UNCERTAINTY EVALUATION IN LARGE-SCALE DYNAMICAL SYSTEMS: THEORY AND APPLICATIONS

Yi Zhou

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

December 2014

APPROVED:

Robert Akl, Major Professor
Yan Wan, Co-Major Professor
Bill Buckles, Committee Member
Yan Huang, Committee Member
Qunfeng Dong, Committee Member
Sandip Roy, Committee Member
Craig Wanke, Committee Member
Barrett Bryant, Chair of the Department of Computer Science and Engineering
Costas Tsatsoulis, Dean of the College of Engineering
Mark Wardell, Dean of the Toulouse Graduate School

Zhou, Yi. Uncertainty Evaluation in Large-Scale Dynamical Systems: Theory and Applications. Doctor of Philosophy (Computer Science and Engineering), December 2014, 269 pp., 4 tables, 64 illustrations, bibliography, 156 numbered references.

Significant research efforts have been devoted to large-scale dynamical systems, with the aim of understanding their complicated behaviors and managing their responses in real-time. One pivotal technological obstacle in this process is the existence of uncertainty. Although many of these large-scale dynamical systems function well in the design stage, they may easily fail when operating in realistic environment, where environmental uncertainties modulate system dynamics and complicate real-time predication and management tasks. This dissertation aims to develop systematic methodologies to evaluate the performance of large-scale dynamical systems under uncertainty, as a step toward real-time decision support.

Two uncertainty evaluation approaches are pursued: the analytical approach and the effective simulation approach. The analytical approach abstracts the dynamics of original stochastic systems, and develops tractable analysis (e.g., jump-linear analysis) for the approximated systems. Despite the potential bias introduced in the approximation process, the analytical approach provides rich insights valuable for evaluating and managing the performance of large-scale dynamical systems under uncertainty. When a system's complexity and scale are beyond tractable analysis, the effective simulation approach becomes very useful. The effective simulation approach aims to use a few smartly selected simulations to quickly evaluate a complex system's statistical performance. This approach was originally developed to evaluate a single uncertain variable. This dissertation extends the approach to be scalable and effective for evaluating large-scale systems under a large-number of uncertain variables. While a large portion of this dissertation focuses on the development of generic methods and theoretical analysis that are applicable to broad large-scale dynamical systems, many results are illustrated through a representative large-scale system application on strategic air traffic management application, which is concerned with designing robust management plans subject to a wide range of weather possibilities at 2-15 hours look-ahead time.

ACKNOWLEDGEMENTS

I have been studying at UNT for almost seven years. I established an important body of research and had unforgettable time with many people. First, I would like to thank my academic advisors, professors Robert Akl and Yan Wan. Without their experience and professional knowledge, I could not devote myself to studying this topic. Robert is a lively and energetic professor, I will not forget the interesting presentations he delivered since they helped me understand and memorize new knowledge. He is kind and helpful to everyone in his group. Yan was very supportive during my graduate study. I appreciate her guidance, understanding, patience and her scientific advice. She always encouraged me that an outstanding engineer is not the one who simply follows protocols, but the one who develops new knowledge and techniques that can benefit the society. I would like to especially thank the MITRE Corporation for the fruitful collaboration and financial support during my Master's and PhD studies and my PhD committee members for their advice and suggestions. I also thank the churches I visited in the Denton area. Even though I am not a Christian, I do appreciate the people who taught me what righteous love was. I believe that love is not self-seeking, is not easily angered, and keeps no records of wrongs. I thank my table tennis partner and my best friends at UNT. We encouraged and supported each other through the difficult times. I thank my labmates for providing useful advice. Finally, I would like to express my deepest love to my parents for their support and patience to share the good and bad. Most importantly, I would express my sincere appreciation to every warm-hearted person's care when I was weak. In general, I am thankful to everything in my life since that helps me become who I really am.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Significant research efforts have been devoted to large-scale dynamical systems, with the aim of understanding their complicated behaviors and managing their responses in real-time. One pivotal technological obstacle in this process is the existence of uncertainty. Although many of these large-scale dynamical systems function well in the design stage, they may easily fail when operating in realistic environment, where environmental uncertainties modulate system dynamics and complicate real-time predication and management tasks. This dissertation aims to develop systematic methodologies to evaluate the performance of large-scale dynamical systems under uncertainty, as a step toward real-time decision support (see also results documented in publications [125, 128, 129, 135, 136, 145, 146, 151–156]).

Two uncertainty evaluation approaches are pursued: the analytical approach and the effective simulation approach. The analytical approach abstracts the dynamics of original stochastic systems, and develops tractable analysis (e.g., jump-linear analysis) for the approximated systems. Despite the potential bias introduced in the approximation process, the analytical approach provides rich insights valuable for evaluating and managing the performance of large-scale dynamical systems under uncertainty. When a system's complexity and scale are beyond tractable analysis, the effective simulation approach becomes very useful. The effective simulation approach aims to use a few smartly selected simulations to quickly evaluate a complex system's statistical performance. This approach was originally developed to evaluate a single uncertain variable. This dissertation extends the approach to be scalable and effective for evaluating large-scale systems under a large-number of uncertain variables. While a large portion of this dissertation focuses on the development of generic methods and theoretical analysis that are applicable to broad large-scale dynamical systems, many results are illustrated through a representative large-scale system application on strategic air traffic management application, which is concerned with designing robust management plans subject to a wide range of weather possibilities at 2-15 hours look-ahead time.

The rest of this dissertation is arranged as follows. In Part I (Chapters 2 and 3), the modeling of large-scale dynamical systems under uncertainty is discussed through two applications, strategic air traffic management and airborne networking. Models that capture complicated network interactions and are succinct enough to meet the real time-requirement are critical to the uncertainty evaluation of large-scale systems. In Part II (Chapters 4 and 5), uncertainty evaluation and decision-making for systems with a single uncertain variable are discussed in detail through the strategic air traffic management application. This part focuses on the integration of system models with uncertain environmental models for both evaluation and decision-making. Both analytical approach and effective uncertainty evaluation approach are introduced. In Part III (Chapters 6 and 7), the single-variable effective simulation approach is generalized to the multivariate case, which is common to large-scale system applications. The properties of the generalized method are theoretically analyzed. In addition, this generalized approach is applied to the uncertainty evaluation and strategy design for the strategic air traffic management application. In Part IV (Chapters 8 and 9), further theoretical development is made to bring the multivariate uncertainty evaluation methods scalable to the number of uncertain variables, critical to its use in large-scale system application. Both effective simulation and jump-linear approaches are developed and their applications to strategic air traffic flow management are discussed.

1.1. Part I: Modeling of Large-Scale Dynamical Systems under Uncertainty

In Chapter 2, we introduce a queuing network model that can comprehensively represent traffic flow dynamics and flow management capabilities in the U.S. National Airspace

System (NAS). We envision this model as a framework for tractably evaluating and designing coordinated flow management capabilities at a multi-Center or even NAS-wide spatial scale and at a strategic (2–15h) temporal horizon. As such, the queuing network model is expected to serve as a critical piece of a strategic flow contingency management solution for the Next Generation Air Traffic System (NextGen). Based on this perspective, we outline, in some detail, the evaluation and design tasks that can be performed using the model, as well as the construction of the flow network underlying the model. Finally, some examples are presented, including one example that replicates traffic in Atlanta Center on an actual bad-weather day, to illustrate simulation of the model and interpretation/use of model outputs.

In Chapter 3, we introduce a novel mobility model for airborne networks (ANs). The design of effective routing protocols in ANs relies on suitable mobility models that capture the movement patterns of airborne vehicles. As airborne vehicles cannot make sharp turns as easily as ground vehicles do, the widely used ground-based mobile ad hoc network (MANET) mobility models are not appropriate to use as the analytical frameworks for airborne networking. The mobility model we developed is called the smooth-turn (ST) mobility model, that captures the correlation of acceleration of airborne vehicles across temporal and spatial coordinates. The proposed model is realistic in capturing the tendency of airborne vehicles toward making straight trajectories and STs with large radii, yet is tractable enough for analysis and design. We first describe the mathematics of this model and then prove that the stationary node distribution is uniform. Furthermore, we introduce a metric to quantify the degree of model randomness, and using this, we compare and classify several mobility models in the literature. We conclude this chapter with several possible variations to the basic ST mobility model.

In Chapter 4, we introduce a promising framework for representing an air traffic flow (stream) and flow-management action operating under weather uncertainty. We propose to use a meshed queuing and Markov-chain model–specifically, a queuing model whose service-rates are modulated by an underlying Markov chain describing weather-impact evolution–to capture traffic management in an uncertain environment. Two techniques for characterizing flow-management performance using the model are developed, namely 1) a master-Markov-chain representation technique that yields accurate results but at relative high computational cost, and 2) a jump-linear system-based approximation that has promising scalability. The model formulation and two analysis techniques are illustrated with numerous examples. Based on this initial study, we believe that the interfaced weather-impact and traffic-flow model analyzed here holds promise to inform strategic flow contingency management in NextGen.

In Chapter 5, we investigate the optimal decision-making under weather and demand uncertainties. Air traffic management at the strategic time frame (with 2-15 hours look-ahead time) is complicated by demand and weather uncertainties. As the Monte Carlo approach to find optimal solutions is time-consuming, we need an effective and systematic approach to quickly 1) assess the impact of uncertain weather, and 2) design optimal management strategies under demand and weather uncertainties. In particular, we investigate a simple strategic flow management scenario: a stream of uncertain flow enters a single weather zone subject to weather uncertainty. We provide an integrated weather-demand-management modeling framework to capture the uncertain dynamics of this scenario. Using this integrated modeling framework, we provide a jump-linear analytical approach to evaluate the first and second moments of weather impact, and a Probabilistic Collocation Method (PCM) based approach for the design of optimal flow management. Possible cost functions for the optimal management is discussed, and examples are shown to demonstrate the performance of the proposed approaches.

1.3. Part III: Evaluation and Decision-Making Subject to High-Dimensional Uncertainty

In Chapter 6, we extend the formal analysis of single-variable PCM to the multivariate case, where multiple uncertain parameters may or may not be independent. Modern large-scale infrastructure systems have typical complicated structure and dynamics, and extensive simulations are required to evaluate their performance. PCM has been developed to effectively simulate a system's performance under parametric uncertainty. In particular, it allows reduced-order representation of the mapping between uncertain parameters and system performance measures/outputs, using only a limited number of simulations; the resultant representation of the original system is provably accurate over the likely range of parameter values. Specifically, we provide conditions that permit multivariate PCM (M-PCM) to precisely predict the mean of original system output. We also explore additional capabilities of the M-PCM, in terms of cross-statistics prediction, relation to the minimum mean-square estimator, computational feasibility for large dimensional parameter sets, and sample-based approximation of the solution. At the end of the chapter, we demonstrate the application of M-PCM in evaluating air traffic system performance under weather uncertainties.

In Chapter 7, we suggest an effective approach to design optimal strategic air traffic management strategies under weather uncertainty. The approach is based on the M-PCM, which constructs a good low-order polynomial approximation of the original system mapping over likely parameter values, from which optimal management strategies can be derived. We illustrate the optimal management design procedure through an example of designing Minutes-in-trail (MINIT) and routing strategies for two correlated weather zones. The performance analysis shows this approach is promising for the real-time design of strategic air traffic management strategies due to its correctness and efficiency.

1.4. Part IV: Further Results on Scalable Uncertainty Evaluation and Decision-Making

In Chapter 8, we introduce a scalable uncertainty evaluation method to significantly reduce number of simulations. Effective uncertainty evaluation is a critical step toward real-time and robust decision-making for complex systems in uncertain environments. The M-PCM was developed to effectively evaluate system uncertainty. The method smartly chooses a limited number of simulations to produce a low-order mapping, which precisely predicts the mean output of the original system mapping up to certain degrees. While the M-PCM significantly reduces the number of simulations, it does not scale with the number of uncertain parameters, making it difficult to use for large-scale applications that typically involve a large number of uncertain parameters. We develop a method to break the curse of dimensionality. The method integrates M-PCM and Orthogonal Fractional Factorial Design (OFFD) to maximally reduce the number of simulations from 2^{2m} to $2^{\lceil log_2(m+1) \rceil}$ for a system mapping of m parameters. The integrated M-PCM-OFFD predicts the correct mean of the original system mapping, and is the most robust to numerical errors among all possible designs of the same number of simulations. The analysis also provides new insightful formal interpretations on the optimality of OFFDs.

In Chapter 9, we study the sensitivity of air traffic flow management performance in the NAS to uncertain disturbances. In particular, we model the air traffic system under weather uncertainty as a discrete jump linear system. As the performance of the system is tractable at each time step, we are able to evaluate the transient traffic congestion caused by weather over a time span. Then, we analyze the sensitivity of the total transient congestion with respect to inflow rate based upon our jump linear modeling framework. Finally, we present our idea of aiding air traffic planning based upon the information obtained from sensitivity analysis. We show that the design of optimal flow management preserves a simple sensitivity structure.

CHAPTER 2

DYNAMIC QUEUING NETWORK MODEL FOR FLOW CONTINGENCY MANAGEMENT

2.1. Introduction

In the Next Generation Air Transportation System (NextGen), traffic flow management (TFM) operations will require better congestion prediction and management at longer look-ahead times (LATs) to more effectively use the limited capacity of the National Airspace System (NAS). Specifically, a decision support capability is envisioned for NextGen, which can account for the uncertainties inherent at these longer LATs while still providing decision makers with effective coordinated strategies for managing congestion. In response to this need, an operational concept for flow contingency management (FCM), which is defined by the Federal Aviation Administration as the process which identifies and resolves congestion or complexity resulting from blocked or constrained airspace or other offnominal conditions [39], has been proposed [123, 127]. The FCM operational concept requires an aggregate flow model that will allow incorporation of dynamic weather impact, design of management strategies for likely weather impact outcomes, and evaluation of these strategies using metrics of interest to multiple decision makers. A critical need in this approach is the ability to capture the types of congestion management controls available now and envisioned in the NextGen environment, for the simulation and evaluation of aggregated traffic flow for diverse management options.

We are developing a new dynamic queuing network model of the NAS to enable the evaluation and design of traffic management strategies under weather uncertainties for FCM. Herein, we describe the model, identify gaps in analysis, and present examples of its application. Our queuing network model has several features that make it promising for

FCM.

(1) It links stochastic weather impact with NAS performance. The model takes stochastic weather impact predictions as input and simulates NAS flow dynamics, allowing calculation of performance metrics such as delay, throughput, and congestion statistics. Given the significant weather uncertainty at longer LATs, such performance analysis under uncertainty is critical for FCM.

(2) It quantitatively captures the impact of realistic traffic management actions at an aggregate-flow level, which is a critical requirement for NAS-wide decision-making and evaluation goals of FCM. Specifically, management actions (five in total) that are used in current practice or are likely to be used in NextGen are comprehensively modeled.

(3) It represents traffic as stochastic flows while maintaining route structures. Because flows rather than individual aircraft are tracked, the dimension of the decision space for FCM is significantly reduced. However, the model allows imposition of route structures specified by an underlying flow network, as needed for the effective simulation and design of the NAS.

(4) It permits parameterization from data and interface with operational practice. The model is constructed to provide an effective evaluation and design of management strategies based upon an aggregated flow structure, demand information, and weather impact data. These constructs readily correspond with physical elements and conceptual paradigms used in the NAS, allowing easy model parameterization from historical data and translation of model results to operational practice.

This chapter is organized as follows: Section 2.2 gives background on queuing network models and motivates their use in FCM. In Section 2.3, we describe the key challenges in developing dynamic flow network (queuing) models for FCM. In Section 2.4, we describe the new dynamic queuing network model in detail, discuss its interface with the FCM framework, and list analytical challenges in using the model for FCM design. In Section 2.5, we present three examples illustrating simulation and performance evaluation of FCM designs using the model; one of the examples replicates a historical convective-weather- constrained traffic scenario in Atlanta Center. In Section 2.6, we briefly introduce an alternative route based dynamic queuing network model. A brief conclusion is given in Section 2.7.

2.2. Background on Dynamic Queuing Network Model

Driven by the need for NAS-wide coordinated management, several research efforts have pursued construction of network models for the NAS, as reviewed in [117] and [53]. Most of these models are deterministic models (e.g., [13, 15, 82, 83, 90, 119]) and, as such, have limited ability to capture uncertainties. Queuing network models provide a natural alternative for modeling traffic systems with uncertainty. A queue comprises 1) a stochastic arriving/upstream flow that is usually modeled as a stochastic process, 2) a flow restriction service capturing limited capability to process flows, and 3) a downstream flow that is shaped by the service acting on it (see Figure 2.1). In a queuing network model, downstream flows from one queue become the arriving flows of other queues, thus forming a network. We believe that queuing network models are promising as a centerpiece for FCM for several reasons.

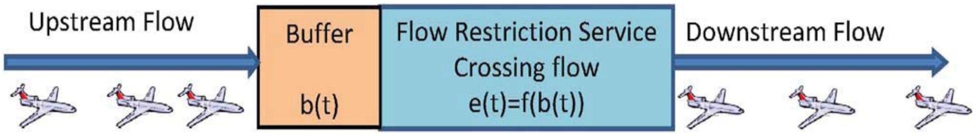


Figure 2.1. Queuing models.

(1) They naturally capture NAS uncertainties. Both schedule and weather uncertainty are prevalent in air traffic at longer LATs (see, e.g., [89] and [104] for analysis/ modeling of traffic uncertainties). Deterministic network models often yield management solutions that are not optimal or implementable under such uncertainty. Queuing network models naturally capture schedule uncertainty in a tractable way and can yield solutions that are robust to schedule and weather uncertainty.

(2) Queuing models naturally capture many current and planned traffic management strategies, which essentially are flow constrictions. Statistical analyses of queuing systems can permit a quantitative evaluation of these restrictions impacts on flows and coordinated design of flow restrictions for FCM.

(3) Queuing network models track flows and management in aggregate, rather than modeling individual aircraft motion. Aggregation reduces computation while providing sufficient fidelity for FCM decision making, which only requires generating high-level coordinated management strategies.

2.2.1. Literature Review: Queuing in Air Transportation

We review queue and queue-network models in air traffic, to provide a context and motivation for the development of a new sophisticated queuing network model for FCM. Queuing models have long been used to analyze uncertain local service constrictions, such as airport delays [60], taxi-out delays in particular [63], and backlog/delay caused by en route restrictions on a single traffic stream [88].

Of importance to our development, queuing network models have been considered in evaluating NAS performance. Some works model a network of airports with departing and arriving traffic represented as queues (see [80] and [115]). In [44], a NAS-wide simulation model, i.e., NASPAC, was developed to evaluate changes such as airport capacity variation, airport improvements, and demand growth. Later, in [77], a static queuing network model incorporating both airports and en route Centers was developed. In [120], an M/M/m Jackson Queuing Network capturing intra-Center flows was used to analyze NAS performance and route selection efficiency. Many of these models adopt Markovian service times and flows, for tractability. In [130], M/D/1 queues were used to model miles- and minutes-in-trail (MIT/MINIT), and linear network abstraction was developed thereof, allowing the design of flow restrictions in a steady-state situation. In [55], the use of models with Erlang-K service distributions for low-precision but flexible/tractable NAS-wide modeling and captured timevarying aspects of traffic was sought out. Analysis of such queue network representation is complicated: An approximate solution can be found in [80].

2.3. Queuing Models for Flow Contingency Management: Challenges

Existing queuing network approaches are not sufficiently sophisticated to achieve the operational goals of the FCM decision support system. Several key advances are needed. First, the model must capture routing options in the NAS (while still modeling traffic in the aggregate). Second, the queuing model must be reasonably easy to parameterize from data and to configure for practical needs. For instance, schedule information needs to be easily reflected in flows. In addition, appropriate aggregation is needed to balance computational costs with ease of transforming strategic decisions to tactical management actions. The most significant modeling challenge is aptly representing real management actions to allow design of effective yet realistic management strategies. Let us discuss this challenge in further detail.

2.3.1. Key Challenge: Modeling Management Actions

We consider five control actions (called traffic management initiatives or TMIs) that are currently used or are planned for use in NextGen: MIT/MINIT, rerouting, time-based metering (TBM), ground delay programs (GDPs), and airspace flow programs (AFPs). Here, we discuss how each can be captured using a queuing model for FCM. We emphasize that our models are abstractions of operational reality. Specifically, except for MIT/MINIT (which are intrinsically flow restrictions), we abstract current flight-specific definitions for TMIs as flow restrictions acting on aggregate flows.

Rate Restrictions (MIT/MINIT): MIT and MINIT specify the minimum allowable separation distance and traveling time, respectively, on a flow segment. An MIT/MINIT restriction can be captured with deterministic service time in a queuing model, e.g., as an M/D/1 or G/D/1 queue. Specifically, each aircraft in the coming flow is modeled as taking a fixed service time to pass a flow restriction point; any additional arriving aircraft will be waiting in the queue and will pass the restriction point on a first-come basis.

Routing: Aircraft are provided alternate routes to avoid convective weather or other adverse situations. To model routing, queuing network models need to distinguish among flows with different destinations, since route choices are destination dependent. Thus, basic queuing network models need to be enhanced to distinguish flow destinations. Then, routing can be abstracted as setting the fraction of flows in each possible direction at flow splitting points. TBM: TBM requires aircraft to be delivered to fixes at precomputed times, to reduce discrepancies between demand and capacity at arrival airports. The metering plan along an aircrafts route is assigned based on the destination airports arrival rate and is achieved via speed adjustment, vectoring, and holding. In contrast to MIT/MINIT, TBM is an aircraft-specific TMI. In aggregate, TBM may be abstracted using a G/M/n model: The multiple-server model captures that any aircraft meeting its metering schedule (as set by the arrival airport) is permitted to pass.

GDP: A GDP delays aircraft at their departure airport so as to resolve demand/capacity imbalances at an arrival airport. The delay assigned to an aircraft at the departure airport is determined by the arrival airports allowed rate. We abstractly capture a GDP as shaping the departure demand or restricting the departure airports rate, according to a queuing model such as G/Ek(t)/n or G/M/n. Because GDP implementation is tied to specific origindestination (OD) pairs, the queuing network model needs to distinguish origindestination pairs to capture GDPs.

AFP: AFPs manage aircraft that are scheduled to pass through constrained airspace through the assignment of estimated departure clearance times (EDCTs). An aircraft-specific EDCT is assigned based on the time required for each aircraft to cross an AFP-constrained area. AFPs can be modeled similarly to GDPs but only for flows intersecting the constrained airspace. Flows again need to be distinguished by their origin airports.

These models for traffic management actions will be integrated into the queuing network model in Section 2.4.

2.4. Queuing Network Model

Let us describe the new queuing network model for traffic flow and management, which is a centerpiece of the FCM solution.We also describe challenges in interfacing the network model with other FCM components and identify the analytical challenges that define major research directions.

2.4.1. Core Model

Let us start with a foundational model that only captures two TMIs, i.e., routing and MIT/MINIT. We discuss two general considerations in modeling and then present model details.

To capture routing (and GDPs), the queuing network model needs to distinguish origin–destination (O-D) pairs as separate subnetworks. Rates for the O-D pair subnetworks can be obtained from schedules, whereupon stochastic process models (e.g., Poisson processes) with these rates can be used to describe the flows. Since regional capacity constraints (and some rate restrictions) act on total flows across O-D pairs, the subnetworks must be overlapped and integrated (see Figure 2.2).

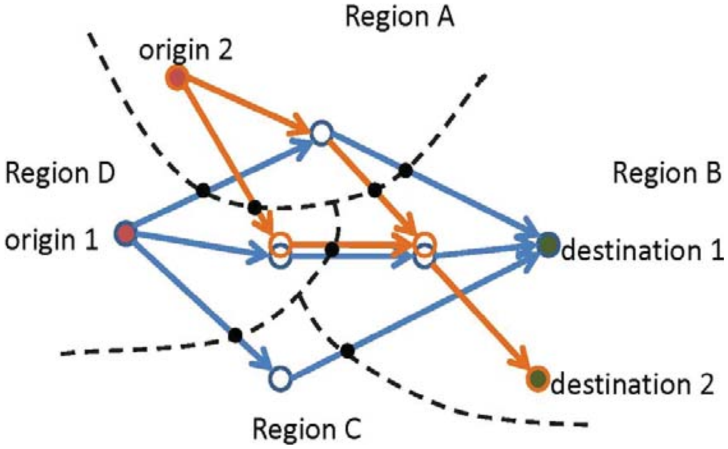


FIGURE 2.2. Preliminary queuing network structure. (Red nodes) Aggregated origin airports. (Green nodes) Aggregated destination airports. (White nodes) Flow splitting/merging points. (Black nodes) Boundary intersection points. (Links) Directed routes. (Dotted curves) Region boundaries. (Orange and blue colors) Subnetworks with different O-D pairs. Note that subnetworks are overlapped to complete a network model.

