2026

A Hierarchical Statistical Engineering Modeling Methodology

Teddy Steven Cotter

Proceedings of the American Society for Engineering Management 2016 International Annual Conference
S. Long, E-H. Ng, C. Downing, & B. Nepal eds.

A HIERARCHICAL STATISTICAL ENGINEERING MODELING METHODOLOGY

Teddy Cotter, Ph.D.
Old Dominion University
tcotter@odu.edu

Abstract

In the ASEM-IAC 2015, Cotter (2015) proposed a systemic joint deterministic-stochastic dynamic causal Bayesian statistical engineering model that addressed the knowledge gap needed to integrate deterministic mathematical engineering models within a stochastic framework. However, Cotter did not specify the modeling methodology through which statistical engineering models could be developed, diagnosed, and applied to predict systemic mission performance. This paper updates research into the development a hierarchical statistical engineering modeling methodology and sets forth the initial theoretical foundation for the methodology.

Keywords

Bayesian causal networks, Statistical Engineering

Introduction

The primary objective in developing a general statistical engineering methodology is to facilitate the construction of hierarchical models of partially observable causal-stochastic socio-technical systems in order to better understand and predict the effects of subsystem, module, and component design or improvement interventions on systemic mission performance. Cotter (2015) addressed the problem integrating deterministic engineering models as system dynamic causal components within general linear models (GLM) by representing them as functional causal hierarchical Bayesian networks within a state-space framework to model joint deterministic-stochastic dynamic causal effects**.** He proposed that **t**he **X** controllable and **Z** noncontrollable covariate input variables become endogenous variables of the form

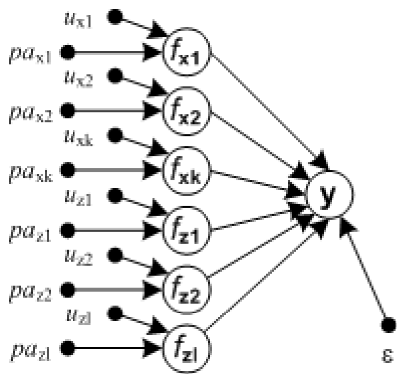
$$\begin{aligned} x_i &= f_i(pa_i, u_{xi}) \qquad i = 1 \text{ to } k \text{ predictors} \\ z_j &= f_j(pa_j, u_{zj}) \qquad j = 1 \text{ to } l \text{ covariates} \end{aligned} \tag{1}$$

where $f_i(\bullet)$ and $f_j(\bullet)$ take on any linear or nonlinear and constant, temporal, instantaneous, or short-term inflection inducing physical or stochastic model that accurately represents the dynamics of the process, pa_i and pa_j are the endogenous parents of **x** and **z** respectively whose functional form and current values determine the *a priori* Bayesian state of each x_i and z_i respectively, and u_{xi} and u_{zj} are the unobserved structural and random errors associated with each x_i predictor and z_j covariate respectively (notation taken from Pearl, 2009). The random component of each u_{xi} and each u_{zi} is not restricted to being $N(0, \sigma^2)$ distributed. Deterministic physical models are incorporated in their functional form as **x** controllable and **z** noncontrollable input variables with respective u_{xi} and u_{zj} error terms to reflect structural and random lack of fit. With this functional notation, the GLM becomes

$$\begin{aligned} &\text{Min } \mathbf{Y}_{\text{Total}} = f(\mathbf{w}'(\mathbf{Y}_{\text{pred}} - \mathbf{T})) \\ &\text{s.t.} \\ &\mathbf{Y} = \mathbf{F}(pa_i, u_{xi})\boldsymbol{\beta} + \mathbf{F}(pa_j, u_{zj})\boldsymbol{\gamma} + \varepsilon \\ &\quad \mathbf{L}_X \le \mathbf{F}(pa_i, u_{xi}) \le \mathbf{U}_X \\ \text{possibly} &\quad \mathbf{L}_Z \le \mathbf{F}(pa_j, u_{zj}) \le \mathbf{U}_Z \end{aligned} \tag{2}$$

