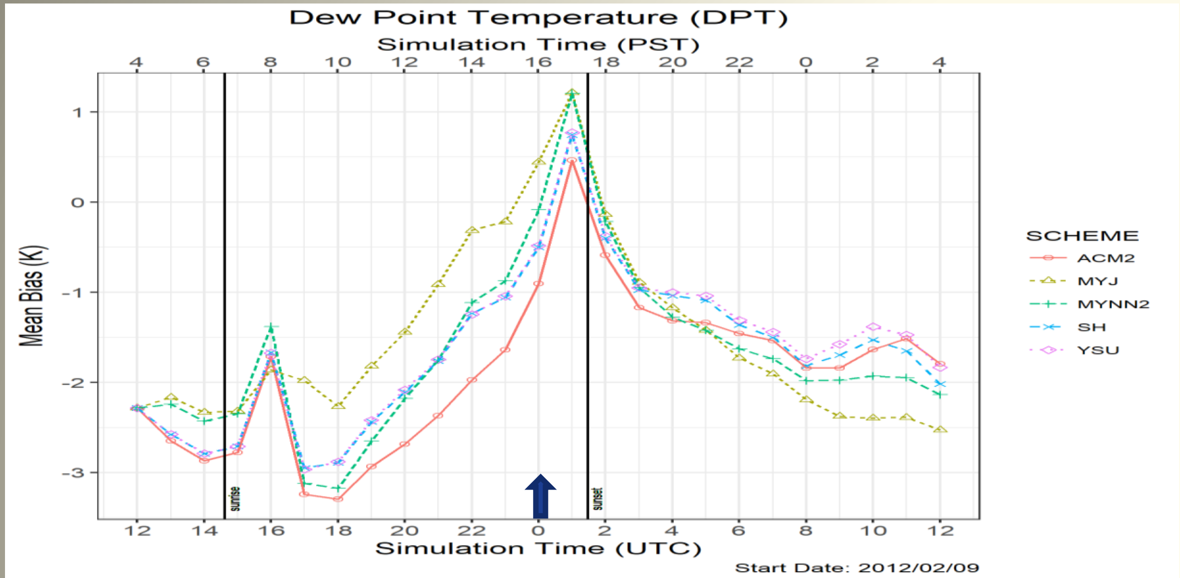


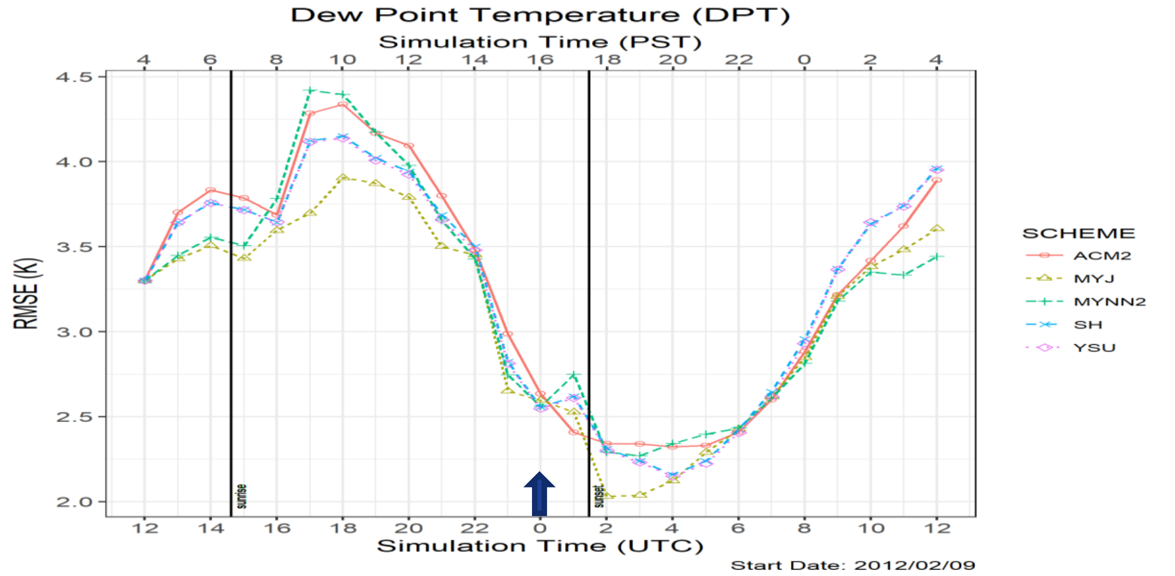
Predicted Mean vs Actual Mean Bias (16L)