In the network model, nodes are classified into three categories: 1) flow merging/splitting points, 2) aggregated origin and destination airports, and 3) boundary intersection points. The definitions for the first two categories are straightforward and in general; we can consider a directed route as beginning at an origin airport (category-2 node), traversing through a number of flow merging/splitting points (category-1 nodes), and terminating at a destination airport (category-2 node). To understand the third category, let us consider the NAS as partitioned into indexed regions (e.g., sectors, clusters of sectors, etc.). The intersections between region boundaries and directed routes are called boundary intersection points. Let us denote a link as the portion of a directed route between two connected nodes belonging to category 1 and/or 2.We note that a boundary intersection point can be uniquely determined by the ends of a link and the region index that the flow enters. As such, when we refer to a node in this chapter, we mean a node belonging to categories 1 and 2 without specific clarification, whereas a link and a region index specify boundary intersection points. Second, we note that we formulate the model in discrete time, both to simplify model parameterization and simulation and to avoid technical complexities of continuous-time queuing models. Thus, traffic flows are captured as numbers of aircraft (on route segments or entering/exiting the airspace), and the queuing models previously described for management actions are discretized. Our previous studies [130], [151] indicate that such discrete-time models are apt approximations for management actions such as MIT/MINIT. To illustrate discretization, let us briefly consider the MIT/MINIT restriction, which we model as a G/D/1 queue. Discrete-time approximation works as follows: At each unit time interval, if backlog b (number of already-waiting aircraft) plus new inflow u (number of aircraft in the impinging flow) is larger than service rate N (maximum number of aircraft that can be served in a unit time), N aircraft pass to the downstream, and the backlog at the next time interval becomes $b-N+u$. If instead, $b+u$ is less than N, then all $b+u$ aircraft can pass to the downstream, and the backlog at the next time interval is 0.

Now, let us formally introduce the core queuing network model. Let us begin with a list of variables.

$f_{od}[k]$: Demand from origin o to destination d at time instance k. This demand is modeled as a stochastic process (e.g., Poisson process).

ijm: Boundary intersection point uniquely determined by link (i, j) and region m.

$f_{odij}[k]$: Flow rate (the number of aircraft per unit time) traveling from origin o to destination d, which are entering j along the link from node i to node j at time k. If this link does not exist in subnetwork o-d, $f_{odij}[k] = 0$. The same rule applies to all other variables.

$g_{odij}[k]$: Time-k flow rate from origin o to destination d, leaving node i along the link from i to j.

$b_{ijm}[k]$: Total backlog of the flow from node i to node j at the boundary of region m at time k.

$u_{ijm}[k]$: Rate of flow from node i to node j, right before entering the boundary of region m at time k.

$d_{ijm}[k]$: Rate of the flow from node i to node j that has just crossed the boundary of region m at time k.

p_{odij}: Fraction of the flow from origin o to destination d entering the link from node i to node j.

N_{ijm}: Number of aircraft allowed to enter region m along path (i, j) in a unit time.

N_{id}: Number of aircraft per unit time allowed to enter destination d from node i through link (i, d).

$k_{i,j}$: Number of time steps for an aircraft to travel from node i to node j (along a direct connection). Here, the nodes may include boundary intersection points.

Γ_{ijm}: Capacity for link (i, j) in region m, i.e., the total number of aircraft allowed to pass through ijm along link (i, j) in a unit time.

Γ_m: Total capacity for region m, i.e., the total number of aircraft allowed to enter region m in a unit time. Defining region capacity based upon entering flights naturally captures controller workload.

Γ_d: Total capacity for destination airport d, i.e., the number of aircraft allowed to enter d in a unit time.

Variables $f_{od}[k]$ represent inflows to the network. Variables 3–7 are state variables of the queuing dynamics. Flow variables f, g, b, u, d are stochastic processes. Variables 8–10 are control variables to capture management strategies, and variables 11–14 are system constraints. We note that variables 8–14 may be slowly time varying; however, we suppress time dependence in our notation, to avoid confusion with state variables.

Using the given notations, the flow network model is described by the following dynamic equations and constraints. At each flow junction point j, the total inflow and the total outflow are equal. Here, the four-character subscript follows the notations 3, 4, and 8, and l denotes a downstream region of j, i.e.,

$$g_{odjl}[k] = p_{odjl} \sum_{i} f_{odij}[k] \tag{1}$$

$$\sum_{l} p_{odjl} = 1 \tag{2}$$

$$1 \geq p_{odjl} \geq 0 \tag{3}$$

As a special case, at origin airport o, we have

$$g_{odoi} = p_{odoi} f_{od}[k] \tag{4}$$

Here, p_{odoi} is subject to two constraints, i.e.,

$$\sum_{i} p_{odoi} = 1 \tag{5}$$

$$1 \geq p_{odoi} \geq 0 \tag{6}$$

Note that fraction variable p_{odjl} models rerouting.

At each boundary intersection point ijm, an MIT restriction (or other flow restriction capability), which is denoted by N_{ij} , shapes flows associated with all source/destination pairs, reducing downstream flow (d) at the cost of increased backlog (b), i.e.,

$$b_{ijm}[k+1] = max(0, b_{ijm}[k+1] + u_{ijm}[k+1] - N_{ijm}) \tag{7}$$

$$d_{ijm}[k+1] = min(N_{ijm}, b_{ijm}[k+1] + u_{ijm}[k+1]) \tag{8}$$

$$N_{ijm} \leq \Gamma_{ijm} \tag{9}$$

$$\sum_{i,j} N_{ijm} \leq \Gamma_m \tag{10}$$

Equations (7) and (8) describe the discrete-time G/D/1 approximation, whereas (9) and (10) enforce capacity constraints on flow restriction variables. Let us next model the flows to a destination airport, which we assume is associated with a particular region. Using the same index d for the airport (as a flow merging point) and region, for convenience, we have

$$b_{idd}[k+1] = max(0, b_{idd}[k] + u_{idd}[k+1] - N_{id}) \tag{11}$$

$$\sum_{i} N_{id} \leq \Gamma_d \tag{12}$$

We also assume that aircraft travel at a constant speed between two connected nodes, including boundary intersection nodes. Thus, network flows evolve as follows:

$$u_{ijm}[k] = \sum_{all(o,d)pairs} g_{odij}[k - k_{i,ijm}] \tag{13}$$

$$\sum_{all(o,d)pairs} f_{odij}[k] = d_{ijm}[k - k_{ijm,j}] \tag{14}$$

$$u_{ijn}[k] = d_{ijm}[k - k_{ijm,ijn}] \tag{15}$$

$$f_{odij}[k] = g_{odij}[k - k_{i,j}] \tag{16}$$

The given network model allows the evaluation and design of routing and MIT/MINIT. Specifically, parameters p_{odjl} capture routing strategies, and N_{ijm} models MIT/MINIT strategies. For simulation, state variables, including the inflow, backlog, and cross-flow, are tracked at all OD specific nodes according to the given recursive equations. When multiple flows are subject to a single constraint or MIT restriction, the restriction on each flow is subject to the following rules, with the consideration that flows with accumulated backlogs should have the high priority to pass: If the total backlog of all flows is greater than the total available constraint/MIT restriction, each flow is assigned a control restriction that is proportional to its backlog; otherwise, all backlogs are allowed to cross, and the remaining control restriction for each flow is assigned as proportional to its inflow.

Our queuing network model differs from existing flow/queuing models in several respects.

- The model captures flows as stochastic processes and, hence, traces stochastic backlogs and downstream flows. Thus, congestion (measured by the backlog) is modeled more realistically than in deterministic models (e.g., [90]). Moreover, time delays (widely used in practice) can be easily calculated. In particular, total delay over a time span equals the interval duration multiplied by the summed backlogs over this span.
- The overlapped OD pair flow network model structure allows performance analysis and the design of management actions. We explicitly incorporate control parameters that reflect in-practice management actions, in contrast with other models that only implicitly capture management (e.g., [25]).
- The model is dynamic and can capture the transient dynamics of the NAS, in contrast with most queueing approaches in air transportation. This is particularly important for understanding the impact of transient weather events.
- Weather uncertainties can be transformed to stochastic descriptions of capacity constraints to facilitate weather driven FCM design (see [151] and [155]).

The queuing network model is flexible, in that it can be configured for different FCM needs. For instance, we can model layered traffic by defining subnetworks for different altitudes. We note that the model abstracts away details such as variable aircraft speeds, aggregates airports/regions, etc. Since FCM is focused on a NAS-wide long-LAT resource distribution, we believe that these abstractions are needed given the high uncertainty. However, extensive tests are needed for validation (see Section 2.5.2 for preliminary verification).

2.4.2. Model Enhancement: Capturing GDP/AFP/TBM

Let us enhance the given core model to incorporate GDPs, AFPs, and TBM to facilitate their analysis and design.

As previously discussed, a GDP at destination airport d can be modeled using G/M/n queues at each impacted origin airport o. Discrete-time approximation is used here. In particular, we replace (4) with the following set of equations:

$$b_{od}[k+1] = max(0, b_{od}[k] + f_{od}[k+1] - M_{od}[k]) \tag{17}$$

$$g_{odoi}[k+1] = p_{odoi} min(b_{od}[k] + f_{od}[k+1], M_{od}[k]) \tag{18}$$

Here,$M_{od}[k]$ is Poisson-distributed with rate N_{od}. $b_{od}[k]$ represents the GDP backlog for the origin-destination pair. A GDP is designed by selecting multiple such destination airports d and specifying M_{od} for all paired (impacted) origin airports o. We note that (17) and (18) describe the flows at impacted origin airports. These restrictions placed do not propagate impact to the destination airport until nominal flight time has passed, as governed by (13)-(16).

AFPs are modeled as reshaping the stochastic departure rates for all origins o that have flows intersecting with the weather (restricted) zone w, using a discretized G/M/n queue. The following two equations describe this process:

$$b_{odoi}[k+1] = max(0, b_{odoi}[k] + p_{odoiw} f_{od}[k+1] - M_{odoiw}[k]) \tag{19}$$

$$g_{odoi}[k+1] = min(b_{odoi}[k] + p_{odoiw} fod[k+1], M_{odoiw}[k]) + (p_{odoi} - p_{odoiw}) f_{od}[k+1] \tag{20}$$

Here, $M_{odoiw}[k]$ is Poisson-distributed with rate N_{odoiw}, $b_{odoi}[k]$ represents the backlog associated with the flow that travels from origin o to node i and then finally to destination airport d, and p_{odoiw} specifies the percentage of flow that intersects with the weather zone among the flows from origin o to node i and then to destination airport d. $p_{odoiw} = \sum_{\forall j,\ldots,lm} p_{odoi} p_{odij} \ldots p_{odlm}$ such that m is located in the weather zone. An AFP is designed by selecting the weather zone and designing the rates Nodoiw for intersecting flows. An alternative modeling scheme that facilitates the modeling of AFP can be found in [125].

To represent TBM, we use discretized G/M/n queues to represent control at metering fixes, which we assume are collocated with boundary intersection points. At a fix, there might be multiple queues, each acting on the flows to a particular destination d. At a fix ijm where TBM is implemented, (7)-(12) are replaced by

$$b_{ijmd}[k+1] = max(0, b_{ijmd}[k] + u_{ijmd}[k+1] - M_{ijmd}[k]) \tag{21}$$

$$d_{ijmd}[k+1] = min(M_{ijmd}[k], b_{ijmd}[k] + u_{ijmd}[k+1]) \tag{22}$$

$$\sum_{d} N_{ijmd} \leq \Gamma_{ijm} \tag{23}$$

$$\sum_{i,j,d} N_{ijmd} \leq \Gamma_{m} \tag{24}$$

Here, $M_{ijmd}[k]$ is Poisson-distributed with rate N_{ijmd}; $b_{ijmd}[k]$, $u_{ijmd}[k]$, and $d_{ijmd}[k]$ represent the backlog and flow rates just prior to and after ijm, for the flows with destination airport d. For TBM at the destination airport, we have

$$b_{iddd}[k+1] = max(0, b_{idd}[k] + u_{iddd}[k+1] - M_{id}[k]) \tag{25}$$

assuming that a destination airport can be viewed as a region and a flow merging point, with the same index. Moreover, $M_{id}[k]$ is Poisson-distributed with rate N_{id}, and

$$\sum_i N_{id} \leq \Gamma_d. \tag{26}$$

$u_{ijmd}[k]$can be obtained through the following equations describing the en route behavior. Thus

$$u_{ijmd}[k] = \sum_{allo} g_{odij}[k - k_{i,ijm}] \tag{27}$$

$$\sum_{allopairs} f_{odij}[k] = d_{ijmd}[k - k_{ijm,j}] \tag{28}$$

$$u_{ijnd}[k] = d_{ijmd}[k - k_{ijm,ijn}] \tag{29}$$

$$f_{odij}[k] = g_{odij}[k - k_{i,j}] \tag{30}$$

For simplicity, we may also assume only one restriction placed at a metering fix that acts on flows to all destinations. In this case, we simply replace the G/D/1 constraint [see (7) and (8)] at each boundary intersection point with the G/M/n constraint. The details are straightforward and, hence, are omitted here.

In sum, the given queuing network models are general and flexible high-level models that define the progression of interacting flows subject to control actions and network constraints. Depending on the problem of interest, the model can be simplified. For instance, p_{odij} can be restricted to take either a 0 or 1 value, if only one route is used for an OD pair. Moreover, flow merging/splitting points and boundary points may be merged. In general, flow restrictions in the network may be considered designable or not: Management actions constitute designable restrictions, whereas restrictions caused by external factors such as weather may not be.

2.4.3. Integrating the Queuing Network into the FCM Framework

A detailed description of our FCM concept and framework is given in [123]. The purpose of our effort is to predict stochastic demand-capacity imbalance and, in turn, to evaluate and design TMIs, using a coarse queuing network framework. We summarize the framework (see Figure 2.3) and describe the queuing network model integration, to explain its use in flow contingency analysis and design.

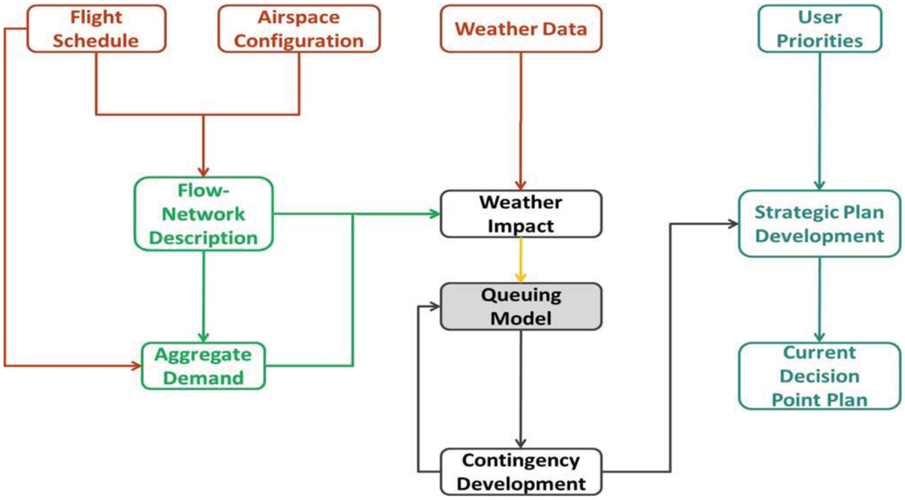


Figure 2.3. Complete FCM framework [39].

In Figure 2.3, FCM framework elements are grouped into four components. Data elements (shown in red) define the information requirements for FCM, namely, flight schedule, airspace configuration, and weather data. Model-building elements, which are shown in green, are concerned with network- and aggregate-demand definition. These model-building elements translate flight schedules and relevant airspace configuration data into forms that can be used for flow contingency planning.

The flow contingency development component, which is shown in black, derives and evaluates flow contingency plans for FCM from the data and built models. It has three pieces, namely, a weather impact predictor, a queuing model, and a flow contingency development tool. The weather impact model analyzes the weather and configuration data to identify evolving weather impacts on NAS parameters (capacities). This information, as well as the aggregated demand, is provided to the queuing model, which simulates traffic under weather impact. This simulation is leveraged to define the flow contingency plans for the different potential weather impact outcomes. These plans are then relayed to the strategic planning component (shown in blue), which consists of user priority inputs, a formal framework for strategic planning, and an incremental decision–planning tool that defines the current point plan.

The queuing network model is integral to the capabilities envisioned for FCM, i.e., the simulation and design of control actions to mitigate performance loss in the presence of weather impact. Let us now discuss issues related to integrating the model and research needs in using the model for the evaluation and design of FCM strategies.

(1) Inputs to the Queuing Network Model: The proposed queuing network model can be adapted to network descriptions with any aggregation level. To use the queuing model for FCM, we need to first parameterize the queuing network to reflect the operational reality of the planning period under consideration. We call the param-

eterized model a queuing network instance. To obtain a queuing network instance, we require the following inputs from the other modules within the FCM framework.

(a) Flow network structure from the flow network description module: Appropriate aggregation levels for queuing model elements (e.g., nodes and routes) are needed to allow the design of management actions at a meaningful scale for FCM, whereupon the linkflowtime parameters can be estimated. The aggregation level for flows should be closely tied to congestion (which may or may not be related to weather severity but does not have to be). Often, multiresolution [122] or even dynamic networks are thus appropriate.

(b) Demand rates (e.g., f_{od}): Stochastic process demand models must be constructed from data. Time-varying Poisson flow models are often used (see, e.g., [88] and [137]) for some justification.

(c) En route, regional, and destination capacity constraints (γ_{ijm}, γ_m, and γ_d): These parameters are provided by the weather impact model (see, e.g., [150] and [112] for the development).

(d) Controls and other flow restrictions (e.g., MIT restriction N_{ijm}, GDP restriction N_{od}, etc.): These parameters may be design variables but may be also impacted by uncertain weather indirectly through capacity variables. These indirect impacts are discussed in [125, 151, 155].

If we initiate the queuing model early in the morning (at a low traffic time), the given inputs are sufficient to develop and simulate a model instance. However, if midday starts are needed, initial conditions for the queuing network models state variables are required for a real-time simulation. Such initial information can be obtained either from historical rates or from a combination of schedules, filed flight plans, and currently airborne flights. Specifically, information on airborne and other flights with precise schedules can be translated to deterministic flow rates, whereupon management actions operate on the deterministic and stochastic flows together. However, if this information already accounts for management actions, care must be taken such that the action is not reapplied.

(2) Simulation and Design Outputs: Direct outputs of the queuing network simulation include the time-series statistics of inbound flow, crossing flow, and backlog at all management nodes in the network. These statistical data capture the transient dynamics of the NAS in response to dynamical weather uncertainties. From these data, important decision-making indicators, such as airport delay, airport throughput, sector count, and sector backlog, can be obtained, which can assist in FCM design. Meanwhile, design outputs from the queuing network model are in the form of aggregated rates at over time and location, for management actions of interest (see Table 2.1 for relevant outputs).

    Other outputs include stochastic weather generated from the integrated flow and weather simulator, prediction of future policies associated with different weather scenarios, performance statistics estimates, and possibly partitioning plans for the tactical time frame. More work is needed to effectively translate the aggregated rates to contingency plans with concrete flow actions that are understandable and implementable at the Air Traffic Control System Command Center and other NAS facilities.

(3) Performance Evaluation and Optimal Design: With the given queuing network model in place, we are developing methodologies to evaluate and design effective FCM strategies with respect to various performance measures of interest to NAS stakeholders. Total cost measures that sum backlog/delay costs across the NAS (or multiple critical regions) are of interest. However, many more complicated issues

Management Actions	Design Output Variables	Notes
MIT/MINIT	N_{ijm}	Flow rate entering each region boundary
	N_{id}	Flow rate entering each destination
Routing	p_{odij}	The flow fraction in each direction at the flow splitting/merging nodes
GDP	N_{od}	Departure rate of flow at each origin that is paired with certain destination
AFP	N_{odoiw}	Departure rate of flow at each origin that is paired with certain destination and intersects with the weather zone
TBM	N_{ijmd}	Rate for flows entering each region boundary and destined to a destination
	N_{id}	Flow rate entering each destination

Table 2.1. Output Variables of the Queuing Network Model

may also influence cost calculation. For instance, en route delays and ground delays will incur different costs. Routing introduces extra costs due to the extra en route time and fuel costs [24]. Each implementation for human controllers and pilots should be accounted for in the cost function, and fairness issues may also need to be considered (see [111] for discussion on fairness/optimality tradeoffs). A full study of the cost definition is left for future work.

(4) Analytical Challenges: We expect to use the queuing network model for simulation, analysis, and design of FCM. As there is significant weather uncertainty at longer LATs, the performance evaluation and design of FCM strategies must take into con-

sideration the range of possible weather scenarios. Queuing models are promising for permitting efficient performance evaluation and design for FCM, without resorting to Monte Carlo simulations. Here, we list critical gaps to such analysis in our case and discuss incipient research to fill these gaps.

(a) Analysis of queuing (network) models (e.g., G/D/1 and G/M/n) with (weather-impacted) stochastic service rates. Noting that weather impact can be captured using Markov models, we are interested in analyzing queues with Markov-modulated service processes. Some steady-state analyses of such queuing processes are available (e.g., Markov-modulated M/M/n queues [14]). Results for other queuing disciplines and queues are needed.

(b) Transient analysis of queuing (network) models. When severe weather events occur, NAS dynamics are usually far from a statistical steady state. Unfortunately, precise transient analyses of queuing models are usually computationally costly, and approximations are needed. In [101], such approximations are studied for transient statistical analysis of landing processes at a network of airports.

(c) Design of multifaceted management strategies under weather uncertainty. Flow restriction design is complicated by the inherent nonlinearity, stochasticity, and the time-varying nature of the FCM queuing model. We are pursuing approximation studies as well as dynamic programming and jump-linear-control-based approaches to address the design problem (see, e.g., [40] and [109] for background on jump linear systems/control).

(d) The robustness of strategic time frame management to tactical time frame implementation. Tactical management actions are used to address short-term local constraints and weather forecast changes. Tactical decision making must

> be easily incorporated within the strategic FCM framework. TFM plans need to be decoupled in such a way that tactical modifications do not introduce huge disturbances to overall NAS performance/management. We believe that a clustering approach can be used to identify/exploit tactical information-sharing needs, to decompose NAS management into semistandalone problems. We will evaluate the impact of stochastic weather events on this decomposition and design a dynamic and rational decomposition algorithm that is weather dependent and identifies partitions of the NAS that require communication.

We are in the process of developing analytical tools to address these new needs. Specifically, for a stochastic flow entering a probabilistic weather zone, we have developed a jump linear approach and a Master Markov approach to effectively predict the statistics of transient delay dynamics. Moreover, we have developed an integrated probabilistic collocation method and a jump linear approach to evaluate and design management strategies under weather and demand uncertainties (see [151] and [155] for the complete study).

2.5. Illustrative Example

We present three examples. The first two examples are small-scale constructed case studies that illustrate analysis and design capabilities using the queuing network model. The third example highlights the models application to evaluate NAS-wide congestion on a historical bad-weather day, i.e., September 26, 2010.

2.5.1. Four OD Pair Network Example

In this example, we evaluate congestion and TMI impact in a four OD pair network (see Figure 2.4). The network has three sectors (S3, S5, and S6), two origin airports (nodes 1 and 2), two destination airports (nodes 11 and 16), and numerous sector boundary points.

The four OD pair subnetworks are indicated by different colors. The demands at departure airports 1 and 2 follow Poisson distributions, with rates shown in Figure 2.5. We have considered several simulation cases to study congestion resulting from sector capacity reduction and its alleviation using TMIs. In each case, flow statistics (including backlogs, sector counts, and arrival/departure throughput and congestion) have been obtained over 24 h at 15-min intervals using 100 Monte Carlo simulations.

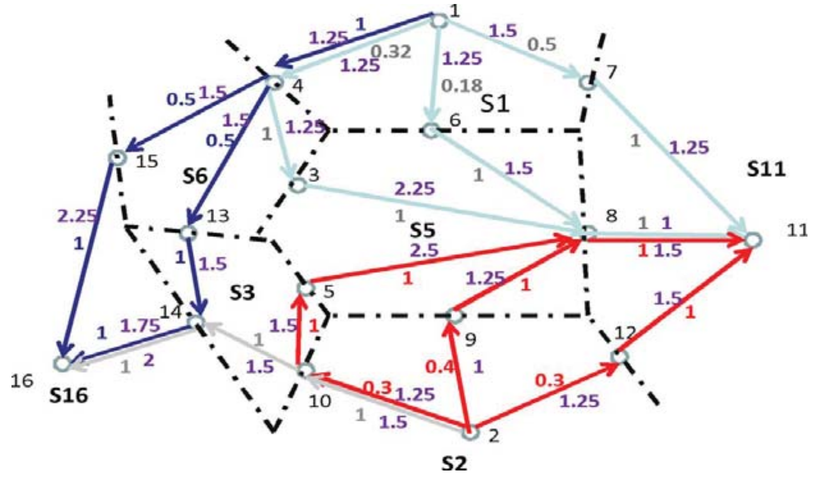


FIGURE 2.4. Four OD subnetwork example. Sector boundaries are marked by dashed lines. The four OD subnetworks are differentiated by the colors of edges. Flow fractions are marked on the edges. Purple numbers marked on edges indicate transit time measured in hours.