where $\mathbf{Y}_{\text{Total}}$ is the vector or matrix of systemic performance variables, $f(\mathbf{w}'(\bullet))$ is a vector or matrix of normalized weighting functions that admit tradeoffs among the $(\mathbf{Y}_{\text{pred}} - \mathbf{T})$ differences, and **T** is the vector or matrix of identified systemic mission performance targets. $F(\bullet)$ is a matrix of functional relationships of the **X** predictors and **Z** within and cross layer covariates respectively (generalization of Sain, Furrer, and Cressie (2011) alternate formulation of a multivariate Markov random field) to the $\mathbf{Y}_{\text{pred}}$ variables performance levels. Where the functional relationship has an unknown form, $f_i(pa_i, u_{xi}) = x_i$ observed data and $f_j(pa_j, u_{zj}) = z_j$ observed covariate values, the residual error accumulates in the ε term. The $\boldsymbol{\beta}$ response parameters of $\mathbf{Y}_{pred}$ to $\mathbf{X}$ and the $\boldsymbol{\gamma}$ response parameters of $\mathbf{Y}_{pred}$ to $\mathbf{Z}$ are constant coefficients to be determined. Under this causal Bayesian network modeling approach, the GLM is represented in Exhibit 1.

Exhibit 1. Functional Causal Bayesian Network GLM DAG.



Although, the above causal Bayesian network form of the GLM will admit integration of deterministic mathematical engineering models within a stochastic framework, it is not in itself sufficient to model systemic performance behavior. The following issues remain to be addressed:

- Decomposing systemic mission performance into functional requirements.
- Systems constraint decomposition.
- Systems boundary interface decomposition.
- Systems performance to functional activation decomposition.
- Systems with nonrecursive directed acyclic graph feedback loops.
- Model synthesis and verification.

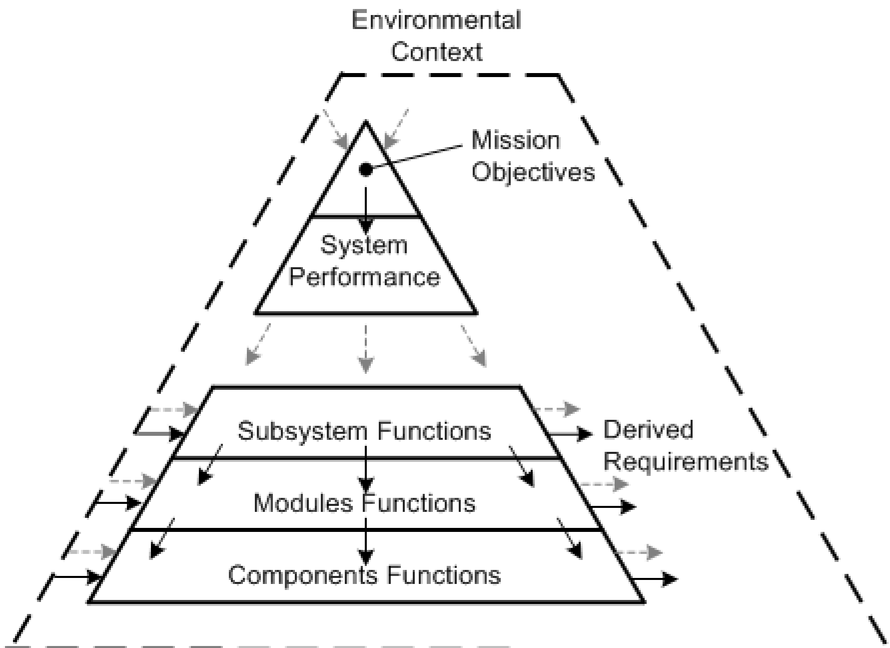
This work reports research into only the first issue of decomposition of systemic mission performance into functional requirements.