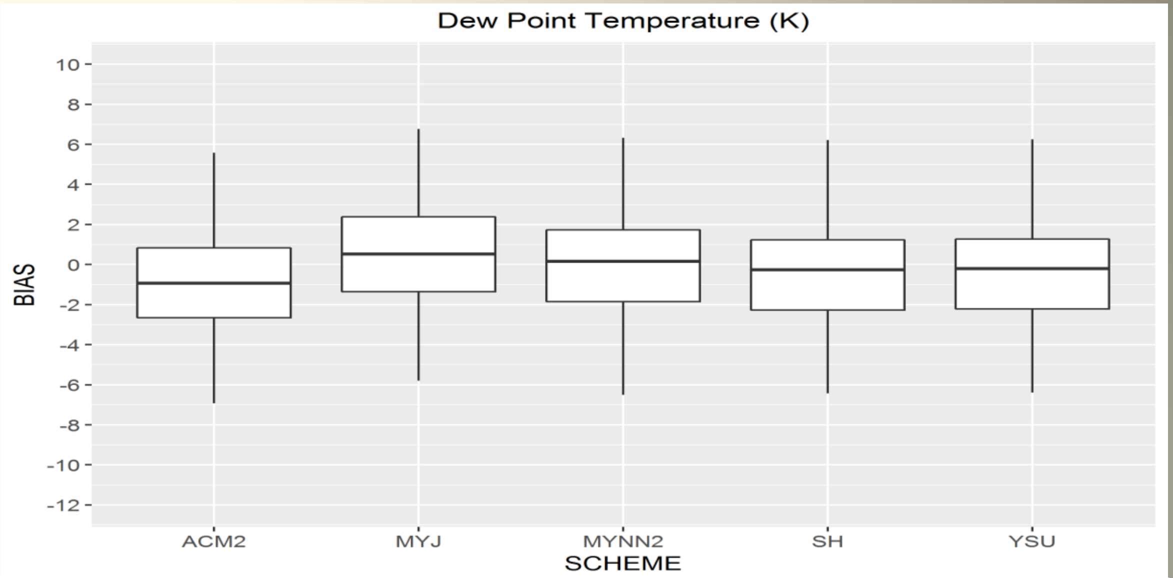


Z2 Dew Points (200 Points) at 00Z (16L)

Data source: Penc et al., Paper 12B.5, 25NWP







Tukey's Honest Significant Difference (0.99)

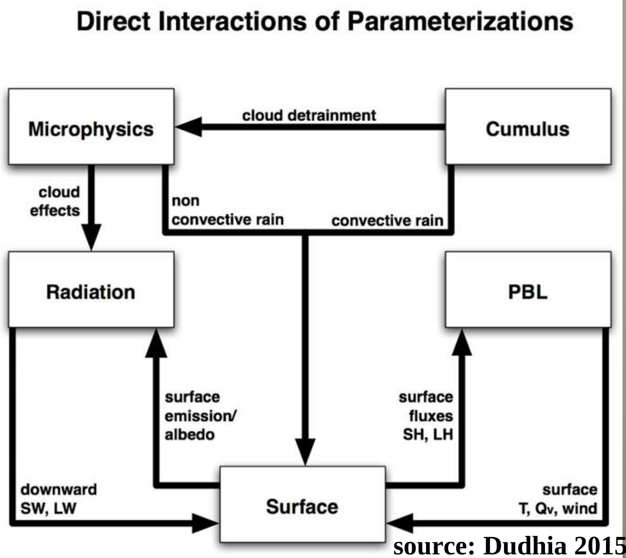
comparison	estimate	conf. low	conf.high	adj.p.value
MYJ-ACM2	1.34775	0.52318	2.17231	0.00000
MYNN2-ACM2	0.81773	-0.00683	1.64229	0.01094
SH-ACM2	0.40435	-0.42021	1.22891	0.49744
YSU-ACM2	0.41977	-0.40479	1.24434	0.45849
MYNN2-MYJ	-0.53002	-1.35458	0.29454	0.22176
SH-MYJ	-0.94340	-1.76796	-0.11883	0.00186
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, as well as managing the configurations and the data is a challenge.
- Using a ‘designed experiment’ allows us to extract quite a bit of information from the data even using a simple linear model.
- Other work suggests that hardware and software considerations can be a nuisance factor in the execution of the experiment.

Summary
- We can observe, at a minimum, Domain and Synoptic contributions to the mean of surface bias errors.
- We can also observe interactions of the parameterization scheme with the surface biases.
- The train/test approach to linear model development is a viable method for analysis; however,
- Additional variables are needed to improve the analysis.

Next Steps
- Employ the train/test approach using a bootstrap regression approach to identify models that are robust across a wide range of permutations of the data.
- Investigate Bayesian methods to improve model evaluations.
- Take advantage of the model/data structure implied by Dudhia’s diagram to expand the diagnostic ability of the model.
- Take advantage of the metrics such as land use and type as well as source of data to add a hierarchical structure to the analysis.
- Investigate approaches that allow us to incorporate forecast time into the analysis.



source: Dudhia 2015

Acknowledgements

The authors wish to thank:

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

- 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.

- Ms. Rivera wrote scripts to automate the scraping the namelist.input files for relevant data.

U.S.ARMY

U.S. ARMY RDECOM ARL RESEARCH LABORATORY

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. Penc, R. S., J. A. Smith, J. W. Raby, B. P. Reen, and R. E. Dumais, Jr., 2018: Intercomparison of seven planetary layer/surface layer physics schemes over complex terrain for battlefield situational awareness applications. *25th Conference on Numerical Weather Prediction, Joint with 29th Conference on Weather and Forecasting*, American Meteorological Society, Paper 12B.15.
12. R Core Team, 2017: R: A language and environment for statistical computing. [Available online at https://www.R-project.org/.]
13. 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.
14. Sacks, J., S. B. Schiller, and W. J. Welch, 1989a: Designs for computer experiments. *Technometrics*, **31,** 41-47. doi: 10.2307/1270363.
15. 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.
16. 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.
17. Santner, T. J., B. J. Williams, and W. I. Notz, 2003: *The design and analysis of computer experiments.* Springer-Verlag.
18. 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.]
19. Skamarock, W. C., and Coauthors, 2008: A description of the advanced research WRF version 3. NCAR Technical Note NCAR/TN-475+STR.
20. 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.
21. Smith, J. A., R. Penc, and J. W. Raby, 2018: Statistical Design of Experiments in Numerical Weather Prediction: Emerging Results. *98th Annual AMS Meeting, Joint with 25th Conference on Probability and Statistics*, Austin, Paper 6.1.
22. ——, 2018b: Statistical Design of Experiments in Numerical Weather Prediction: Emerging Results. *25th Conference on Numerical Weather Prediction, Joint with 29th Conference on Weather and Forecasting*, Denver, CO, American Meteorological Society, Paper 13B.6.
23. 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.
24. 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.
25. 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.]
26. 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.