Nominal operations are simulated first. In this case, airports 1 and 2 have departure rate constraints of 8 and 7 aircraft per time interval, respectively. Airports 11 and 16 are subject to arrival rate constraints of 3 and 4 per interval, respectively. Sectors 3, 5, and 6 have capacity constraints of 4, 3, and 3 aircraft, respectively (where capacity represents the maximum allowed inflow). We also compare the nominal operations scenario with a reduced-capacity scenario, where the capacity for sector 5 is reduced to one aircraft per

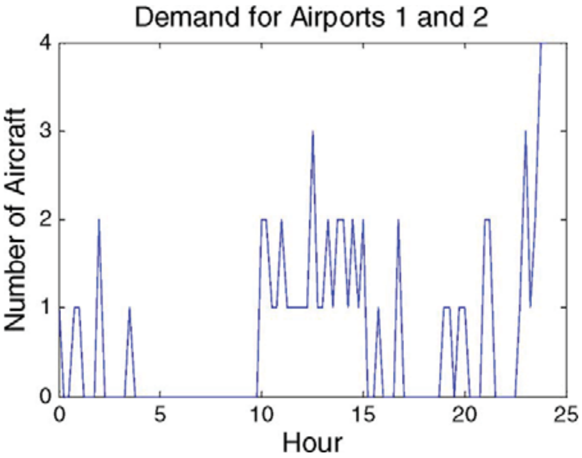


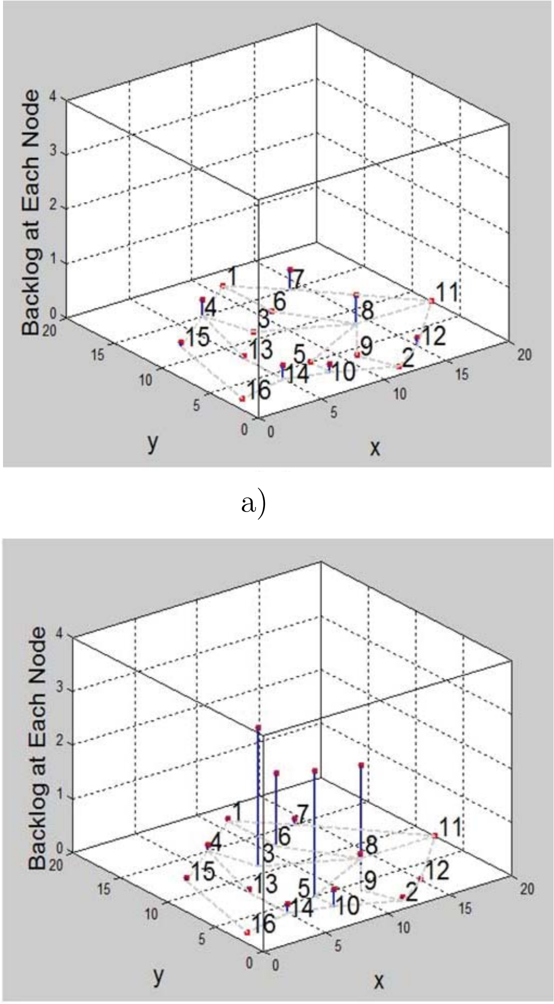
Figure 2.5. Mean demand rates for airports 1 and 2 in example 1.

interval during the time span 14:00–17:00 (e.g., due to bad weather).

Figure 2.6 shows snapshots of a mean-node-backlog video for nominal and reduced-capacity cases. Such 3-D videos provide direct visualization of spatiotemporal impact of capacity reduction. In particular, capacity reduction causes congestion, as reflected by the backlog accumulated at sector-entry nodes 3, 5, 6, and 8. Congestion in sector 5 is also indicated by the saturation of the mean sector count in the reduced-capacity case during 14:00–17:00 (see Figure 2.7). Congestion of this magnitude would require tactical resolution, which can be very costly. To avoid tactical mitigation responses, we consider three possible strategic management strategies that are promising for alleviating the congestion.

1) Routing: We apply rerouting to reduce the flows into sector 5. Specifically, during 13:00–16:00, the flow fractions from nodes 1–4, 1–6, and 1–7 in the OD subnetwork 1–11 are changed to 0.1, 0.1, and 0.8. In addition, during the same time span, the flow fractions from nodes 2–9, 2–10, and 2–12 in the OD subnetwork 2–11 are changed to 0.1, 0.2, and 0.7. Figure 2.7 shows that rerouting is effective: The aircraft count in sector 5 remains below the threshold during the capacity–reduction event (see red curve).

2) GDP: As an alternative control, GDPs are enforced on the OD subnetworks 1–11



a)

b)

Figure 2.6. Snapshots from the videos of backlogs at all nodes. x and y dimensions show the location of the nodes, and z dimension shows the mean backlog for 100 sample runs. (a) Case 1: Normal condition. (b) Case 2: Reduced-capacity condition.

and 2–11 (whose flows traverse sector 5), with a reduction in both departure rates to 2 during 12:00–16:00. Comparing the arrival backlog and departure backlog in Figure 2.8 (a) and Figure 2.8 (b), we see that the GDP delays flows arriving at airport 11 to reduce the demand–capacity imbalance, albeit at a cost of introducing departure backlog.

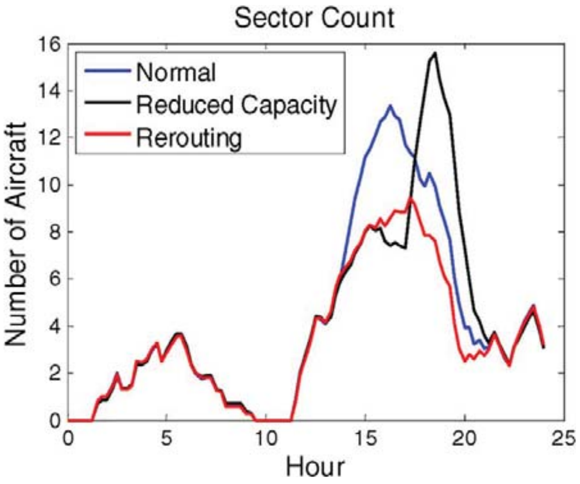


FIGURE 2.7. Comparison of time-course mean sector counts sector 5 among case 1 (normal condition), case 2 (reduced-capacity condition), and case 5 (rerouting in action), respectively.

3) MIT: In this case, we apply MITs to several flows upstream of sector 5. Nodes 4, 10, 6, and 9 are assigned MIT rates of 2, 2, 0.2, and 0.2 aircraft per time interval from 13:00 to 16:00. The arrival backlogs in Figure 2.8 (c) indicate that appropriately placed MITs delay traffic to both destination airports but avoid tactical impacts due to bad weather.

2.5.2. Managing Multiple Flows Under Stochastic Weather

This example is concerned with traffic flows to a cluster of nearby airports that will be impacted by a stochastic winter-weather event. The majority of the traffic arriving at this cluster of airports can be viewed as coming from four different origins (i.e., for the purpose of flow management design, the source airports can be clustered into four groups with traffic flow originating from each). In this case, we will model the demand for each of these flows as Poisson processes with average rates of 25 aircraft per hour (major flow), 12 aircraft per hour (moderate flow 1), 11 aircraft per hour (moderate flow 2), and two aircraft per hour (minor flow of very-long-distance traffic), respectively, for the duration of the planning horizon (see

Figure 2.9). We note that the average demand for the destination cluster is thus 50 planes per hour. During good weather conditions, the capacity of the airspace near the destination airport is more than sufficient to handle this demand. Note that in real operations, the flow rates would be time varying, and the queuing model handles this case as well; however, for simplicity, this example uses time-invariant flow rates.

A winter-weather event is anticipated in the destination airspace over a 7-10-h time period. During the winter-weather event, terminal area and en route constraints will reduce the capacity of the destination airspace to 35 aircraft per hour. Thus, a flow contingency plan is needed to address the demand–capacity imbalance. The exact duration of the winter weather is uncertain; however, a 10-h event (e.g., between noon and 10 P.M.) is a conservative estimate. In particular, the end time of the event is uncertain, and the winter weather may cease as early as 7 P.M. but no later than 10 P.M. As such, the uncertain end time is modeled as uniformly distributed between 7 P.M. and 10 P.M.

Several management capabilities are available to handle the capacity-and-demand imbalance. First, a GDP can be implemented on all or a subset of the major flow and two moderate flows. In the case where only a GDP is used, it must be implemented for the entire potential 10-h weather event, since the distance of the origin clusters to the destination does not permit reaction to a shorter-duration weather event. We note that the GDP will tend to reduce flow rates proportionally (in an average sense) on the impacted flows. We also assume that another rate-reduction capability is available for the major flow (specifically, an AFP or perhaps a future airspace construct) that allows us to tune the rate on the major flow compared with the other flows. This capability must be also implemented for the maximum possible duration of the event if used as the sole management response. Finally, an en route flow restriction can be implemented, which acts together on the major flow and moderate flow 1 (in the form of a MINIT restriction). In particular, given the uncertainty in end time of the weather event, we also have the option to implement a GDP for flows arriving over a shorter time interval (e.g., 7 h) and then use the MINIT restriction to modulate the rates thereafter; the en route restriction has the advantage that it can be flexibly implemented based on the weather condition (although at the cost of accumulating en route backlog, as opposed to ground delays at the origin cluster, if insufficient capacity exists).

This example can be directly formulated in the queuing network framework previously introduced, and we can compare several flow management strategies. In particular, we will use the queuing model to easily compute the delays and backlogs incurred by the implementation of each strategy to enable proper strategy selection. Specifically, we consider the following five possible strategies for flow management:

Strategy 1: Use a GDP that acts on the major flow.

Strategy 2: Use a GDP that acts on the two moderate flows.

Strategy 3: Use a GDP that acts on the major flow and the two moderate flows.

Strategy 4: Use a GDP that acts on the moderate flow and an AFP on the major flow to tune the flow rate on the major flow compared with the others.

Strategy 5: Use a shorter-duration GDP (for all three flows) together with an en route restriction.

We will also compare these strategies with a nominal case where strategic flow management is not used or is put in place in an ad hoc fashion.

Let us compare the performance of the strategies. Figure 2.10 - Figure 2.12 show statistics of the numbers of aircraft delayed at each source airport and statistics of the total number of aircraft delayed as a function of time. In particular, for each of the five GDP strategies, the mean numbers of aircraft delayed (waiting) at each source airport is plotted versus time, as is the mean total number of waiting aircraft. All aircraft can be delivered to the destination, given long enough time, as management strategies impact on flows is transient. We note that these statistics can be obtained through a jump linear approach or a Master Markov approach, as given in [151] and [155], or through a Monte Carlo simulation.

These preliminary results demonstrate that delays can be effectively distributed among source airports through the use of GDPs that concurrently delay multiple flows, albeit at a cost of increased complexity for traffic managers. The proper employment of MIT reduces ground delay at the origin airport but introduces possible large en route delay if the winter weather lasts longer than expected. This can be seen from the arrival backlog for Strategy 5 [see Figure 2.12(b)], assuming that the airport arrival rate (AAR) is 35 aircraft per hour. Noting that the MIT smoothens flows so that the maximum flow rate meets the AAR (rather than only enforcing that the average arrival flow rate is correct, as for the other strategies), the comparison of the MIT with the other strategies is perhaps not quite fair. This also suggests that a rigorous study of performance metrics is crucial. We leave the details to future work.

Finally, let us briefly compare the performance of the flow management strategies with the nominal strategy that does not initiate management actions and a strategy that mimics current operating practice in the NAS. If no FCM strategy is implemented, a significant capacity imbalance would result at the destination, with up to 150 aircraft being forced to wait in the air to enter the capacity-constrained airspace. In practice, current operators would not permit such an imbalance, since many flights would be forced to divert to alternate airports. Instead, a ground stop or GDP would be imposed. However, current operators would not have the benefit of comparing different possible strategies in terms of delays/backlogs incurred. Furthermore, it is quite possible that their enacted strategy would be reactive to the weather conditions-possibly leading to excess delay and unnecessary cancellations. A primary benefit of queuing model representation is its use in quickly estimating costs associated with different possible flow management strategies, thus facilitating contingency selection.

2.5.3. Atlanta Example

Our final example highlights the extensibility of this methodology to realistic problems of interest. We begin by verifying the queuing model simulation behavior for a clear-weather day (August 30, 2010) and then examine how the model captures capacity-demand imbalances on a bad-weather day (September 26, 2010) for traffic between 5:00 A.M. and midnight. Using a probabilistic forecast of weather for September 26, 2010, we generated different representative weather scenarios using the influence model [150] and simulated the NAS behavior with the traffic demand from August 30, 2010. The overall research is exploring how to define management strategies for these conditions, as described in the previous examples; however, as the focus of this chapter is not on the simulation but on the introduction of the dynamic queuing network model, we only illustrate the extensibility of the model to real-world problems.

On a bad-weather day in question, there was significant convection and low ceilings within and around ZTL, and as ZTL is the area most affected by the weather event, we focused on the traffic in ZTL. Defining ZTL as the area of interest, we abstracted the airspace outside ZTL at the Center level and the airspace in ZTL at the sector level. This abstraction significantly improves tractability without causing much loss of relevant information. Similarly, airports outside ZTL are aggregated, as described in [137], and airports within ZTL are individually represented. This resulted in a NAS-wide network with 68 airport nodes (with 22 within ZTL); 260 regions (including 68 departure airports, 68 arrival airports, all sectors within ZTL, and centers outside ZTL), which produced a network with 1074 nodes (including 68 departure airports, 68 arrival airports, and all of the sector boundary points); and 16 243 OD specific arcs.

Using this network, we first compare the accuracy of the queuing network model for simulating traffic flow. Comparing the actual arrivals at ATL on August 30, 2010 and the simulated arrival rate, Figure 2.13 shows that despite the significant abstraction in network formulation, the queuing model provides a reasonable estimation of traffic flow throughout the NAS.

We now compare the impact of the weather-reduced capacities generated for ZTL with the nominal conditions. Figure 2.14 shows the snapshots of the sector backlog (the summation of all backlogs accumulated at the nodes on a sectors boundary) for the normal weather and one bad-weather scenario, at a particular time. We see from the snapshots that there is an increasing backlog in ZTL due to the occurrence of bad weather. We will leave systematic model validation to future work.

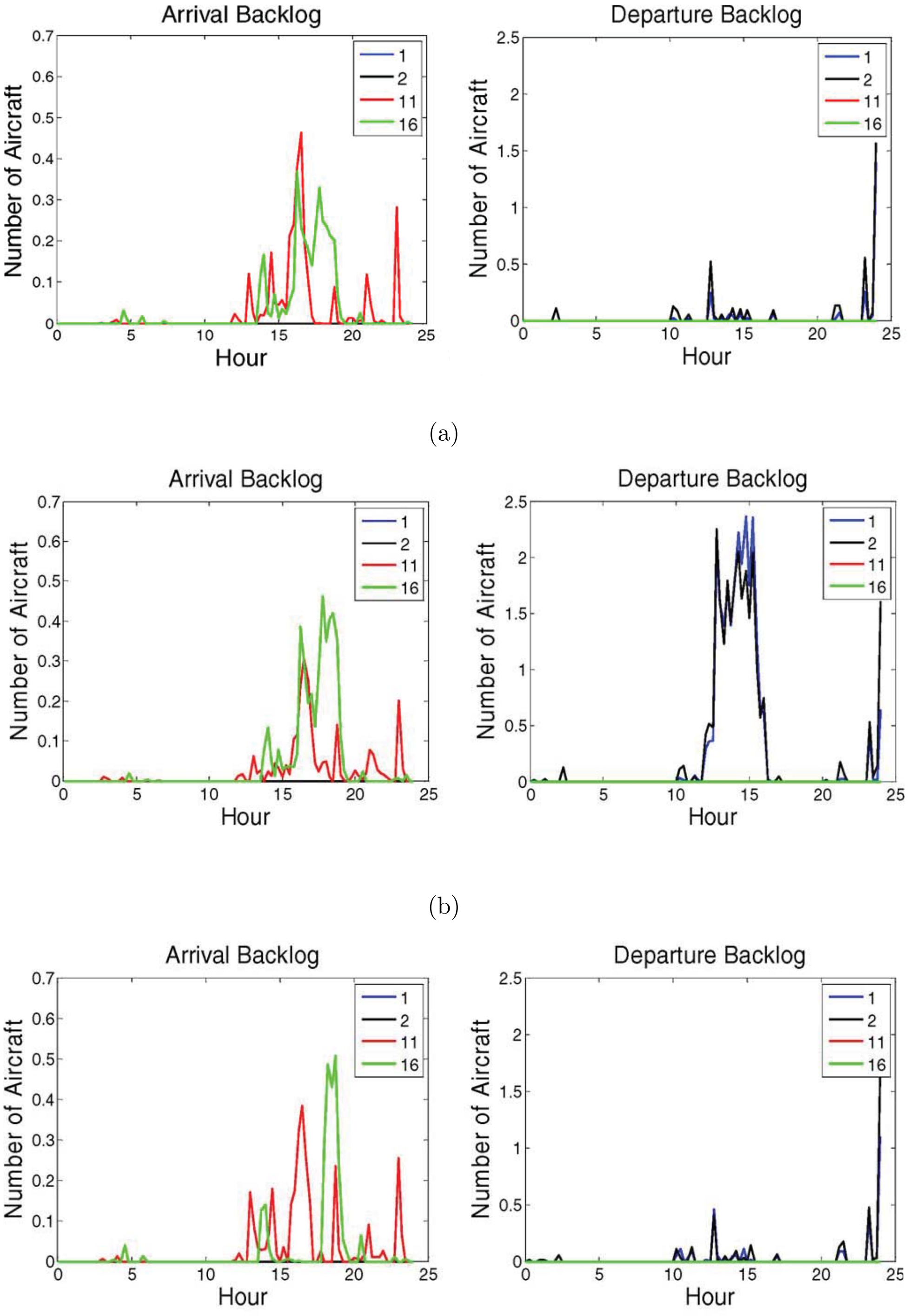
Finally, we note that queuing model simulation is computationally efficient as only the flow dynamics at management locations are tracked. For this example simulation, which was implemented in MATLAB, a single sample run takes less than 1 min on a common university desktop computer that is typically used for teaching purposes. We leave more rigorous estimation of computational times and modification of the algorithm to maximize computational efficiency to future work.

2.6. Alternative: A Route-Based Dynamic Queuing Network Model

The dynamic queuing network model is the core piece for the framework of the Flow Contingency Management. Its key idea is to decompose the whole traffic network in the NAS into sub-networks identified by different O-D pairs. For the traffic demand propagating from one particular original to the destination, it is naturally to understand that there are multiple routes available. As such, we also propose a concept of route assignment for managing aggregate air traffic demand and evaluating the contingency plan at strategic time-frame. The idea of the so-called route-based dynamic queuing network model is similar to that of the O-D specific one. The whole network is decomposed into route specified sub-networks. The traffic demand feeds into one particular O-D pair will be assigned to multiple available routes. Each route contains original node, a sequence of intermediate nodes and destination node. Each node is associated with predefined sector with capacity value. The route-based dynamic queuing network also permits us to simulate the traffic propagation behavior in the NAS at flow level and to analyze possible strategic management plans. The advantage of this modeling framework is to facilitate the re-routing management action since all of its components are route specifically defined. We also use realistic data to validate the feasibility of the this model for the sake of our FCM design. The system-wide performance of this model is equivalent to that of our original OD specified dynamic queuing network model. We envision those two models will complement each other to effectively deal with the FCM tasks.

2.7. Concluding Remarks and Future Work

We have introduced a queuing network model for air TFM. The developed model is an effort to comprehensively capture the dynamics of air traffic and traffic management at the level of traffic flows to permit evaluation and design of management strategies at a NAS-wide scale and strategic time horizon. As such, our queuing network model enhances existing modeling capabilities significantly in several directions, including by permitting 1) representation of a wide range of current and potential future traffic management actions, 2) analysis of traffic flows under weather uncertainty, and 3) parameterization of the underlying flow network from data. While our queuing network model is under various enhancements, it plays a central role as a prediction and design tool in our proposed FCM method, and we envision that the concepts introduced in this chapter will produce significant operational impact, toward producing a decision support tool at the long strategic LAT.



(a)

(b)

(c)

Figure 2.8. Time-course airport usage statistics. (a) Case 2: Reduced capacity. (b) Case 3: GDP. (c) Case 4: MIT.

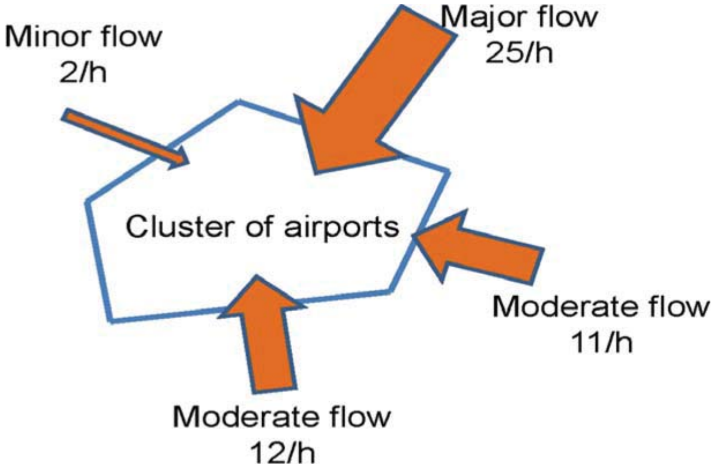


FIGURE 2.9. Cluster of airports with four incoming flows: one major flow, two moderate flows, and one minor flow.

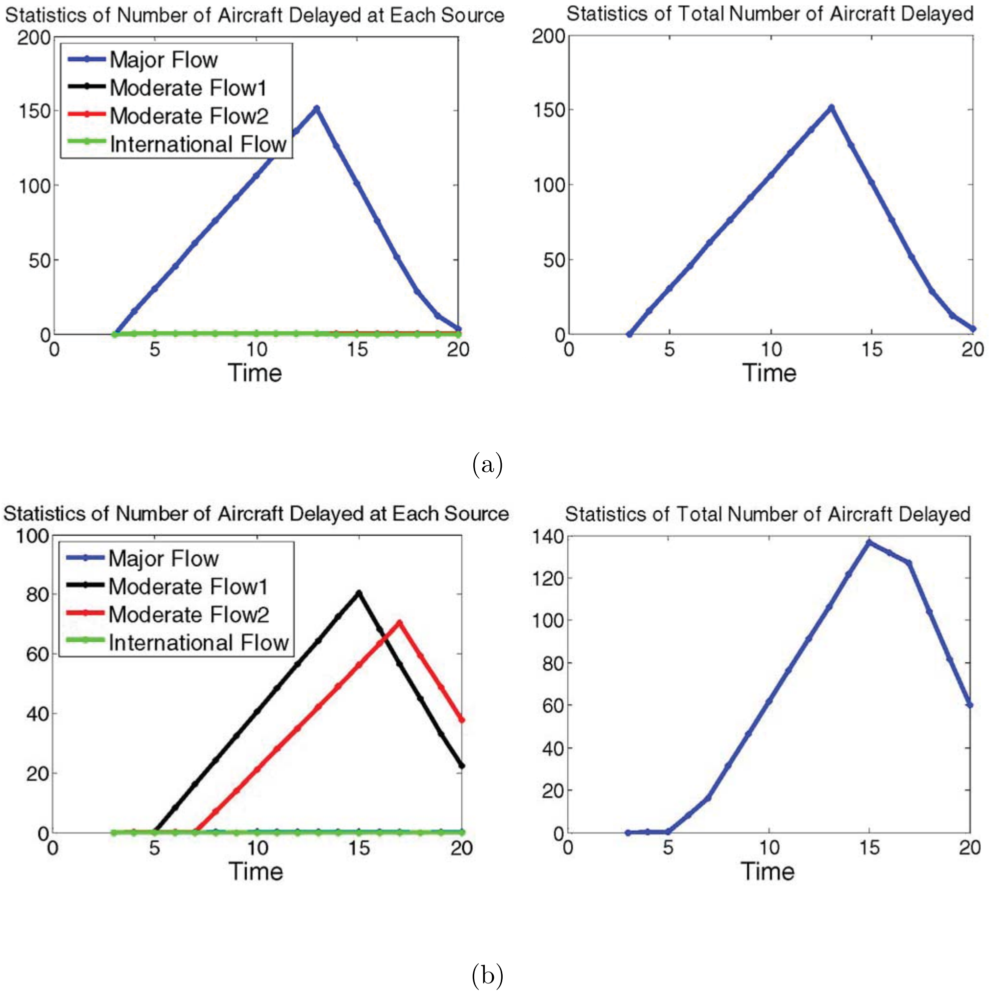
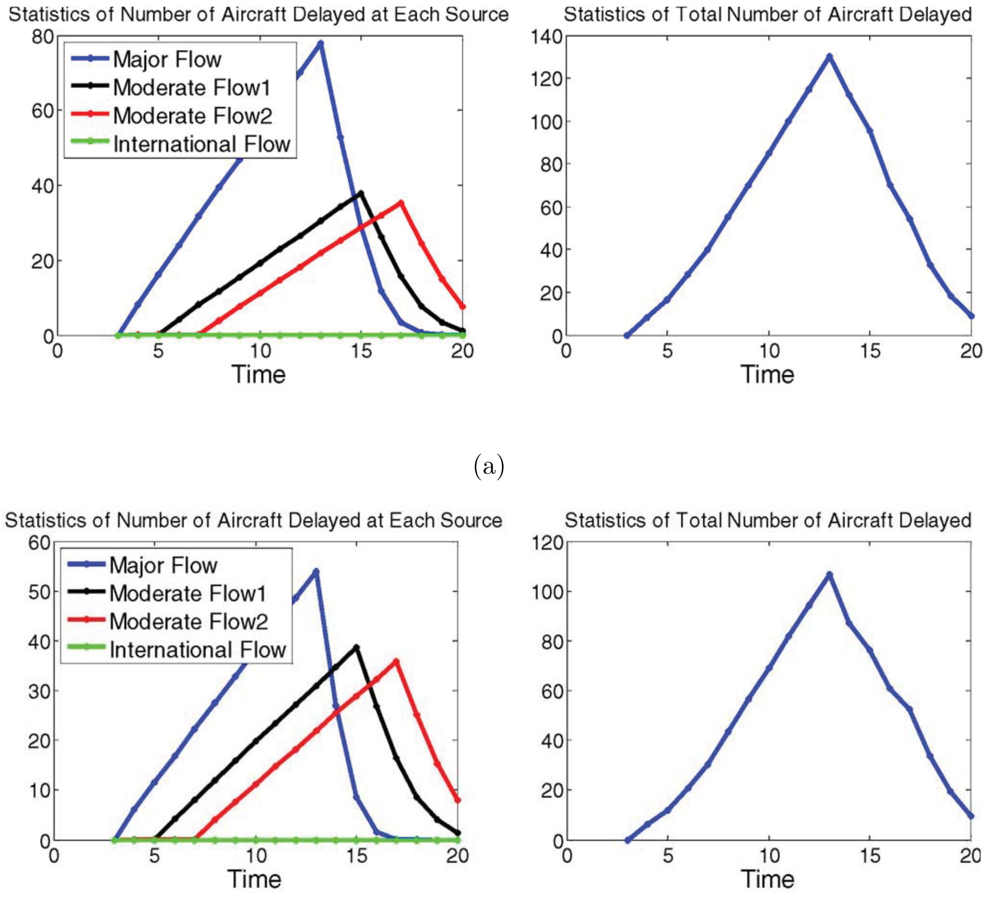