Problems in Decomposing Mission Performance into Functional Requirements

Buede (2000) specifies decomposition as the "… top-down structuring, begins with the top-level system function and *partitions* that function into several subfunctions. This decomposition process must conserve all of the inputs to and outputs from the system's top or zero-level function" (p. 182). The key terms in Buede's specification are "top-down structuring" and "conserve all of the inputs to and outputs from…" implying some type of mapping. The top-level system function to which Buede refers is specified in mission requirements, which he defines as "… the interaction of several systems, … which include individuals or groups of people …. Mission requirements represent stakeholder preferences …." In its definition of *Measures of Effectiveness*, the INCOSE *Systems Engineering Handbook* (2010) adds the concepts of "…operational objective ... in the intended operational environment under a specified set of conditions." In his ***Mission Axiom***, Cotter (2015) integrated these concepts into "…the minimal set of tasks mapped to associated outcomes that the product, service, or system is charged with accomplishing in order to produce specified useful mission outcomes under specified environmental conditions." Thus, decomposing mission performance into internal functional requirements is a mapping of systemic mission outcome objectives to the system level performance functions (action transformation tasks) to internal subsystems, modules, and components functions (input-output transformation tasks) needed to accomplish the top level systemic outcome objectives under specified environmental conditions.

In his subsequent work toward defining requirements, Buede indirectly identifies the central problem in mapping mission objectives to internal functionality as the "… result of the design process creates two hierarchies of requirements …," which Buede defines as originating requirements and derived requirements. Specifically, as shown in Exhibit 2, there are two mapping discontinuities in all existing design methods; from the environmental context to system outcome objectives and performance functions and from the system level performance functions to the internal functional requirements.

Exhibit 2. Functional Mapping Discontinuities of Existing Design Methods.



In the figure, the functional mapping discontinuities are represented by the grayed dashed arrows, and the realized mappings are represented by the solid black arrows. Additionally, dashed arrows represent unknown or undefined required inputs from and outputs to the environment, and solid arrows represent known internal functional decomposition mappings and known required inputs from and outputs to the environment. The functional mapping discontinuities result from: (1) typical product environmental specifications being set as functional expectations within environmental expectations or constraints (i.e. minimum, minimum-maximum, or maximum conditions); (2) service specifications being set on some concept of delivering expected value within a qualitatively specified informational or organizational context; and (3) systems specifications being set on some combination of product and services specifications. All three methods rely primarily on a general concept of transforming stakeholder and user views of desired product and service features into a product or service that can deliver those features. The problem of relying on stakeholder and user views is that they do adequately specify causal-stochastic product, service, or system mission performance topologies constrained within causal-stochastic environmental topologies. That is, current design methods result in incomplete environmental topological conditions and constraints specifications and in incomplete topological product, service, or system mission performance specifications at the systems level. The subsequent gap in translating systemic performance topological information into subsystems functions further diffuses incomplete systemic performance information.