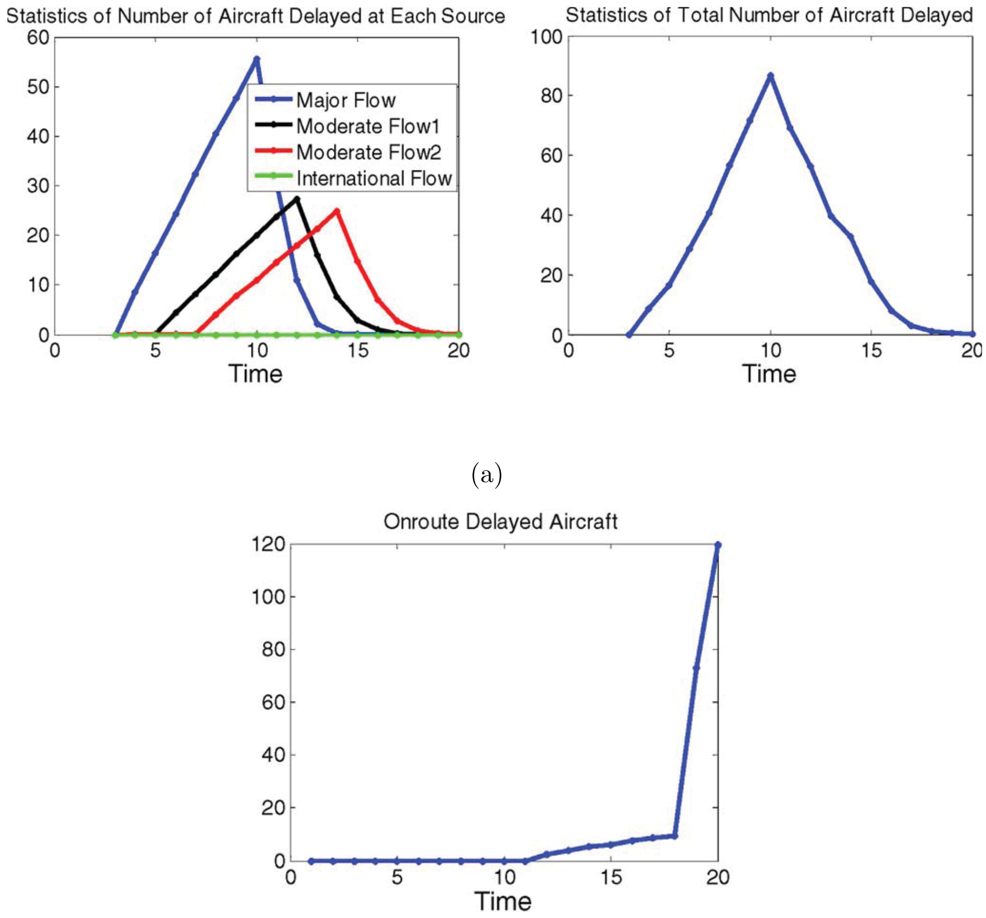
FIGURE 2.10. (a) Aircraft delay statistics when a GDP is applied to the major flow (Strategy 1). (b) Aircraft delay statistics when a GDP is applied to the two moderate flows (Strategy 2). (Left plots) Mean number of aircraft that have been delayed (i.e., are waiting) at each source airport, as a function of time. (Right plots) Mean of the total number of delayed aircraft versus time.



(a)

(b)

FIGURE 2.11. (a) Aircraft delay statistics when a GDP is applied to all three flows (Strategy 3). (b) Aircraft delay statistics when a GDP is applied to two moderate flows and AFP is applied to the major flow (Strategy 4). (Left plots) Mean number of aircraft that have been delayed (i.e., are waiting) at each source airport, as a function of time. (Right plots) Mean of the total number of delayed aircraft versus time.



(a)

(b)

Figure 2.12. (a) Aircraft delay statistics when a shorter-span GDP and an MIT are applied to all three flows (Strategy 5). (Left plot) Mean number of aircraft that have been delayed (i.e., are waiting) at each source airport, as a function of time. (Right plot) Mean of the total number of delayed aircraft versus time. (b) Mean of the number of aircraft delayed in flight.

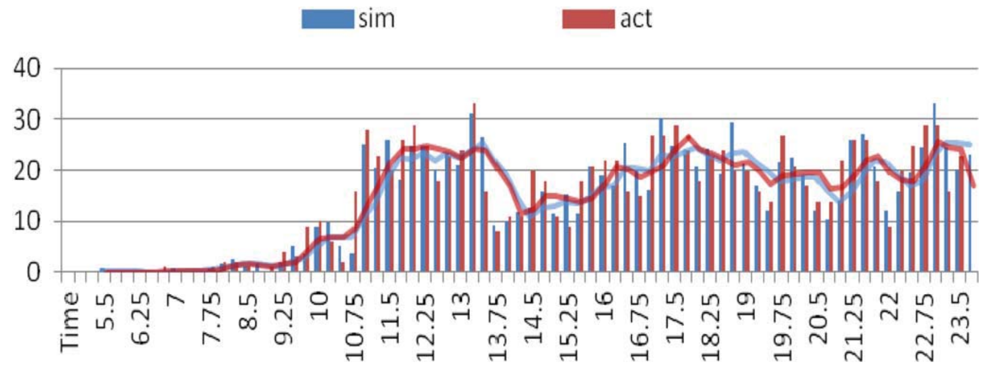
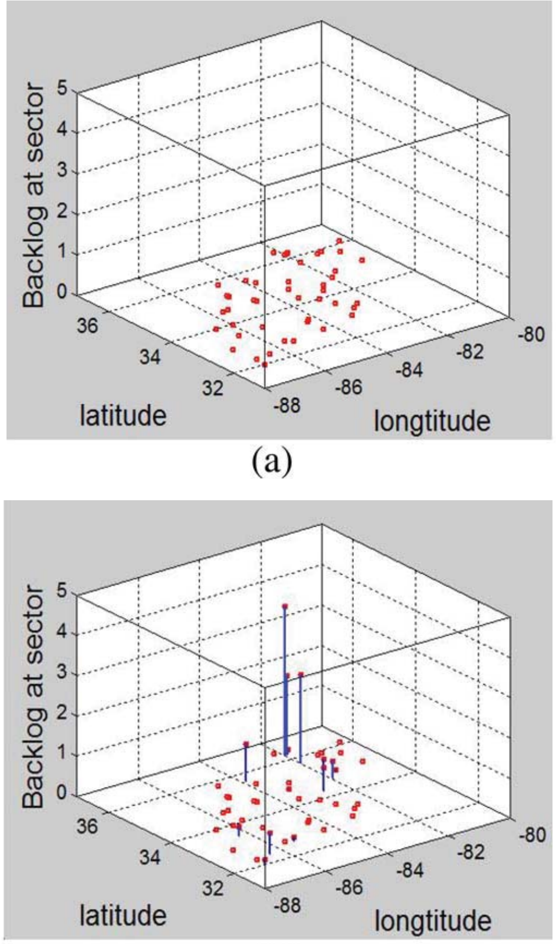


Figure 2.13. Comparison of simulated and actual arrivals at the ATL airport on August 30, 2010. The histogram shows the number of arrivals per 15 min, and the trend shows the average number of arrivals per hour



(a)

(b)

Figure 2.14. Snapshots from the videos of sector backlogs (in the ZTL center) for (a) a normal day and (b) a bad-weather day. The locations of the sectors are plotted according to their latitude and longitude. The z dimension shows the total backlogs of all flows entering each sector.

CHAPTER 3

A SMOOTH-TURN MOBILITY MODEL FOR AIRBORNE NETWORKS

3.1. Introduction

The wide variety of military and civilian applications of airborne networking have resulted in dramatically growing research efforts in airborne networks (ANs) over the past few years. Supported by the advances in sensing and wireless communication technologies, ANs hold promise in providing effective, wide-applicable, low-cost, and secure information exchange among airborne vehicles. For instance, the in-flight communication among commercial airlines can allow the sharing of adverse weather conditions and emergency situations, which are of significant value, especially, when the flights are in areas outside the reach of ground control stations. Similarly, unmanned airborne vehicles (UAVs) may rely on fast communication and networking schemes for safe maneuvering. It is anticipated that ANs will be the platform for information exchange among airborne vehicles and connect with space and ground networks to complete the future multiple-domain communication networks [114, 126].

In the study of ANs, significant efforts have been focused on the development of reliable routing protocols that minimize the number of packets lost due to link and path failures [45, 71, 93, 100, 107, 126]. Designing robust routing strategies is challenging considering the unique attributes of ANs such as high node mobility and frequent topology changes. For example, several widely-used routing algorithms that are based on the principle of shortest path tend to find paths with relay nodes at the edges of transmission radius [75], leading to what is known as “edge effect”. In such a scenario, even a slight movement of nodes can lead to link failures. This edge effect prominently occurs in highly varying networks such as ANs. Therefore, we anticipate that designing reliable routing strategies with minimal impact of edge effect should take into account the statistically varying AN structure. The statistics of interest include node distribution, movement and connectivity patterns, etc [27].

Mobility models commonly serve as the fundamental mathematical framework for network connectivity analysis, network performance evaluation, and eventually the design of reliable routing protocols [48]. In particular, mobility models capture the random movement pattern of each network agent, based on which rich information related to the varying network structure can be derived, such as node distribution, link statistics and path lifetime. Some mobility models have been extensively studied in the literature; the most well-known among them include random direction (RD), and random waypoint (RWP) [12, 18, 21, 62]. The RWP model assumes that an agent chooses a random destination (waypoint) and traveling speed. Upon reaching the waypoint, it pauses, and then travels to the next waypoint. The basic RD model assumes that nodes travel between endpoints located at region boundaries [59]. The extended version allows an agent to randomly select a speed and direction after the completion of a randomly chosen traveling time [49, 51]. The stochastic properties of these common models such as their spatial distributions can be found in e.g., [12, 18, 21, 62, 92].

Developing suitable mobility models for ANs is, undoubtedly, the foundation for designing realistic AN networking strategies. We note that the widely-used RWP and RD models are well suited to describe the random activity of mobile nodes in Mobile Ad Hoc Networks (MANETs); however, they lack the ability to capture the unique features of airborne mobility. In particular, mobile nodes on ground can easily slow down, make sharp turns, and travel in an opposite direction (see an enhanced random mobility model that captures such movement [16]). However, airborne vehicles tend to maintain the same heading speed and change direction through making turns with large radii. This unique feature is caused by the mechanical and aerodynamic constraints for airborne vehicles and reflected in the *correlation in acceleration along spatial and temporal dimensions.* Our aim here is to develop *realistic* models that capture such features unique to airborne networks, yet are *simple* and *tractable* enough to facilitate connectivity analysis and routing design.

Let us relate our modeling efforts with the very limited existing AN mobility models [106, 126, 138]. We believe that AN mobility models need to be *application-specific*, because of their wide variety of applications, and the associated different movement patterns. Within this framework, let us summarize three types of AN models in the literature including our proposed model, by focusing on specific applications and the movement patterns associated with each application.

1) Semi-random Circular Movement (SRCM) Mobility Model for Search and Rescue Applications: In this model, each UAV moves around a fixed center with a randomly selected radius; after it completes a round, it chooses another radius and circles around the *same* center [138]. Although this model seems to be limited as the movement is constrained by the location of the fixed circling center, it captures well the mobility of UAVs in *search and rescue* applications, in which the potential location of search target is usually available and chosen as the fixed center, and UAVs are dispatched to hover around the center to pinpoint the exact target location. The knowledge about the potential target location (used as the fixed center) provides extra information for predicting trajectories and connectivity structures of the AN.

2) Flight Plan-based (FP) Mobility Model for Cargo and Transportation Applications: In this mobility model, a mobility file is created using the pre-defined flight plan, and is then converted into a time-dependent network topology map (TDNT) for the design and update of routing protocols [126]. If the actual flight status deviates from the pre-described plan, the TDNT and the relevant routes are updated. The flight plan-based model is well suited for *cargo and transportation* purposes, in which the entire trajectory is usually planned in advance. Although various *uncertainties* such as weather events, departure delays, etc., may effect the adherence to the flight plans [130, 132], the existence of a plan allows for a more accurate prediction of flight trajectories and hence the varying network topology beforehand [19, 143]. The Ground-station (GS) mode of the AeroRP protocol [93] is similar in the sense that the GS broadcasts the update of AN topology periodically and also when a change in topology is sensed.

3) Smooth-Turn (ST) Mobility Model for Patrolling Applications: The above two AN mobility models assume the availability of abundant trajectory information. However, in AN applications such as *patrolling*, a predefined trajectory or a potential target location might not be available; instead, airborne vehicles simply *swarm* in a certain defined region in the airspace. Such flexible movement resembles the highly random RD model for MANETs. In this chapter, we present a novel mobility model named the smooth-turn mobility model, which allows *flexible* trajectories while also takes into account the features unique to airborne vehicles, e.g., the preference toward *smooth* rather than sharp turns. Capturing such smooth-turn features in mobility models can better represent realistic maneuvering of airborne vehicles, and improve the capability of path estimation and connectivity analysis for ANs. This new model is realistic in capturing the random movement of airborne vehicles in favor of smooth turns, and yet is *analyzable* for node distribution and connectivity analysis. We note that Gaussian Markov models (as described in [12, 74, 106]) may also be suited for patrolling applications, as they describe the memory-equipped movement of airborne vehicles. However, as we will discuss in Section 3.2.2 and was also presented in [73, 91], these models may not directly capture the kinematics of *turning* aerial objects.

This chapter contributes to the existing literature on mobility models in the following aspects:

- *A novel AN mobility model that captures smooth turns:* This mobility model resembles the traditional RD model in terms of the flexibility of trajectories, but captures the temporal and spacial correlation specific to the movement of airborne vehicles. The model is well suited for patrolling applications, in complementary to limited existing AN models in the literature. Also, a significant feature of the model is that it is simple enough to serve as the framework for not only simulation studies, but also tractable theoretical analysis.

- *The stationary analysis and preliminary connectivity analysis of the model:* We prove that the stationary node distribution of this ST model is *uniform*. The nice uniformity directly leads to a series of closed-form results for connectivity.
- *A metric to quantify the degree of randomness for mobility models:* We envision that the formal analysis of trajectory *predictability/randomness* can help better understand the difference and applicability of AN (and more general) mobility models, and more importantly, help design smart predictability-based routing algorithms. As such, we introduce a metric to quantify the randomness of future trajectory using the concept of *entropy*, and compute this metric for four mobility models.
- *The classification and comparison of different types of AN mobility models:* We identify the need to use different mobility models for different applications, and group AN mobility models according to application categories and entropy-based randomness levels.

The chapter is organized as follows. In Section 3.2, we describe the ST mobility model, and also present the basic analysis of its dynamics. In Section 3.3, we investigate the stationary distribution of the model using theoretical analysis as well as simulations. In Section 3.4, we motivate and formulate the concept of predictability/randomness, and provide a comparison of our model with three other mobility models in the literature in terms of the degree of randomness. In Section 3.5, multiple variants and enhanced versions of the ST mobility model are discussed and simulated. Finally, a brief conclusion and discussion about future works are provided in Section 3.6.

3.2. Smooth-Turn Mobility Model

In this section, we first describe the basic mathematical ST mobility model in Section 3.2.1. We then discuss the roles of model parameters, and relate the model with relevant models in the literature in Section 3.2.2. We also describe the models capturing how nodes move at boundaries. Finally, in Section 3.2.3, we present the basic analysis of the model dynamics.

3.2.1. Basic Model Description

We introduce the ST mobility model to capture the movement of airborne vehicles in highly random ANs. The model captures the unique feature of airborne vehicles—the tight spatiotemporal correlation of acceleration. Incorporating this special feature into mobility models increases the predictability of a vehicle's trajectory, which in turn, facilitates connectivity analysis and the design of networking strategies.

The idea behind the ST random mobility model is simple. An airborne vehicle selects a point in the space along the line *perpendicular* to its heading direction and circles around it until the vehicle chooses another turning center. This perpendicularity ensures smooth flight trajectories. Besides that, we assume the waiting time for the change of turning centers to be *memoryless*, i.e., the timing of the center change does not depend on the duration for which the UAV has maintained its current center. This memoryless model is typically used to abstract the waiting time for the occurrence of random events, as it brings in the nice tractability of renewable processes [97]. For instance, connectivity analysis can be taken at any time instance without prior knowledge of how long a vehicle has kept its current mobility status. Furthermore, since a vehicle commonly favors straight trajectories and slight turns than very sharp turns, we model the inverse length of the turning radius to be Gaussian distributed.

Now let us describe the mathematics of the model. We use $l_x(t)$, $l_y(t)$, $v_x(t)$, $v_y(t)$, $w(t)$, and $\Phi(t)$ to describe the X and Y coordinates, velocities in X and Y directions, angular velocity, and the heading angle at time t. For simplicity and realistic considerations, we assume a constant forward speed V in a 2-D plane; therefore, the tangential acceleration $a_t(t)$ is 0 (see Equation (31)). This assumption is reasonable for airborne vehicles, especially for jets and gliders, as they tend to maintain the same speed in flight and "reduce speed" through zigzagging and circling [3].

Furthermore, the vehicle changes its centripetal acceleration $a_n(t)$ at randomly selected time points T_0, T_1, T_2, ..., where $0 = T_0 < T_1 < T_2 < ....$ The duration for the vehicle to maintain its current centripetal acceleration $\tau(T_i) = T_{i+1} - T_i$ follows exponential distribution as motivated by its memoryless property [97]. In particular, the probability density function $f(\tau(T_i)) = \lambda e^{-\lambda\tau(T_i)}$, where $\frac{1}{\lambda}$ is the mean of $\tau(T_i)$.

Next, we describe how the new centripetal acceleration $a_n(t)$ is selected at each time point T_i. $a_n(T_i)$ is determined by the randomly selected turning radius $r(T_i)$ according to $a_n(T_i) = \frac{V^2}{r(T_i)}$ (Equation (32)). The selection of $r(T_i)$—the distance between the vehicle's current location $(l_x(T_i), l_y(T_i))$ —also determines the new turning center with coordinates $(c_x(T_i)), c_y(T_i))$ (see Figure 3.1b and the details in Section 3.2.3). It is important to note that the new turning center $(c_x(T_i)), c_y(T_i))$ resides along the line perpendicular to the heading of the vehicle at time T_i (denoted as $\Phi(T_i)$), to guarantee smoothness. Also of note, the random variable $r(T_i) \in R$ allows turns to both left and right, with $r(T_i) > 0$ representing the right turns. $\frac{1}{r(T_i)}$ is a Gaussian random variable with zero mean and variance σ^2. This distribution is selected so that straight trajectories and large-radius turns are favorable than sharp turns with small radii.

Finally, Equations (33)-(35) describe the relationships between location, heading angle, velocity, and angular velocity (see [73] for a review of dynamic models for moving aerial objects). In summary, the dynamics of the basic ST mobility model during the time interval $T_i \leq t < T_{i+1}$ is shown in Equation (31)-(35), where the "." symbol represents the first-order derivative with respect to time. Because a vehicle keeps its centripetal acceleration for a duration of $\tau(T_i)$ before changing its centripetal acceleration, it is easy to see that during the interval $T_i \leq t < T_{i+1}$, $r(t) = r(T_i)$, $a_n(t) = a_n(T_i)$, $c_x(t) = c_x(T_i)$, $c_y(t) = c_y(T_i)$, and $\tau(t) = \tau(T_i)$. A typical trajectory of the model is shown in Figure 3.1. The simulation is written in Matlab and the code is available at [1].

$$a_t(t) = 0 \tag{31}$$

$$a_n(t) = \frac{V^2}{r(T_i)} \tag{32}$$

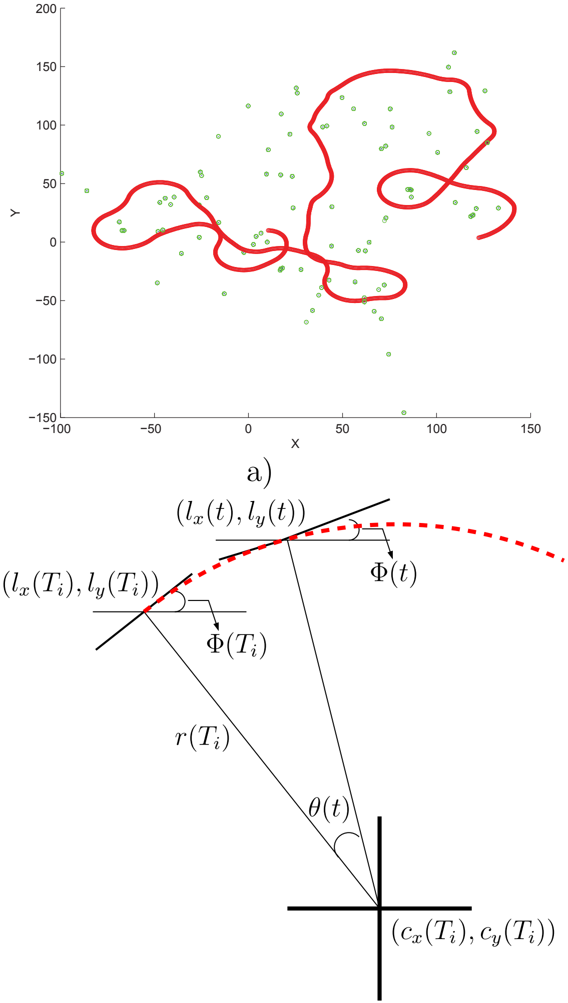
$$\dot{\Phi}(t) = -w(t) = -\frac{V}{r(T_i)} \tag{33}$$

$$\dot{l}_x(t) = v_x(t) = Vcos(\Phi(t)) \tag{34}$$

$$\dot{l}_y(t) = v_y(t) = Vsin(\Phi(t)) \tag{35}$$

3.2.2. Further Discussions of the Model

The ST mobility model naturally captures the highly random movement patterns of ANs, and the preference toward straight trajectories and large smooth turns with constant speed and turn rate. In this section, let us first comment on the three parameters in the model, and then connect the model with the RD model and target tracking models. Finally, we discuss two models capturing the movement at boundaries.



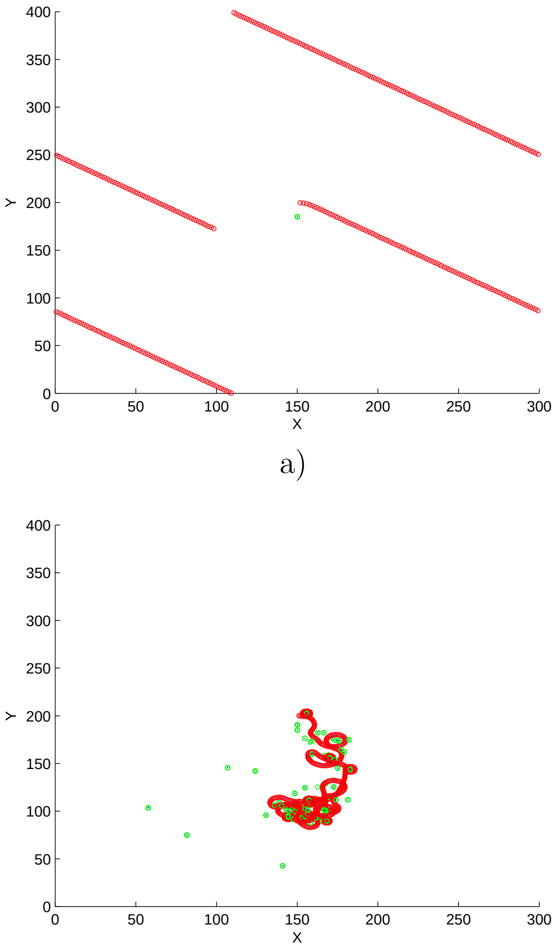
a)

b)

FIGURE 3.1. a) A simulation of the trajectory (shown in red) of an UAV in a 2-D domain. Green spots are the randomly chosen turning centers. b) The trajectory analysis diagram to predict the location of an airborne vehicle at any $T_{i+1} \geq t \geq T_i$. The dashed red curve represents the trajectory.

The first parameter is the vehicle speed V. ANs typically have high vehicle speed (in the range of $50 - 500$ miles per hour or even more), which causes highly varying connectivity structures. The second parameter is the inverse of the mean of the exponential random variable $\tau(T_i)$, λ. A large λ indicates that the airborne vehicle changes its turning center frequently, and thus results in a more wavy trajectory. The last parameter is the variance of the Gaussian variable $\frac{1}{r(T_i)}$, σ^2, which determines the preference for straight trajectory versus turns. In particular, a small variance denotes the high possibility of a very large turning radius, and therefore a more straight trajectory. At one extreme, if the variance is close to 0, the ST mobility model has very straight trajectories which resemble those of the RD model without directional change, as shown in Figure 3.2a. At the other extreme, a large σ^2, large V and large λ result in more curvy trajectories (Figure 3.2b). Through choosing proper combinations of the parameters V, λ, and σ^2, the model can capture a wide range of AN mobility patterns.

It is worthwhile to relate our ST mobility model with two relevant categories of models in the literature. First, we can view the ST model as an *RD model equipped with smooth trajectory* because of their similarities. In the RD model, an agent chooses a random *straight direction* and follows it before choosing the next direction. Similarly, in the ST model, an agent chooses a random *turning center* and circles around it before choosing the next center. We will see in Section 3.2.3 that the node distribution of our model also resembles that of the RD. Second, our model is built upon the abundant literature in the context of aerial target tracking (see e.g., [73] for a thorough review). Let us briefly discuss the works in this field, as they thoroughly studied the kinematics of aerial objects and laid out the theoretical foundation for our model. Early models in this field assume that the acceleration is *uncorrelated* in 2-D or 3-D space, and abstract acceleration in each coordinate as a Markov process (e.g., random noise passing through a linear system) [57, 73, 116]. The Gaussian-Markov model adopted in [106] is an extension of these works. The latter models, known as the *coordinated turn models*, reflect the *physical laws* of airborne objects, see e.g., [73, 91], and therefore better capture the correlation of acceleration among coordinates. However, because these models are built for target-tracking purposes, they focus on the high-precision prediction of the acceleration and path of an individual aircraft; and therefore their motion



a)

b)

FIGURE 3.2. Simulations of the ST mobility model to show the impact of parameters on the trajectories: a) when σ^2 is close to 0; b) when σ^2 and λ are large and the ratio between V and them is also large.

dynamics are more complex than necessary for our purpose. We capture the correlation across spatiotemporal coordinates through a simple parameter—the turning radius r, making the model at a coarse *statistical-group level*, and simple enough for mobility analysis.

Another topic to discuss is the modeling of node movements at boundaries. In this chapter, we adopt the boundary models typically used for the *RD* model, namely the "wrap-around" and the "reflection" models [17, 92]. In the "wrap-around" model, after an airborne vehicle hits the boundary, it wraps around and appears at the opposite side of the region. Alternatively, in the "reflection" model, the trajectory is mirrored against that boundary. Typical trajectories of these two boundary models are shown in Figure 3.3. Although these movement patterns may not be typical in reality, they provide rich analyzability and permit us to focus more on the mobility itself instead of the impact of boundaries. In the rest of this chapter, we largely focus on the "wrap-around" model, but will also briefly discuss the "reflection" model.

3.2.3. Basic Model Analysis

We consider an airborne vehicle flying within an rectangular airspace $[0, L] \times [0, W]$. Assuming the wrap-around boundary model, the mobility state of the vehicle at any $T_i \leq t < T_{i+1}$ can be obtained from the state at time T_i, as shown below:

$$c_x(T_i) = l_x(T_i) + r(T_i)sin(\Phi(T_i)) \tag{36}$$

$$c_y(T_i) = l_y(T_i) - r(T_i)cos(\Phi(T_i)) \tag{37}$$

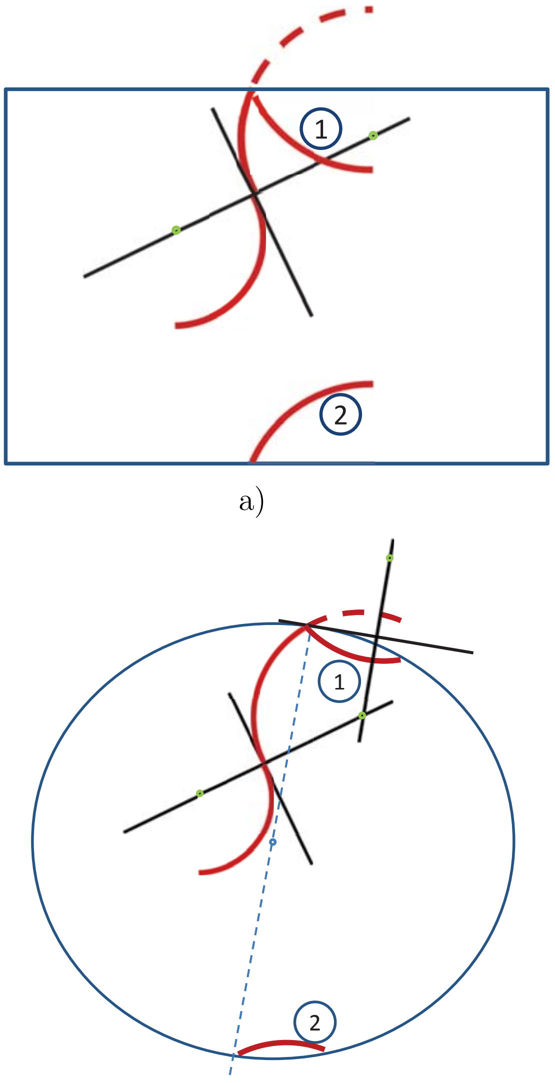
$$\theta(t) = \frac{V}{r(T_i)}(t - T_i) \tag{38}$$

$$\Phi(t) = \Phi(T_i) - \theta(t) - 2\pi \left\lfloor \frac{\Phi(T_i) - \theta(t)}{2\pi} \right\rfloor \tag{39}$$

$$l_x(t) = \left| c_x(T_i) - r(T_i)sin(\Phi(t)) - W\lfloor \frac{c_x(T_i) - r(T_i)sin(\Phi(t))}{W} \rfloor \right| \tag{40}$$

$$l_y(t) = \left| c_y(T_i) + r(T_i)cos(\Phi(t)) - L\lfloor \frac{c_y(T_i) + r(T_i)cos(\Phi(t))}{L} \rfloor \right| \tag{41}$$

These equations can be easily derived from (31)-(35) by observing the relationship between node locations and the turning center as shown in Figure 3.1b. In particular, as the turning center is along the line perpendicular to the heading direction, its x and y locations $c_x(T_i)$ and $c_x(T_i)$ can be expressed in terms of node locations at time T_i using the



a)

b)

FIGURE 3.3. Illustration of a) the reflection (1) and the wrap-around (2) boundary models in a rectangular region, b) the reflection (1) and the wrap-around (2) models in a circular region. Red bold curves represent the real trajectories.

trigonometric functions as shown in Equations (36) and (37). The turning angle $\theta(t)$ is the difference between $\Phi(T_i)$ and $\Phi(t)$. As the velocity V is fixed, $\theta(t)$ equals the arc length divided by the turning radius $r(T_i)$ as shown in Equation (38). The floor function (denoted by "$\lfloor\rfloor$") in Equation (39) guarantees that $\Phi(t)$ is within 0 and 2π (see [92] for the detailed illustration). Once the heading direction at time t, $\Phi(t)$, is calculated, the node location at time t can be expressed in terms of the turning center, again using trigonometric functions (see Equations (40) and (41)). Similar to Equation (39), these floor functions implement the wrap-around boundary model: if the location is beyond (W, L) by (w, l), it is shifted to (w, l).

With regard to the reflection boundary model, the only changes are to replace Equations (40) and (41) with:

$$l_x(t) = \left| c_x(T_i) - r(T_i)sin(\Phi(t)) - 2W \lfloor \frac{c_x(T_i) - r(T_i)sin(\Phi(t))}{2W} + \frac{1}{2} \rfloor \right| \tag{42}$$

$$l_y(t) = \left| c_y(T_i) + r(T_i)cos(\Phi(t)) - 2L \lfloor \frac{c_y(T_i) + r(T_i)cos(\Phi(t))}{2L} + \frac{1}{2} \rfloor \right| . \tag{43}$$

The floor functions and the addition of 0.5 guarantee that the trajectory is reflected back into the region when an agent moves to the boundary (as motivated by triangular wave forms [2]). The above motion analysis (Equations (36) to (43)) will be used to derive a variety of properties of the ST mobility model in the rest of this chapter.

3.3. Node Distribution and Connectivity

In this section, we consider multiple airborne vehicles following the ST mobility model. We analyze the distribution of node locations and heading angles in Section 3.3.1. The analysis is based upon that of the RD model [92], but here for the mobility model with smooth trajectory. The uniform node distribution gives rise to interesting properties in terms of network connectivity, which we will briefly summarize.

3.3.1. Node Distribution

Lemma 3.1. *$\int_{u=0}^{b} \mathbf{1}\left\{u + a - b\lfloor \frac{u+a}{b} \rfloor < x\right\} du = x$ holds for any $x \in [0, b]$, where $a \in R$, $b \in R^+$, and $\mathbf{1}\{\}$ is 1 if $\{\}$ is true and 0 if $\{\}$ is false.*

Proof. Introduce $u' = \frac{u}{b}$. According to the property of floor operations [92], we have

$$
\begin{aligned}
(44) \qquad & \int_{u=0}^{b} \mathbf{1}\left\{u + a - b\lfloor \frac{u+a}{b} \rfloor < x\right\} du \\
& = b\int_{u'=0}^{1} \mathbf{1}\left\{u' + \frac{a}{b} - \lfloor u' + \frac{a}{b} \rfloor < \frac{x}{b}\right\} du' = b\frac{x}{b} = x.
\end{aligned}
$$

□

Theorem 3.2. *N airborne vehicles move independently in the space $[0, L) \times [0, W)$ according to the ST mobility model associated with wrap-around boundary model. If the initial locations of these vehicles are uniformly distributed in $[0, L] \times [0, W]$, and the heading angles are also initially uniformly distributed in $[0, 2\pi)$, then the node locations and heading angles remain uniformly distributed for any $t > 0$.*

Proof. Let us first examine a single vehicle and show that if its position and heading angle are uniformly distributed initially, they remain uniformly distributed. Then, because the vehicles move independently, we can show the uniform distribution of node locations and heading angles for all vehicles.

For a moment, we consider the fixed movement pattern of a vehicle. In particular, the time sequence to change the turning center $0 = T_0 \leq T_1 \leq T_2, ...$ and the corresponding radii $r(T_0)$, $r(T_1)$, ..., are all fixed. Let us show that for each fixed movement pattern, uniform distribution remains for any $t > 0$.

We start with examining any time $t \in [T_0, T_1)$. At initial time T_0, the joint probability distribution function (PDF) satisfies $P(l_x(T_0) < x_0, l_y(T_0) < y_0, \Phi(T_0) < \Phi_0) = \frac{x_0}{L}\frac{y_0}{W}\frac{\Phi_0}{2\pi}$, for any $0 \leq x_0 < L$, $0 \leq y_0 < W$, and $0 \leq \Phi < 2\pi$. Let us prove that $P(l_x(t) < x, l_y(t) < y, \Phi(t) < \Phi) = \frac{x}{L}\frac{y}{W}\frac{\Phi}{2\pi}$ for any $0 \leq x < L$, $0 \leq y < W$, and $0 \leq \Phi < 2\pi$.

Equations (36)-(41) inform that $l_x(t)$, $l_y(t)$, and $\Phi(t)$ are functions of $l_x(T_0)$, $l_y(T_0)$, $\Phi(T_0)$, $r(T_0)$, and T_0:

$$\begin{aligned}
\Phi(t) &= \Phi(T_0) - \frac{V}{r(T_0)}(t - T_0) - 2\pi \left\lfloor \frac{\Phi(T_0) - \frac{V}{r(T_0)}(t - T_0)}{2\pi} \right\rfloor \\
l_x(t) &= l_x(T_0) + r(T_0)sin(\Phi(T_0)) - r(T_0)sin(\Phi(t)) \\
&- W\lfloor \frac{l_x(T_0) + r(T_0)sin(\Phi(T_0)) - r(T_0)sin(\Phi(t))}{W} \rfloor \\
l_y(t) &= l_y(T_0) - r(T_0)cos(\Phi(T_0)) + r(T_0)cos(\Phi(t)) \\
&- L\lfloor \frac{l_y(T_0) - r(T_0)cos(\Phi(T_0)) + r(T_0)cos(\Phi(t))}{L} \rfloor .
\end{aligned} \tag{45}$$

For the convenience of presentation, we denote the expressions in the right of the above three equations as $\Psi(\Phi(T_0), r(T_0), t, T_0)$, $\Lambda_x(l_x(T_0), \Phi(T_0), r(T_0), t, T_0)$, and $\Lambda_y(l_y(T_0), \Phi(T_0), r(T_0), t, T_0)$, respectively. Using the abbreviated notations and according to Equation (45), we can find the joint PDF of $P(l_x(t) < x, l_y(t) < y, \Phi(t) < \Phi)$ as:

$$\begin{aligned}
&P(l_x(t) < x, l_y(t) < y, \Phi(t) < \Phi) = \\
&P(\Lambda_x(l_x(T_0), \Phi(T_0), r(T_0), t, T_0) < x, \\
&\Lambda_y(l_y(T_0), \Phi(T_0), r(T_0), t, T_0) < y, \Psi(\Phi(T_0), r(T_0), t, T_0) < \Phi) \\
&= \int_{\Phi(T_0)=0}^{2\pi} \frac{1}{2\pi} \int_{l_x(T_0)=0}^{L} \frac{1}{L} \int_{l_y(T_0)=0}^{W} \frac{1}{W}
\end{aligned} \tag{46}$$

$$\mathbf{1}\left\{\Psi(\Phi(T_0), r(T_0), t, T_0) < \Phi\right\} \mathbf{1}\left\{\Lambda_x(l_x(T_0), \Phi(T_0), r(T_0), t, T_0) < x\right\}$$

$$\mathbf{1}\left\{\Lambda_y(l_y(T_0), \Phi(T_0), r(T_0), t, T_0) < y\right\} dl_x(T_0) dl_y(T_0) d\Phi(T_0),$$

$$= \frac{1}{2\pi WL} \int_{\Phi(T_0)=0}^{2\pi} \mathbf{1}\left\{\Psi(\Phi(T_0), r(T_0), t, T_0) < \Phi\right\}$$

$$\int_{l_x(T_0)=0}^{W} \mathbf{1}\left\{\Lambda_x(l_x(T_0), \Phi(T_0), r(T_0), t, T_0) < x\right\} dl_x(T_0)$$

$$\int_{l_y(T_0)=0}^{L} \left\{\Lambda_y(l_y(T_0), \Phi(T_0), r(T_0), t, T_0) < y\right\} dl_y(T_0) d\Phi(T_0).$$

The last equality is due to the independence of l_x and l_y.

According to Lemma 3.1, we can easily see that $\int_{l_x(T_0)=0}^{W} \mathbf{1}\big\{\Lambda_x(l_x(T_0), \Phi(T_0), r(T_0), t, T_0) < x\big\} dl_x(T_0) = x$. Similar relationship holds for l_y. As such, Equation (46) is further simplified to

$$P(l_x(t) < x, l_y(t) < y, \Phi(t) < \Phi) = \frac{xy}{2\pi WL} \int_{\Phi(T_0)=0}^{2\pi} \mathbf{1}\left\{\Psi(\Phi(T_0), r(T_0), t, T_0) < \Phi\right\} d\Phi(T_0) = \frac{x}{L}\frac{y}{W}\frac{\Phi}{2\pi}. \tag{47}$$

The above proof shows that the uniform distribution remains in the time interval $[T_0, T_1)$ for a particular movement pattern. Furthermore, let us denote the time right before t as t^-. We then easily observe that $\Phi(T_1) = \Phi(T_1^-)$, because choosing a new center $(c_x(T_1), c_y(T_1))$ along the line perpendicular to the heading angle $\Phi(T_1^-)$ at time T_1 does not change the heading angle at T_1. Combining the facts that $l_x(T_1) = l_x(T_1^-)$, $l_y(T_1) = l_y(T_1^-)$, we conclude that the uniform distribution also holds for the closed time interval $[T_0, T_1]$.

The proof to show that the uniform distribution remains during $[T_1, T_2]$ and any $[T_i, T_{i+1}]$ follows exactly the same procedure. Therefore, uniform distribution remains for any time $t \geq 0$ for each particular movement pattern, and therefore generally for a particular vehicle.

Because the N vehicles move independently, the joint distribution of node locations and heading angles is the multiplication of individual distributions. As each individual distribution is uniform, we can conclude that the N vehicles' node locations and heading angles remain uniformly distributed at any time $t \geq 0$. The proof is complete. $\square$

Theorem 3.2 informs that the uniform distribution at the initial time is preserved. The next theorem says that the steady-state distribution is uniform, independent from the initial distribution.

THEOREM 3.3. *N airborne vehicles move independently in the space $[0, L] \times [0, W]$ according to the ST mobility model associated with wrap-around boundary model. Assuming that λ and σ are finite and not equal to 0, the distributions of node locations and heading angles are uniform in $[0, L) \times [0, W)$ and $[0, 2\pi)$, respectively, in the limit of large time, regardless of the distribution at the initial time.*

PROOF. Let us first sketch the structure of the proof. We first construct a Markov process with states $S(t) = \big(l_x(t), l_y(t), \Phi(t), \frac{1}{r}(t), \tau(t)\big)$ and find the probability transition kernel for the Markov chain defined at the time sequence T_i, namely $S(T_i)$. We then find the invariant distribution of $S(T_i)$. The Palm Formula [11, 92] is then used to find the limiting probability distribution of the Markov process $S(t)$.

First, we note that $S(t)$ is a Markov process, because $S(t + \Delta t)$ is only dependent upon $S(t)$, but not on any state before time t. $S(T_i)$ for $i = 0, 1, ...$ form a discrete-time Markov chain, with the transition probability kernel:

$$f(S(T_{i+1})|S(T_i)) = f(l_x(T_{i+1}) = \Lambda_x(l_x(T_i), \Phi(T_i), r(T_i), T_{i+1}, T_i),$$
$$l_y(T_{i+1}) = \Lambda_y(l_y(T_i), \Phi(T_i), r(T_i), T_{i+1}, T_i),$$
$$\Phi(T_{i+1}) = \Psi(\Phi(T_i), r(T_i), T_{i+1}, T_i), \frac{1}{r}(T_{i+1}), \tau(T_{i+1})|l_x(T_i),$$
$$l_y(T_i), \Phi(T_i), \frac{1}{r}(T_i), \tau(T_i))$$
$$= \mathbf{1}\{l_x(T_{i+1}) = \Lambda_x(l_x(T_i), \Phi(T_i), r(T_i), T_{i+1}, T_i)\}$$
$$\mathbf{1}\{l_y(T_{i+1}) = \Lambda_y(l_y(T_i), \Phi(T_i), r(T_i), T_{i+1}, T_i)\}$$
$$\mathbf{1}\{\Phi(T_{i+1}) = \Psi(\Phi(T_i), r(T_i), T_{i+1}, T_i)\}$$
$$f(\frac{1}{r}(T_{i+1}), \tau(T_{i+1})|l_x(T_i), l_y(T_i), \Phi(T_i), \frac{1}{r}(T_i), \tau(T_i)). \tag{48}$$

This is because $l_x(T_{i+1})$, $l_y(T_{i+1})$, and $\Phi(T_{i+1})$ are fully determined by $S(T_i)$ according to Equation (45). Furthermore, since $\frac{1}{r}(T_{i+1})$ and $\tau(T_{i+1})$ are independently and identically distributed (i.i.d.) normal and exponential random variables selected at time T_{i+1}, they are independent from $S(T_i)$. Therefore, we can simplify Equation (48) to

$$f(S(T_{i+1})|S(T_i)) = \mathbf{1}\{l_x(T_{i+1}) = \Lambda_x(l_x(T_i), \Phi(T_i), r(T_i), T_{i+1}, T_i)\}$$
$$\mathbf{1}\{l_y(T_{i+1}) = \Lambda_y(l_y(T_i), \Phi(T_i), r(T_i), T_{i+1}, T_i)\}$$
$$\mathbf{1}\{\Phi(T_{i+1}) = \Psi(\Phi(T_i), r(T_i), T_{i+1}, T_i)f(\frac{1}{r}(T_{i+1}), \tau(T_{i+1}))$$
$$= \mathbf{1}\{l_x(T_{i+1}) = \Lambda_x(l_x(T_i), \Phi(T_i), r(T_i), T_{i+1}, T_i)\}$$
$$\mathbf{1}\{l_y(T_{i+1}) = \Lambda_y(l_y(T_i), \Phi(T_i), r(T_i), T_{i+1}, T_i)\}$$
$$\mathbf{1}\{\Phi(T_{i+1}) = \Psi(\Phi(T_i), r(T_i), T_{i+1}, T_i)\} \tag{49}$$

$$\lambda e^{-\lambda\tau(T_{i+1})}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2r^2(T_{i+1})\sigma^2}}.$$

The Markov chain $S(T_i)$ is aperiodic, Φ-irreducible, and Harris recurrent, when λ and σ are finite and not equal to 0. Hence there exists a unique invariant distribution measure, which is the stationary distribution [85]. Let us prove that the invariant distribution takes the following form:

$$\lim_{i\to\infty} f(S(T_i)) = \frac{1}{L}\frac{1}{W}\frac{1}{2\pi}\lambda e^{-\lambda\tau(T_i)}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2r^2(T_i)\sigma^2}}. \tag{50}$$

To prove it, we only need to show that $\lim_{i\to\infty} f(S(T_i))$ as demonstrated in Equation (50) satisfies Equation (51).

$$\begin{aligned}\lim_{i\to\infty} f(S(T_i)) &= \lim_{i\to\infty} f(S(T_{i+1}))\\ &= \lim_{i\to\infty}\int_{S(T_i)\in\Omega} f(S(T_i))f(S(T_{i+1})|S(T_i))dS(T_i),\end{aligned} \tag{51}$$

where Ω is the sample space of $S(T_i)$. The first equality is straightforward as $\tau(T_i)$ and $\tau(T_{i+1})$ are i.i.d. random variables, and $r(T_i)$ and $r(T_{i+1})$ are also i.i.d. random variables.

To show the second equality in Equation (51), we substitute Equations (49) and (50) into the right side of (51). Noticing that $\int_{l_x(T_i)=0}^{W}\mathbf{1}\{l_x(T_{i+1}) = \Lambda_x(l_x(T_i),\Phi(T_i),r(T_i),T_{i+1},T_i)\}$ $dl_x(T_i) = 1$ according to Lemma 3.1 and also similar relationships hold for $l_y(T_i)$ and $\Phi(T_i)$, we obtain

$$\begin{aligned}&\int_{S(T_i)\in\Omega} f(S(T_i))f(S(T_{i+1})|S(T_i))dS(T_i) =\\ &\int_{S(T_i)\in\Omega}\mathbf{1}\{l_x(T_{i+1}) = \Lambda_x(l_x(T_i),\Phi(T_i),r(T_i),T_{i+1},T_i)\}\end{aligned} \tag{52}$$

$$
\begin{aligned}
&\mathbf{1}\{l_y(T_{i+1}) = \Lambda_y(l_y(T_i), \Phi(T_i), r(T_i), T_{i+1}, T_i)\} \\
&\mathbf{1}\{\Phi(T_{i+1}) = \Psi(\Phi(T_i), r(T_i), T_{i+1}, T_i)\}\lambda e^{-\lambda\tau(T_{i+1})} \\
&\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2r^2(T_{i+1})\sigma^2}}\frac{1}{L}\frac{1}{W}\frac{1}{2\pi}\lambda e^{-\lambda\tau(T_i)}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2r^2(T_i)\sigma^2}}d(S(T_i)) \\
&= \frac{1}{L}\frac{1}{W}\frac{1}{2\pi}\lambda e^{-\lambda\tau(T_{i+1})}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2r^2(T_{i+1})\sigma^2}} = f(S(T_{i+1})).
\end{aligned}
$$

Finally, let us find the limiting probability distribution of the Markov process $S(t)$. According to the Palm formula [11, 92], the limiting distribution can be found by conditioning upon the stationary distribution of $S(T_i)$, where $T_i \to \infty$. In particular,

$$
\begin{aligned}
(53) \qquad \lim_{t\to\infty} f(S(t)) &= \frac{1}{E^0[\tau(T_i)]}E^0\left[\int_{T_i}^{T_i+\tau(T_i)} \mathbf{1}(S(t))dt\right] \\
&= \lambda E^0\left[\int_{T_i}^{T_i+\tau(T_i)} \mathbf{1}(S(t))dt\right] \\
&= \lambda \int_{S(T_i)\in\Omega}\int_{t=T_i}^{T_i+\tau(T_i)} f(S(T_i))f(S(t)|S(T_i))dtdS(T_i),
\end{aligned}
$$

where E^0 represents the empirical average. As shown in Equation (53), $E^0[\tau(T_i)] = \frac{1}{\lambda}$ because $\tau(T_i)$ is independently exponentially distributed with a finite mean $\frac{1}{\lambda}$.