Decomposing Systems Mission Causal-Bayesian Topologies

This work proposes a refocus to specifying prior environmental topologies $g(pa_{gi}, u_{xgi})$ as constraints on the conditional systemic performance topologies $\mathbf{Y}_{pred} = \mathbf{F}(pa_{si}, u_{xsi} \mid g)\boldsymbol{\beta}_s + \mathbf{F}(pa_{sj}, u_{zsj} \mid g)\boldsymbol{\gamma}_s + \varepsilon_s$ necessary and sufficient for mission accomplishment. In these formulations, pa_{gi} are the system and local endogenous parents of $\mathbf{z}_{gi}$ noncontrollable environmental variables, u_{xgi} are the temporally exchangeable unobserved system of systems environmental structural and random errors, pa_{si} are the system-of-interest level endogenous parents of the $\mathbf{x}_s$ predictors of $\mathbf{Y}_{pred}$ variable performance levels, u_{xsi} are the unobserved structural and random errors associated with each x_{si}, pa_{sj} are the system-of-interest level endogenous parents of the $\mathbf{z}_s$ covariates of $\mathbf{Y}_{pred}$ variable performance levels, u_{xsj} are the unobserved structural and random errors associated with each z_{sj}, and the $\boldsymbol{\beta}_s$ response parameters of $\mathbf{Y}_{pred}$ to $\mathbf{x}_s$ and the $\boldsymbol{\gamma}$ response parameters of $\mathbf{Y}_{pred}$ to $\mathbf{z}_s$ are constant coefficients to be determined from subsystem-modules-components configurations. For any given subsystem-module-components configuration, residual error accumulates in the ε_s term. In the limiting ideal configuration, ε_s is N($\mathbf{0}$, $\sigma^2\mathbf{I}$) distributed yielding an efficient, consistent, and sufficient MINVUE estimate of $\mathbf{Y}_{pred}$.

The primary issues to be address in specifying prior environmental topologies and conditional systemic performance topologies are non-informative versus informative, conjugate versus nonconjugate, and proper versus improper prior environmental distributions and diffusing or concentrating decomposition. The following axioms are proposed as criteria for selecting the appropriate form of prior environment and hierarchical systemic conditional decomposition.

> *Environmental Systems of Systems Axiom*: All local environments are proper subsets of parental system of systems environments and are determined by processes that exist at the system of systems, system, and local spatial scales
>
> *Requisite Variety Axiom*: Per Ashby's Law of Requisite Variety, the system of interest can achieve stated mission performance outcomes if and only if systemic variety is greater than or equal to the variety of the environment in which it must conduct the mission.
>
> Proof: System(information) >= environmental(information) as a lower bound. If system(information) < environment(information), then system(information) < environment(information *s*) + environment(information *u*) where system(information) = environment(information *s*), and the system is incapable of responding to environment(information *u*). Conversely, if system(information) > environmental(information), then system(information *g*) + system(information *u*) > environmental(information) where system(information *g*) = environment(information), and the system is capable of fully responding to environment(information) with excess system(information *u*).
>
> *Requirements Axiom*: Only systemic mission performance information conditional on environmental information $\mathbf{F}(pa_{si}, u_{xsi} \mid g)\boldsymbol{\beta}_s$ given $\mathbf{F}(pa_{sj}, u_{zsj} \mid g)\boldsymbol{\gamma}_s$ covariate knowledge is necessary and sufficient to specify aggregate sub-systems, modules, and components functionality for systemic mission accomplishment
>
> Proof: Recursive hierarchical application of the Requisite Variety Axiom to decompose system(information) to subsystem(information) to module(information) to component(information).
>
> *Stakeholder State of Mission Knowledge Axiom*: All human knowledge, including stakeholders of potential environmental topologies and systemic mission performance, is incomplete and subjective to varying degrees of freedom. Stakeholder knowledge must induce some variety in the form of bias and additional variance in specifications of expected environmental conditions and systemic mission performance.
>
> Proof: In the Requisite Variety Axiom, set Stakeholder-System(information) $\neq$ System(information) $\Rightarrow$ Stakeholder-System(information) = E[Stakeholder-System(information)] – E[System(information)] + $\tau^2_{stakeholder}$ + σ^2_{system}, where E[Stakeholder-System(information)] is stakeholders expected systemic performance outcomes, E[System(information)] is the realizable systemic performance centroid outcome, $\tau^2_{stakeholder}$ is residual structural and random error due to incomplete and subjective stakeholder knowledge, and σ^2_{system} is residual system error.
>
> *Corollary – Stakeholder State of Mission Knowledge Axiom*: Stakeholder uncertainty or imprecision in statements of expected systemic performance outcomes only adds to variety to systemic mission specifications.