Furthermore, noticing that when t is between T_i and $T_i + \tau(T_i)$, $f(S(t)|S(T_i))$ can be found in a similar way to obtain Equation (49). Substituting the expression of $f(S(t)|S(T_i))$ and Equation (50) into Equation (53) and using the same reasoning that derives Equation (52), we obtain

$$
\begin{aligned}
(54) \qquad &\int_{S(T_i)\in\Omega}\int_{t=T_i}^{T_i+\tau(T_i)} f(S(T_i))f(S(t)|S(T_i))dtdS(T_i) \\
&= \frac{1}{L}\frac{1}{W}\frac{1}{2\pi}\lambda e^{-\lambda\tau(t)}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2r^2(t)\sigma^2}}
\end{aligned}
$$

$$\int_{\tau(T_i)=0}^{\infty}\int_{t=T_i}^{T_i+\tau(T_i)}\lambda e^{-\lambda\tau(T_i)}dtd\tau(T_i)$$
$$=\frac{1}{L}\frac{1}{W}\frac{1}{2\pi}\lambda e^{-\lambda\tau(t)}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2r^2(t)\sigma^2}}$$
$$\int_{\tau(T_i)=0}^{\infty}\tau(T_i)\lambda e^{-\lambda\tau(T_i)}d\tau(T_i)$$
$$=\frac{1}{L}\frac{1}{W}\frac{1}{2\pi}\lambda e^{-\lambda\tau(t)}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2r^2(t)\sigma^2}}\frac{1}{\lambda}.$$

Substituting Equation (54) to Equation (53) leads to $lim_{t\to\infty}f(S(t))=\frac{1}{L}\frac{1}{W}\frac{1}{2\pi}\lambda e^{-\lambda\tau(t)}$ $\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2r^2(t)\sigma^2}}$. Integrating with respect to $\tau(t)$ and $\frac{1}{r(t)}$, we obtain that $f(l_x(t),l_y(t),\Phi(t))=$ $\frac{1}{L}\frac{1}{W}\frac{1}{2\pi}$ as $t\to\infty$. The proof is complete. $\square$

Theorems 3.2 and 3.3 demonstrate the uniform distribution of node locations and heading angles. The results also suggest the close analogy between the ST and RD models. Imposing the smooth trajectory requirement in the ST mobility model does not change the stationary uniform distribution of the RD model. This is because the wrap-around model avoids boundary impact, a key reason for non-uniform node distribution.

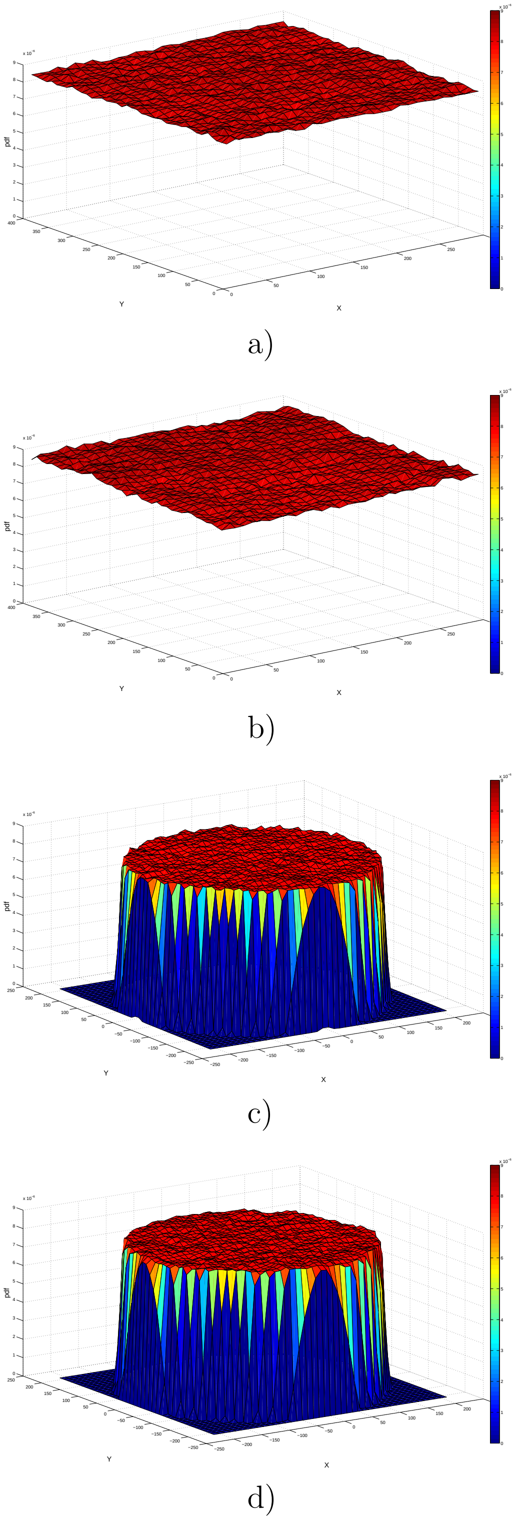
A Monte Carlo simulation of the node distribution is shown in Figure 3.4a. The simulation verifies the uniform node distribution proved in Theorem 3.3. In the simulation, an aerial vehicle is initially randomly placed in a $300\times 400km^2$ simulation area divided into grids of size $10\times 10km^2$. The vehicle then moves within the area following the ST mobility model (with $V=100m/s$, $\sigma=5\times 10^{-5}$, $\Delta t=1s$, and $\lambda=0.01/s$) and the wrap-around boundary model. The number of times that the vehicle falls in each grid is tracked. After sufficiently long simulation time, the counts are used to produce the node distribution at steady state. We note that the uniform distribution also applies to the reflection boundary model, as shown in Figure 3.4b. The proofs can be easily adapted from the proofs for

Theorems 3.2 and 3.3, by noticing the equivalence between these two boundary models [92]. The results also hold for the circular area (see simulations in Figure 3.4c and 3.4d).

The uniform stationary node distribution directly leads to a series of network connectivity results, such as the distribution of the number of neighbors for an individual node, the expected number of neighbors, the probability for a network to be connected, the probability for a network to be k-connected, and the transition range and number of neighbors required for the network to be connected with probability 1. Please refer to [17, 43, 148] for the details.

3.4. Exploring Randomness

We envision that the *randomness/predictability* of mobility models is a crucial factor for the design of effective routing schemes, but has not received much attention in the literature (see [16] for a very brief discussion). If a mobility model captures some degree of *predictability* for future trajectories, routing design could be significantly more effective by smartly taking into account this information. At one extreme, the routing design is fairly simple for a network of agents with deterministic trajectories. As relative locations of agents at future times are available beforehand to each agent, global optimization can be enacted to find the best routing design. At the other extreme, in completely random networks without any predictive information about future movement, it is highly possible that a relay node located at the boundary of transmission range (selected by the routing algorithms to minimize the number of hops) is moving out of the transmission range, leading to the loss of data transmission. There has not been quantitative studies on capturing the degree of randomness/predictability for mobility models, per our knowledge. In this section, we provide an *entropy*-based randomness measure for mobility models. As the focus of this chapter is on modeling, we leave the utilization of randomness/predictability in robust



a)

b)

c)

d)

FIGURE 3.4. Simulation of node distribution with: a) wrap-around model in a rectangular region, b) reflection model in a rectangular region, c) wrap-around model in a circular region, and d) reflection model in a circular region.

routing design to the future work.

Another motivation to the study of randomness is concerned with better understanding of the existing AN mobility models. We observe that the three AN mobility models suitable for different applications are also aligned with different degrees of randomness. For instance, mobility patterns of UAVs for security and patrol purposes (captured by the ST mobility model) may be highly random. However, the ST model may be more deterministic than the RD model, because it captures the spatiotempoeral correlation of movement. UAVs used for search and rescue purposes (captured by the SCRM mobility model) are usually provided with certain target location information to start with, and hence their mobility patterns are less random. Commercial aircraft and UAVs envisioned for NextGen cargo transportation have pre-planned trajectory information and hence their mobility patterns (captured by the FP mobility model) are almost deterministic. Because of the significance of mobility models in routing design, there is a need to understand and differentiate the various AN mobility models in great depth, and explore routing strategies to enhance network connectivity in each. Randomness/predictability provides a measure for such investigation.

3.4.1. Definition of Randomness/Predictability

A natural metric for randomness is the entropy measuring the predictability of future trajectory conditioned upon the current information. We note that the entropy defined on a long look-ahead time does not help. For instance, RD, ST and the SRCM mobility models all result in a uniform stationary node distribution, which does not differentiate among themselves. Therefore, we are motivated to study an *immediate* entropy measure defined at a very short time-frame, using a Markov chain that describes the trajectory dynamics.

Specifically, the state of the Markov chain represents the vehicle status, such as position and direction. Randomness is defined based upon the **entropy rate** H [37]:

$H = -\int_i \int_j p_i Q_{ij} ln Q_{ij}$, where p_i is the probability to stay at state i, and Q_{ij} denotes the transition probability from state i to j.

As a special case, if the states are uniformly distributed (e.g., when $t \to \infty$ for the RD and ST mobility models) and therefore p_is are all equal, and also the pattern of the transition probabilities Q_{ij} associated with different state i is the same, the calculation of H can be simplified to $H' = -\int_j Q_{ij} ln Q_{ij}$, for any i.

To facilitate analysis, we consider a discretization of time, and assume that the transition from one state to the other takes a unit time Δt, where Δt is sufficiently small.

3.4.2. Comparison of Randomness for AN Mobility Models

In this section, we estimate the degree of randomness for four mobility models, including RD, SRCM, ST and FP, using the entropy rate-based metric. The quantitative analysis also allows us to evaluate the impact of model parameters on randomness, and compare the four different mobility models. For a fair comparison, we assume that every model takes the same fixed forward speed V, and also ignore boundary effects.

3.4.2.1. RD Mobility Model

Assume that a vehicle keeps its direction for an exponentially distributed duration (with mean $\frac{1}{\lambda}$) before choosing its new direction uniformly distributed between 0 and 2π. Because of the uniform stationary distribution and the memoryless property, we can use the simplified randomness measure H'. Without loss of generality, assuming that the vehicle is moving from location 0 to the right, let us examine the probability of location and direction at time Δt. The probability for changing direction k times within Δt thus follows the Poisson distribution $P(n = k) = \frac{(\lambda \Delta t)^k e^{-\lambda \Delta t}}{k!}$. As Δt is sufficiently small, we assume that the change of direction occurs at most once within Δt. Therefore, $P(n = 1) \approx \lambda \Delta t$, and $P(n = 0) \approx 1 - \lambda \Delta t$. If $n = 0$, the vehicle moves to the right and ends up at the location

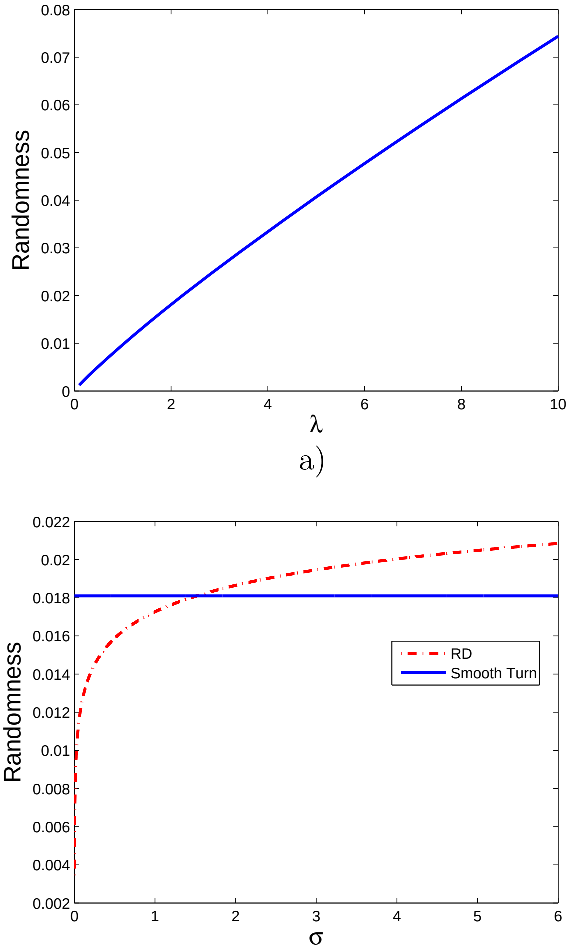
$V\Delta t$ with direction to the right. If $n = 1$, we assume that the change of direction occurs at the very beginning, as Δt is very small. In this simplified case, the vehicle will locate uniformly on a circle centered at the starting location with heading directions pointing outwards. As the direction is completely correlated with the location at Δt, we can find the probability associated with the ending location/direction as $\frac{1}{2\pi}$. Therefore, we can compute the randomness as

$$
\begin{aligned}
(55) \qquad H_{RD} &= -P(n=0)lnP(n=0) - \int_{\Phi=0}^{2\pi} \frac{P(n=1)}{2\pi} ln \frac{P(n=1)}{2\pi} d\Phi \\
&= -(1-\lambda\Delta t)ln(1-\lambda\Delta t) - \int_{\Phi=0}^{2\pi} \frac{\lambda\Delta t}{2\pi} ln \frac{\lambda\Delta t}{2\pi} d\Phi \\
&= -(1-\lambda\Delta t)ln(1-\lambda\Delta t) - \lambda\Delta t ln \frac{\lambda\Delta t}{2\pi}.
\end{aligned}
$$

Interestingly, the speed of a vehicle does not affect the randomness of the model. Moreover, the degree of randomness increases with the increase of the parameter λ as suggested by Figure 3.5a. This result is reasonable, as λ represents how frequently a random direction is selected. The more frequently a random direction is selected, the more random the future trajectory becomes.

3.4.2.2. ST Mobility Model

As defined earlier in this chapter, the basic ST mobility model assumes that a vehicle circles around for an exponentially distributed duration with mean λ, before selecting a new turning radius with its inverse normally distributed with mean 0 and variance σ^2. Similar to the RD model, if no change of turning center occurs within Δt, the vehicle will travel around its original turning center for a duration Δt. Otherwise, the vehicle will end up at a random location/heading direction by following a curve with the inverse of its turning radius



a)

b)

FIGURE 3.5. a) Randomness against λ in the RD model ($\Delta t = 0.001s$), b) Comparison of the randomness between the RD and ST mobility models ($\Delta t = 0.001s$, $\lambda = 2/s$).

r normally distributed with probability $\frac{e^{-\frac{1}{2r^2\sigma^2}}}{\sqrt{2\pi}\sigma}$. We thus can find the randomness for the ST model, in a way similar to that of the RD model, as follows:

$$H_{SR} = -P(n=0)lnP(n=0) - \tag{56}$$

$$\int_{\frac{1}{r}=-\infty}^{\infty} \frac{P(n=1)e^{-\frac{1}{2r^2\sigma^2}}}{\sqrt{2\pi}\sigma} ln \frac{P(n=1)e^{-\frac{1}{2r^2\sigma^2}}}{\sqrt{2\pi}\sigma} d\frac{1}{r}$$

$$= -(1 - \lambda\Delta t)ln(1 - \lambda\Delta t) - \int_{\frac{1}{r}=-\infty}^{\infty} \frac{\lambda\Delta t e^{-\frac{1}{2r^2\sigma^2}}}{\sqrt{2\pi}\sigma} ln \frac{\lambda\Delta t e^{-\frac{1}{2r^2\sigma^2}}}{\sqrt{2\pi}\sigma} d\frac{1}{r}$$

$$= -(1 - \lambda\Delta t)ln(1 - \lambda\Delta t) - \lambda\Delta t ln \frac{\lambda\Delta t}{\sqrt{2\pi e}\sigma}.$$

The result suggests that V also does not impact the randomness of the ST mobility model. However, both λ and σ play a role in the degree of randomness. In particular, the randomness is less with smaller σ, which suggests more straight trajectory. For a large σ, denoting high variability of turning radius, the location and direction of the vehicle can be very uncertain. In summary, the randomness of the ST mobility model is less than that of the RD model when σ is less than a threshold. The comparison between Equations (55) and (56) suggests that the threshold $\sigma = \frac{2\pi}{e}$ (also see Figure 3.5b).

3.4.2.3. Semi-Random Circular Movement Mobility Model

We consider a SRCM model that is slightly different from the one described in [138]. Again, we assume a constant forward speed V and a fixed center. Upon the completion of one round, the vehicle chooses a radius r *uniformly* distributed between r_I and r_O (where $V\Delta t << r_I < r_O$). We also assume that the transition time from one radius to the other is neglected, and we only consider the circular movement as suggested in [138]. The resulting node distribution is uniform as shown in Figure 3.6a. We note that the analysis of randomness for the SRCM model is different from that of the RD and ST models, in the sense that the original randomness measure H needs to be used, instead of the simplified H'. This is because the SRCM model does not have the memoryless property; as such, how long the vehicle has traveled along a circle matters, and hence the pattern of Q_{ij} is different for different i. In a small time interval Δt, if the vehicle has traveled along a circle with radius r for a distance less than $2\pi r$, the vehicle will continue moving along the circle with probability $Q_{ij} = 1$; otherwise, the vehicle will end up at a circle with the radius uniformly distributed with probability $Q_{ij} = \frac{1}{r_O - r_I}$.

To facilitate the analysis, we first obtain the entropy rate conditioned upon the current radius r, and then integrate over all possible radii. As the randomness associated with any radius r only occurs when the vehicle changes its radius within Δt (with probability $\frac{V\Delta t}{2\pi r}$), we find $H_{SC}|r = -\frac{V\Delta t}{2\pi r}\int_{r=r_I}^{r=r_O} \frac{1}{r_O - r_I} ln \frac{1}{r_O - r_I} dr = -\frac{V\Delta t}{2\pi r} ln \frac{1}{r_O - r_I}$.

Integrating over the range of r, we obtain

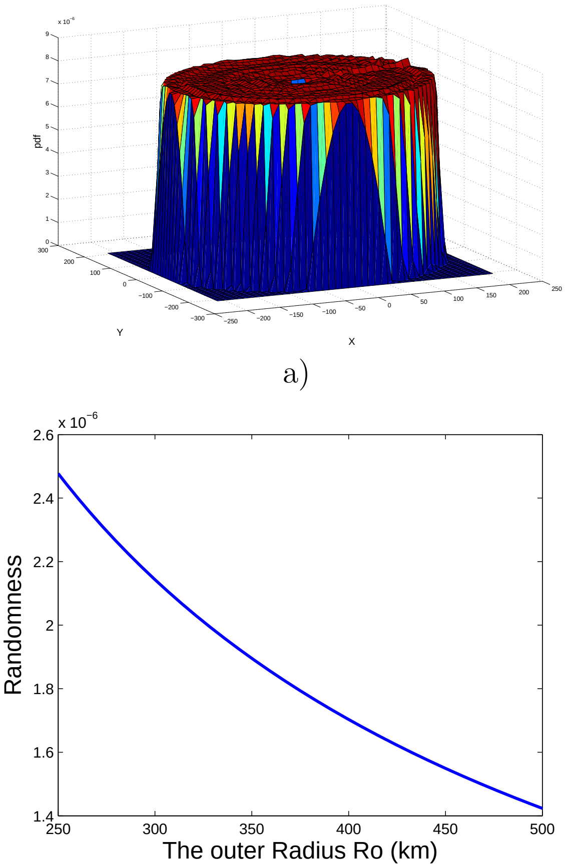
$$
\begin{aligned}
H_{SC} &= \int_{r=r_I}^{r=r_O} -\frac{V\Delta t}{2\pi r} ln \frac{1}{r_O - r_I} \frac{1}{r_O - r_I} dr \\
&= -\frac{1}{r_O - r_I} ln \frac{1}{r_O - r_I} \frac{V\Delta t}{2\pi} \int_{r=r_I}^{r=r_O} \frac{1}{r} dr \\
&= -\frac{1}{r_O - r_I} ln \frac{1}{r_O - r_I} \frac{V\Delta t}{2\pi} (lnr_O - lnr_I).
\end{aligned} \tag{57}
$$

In the ST mobility model, both the velocity V and the area defined by the radii R_O and R_I affect the randomness. Clearly, the degree of randomness increases proportionally with the increase of V; however it typically decreases with the increase of R_O (as shown in Figure 3.6b). It is also observed that under typical maneuvering conditions, the randomness of the ST model is larger than that of the SRCM model.

3.4.2.4. Flight Plan-Based Mobility Model

With the simplest assumption that a trajectory follows a pre-planned trajectory with a tiny variation modeled by a Gaussian noise (with mean 0 and variance $\hat{\sigma}^2$), we have $H_{FP} = ln(\sqrt{2\pi e}\hat{\sigma})$. When $\hat{\sigma} < \frac{1}{\sqrt{2\pi e}}$, the entropy is negative, representing the randomness to be less than a uniform distribution in $[0, 1]$ [31].

In summary, the randomness and application categories of the RD model and three AN mobility models discussed in this section are listed in Figure 3.7.



a)

b)

FIGURE 3.6. a) The uniform distribution of the SRCM model, b) The degree of randomness decreases with the increase of the outer radius R_O ($\Delta t = 0.001s$, $V = 40m/s$, and $R_I = 100m$).

Mobility Model	RD	ST	SRCM	FP
Application	MANET	Patrolling	Search and rescue	Cargo and commercial
Degree of Randomness	Highest	Medium	Low	Lowest

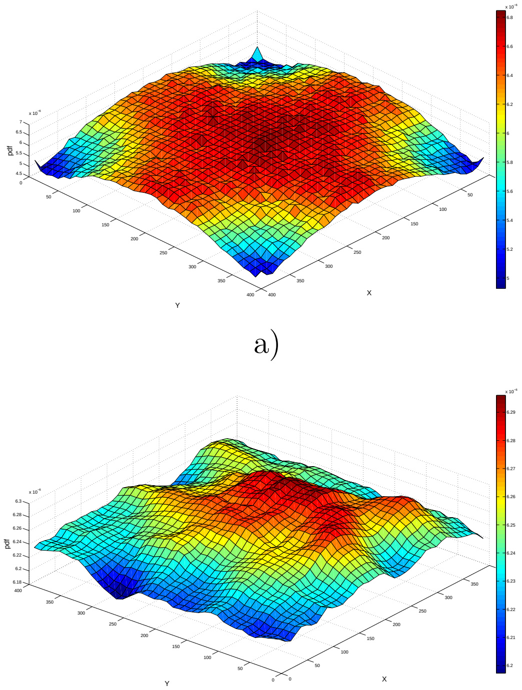
FIGURE 3.7. Comparison of randomness among the four mobility models.

Finally, let us also discuss possible variants and enhancements of the ST mobility model.

Enhanced Modeling of Model Parameters This basic ST mobility model can be easily generalized to include varying forward speed and 3D movement. Also, as an airborne vehicle typically has certain minimum safe turning radius, instead of roughly modeling $\frac{1}{r}$ as a Gaussian variable, we can provide a more detailed model for r, and require r to reside in the safe range. All of the above enhancements maintain the uniform distribution of node locations and directions. We also note that the random model in [71] is a variant of the ST model, e.g., with the assumption that the radius r takes two fixed constants and ∞. An enhanced pheromone repel model that maximizes coverage is also introduced therein. Geographical routing is developed for this mobility model[70].

Collision Avoidance In the current mobility model, we assume that each vehicle moves independently. In reality, neighboring airborne vehicles need to satisfy a safe separation distance, and therefore proper collision avoidance mechanisms may be included. As the centripetal and tangential accelerations are directly captured in the mobility model, control mechanisms for collision avoidance can be easily added (see [20] for a related implementation) if desired.

RWP-like ST Mobility Model The current ST mobility model resembles the RD model equipped with smooth trajectory. We can similarly develop RWP-like ST mobility models. Possible strategies include: 1) randomly choosing a center which satisfies the smooth trajectory requirement and is uniformly distributed in the region, and circling around it for an exponential duration, before choosing another center; 2) randomly choosing a destination uniformly distributed in the region and reaching it along a smooth trajectory, before choosing another destination. Similar to the RWP models, we also observe non-uniform node distributions (see Figure 3.8).



a)

b)

FIGURE 3.8. Simulation of node distribution for a) the random-center RWP-alike ST mobility model b) the random-destination RWP-alike ST model after smoothening.

3.6. Concluding Remarks and Future Work

In this chapter, we present a novel ST random mobility model for highly random ANs. The ST mobility model captures the tendency of airborne vehicles toward making straight trajectories and smooth turns. It is developed based upon the physical laws and aerodynamic constraints governing moving aerial objects, whereas is simple enough for tractable analysis. We prove that, similar to the RD model, the stationary node distribution is uniform. This result permits a series of closed-form connectivity properties.

The randomness is an important characteristic of mobility models in that 1) it is a natural metric to characterize mobility models, and 2) it serves as an important factor for the design/selection of robust routing algorithms. As such, we have developed a quantified randomness measure using entropy rate. We then compute and compare the degrees of randomness for four mobility models. Compared to the FP and SRCM mobility models, the ST model represents the highest degree of randomness, but is less random than RD because of the constraint on smooth trajectory specific to airborne vehicles. The ST mobility model can be viewed as an RD model equipped with smooth trajectory, with the only difference that it randomly chooses a turning radius instead of a heading direction. The classification of AN mobility models based upon the degree of randomness aligns with the classification based on applications. Different AN mobility models are needed for different applications (e.g., transportation, search and rescue, and patrolling) due to their diversity.

We will investigate various enhanced versions of the ST mobility model as suggested in Section 3.5 in the future. We will also fit model parameters using real UAV flight data for model validation. Moreover, we will investigate area coverage, as the full coverage and the time needed are also important characteristics of ANs [76, 99]. Finally, we will further our investigation with more advanced connectivity properties such as path duration and link duration, fully investigate the impact of randomness on the performance of routing protocols, and design effective routing protocols that utilize this information.

CHAPTER 4

A STOCHASTIC MODELING AND ANALYSIS APPROACH TO STRATEGIC TRAFFIC FLOW MANAGEMENT IN A WEATHER IMPACTED REGION

4.1. Background

Let us start with a review of the existing literature relevant to management at the strategic time-frame, with the aim of summarizing the current research status, stressing research needs, and also motivating our approach.

Because deterministic traffic models lack the capability to capture uncertainty at the strategic time-frame, stochastic flow network models (and in particular queuing network models) are considered to be valuable for strategic decision-making. Some recent efforts, including our group's, in using queuing models to capture uncertain traffic and obtain insights for planning can be found in e.g., [54, 68, 88, 111, 120, 130, 131, 135]. Of most interest to us, the article [120] used a center-level open Jackson network model to evaluate path efficiency. In [130, 134], with the aim of designing network-level en route flow rates, a M/D/1 queuing network model was constructed, and various abstractions of it were sought to facilitate design. Articles [130, 134] were extended to provide insights into optimal routing design through a sensitivity study on queuing models [131]. These modeling and analysis efforts using queuing-network models provide a natural framework for the evaluation and design of management actions at the strategic time-frame, because they consider traffic as flows and ignore the schedule details of individual aircraft. However, these studies do not address the performance of the NAS in response to dynamic and uncertain weather events.