Given the Environmental Systems of Systems Axiom, the local spatial environment and the system of interest must be modeled within the context of its parental system of systems environment. Currently, the most comprehensive environmental hierarchical Bayesian models are those seeking to model global climate change or weather patterns. Advances in computational speed and numerical methods over the last 20 years have permitted the development of hierarchical Bayesian space-time dynamic environmental models. Hughes and Guttorp (1994) modeled space-time atmospheric precipitation using unobserved weather states in hidden Markov models. Brown, Le, and Zidek (1994) developed a hierarchical Bayesian interpolator model for multivariate Gaussian random fields for a fixed set of environmental monitoring sites. Handcock and Wallis (1994) applied a Bayesian best linear unbiased prediction (BLUP) procedure to model the space-time mean temperature meteorological fields in the Northwest United States. Huang and Cressie (1996) introduced a spatio-temporal model that incorporated past snow-water equivalent data resulting in a Kalman-filter prediction algorithm. Wilke (1998) developed a three-stage hierarchical Bayesian modeling process in which the first stage specified a measurement error process for the observational data, the second stage allowed for site-specific time series models, and the third stage estimated model parameters of these time series models as Markov random fields with spatial dependence. Prior to 2000, climate models focused on predicting long term distributions and simulated decades of climate to estimate the long-term parameters.

Since 2000, focus has shifted to climate models with higher regional and local spatial resolution. Regional climate models apply dynamical or statistical downsizing to a limited spatial domain within grid boxes on the scale of 20 to 100 km and time-dependent lateral boundary conditions from a general circulation model. Tebaldi, Smith, Nychka, and Mearns (2005) proposed a hierarchical Bayesian model that combined information from a multimodel ensemble of atmosphere–ocean general circulation models and observations to determine probability distributions of future temperature change on a regional scale. Furrer et al. (2007a and 2007b) developed a univariate hierarchical Bayesian model that separates atmosphere–ocean general circulation spatial response into a large scale climate change signal and an isotropic process representing small scale variability. Berliner and Kim (2008) develop hierarchical Bayesian models from general multiple climate models in a formulation that enables treatment of model specific means, biases, and covariance matrices in a manner that estimates true states of nature. Smith, Tebaldi, Nychka, and Mearns (2009) proposed a Bayesian analysis that allowed the combination of different univariate climate models into a posterior distribution of future temperature increases over different regions around the globe. Tebaldi and Sanso (2009) extend these approaches to estimating posterior distributions of bivariate models of joint temperature and precipitation by applying a hierarchical Bayesian model to data sets of simulated climate from general circulation models. Cooley and Sain (2010) developed a hierarchical Bayesian extreme value model of extreme precipitation events. Sain, Furrer, and Cressis (2011) developed a hierarchical Bayesian modeling approach that allows for flexible modeling of multivariate dependencies.

Parallel to statistical engineering's goal of constructing hierarchical models of causal-stochastic systems in order to better understand design or improvement intervention effects on mission performance, climate scientists' goal in creating hierarchical climate models is to study the effects of anthropogenic forcings on global climate. Combining the Requisite Variety and Requirements axioms with the Environmental Systems of Systems Axiom, this research's current focus is on transforming existing hierarchical Bayesian climate modeling knowledge into the socio-technical domain. Current knowledge in the climate domain that is applicable to the socio-technical domain includes:

- Integrating information from a multimodel ensemble of atmosphere–ocean general circulation models into a joint hierarchical Bayesian model. The primary stated goal of statistical engineering is to integrate deterministic engineering and stochastic models from multiple disciplines within GLM causal-stochastic Bayesian hierarchical, state-space networks.
- Estimating posterior distributions derived from statistical assumptions incorporating bias among environmental models with convergence that allows estimates of correlation between present and future temperature responses, and testing alternative forms of probability distributions for the model error terms. In the socio-technical domain, the equivalent problem is estimating posterior topologies adjusted for bias between deterministic causal and stochastic terms with convergence that allows estimates of correlation between current and future mission performance and tests for alternative mission performance topologies from lower order cumulative error terms.
- In climate models, the posterior distributions of regional temperature change in many region–season combinations have been found to differ both in variance and shape, depending on the statistical model adopted. This suggests that hierarchical causal Bayesian systems, including socio-technical, will be dominated by environmental and mission performance informative prior topologies and nonconjugate but proper decomposition topologies.
- Climate hierarchical Bayesian models include time-varying parameters that are modeled as random variables with distributions depending in part on atmospheric pollution levels. Similarly, socio-technical models must include time-varying parameters that model systemic mission performance reliability and changes in performance and reliability due to future new technology insertions and upgrades.
- Uncertainties in climate change projections are modeled as arising from: (1) natural climate variability, (2) uncertainties in the responses to climate forcing factors, and (3) uncertainties in future forcing factors and other factors that could influence climate. Similarly, socio-technical systems topologies must account for uncertainties in: (1) natural environmental variability, (2) cumulative systemic performance variability, and (3) design or improvement interventions variability.

- Extreme value hierarchical Bayesian analysis has been applied to characterize the tails of regional and location marginal precipitation and wind distributions. Causal hierarchical Bayesian models of socio-technical systemic mission performance must also incorporate extreme value Bayesian analyses to model environmental minimum and maximum extremes rather than setting deterministic minimum and maximum limits as is the current common practice.
- The challenging aspect of decomposing and modeling the spatial processes of the global climate system, and any socio-technical system, is the number of parameters and the amount of data involved. As the number of parameters increase, either model precision or bias must increase. Thus, statistical engineering must be built on a sound theoretical body of the most recent causal Bayesian hierarchical modeling knowledge.

Decomposition knowledge in the climate domain that is not directly applicable to the socio-technical domain includes:

- Hierarchical Bayesian climate models use lattice decomposition with defined uniform local boundary input-output points on spatial scales of 20 to 500 km^2. Open socio-technical systems, sub-systems, modules, and components boundaries are, by definition, permeable and not crisply defined, and they have feedforward and feedback inputs that exist at points determined by the physics and economics of the processes.
- Not adequately addressed in climate models, from causal Bayesian theory a model's boundary is determined by the modeler's partition between the pa_i and pa_j parents of x_i predictors and z_i covariates and their respective u_{xi} and u_{zj} unobserved structural and random errors. Thus, hierarchical causal Bayesian models' boundaries will rarely equal the true boundaries of a socio-technical system; however, boundary conditions must be accurately considered in model development.
- Decomposition of climate models proceed from reasonably bounded and identified global climate patterns to regional and local climate patterns. Such may not be the case for socio-technical systems hierarchical models. In design or improvement interventions of socio-technical systems, only relevant decomposition paths may be necessary to understand the effects of changes within subsystems, modules, or components on systemic mission performance. In these modeling efforts, it will be necessary build accurate scaffolds of the unmodeled subsystems, modules, and components in order to assess the effects of design or improvement interventions on system mission performance.
- To date climate models have not sought to integrate deterministic engineering models as system dynamic causal components, which are essential in modeling socio-technical systemic performance. Anthropogenic forcings on global climate are modeled as dynamic random variables. Inclusion of deterministic functions will require integration of Pearl's (2009) $P(y \mid do(x))$ operators as $f(y_t \mid do(x_t))$ functions to dynamically set deterministic and feedback functional states prior to updating hierarchical expectations and variances.

Conclusion

Climate hierarchical Bayesian models and modeling methodologies articles have been gathered into a corpus. Current research is oriented toward extracting and integrating models and methods relevant to the problem of developing hierarchical causal Bayesian models of socio-technical systemic mission performance. Once a hierarchical causal Bayesian socio-technical modeling body of knowledge has been developed, validated, and peer reviewed, research will move to systemic constraint decomposition to assure that imposed economic, environmental, legal, political, social, and technical constraints are consistently jointly decomposed to subsystems, modules, and components levels.