In a parallel vein, there have been advances in the evaluation and design of management actions under dynamic and uncertain weather. One straight-forward approach to address this problem is Monte-Carlo simulations, i.e., using ensembles of uncertain weather to evaluate management actions and select optimal ones. An intelligent way of selecting a minimum set of ensembles was given in [132]; these numerical methods however are not appropriate for robustly solving large-scale design problems. In order to improve efficiency, systematic analysis and design has been sought using stochastic programming approaches (see e.g. [30, 94, 95]). For instance, in [95], weather was modeled using Markov chain models, and routing design was formulated as a Markovian Decision Processes. These studies are valuable in providing systematic designs that take into consideration of uncertain weather. However, they address the management of individual aircraft instead of flows, while flow-level designs are necessary at the strategic horizon considering the large dimension of the decision space at this horizon.

Because of the significant role that weather plays in strategic decision-planning, analytical tools that permit the evaluation and design of strategic management actions under uncertain weather are urgently needed. To meet this need, I take the perspective that uncertain weather-impact models must be seamlessly interfaced with aggregated traffic flow models, and the analysis of the integrated models must be completed. Pursuing this direction, in this chapter I model traffic as stochastic flows, and model management actions as flow restrictions or *queues* that shape downstream flows (to comply with capacity constraints) at the cost of delay/backlog upstream. Moreover, I model dynamic and uncertain weather impact using Markov chains. To capture the impact of uncertain weather events on flow, I consider the parameters of the queuing models as being modulated by uncertain weather. In particular, the service rates of the queuing models are viewed as being randomly reduced by convective weather impact. Such an integrated weather and flow modeling perspective lays is the modeling foundation for the performance analysis under weather uncertainty that I pursue here.

The performance analysis of queuing systems with random service rate reduction, such as I am seeking in this work, is mostly studied outside the air traffic management domain. In the field of road traffic planning, random service rate reduction is instead caused by uncertain events such as traffic accidents, vehicle failures, and other emergent road conditions. In paper [14] and the references therein, random service rate reduction was modeled as a Markovian process, and queuing analysis with markovian modulated services was sought. These studies are mostly focused on the analysis of steady-state performance, i.e., the statistics of queuing performance in a long run. However, in air traffic planning, transient performance is typically of significant value. For instance, a predicted temporary surge of traffic due to a severe storm, and the resultant congestion, plays a key role in strategic planning. To some extent, I can say strategic planing is in essence redistributing resources in advance to alleviate temporary congestion caused by uncertain transient weather. In this chapter, I aim to provide novel analytical tools that allow the evaluation of transient queuing performance.

Our main contribution in this chapter is the development of *analytical* tools that allow the evaluation of the impact of transient convective weather on uncertain flows, and thus give insight toward optimal management strategy design under weather uncertainty at the strategic time-frame. Let us briefly discuss the specific analyses of the integrated stochastic weather and flow model that I complete here. In particular, I consider two approaches that allow the transient performance analysis for queuing models driven by uncertain weather. In the first approach, I track both weather and flow dynamics using Markov chains, and investigate the analysis of steady-state and transient statistics of traffic delay under weather uncertainty. Though accurate, the computational complexities associated with intensive (transient) Markov chain analysis makes it hard to generalize this approach to queuing network models at a broader spatial scale. I then suggest a novel scalable jump-linear approach to analyze the integrated weather and flow models. I will show that the jump-linear approach is capable of effectively evaluating and comparing management actions under uncertainty. I envision that the jump-linear approach developed in this chapter is promising to allow the evaluation and design of optimal flow contingency plans at a broad spatial scale, and is robust to likely weather scenarios.

It is worthwhile to note that this work is part of our ongoing effort in developing flow contingency management (FCM) framework for the NextGen [127]. In [112, 150], I provided the tool for the modeling and prediction of uncertain weather impact, using a stochastic automaton; in [135], I established a queuing network framework that allows the design of several management actions in practice or potential for use in NextGen. This chapter discusses our efforts in integrating the above two directions, by developing systematic analytical and design tools for management actions under weather uncertainty.

The remainder of this chapter is organized as follows. In Section 4.2, I overview the use of stochastic models to represent weather impact, and the use of queuing models to capture management actions. I thus formulate the problem of analyzing the integrated two models. In particular, a stochastic automaton known as the influence model is used to capture the dynamics of stochastic weather at a broad spatio-temporal scale. Weather dynamics in a single region can be predicted from the model, and approximated using a low-order Markov chain model. Moreover, queuing models are used to represent management restrictions (e.g., miles-in-trail (MIT) or minute-in-trail (MINIT)) acting on flows. In Section 4.3, I provide an extended Markov modeling approach to analyze the performance of management actions on flows in the presence of uncertain weather. Specifically, the statistics concerning the backlog of traffic can be obtained from the Markov analysis of the integrated flow and weather model. In Section 4.4, I re-formulate the model as a jump-linear system, and show that this formulation permits an efficient performance analysis. Moreover, I use an example to show a key insight that the jump-linear modeling approach provides, concerning the role of detailed stochastic weather information in the performance of management actions. Finally, I discuss

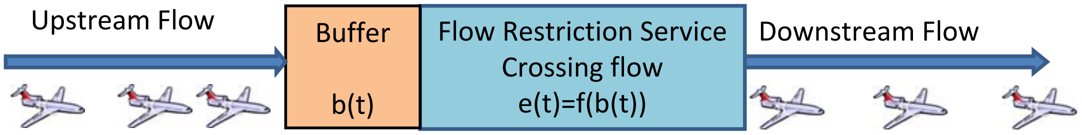


Figure 4.1. A stochastic entering a single weather zone.

several features of the jump-linear approach. In Section 4.5, a brief conclusion is provided.

4.2. Stochastic Weather and Flow Restriction Modeling: Overview and Problem Formulation

In this section, I provide an overview of the stochastic modeling of uncertain weather/ weather-impact and of traffic flows/restrictions. I then motivate the problem of evaluating delay and backlog under uncertain weather, using Monte Carlo simulations to illustrate our motivation.

4.2.1. Modeling Restriction's Impact on Stochastic Flow

Queuing models are widely used to capture management actions' impact on air traffic flows. In [135], I discussed the use of queuing models to capture various management actions, including MINIT/MIT, rerouting, time-based metering (TBM), ground-delay programs (GDP), and airspace flow programs (AFP), as part of a comprehensive network model for air traffic. In our work, I focus on a single traffic flow entering a weather zone (see Figure 4.1).

To begin, I recall that a single en route restriction's impact (e.g., an MIT/MINIT restriction's impact) under fixed weather conditions can be modeled using an M/D/1 queue. Here let us briefly review the modeling of a single restriction under fixed weather, and then describe the M/D/1 model and the approximation of it using a saturation model (please see [130] for more illustration).

I assume that a flow enters the boundary of a weather zone with an **inflow rate** λ (the number of coming aircraft per unit time). In this chapter, I focus on Poisson flows (see

[88] for motivation), i.e., the distribution of the number of aircraft coming to the restriction per unit time is $P(\lambda, k) = \frac{\lambda^k e^\lambda}{\lambda!}$. However, the study developed in our work can be generalized to other stochastic flows. In those cases, further parameters in addition to flow rate may be required to describe the flow. As a flow is approaching a boundary, it is considered as entering an imaginary buffer. The number of aircraft in the buffer at time t is denoted as **buffer length** $b(t)$. Because of the en route rate restriction set either by management actions or by weather-impact capacity constraints, only a portion of the aircraft in the buffer is allowed to cross. The relationship between the **crossing flow** $e(t)$ and buffer length is denoted by $e(t) = f(b(t))$. In general, $f()$ can be either a deterministic or a stochastic function, depending on the nature of the restriction. The **Backlog** $B(t)$, which captures the number of aircraft being delayed at time t, is defined as the number of aircraft in the buffer excluding the ones crossing the boundary at the current time.

In the case of an $M/D/1$ queue, each aircraft takes a fixed **service time** (denoted as T_c) to cross the boundary. The **service rate** u_c is defined as $\frac{1}{T_c}$. Specifically, if the buffer is not completely empty within T_c time units after an aircraft leaves the boundary, the first aircraft in the buffer cannot cross the boundary until the T_c duration is completed. The deterministic service time in the $M/D/1$ model forces a minimum separation distance/time between successive aircraft, and as such the M/D/1 model is natural to capture MINIT/MIT en route restrictions. Using standard queuing analysis (see [50]), some steady-state backlog and delay statistics for the $M/D/1$ queue can be calculated. For instance, the mean backlog in steady-state can be calculated as

$$E(B) = \frac{\lambda^2}{2(u_c^2 - \lambda u_c)} \tag{58}$$

Unfortunately, it is not straightforward to find higher-order steady-state statistics or characterize transient dynamics for the M/D/1 model using standard queuing analysis, hence a **saturation model** was developed to approximate the M/D/1 model to permit richer analysis (especially for networks of restrictions) [130]. The saturation approximation is a discrete-time model that assumes the following: during any time interval Δt, a maximum number of $N_c = u_c \Delta t$ aircraft (denoted as **saturation restriction**) is allowed to cross the boundary. The saturation model can be mathematically described as:

$$\begin{aligned} e[k] &= \begin{cases} b[k-1], & (b[k-1] \leq N_c)) \\ N_c, & (b[k-1] \geq N_c)) \end{cases} \\ b[k] &= b[k-1] + x[k] - e[k] \\ B[k] &= b[k-1] - e[k] \end{aligned} \tag{59}$$

where $e[k]$, $b[k]$ and $B[k]$ represent the crossing flow, the buffer length, and the backlog at time interval k. Clearly, the saturation model is a discrete-time version of the M/D/1 model. In the limit as Δt is made small, the saturation model approaches the M/D/1 model. Let us briefly discuss validation of the model. In Table 4.1, I approximate the steady-state mean backlog of an M/D/1 queue through Monte-Carlo simulation of the saturation model, when the inflow rate $\lambda = 9.5$ and service rate $u_c = 10$. As seen from the simulation results, as the time interval of the approximation is made smaller, the saturation model approaches the M/D/1 queuing model in predicting steady-state mean backlog, which is 9.025 according to Equation 58.

In this chapter, I use the saturation model to capture the impact of restrictions on flows. I will show that this discrete-time recursive description of an M/D/1 model permits systematic analysis of the queue's transient dynamics, even under weather uncertainty.

TABLE 4.1. Steady State Mean Backlog Obtained from Simulating the Saturation Model with Different Time Intervals (λ=9.5, $u_c = 10$).

Time Interval Δt	3 hour	1 hour	0.1 hour
Mean Backlog	6.87	7.90	9.025

4.2.2. Modeling Stochastic Weather-Impact

Properly modeling and predicting weather impact is significant to decision-making at the strategic time-frame. Existing ensemble and probabilistic forecast products are not suitable to be directly employed in strategic planning because of 1) they focus on describing weather rather than weather impact, 2) they lack spatio-temporal descriptions of weather dynamic, and 3) they are computationally incredibly complex [150]. These limitations motivated us to develop a spatio-temporal weather-impact model using a stochastic automaton called the **influence model** [112, 150].

Analysis of the spatio-temporal stochastic weather model can provide various weather-impact statistics of interest, e.g., the statistics of weather dynamics in a single weather zone. Such information may be of particular interest, for instance when a critical weather zone plays a significant role in delay performance (see [112] for more illustration). When the statistics/pdf of the dynamical weather impact at a single zone is obtained from the stochastic weather model, continuous-time Markov chain models can be constructed to approximate the statistics (see e.g. [47] for a technique for parameterizing Markov models so that state transition durations match desired pdfs). I am particularly interested in these Markov models for local weather impact,since I would like to study impact of weather on particular traffic flows; let us thus discuss these models in further detail. Specifically, in the Markov chain model, states represent different stages of weather or weather-impact evolution, and the weather-state probabilities are governed by

$$\dot{\mathbf{p}}_w(t) = \mathbf{p}_w(t) Q_w \tag{60}$$

where $\mathbf{p}_w(t) = [p_{w_1}(t), ..., p_{w_i}(t), ..., p_{w_n}(t)]$, $p_{w_i}(t)$ represents the probability of weather being at state i at time t, n is the number of states in the Markov chain, and $Q_w \in R^{n \times n}$ is the continuous-time transition matrix.

In this development, I examine two classes of weather events separately, namely an extended-duration severe weather event (e.g. repeated occurrence of storms during a busy-traffic period, or a long-duration winter-storm) and a transient weather event that lasts for a while and then disappears (e.g. morning fog in San Francisco). In the first case, convective weather can be modeled using a **recurrent** Markov chain. In the second case, the weather is represented using a **non-recurrent** Markov chain[1].

For instance, consider the case that bad weather is temporarily present in an airspace region, causing a decrease in the region's capacity, but then disappears. The pdf of the duration of the capacity reduction at the single region due to the convective weather as shown in Figure 4.2a, as generated by the full spatio-temporal weather-impact model. The pdf can be well approximated as being generated from a non-recurrent four-state continuous-time Markov chain with $Q_w = \begin{bmatrix} -0.95 & 0.85 & 0 & 0.1 \\ 0 & -0.85 & 0.85 & 0 \\ 0 & 0 & -0.85 & 0.85 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ (see Figure 4.2b). Specifically in this example, state 1 to 3 represent bad weather (the zone is at reduced capacity), and state 4 represents that the bad weather is gone (the zone is at normal capacity). Once

---

[1] A Markov chain is recurrent, if starting from any state in the Markov chain, there is a non-zero probability that the Markov chain will return to the starting state; otherwise, the Markov chain is non-recurrent.

the Markov chain jumps into state 4, it stays at state 4 forever. The duration of capacity reduction is the length of time that takes the Markov chain to reach state 4 for the first time.

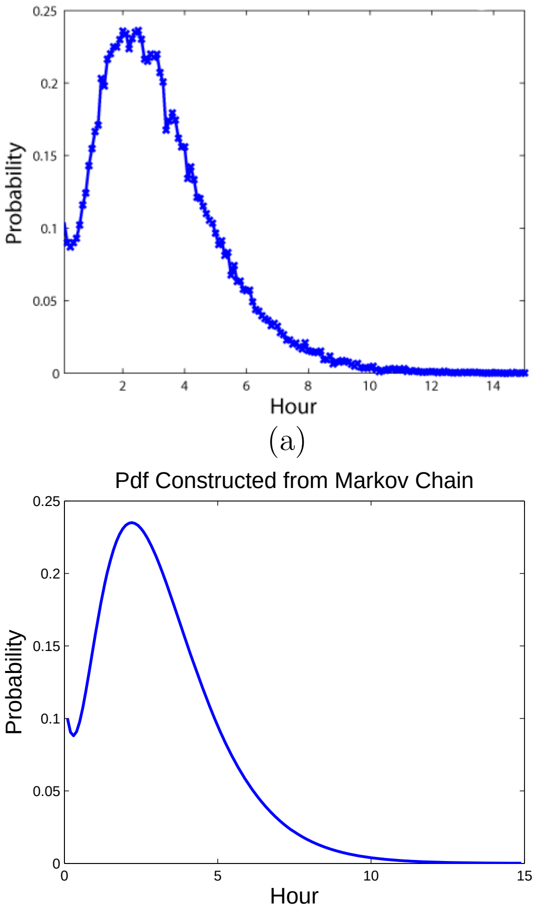
To track weather-state probabilities described by the Markov chain, it is simpler and more apt for air traffic decision-making to use a discrete-time approximation of the Markov chain. It is not difficult to obtain an accurate discretization, simply through choice of a small discretization interval, as shown in Equation 61 (see Figure 4.3 for a discrete time version of the example, with discretization time-step $\Delta t = 20$ min).

$$\begin{aligned} \mathbf{p}_w[k+1] &= \mathbf{p}_w[k] P_w \\ &= \mathbf{p}_w[k](Q_w \Delta t + I), \end{aligned} \tag{61}$$

where $\mathbf{p}_w[k] = [p_{w_1}[k], ..., p_{w_i}[k], ..., p_{w_n}[k]]$, $p_{w_i}[k]$ represents the probability of weather state i at time interval $k\Delta t$, $P_w \in R^{n \times n}$ is the transition matrix, and $P_{w_{I,J}}$ represents the conditional probability that the Markov chain is at state J given that the state is I at the previous time step. For the weather model example discussed in this section, the probability that the duration of capacity reduction takes k time steps can be obtained by recording $p_{w_4}[k]$ (the probability of being at state 4), and subtracting $p_{w_4}[k]$ from the probability being at state 4 at the previous time step $p_{w_4}[k-1]$.

4.2.3. Problem Formulation

The variability in weather events, and especially convective weather, creates significant difficulties in defining strategic traffic management actions as these uncertainties must be accounted for in the models developed. I take the perspective that by integrating the stochastic weather model (using the influence model) and the flow restriction model (using the queuing network model) I can analyze and design strategic management under weather uncertainty. In this development, I begin the investigation by addressing the backlog analysis



(a)

(b)

FIGURE 4.2. (a) Pdf of the duration of capacity reduction at a single region (generated from the simulation of the influence model, (b) Pdf reconstructed from a 4-state Markov chain model.

when a single stream of flow intersects with a weather zone. Here, the bad weather reduces the service rate of the traffic restriction acting on the flow, to reflect deliberate flow management or intrinsic rate reduction due to the weather. In Figure 4.4, I show the dynamics of backlog statistics using Monte Carlo simulations. The comparison between Figure 4.4c and 4.4d reveals that the mean value and the mean duration of abnormal backlog caused by bad weather are enlarged if there is a greater chance of a prolonged bad weather event. The comparison between Figure 4.4d and 4.4e shows that larger inflow rates (representing

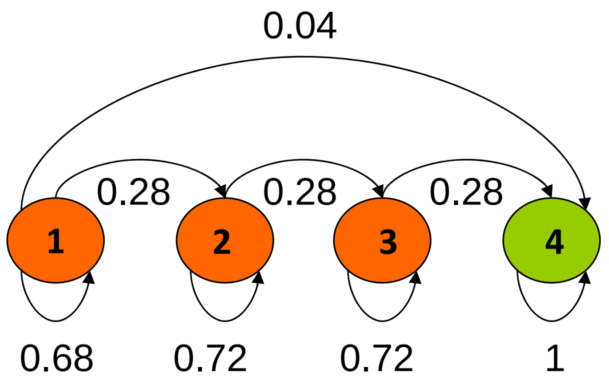


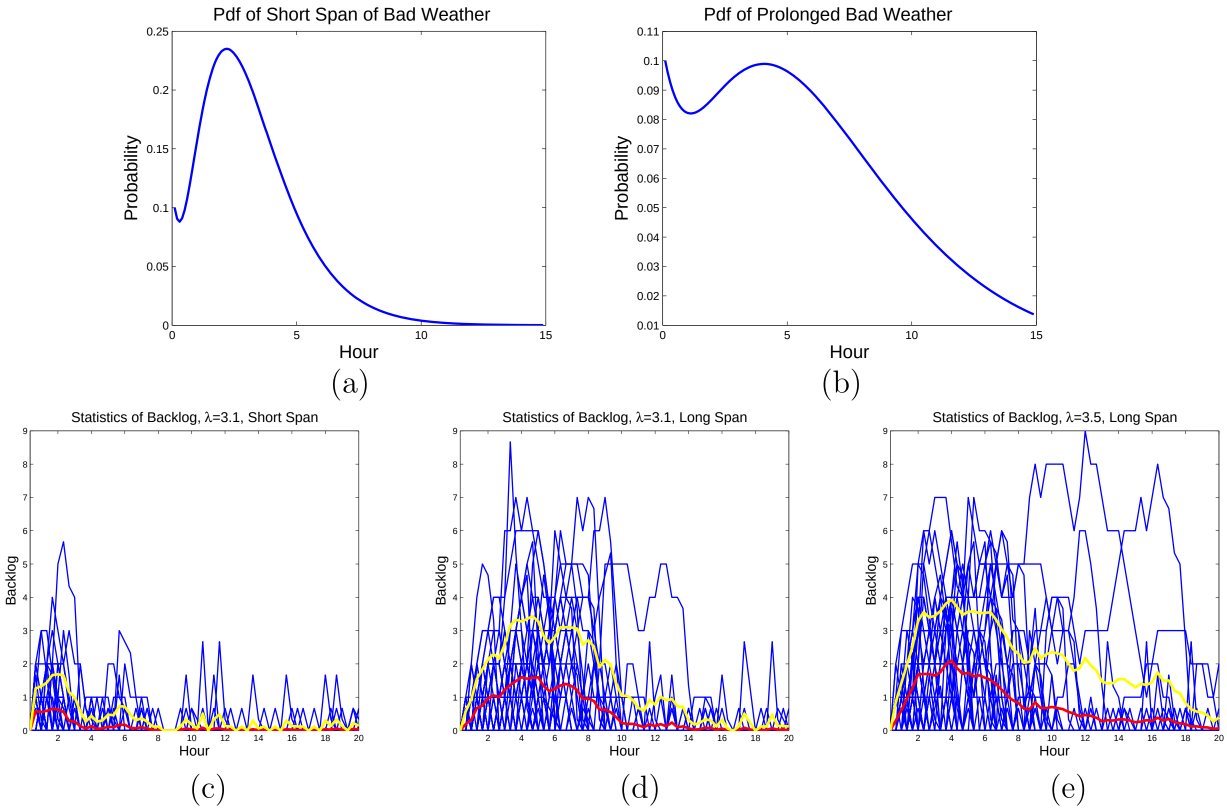
Figure 4.3. Discrete time 4-state Markov model to generate the pdf of the duration of bad weather ($\Delta t = 20$ min).

more demand) also increases the mean value and the mean duration of the abnormal backlog. The Monte Carlo simulations demonstrate that both weather and inflow uncertainties produce significant impact on the number of aircraft delayed. Due to the computational cost of Monte Carlo simulations, and its limitations as a design tool for management actions, it is important to develop efficient analytical tools that permit the prediction of backlog dynamics with uncertain weather and flow.

In the next two sections, I investigate the prediction of dynamical backlog statistics for a stochastic flow under the impact of uncertain weather. Specifically, uncertain weather is modeled using a discrete-time Markov chain model, and the weather-impact-modulated restriction is modeled using a saturation model that approximates a M/D/1 queue with time-varying service rates. The number of aircraft coming during each time interval follows a Poisson distribution, with the probability at time step k represented as

$$P_\lambda(x[k] = c) = \frac{(\lambda \Delta t)^c e^{(-\lambda \Delta t)}}{c!}, c \geq 0. \tag{62}$$

Note: In all examples presented in our work, I assume that the discretization interval $\Delta t = 20min$ for both the saturation model and the weather Markov model, considering that $20min$



(a) (b)

(c) (d) (e)

Figure 4.4. Illustration of the impact of stochastic weather and inflow on backlog. Capacity/service rate (number of aircraft allowed to pass in a hour) is 3 under bad weather conditions, and 10 under good weather conditions. Plot (a) shows the pdf of the duration for a short span of bad weather. (b) shows the pdf of the duration for long span of bad weather. (c),(d), and (e) are the performance analyses for the two-weather-impact-distribution models, under two possible inflow conditions. Blue lines represent the backlog dynamics obtained from Monte Carlo sample runs, red lines represent the mean backlog, and yellow lines represent mean backlog plus the standard deviation of backlog. (c) shows the backlog corresponding to the inflow rate 3.1 and short span of bad weather. (d) is corresponding to the inflow rate 3.1 and prolonged span of bad weather. (e) is corresponding to the inflow rate 3.5 and long span of bad weather.

allows a reasonable approximation and also that it is a reasonable time interval for planning at the strategic time frame.

4.3. Integration of Weather and Flow Models for Performance Evaluation: A Markov Approach

To predict the statistics of backlog/delay caused by flow restrictions in the presence of uncertain weather without using time-consuming Monte Carlo studies, I develop a discrete-time Markov approach to the modeling and analysis of the integrated weather and flow restriction models. Specifically, I use a Markov chain to track the dynamics of buffer length as described by the saturation model. I then construct a master Markov chain, whose states are the combined pairs of weather states and buffer lengths, so as to track the dynamics of buffer length under uncertain weather. Markov chain analysis permits the prediction of steady-state and transient backlog. Finally, I pursue approximation studies of the transient analysis for extreme weather/flow scenarios.

4.3.1. Integrated Markov Model for Weather and Poisson Flow

Let us first consider tracking the dynamics of a saturation model with a fixed service rate u_c. (or equivalently, with saturation restriction $N_c = u_c\Delta t$ during a time interval Δt). The dynamics can be tracked using a Markov chain. Specifically, I construct an infinite-state Markov chain with each state $i \in \{0, 1, 2, ..., \infty\}$ representing the buffer length, i.e., the number of aircraft in the buffer. The transition probability $P_{Q_{i,j}} = P(s[k+1] = j | s[k] = i)$ representing the probability of transiting from buffer length i to buffer length j can be calculated as

$$P_{Q_{i,j}}(N_c) = \begin{cases} P_\lambda(j), & 0 \leq i \leq N_c;\ j \geq 0 \\ P_\lambda(j-i+N_c), & i > N_c;\ j \geq i - N_c \\ 0, & i > N_c;\ j < i - N_c. \end{cases} \tag{63}$$

I note that in simulation or analytical studies, I usually use a truncated finite-state Markov chain to approximate the infinite-state Markov chain. As long as the dimension of the finite-state Markov chain is sufficiently large, the approximation is accurate.