References

Ashby, W. R. (1964). *An Introduction to Cybernetics*. London: Methuen & Co. Ltd.

Berliner, L. and Kim, Y. (2008). Bayesian design and analysis for superensemble based climate forecasting. *Journal of Climate*, *21*(9), 1891-1910

Brown, P., Le, N., and Zidek, J. (1994). Multivariate spatial interpolation and exposure to air pollutants. *The Canadian Journal of Statistics*, *22*(4), 489-509.

Buede, D. (2000). *The Engineering Design of Systems: Models and Methods*. New York: Wiley.

Cooley, D. and Sain, S. (2010). Spatial hierarchical modeling of precipitation extremes from a regional climate model. *Journal of Agricultural, Biological, and Environmental Statistics*, *15*(3), 381-402.

Cotter, T. (2015). Engineering Analytics: Research into the Governance Structure Needed to Integrate the Dominant Design Methodologies. *Proceedings of the ASEM 2015 International Conference*. Indianapolis.

Cotter, T. (2015). Statistical Engineering: A Causal-Stochastic Modeling Research Update. *Proceedings of the ASEM 2015 International Conference*. Indianapolis.

Furrer, R., Knutti, R., Sain, S., Nychka, D., and Meehl, G. (2007a). Spatial patterns of probabilistic temperature change projections from a multivariate Bayesian analysis. *Geophysical Research Letters 34, L06711. DOI: 10.1029/2006GL027754*.

Furrer, R., Sain, S., Nychka, D., and Meehl, G. (2007b). Multivariate Bayesian analysis of atmosphere-ocean general circulation models. *Environmental and Ecological Statistics*, *14*(3), 249-266.

Handcock, M. and Wallis, J. (1994). An approach to statistical spatial-temporal modelling of meteorological fields. *Journal of the American Statistical Association*, *86*(426), 368-378.

Huang, H. and Cressie, N. (1996). Spatio-temporal prediction of snow water equivalent using the Kalman filter. *Computational Statistics and Data Analysis*, *22*(2), 159-175.

Hughes, J. and Guttorp, P. (1994). Incorporating spatial dependence and atmospheric data in a model of precipitation. *Journal of Applied Meteorology*, *33*(12), 1503-1515.

INCOSE. (2010). *Systems Engineering Handbook:A Guide for System Life Cycle Processes and Activities*. New York: Wiley.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press.

Sain, S., Furrer, R., and Cressie, N. (2011). A Spatial Analysis of Multivariate Output from Regional Climate Models. *The Annals of Applied Statistics*, *5*(1), 150-175.

Smith, R., Tebaldi, C., Nychka, D., and Mearns, L. (2009). Bayesian modeling of uncertainty in ensembles of climate models. *Journal of the American Statistical Association*, *104*(485), 97-116.

Tebaldi, C. and Sans6, B. (2009). Joint projections of temperature and precipitation change from multiple climate models: A hierarchical Bayes approach. *Journal of the Royal Statistical Society, Series A*, *172*(1), 83-106.

Tebaldi, C., Smith, R., Nychka, D., and Mearns, L. (2005). Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multimodel ensembles. *Journal of Climate*, *18*(10), 1524-1540.

Wilke, C. (1998). Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics*, *5*(2), 117-154.

About the Author

T. Steven Cotter is a Lecturer with the Engineering Management and Systems Engineering department at Old Dominion University. He earned a Ph.D. in Engineering Management and Systems Engineering from Old Dominion University, a Master of Science in Engineering Management with a concentration in quality/reliability engineering from the University of Massachusetts at Amherst, a Master of Business Administration with a concentration in finance and a Bachelor of Science both from the University of South Carolina, and a diploma in Electronic Technology from Graff Area Vocational and Technical School (now Ozarks Technical Community College). He is a certified Quality Engineer and certified Reliability Engineer with the American Society for Quality. His research interests are in engineering design analytics, human-machine intelligent decision governance, quality systems design, and statistical engineering.