Now let us consider the full model with overlayed stochastic weather. When the flow enters a weather zone, I model the service rate of the flow-restriction as varying because of changing weather impact, and hence the transition probabilities in a Markov model for the queue will also vary. Recall from Section 4.2.2 that weather impact in a single region can be represented by a discrete-time Markov chain with transition probability $P_{w_{I,J}}$, where I and J represent the states of weather as described by the Markov chain. For each weather state I, I model the flow-restriction as having an (in general) different service rate. Therefore, when the weather state is I, the queue length transitions according to the probability $P_{Q_{i,j}}(N_I)$; where N_I is the saturation restriction value (which reflects the service rate) at weather impact state I.

Now I integrate the queuing and weather Markov models by modeling flow restrictions (specifically, queue service rates) as being driven by the stochastic weather model. Specifically, I construct a larger-size master Markov chain, whose states are each defined as a combination of both a weather state I and a buffer length i (which I denote as the pair Ii). The states transition in the master Markov chain according to both weather propagation and incoming flow. Since, given a current weather state and buffer length, weather and buffer length evolve independently, the transition probability of the master Markov chain is equal to the multiplication of the transition probabilities of the two individual Markov chains. Specifically, the transition probability of beginning at weather state I when the number of aircraft in the buffer is i (i.e., state Ii) and going to weather state J and buffer length j (i.e., state Jj) is represented as

$$\begin{aligned} P_{M_{Ii,Jj}} &= P(s[k+1] = Jj | s[k] = Ii) \\ &= P_w(s[k+1] = J | s[k] = I) P_Q(s[k+1] = j | s_w[k] = I, s_Q[k] = i) \\ &= P_{w_{I,J}} P_{Q_{i,j}}(N_I) \end{aligned} \tag{64}$$

Let us denote that the transition matrix for the master Markov chain as P_M.

Let us illustrate the construction of the master Markov chain using the weather impact model discussed in Section 4.2.2. In the weather-impact Markov model, weather impact takes two states: for the good weather condition, the maximum number of aircraft allowed to pass during a time interval Δt is N_{c1}; for the bad weather condition, the maximum number of aircraft allowed to pass during a time interval Δt is N_{c2}. The integrated Markov chain for the 4-state weather model and M/D/1 queue model I considered is shown in Figure 4.5. In this case, $I = 1, 2, 3, 4$, $j = 1, 2, ...\infty$, and $N_I = N_{c2}$ for $I = 1, 2, 3$, and $N_4 = N_{c1}$. When $I = 1, 2, 3$, the transition probability $P_{M_{Ii,J,j}} = P_{w_{I,J}} P_{Q_{i,j}}(N_{c2})$; when $I = 4$, $P_{M_{Ii,J,j}} = P_{w_{I,J}} P_{Q_{i,j}}(N_{c1})$ For the good weather circumstance, the queue length transitions according to the probability $P_{Q_{i,j}}(N_{c1})$; and for the bad weather circumstance, the queuing length transitions instead according to $P_{Q_{i,j}}(N_{c2})$.

4.3.2. Steady-State and Transient Analysis

The construction of the integrated Markov chain allows us to obtain statistics of backlog in a systematic fashion, instead of using intensive Monte-Carlo simulation. In this section, I summarize the steady-state and transient analysis of backlog statistics using the

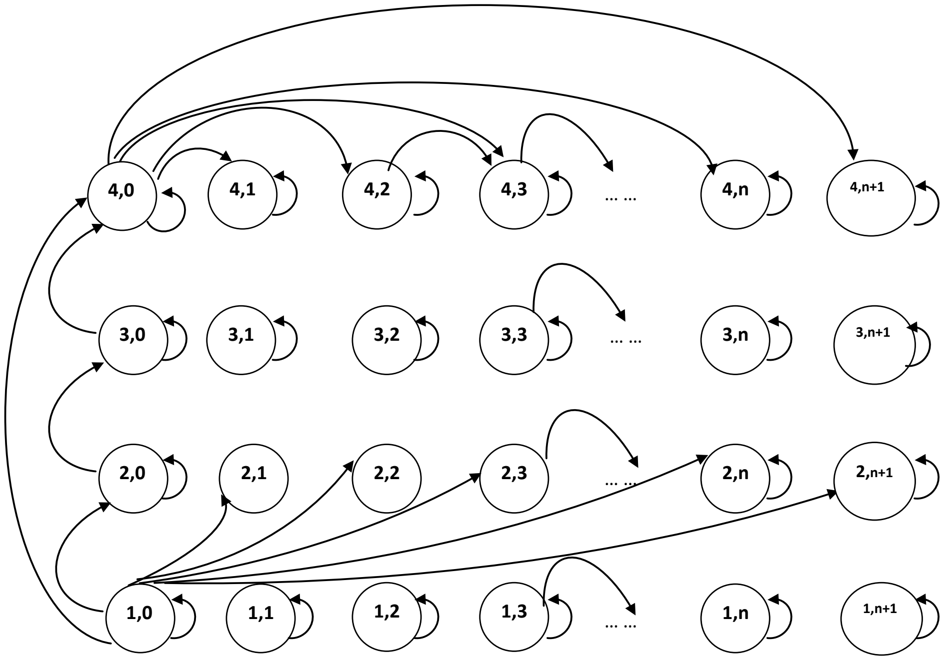


FIGURE 4.5. Integrated master Markov chain

Markov chain approach.

Steady-state statistics can allow performance analysis over a long time horizon, e.g., computation of the average backlog or delay over a long time-span. For extended-duration severe weather, steady-state statistics of traffic backlog under weather uncertainty gives a valuable performance measure. However, for transient weather events, transient dynamics of the flow statistics is often of more interest, because of the short span of severe weather. I now examine both steady-state and transient analysis in greater detail.

In order to obtain the steady-state statistics of performance in terms of backlog, I need to first identify the steady state distribution of the integrated master Markov chain $p_M(s[k \to \infty])_{Ii}$ for $I = 1, ..., n$ and $i = 1, ..., \infty$, i.e., the steady-state probability of state Ii (where i is the queue length and I represents weather state). This steady-state probability can be found from the probabilistic recursion of the master Markov chain (or from one sample run, according to the ergodicity theory [97]). From the steady-state probability distribution, the lth-order moment of backlog can be found as

$$E(B^l) = \sum_{\forall i} \sum_{\forall I} max(i - N_I, 0)^l p_M(s[k \to \infty])_{Ii}. \tag{65}$$

where function $max(a, b)$ takes the maximum value of a and b. I note that statistics of other performance metrics, such as delays, can be found in a very similar way.

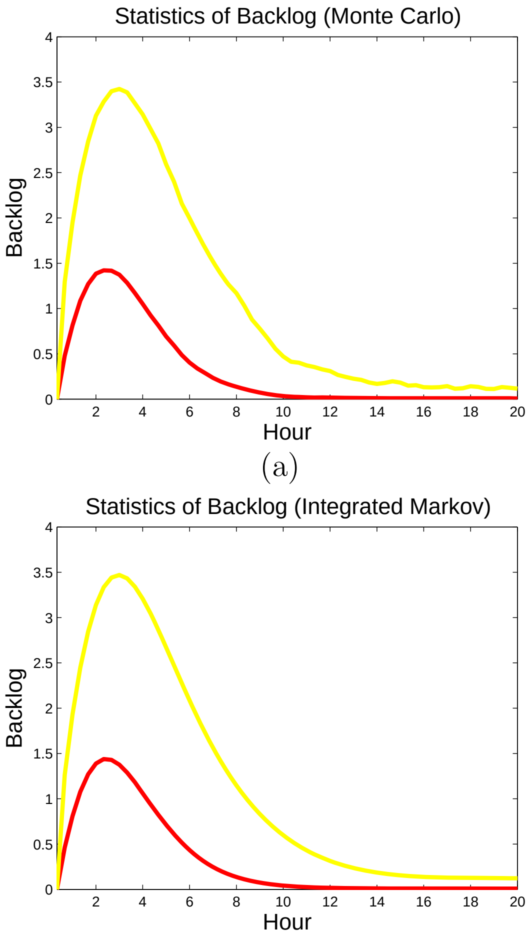
The transient backlog at any time instance can be found by invoking the probability recursion of the combined Markov chain. Upon doing so, the lth-order moment of the backlog at time k can be found as

$$\sum_{\forall i} \sum_{\forall I} max(i - N_I, 0)^l p_M[k-1]_{Ii}, \tag{66}$$

where $p_M[k-1]_{Ii}$ is the probability of the master Markov chain being at state Ii at time step $k - 1$.

Figure 4.6 shows that the mean and variance calculated from the the integrated Markov chain match those calculated from the Monte Carlo simulation. In this example, $N_{c1} = 4$,$N_{c2} = 1$, $\Delta t = 20min$, and $\lambda = 3.5$.

The Markov approach allows the prediction of dynamical backlog statistics for both extended duration severe weather and transient weather. However, from the analysis, I see that the computation is not effective because of the use of infinite-state Markov chain (or a large truncated finite-state Markov chain for approximation) to track the saturation model. In order to obtain a good approximation, even for a single region, the dimension of the transition matrix is high. This limitation makes this approach hard to generalize to the evaluation of backlog for a network of regions, since the dimension of the Markov chain grows exponentially with the number of regions. Next, I seek for some lower-computation approximations of key performance metrics.



(a)

(b)

Figure 4.6. Comparison between the mean and variance calculated from the integrated Markov chain (b) and the Monte Carlo approach (a). The red curve shows the mean of backlog, and the yellow line shows the mean of backlog plus standard deviation.

4.3.3. Approximations for Transient Statistics

In many cases, I may not care about the complete transient dynamics, but only care about a few characteristics of the transient performance, such as the time that the maximum delay occurs, the maximum backlog, and the duration for the excessive backlog to vanish after a severe weather event passes. Our approach has been to use abstractions of the

Markov-chain analysis to obtain simple approximations to the above quantities.

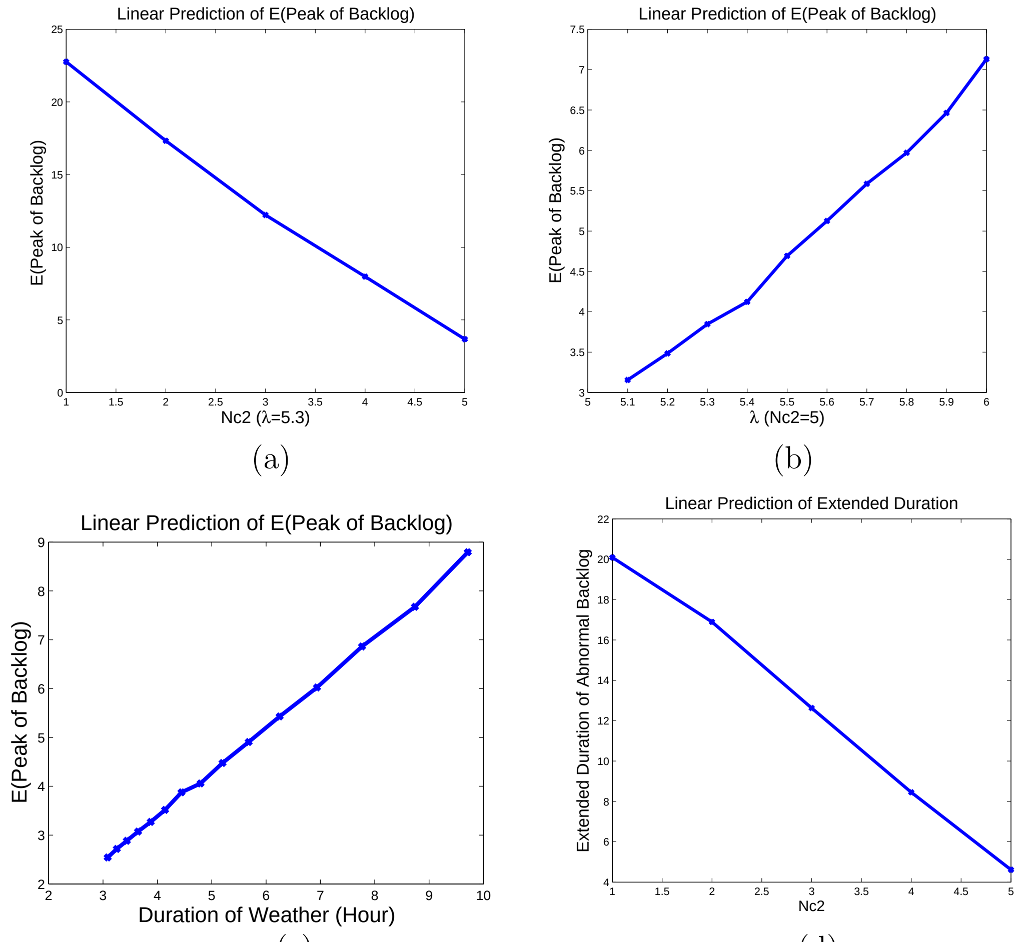
For instance, if I plot the peak of backlog with respect to a small range of varying parameters such as inflow rate, service rates, and the mean of bad weather duration (as shown in Figure 4.7a-c), I see that the peak of backlog can be predicted from these parameters using linear relationships. Similarly, the extended duration for the excessive backlog to vanish after a severe weather event passes can also be predicted using a linear relationship as shown in Figure 4.7d. In fact, under extreme conditions (i.e., inflow rate is much greater than service rate at bad weather, and much smaller than the service rate at good weather) and stochastic weather duration with small variance, I can verify that the peak of backlog (denoted as B_p) and the extended duration (denoted as T_s) can be roughly calculated from the following simple equations:

$$B_p = (\lambda - \frac{N_{c2}}{\Delta t})T_d \tag{67}$$

$$T_s = \frac{B_p - B_s}{\frac{N_{c1}}{\Delta t} - \lambda} \tag{68}$$

where B_s is the steady state backlog after the bad weather is gone and T_d is the mean duration of bad weather. Such measures of backlog and excessive delay predicted using the linear relationships can be used to assist in the design of management actions to reach performance goals, under uncertain severe weather conditions.

For non-extreme conditions, the above linear relationships do not yield good predictions. I instead introduce a jump-linear approach to analyze the transient congestion in Section 4.4. The jump-linear approach allows novel and effective evaluation and possibly design of strategic management plans that take into account of all probable weather scenarios.



(a) (b)

(c) (d)

Figure 4.7. Linear prediction of the characteristics of transient dynamics: (a) The linear relationship between mean value of peak backlog and the saturation restriction under bad weather ($N_{c1} = 12$, $\lambda = 5.3$, prolonged bad weather (see Figure 5.3b)); (b) The linear relationship between mean value of peak backlog and inflow rate ($N_{c1} = 12$, $N_{c2} = 5$, prolonged bad weather (see Figure 5.3b); (c) The linear relationship between mean value of peak backlog and the mean of bad weather duration ($N_{c2} = 4$, $N_{c1} = 7$, and $\lambda = 5.3$)); (d) The linear relationship between the extended duration of backlog and the saturation restriction under bad weather ($N_{c1} = 7$, $\lambda = 5.3$, prolonged bad weather (see Figure 5.3b)).

4.4. Integration of Weather and Flow Models for Performance Evaluation: A Jump-Linear Approach

To evaluate the dynamical impact of convective weather on uncertain flows at a broad spatial scale, I need analytical tools that are computationally-efficient. In this section, I develop a jump-linear approach to the modeling and analysis of the effect of uncertain weather's impact on flows.

Markovian Jump-linear systems—i.e., linear systems whose parameters are modulated by an underlying Markov chain with finite state-space—are a broad class of stochastic hybrid models which have nice tractabilities (e.g., [32, 110, 113]). In this section, I first introduce the jump-linear modeling of the integrated stochastic weather and flow restriction models, then present the prediction of impact statistics using this model, and finally discuss some features and benefits of the jump-linear approach.

4.4.1. Formulation of the Dynamics as a Jump-Linear System

The jump-linear modeling of the integrated flow and weather model is based upon a linear abstraction of the impact of a restriction's impact on flows. Linear abstractions are appealing for large-scale traffic flow modeling because of their tractability and scalability. Let us first describe the principles for developing a linear abstraction, and then the formulation of the integrated weather and flow restriction model into a jump-linear system based upon the linear abstraction [130].

A linear restriction model approximates the relationship between the crossing flow and buffer length as a linear (actually, affine) function [130]. Specifically, in a unit time, the crossing flow $e[k]$ is modeled as a fraction of buffer length $b[k]$ plus a constant, as shown in Equation 69.

$$e[k] = ab[k] + c \tag{69}$$

I note that the linear restriction can be made to resemble the saturation restriction well, through proper choice of the parameters a and c. In essence, the linear restriction is a stochastic linearization of the nonlinear saturation model. For stringent saturation restrictions (i.e., $\lambda\Delta t$ is close to N_c), a is typically small while c is moderate. Meanwhile, for loose saturation restrictions (i.e., $\lambda\Delta t$ is much small than N_c), a is close to 1 and c is close to 0. The parameters a and c can be found by matching the statistical impact of the linear and saturation restrictions on flows. For instance, a procedure to find a and c by matching the steady-state mean backlog and downstream flow variance for Poisson flows was presented in [130]. In cases where transient dynamics need to be matched, I can apply curve fitting tools to find the corresponding a and c for a particular combination of saturation restriction and incoming flow.

Now consider the case that the flow restriction is subject to modulation by stochastic weather. In this case, I model the linear restriction's parameters as being modulated by the stochastic weather. That is, at different weather severities, the restriction strength and hence linear restriction parameters will be different. Since weather is modeled using a Markov chain, the parameters of the restriction are changing according to the weather Markov chain. The dynamics of the integrated model can thus be represented as a jump-linear system:

$$\begin{aligned} e[k] &= a([q[k])b[k-1] + c([q[k]) \\ b[k] &= b[k-1] + x[k] - e[k] \\ B[k] &= b[k-1] - e[k] \end{aligned} \tag{70}$$

where $q[k] \in R^{n\times 1}$ has only one entry as 1 and all other entries as 0, representing the state of the weather Markov chain at time step k, and $a(q[k])$ and $c(q[k])$ represent the values of the parameters a and c associated with the state of the Markov chain $q[k]$. For instance,

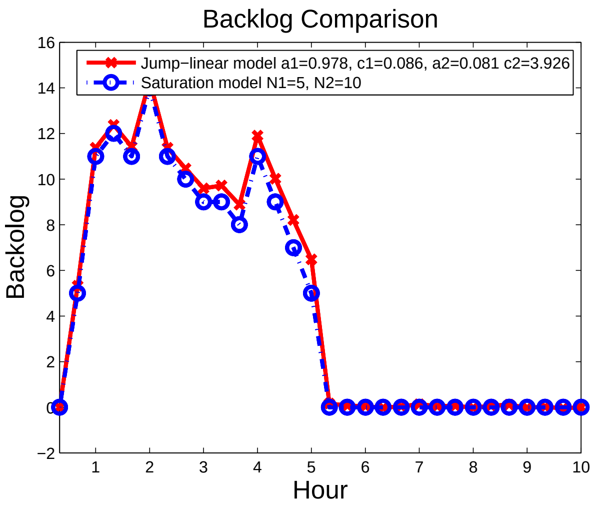


FIGURE 4.8. Comparison between the jump-linear model and the saturation model for a specific inflow sequence and weather sample.

for the weather Markovian model shown in Figure 4.3, a and c take the same value when the Markov chain is in states $1, 2, 3$ representing a bad weather, and takes a different value when the chain is in state 4 representing good weather. The parameters $a(q[k])$ and $c(q[k])$ change values as the Markov chain jumps among states $q[k]$.

The jump-linear formulation presented above resembles the original saturation model integrated with weather, but allows nice tractability as shown in the next subsection. Here in Figure 4.8, I show the simulation of the jump-linear representation and the saturation representation for a particular Poisson flow with $\lambda\Delta t = 4.9$, $\Delta t = 20min$, and a particular weather ensemble where bad weather lasts for 5 hours on average and then disappears. The plots in Figure 4.8 demonstrate that the jump-linear abstraction captures the saturation restriction well.

4.4.2. Statistical Analysis of the Jump-Linear Model

In this section, let us demonstrate the prediction of backlog statistics using the jump-linear model. To do so, I write recursions for the moments the jump-linear model (Equation

71) into a moment-linear representation [110], which allows us to trace the statistics of a jump-linear model.

Specifically, as an illustration, let us trace statistics of the backlog. To begin from Equation 71, the dynamics of backlog $B[k]$ can be represented as

$$B[k+1] = (1 - a(q[k]))(B[k] + x[k]) - c(q[k]) \tag{71}$$

Now let us introduce the vector $\sigma[k]$, which is defined as

$$\sigma[k] = q[k] \otimes \begin{bmatrix} B[k] \\ 1 \end{bmatrix}. \tag{72}$$

I note that the conditional expectation $E[\sigma[k+1]|\sigma[k])$ can be written as

$$\begin{aligned} E[\sigma[k+1]|\sigma[k]] &= E\left[q[k+1] \otimes \begin{bmatrix} B[k+1] \\ 1 \end{bmatrix} |B[k], q[k]\right] \\ &= E[q[k+1]|q[k]] \otimes E\left[\begin{bmatrix} B[k+1] \\ 1 \end{bmatrix} |B[k], q[k]\right] \\ &= {P_w}'q[k] \otimes \left(\begin{bmatrix} 1 - a[q[k]] & c[q[k]] \\ 0 & 1 \end{bmatrix} \begin{bmatrix} B[k] \\ 1 \end{bmatrix} + \begin{bmatrix} \lambda\Delta t \\ 0 \end{bmatrix}\right) \\ &= {P_w}'q[k] \otimes \begin{bmatrix} \lambda\Delta t \\ 0 \end{bmatrix} + {P_w}'q[k] \otimes \begin{bmatrix} 1 - a[q[k]] & c[q[k]] \\ 0 & 1 \end{bmatrix} \begin{bmatrix} B[k] \\ 1 \end{bmatrix} \\ &= {P_w}'q[k] \otimes \begin{bmatrix} \lambda\Delta t \\ 0 \end{bmatrix} + P_w' \otimes \begin{bmatrix} 1 - a[q[k]] & c[q[k]] \\ 0 & 1 \end{bmatrix} \sigma[k] \end{aligned} \tag{73}$$

From Equation 74, I can find the mean of $\sigma[k+1]$ as

$$E[\sigma[k+1]] = E[E[\sigma[k+1]|\sigma[k]]] \tag{74}$$

$$= \quad {P_w}' E[q[k]] \otimes \begin{bmatrix} \lambda \Delta t \\ 0 \end{bmatrix} + P_w' \otimes \begin{bmatrix} 1 - a[q[k]] & c[q[k]] \\ 0 & 1 \end{bmatrix} E[\sigma[k]]$$

Since all quantities in Equation 75 except the variables $E[\sigma[k]]$ are known, this equation allows us to calculate the dynamical mean of $\sigma[k]$ through an effective recursive fashion. In fact, if I define $A = {P_w}' E[q[k]] \otimes \begin{bmatrix} \lambda \Delta t \\ 0 \end{bmatrix}$ and $B = P_w' \otimes \begin{bmatrix} 1 - a[q[k]] & c[q[k]] \\ 0 & 1 \end{bmatrix}$, $E[\sigma[k+1]]$ can be found using

$$E[\sigma[k+1]] \quad = \quad B^k(E[\sigma[0]] + (B - I)^{-1}A) - (B - I)^{-1}A. \tag{75}$$

Moreover, from Equation 72, I can easily derive that $E[B[k]]$ can be calculated from $E[\sigma[k]]$ using

$$E[B[k]] = E[\sigma[k]] - 1. \tag{76}$$

In Figure 4.9, I show the prediction of mean backlog using the integrated Markov chain approach and the jump-linear approach. In this example, stochastic weather model is shown in Figure 4.2 and 4.3, $\lambda \Delta t = 4.9$, $\Delta t = 20 min$, $N_{c1} = 10$ and $N_{c2} = 5$. The comparison between the two plots shows that the jump-linear approach allows a good prediction of mean backlog.

I note that higher-order statistics of the backlog, as well as statistics of other performance metrics, can be computed in similar fashion.

4.4.3. Example and Discussions

The major contribution of this section is the introduction of the jump-linear approach to the evaluation of air traffic system performance in the presence of uncertain weather.

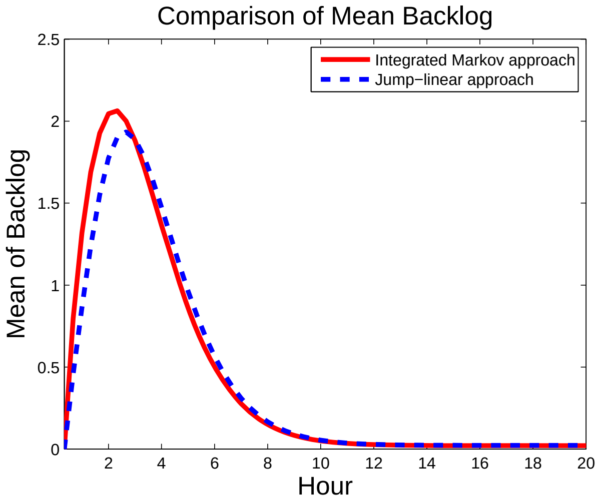


Figure 4.9. Prediction of mean backlog: comparison between the jump-linear and the integrated Markov chain approaches.

In this section, let us first use an example to show the insights the jump-linear approach provides, and then discuss some features/benefits of this approach.

It is very efficient to obtain various insights using the jump-linear approach. As an example, let us answer whether the mean of weather duration is sufficient for performance prediction. In Figure 4.10, I compare backlog predicted from the full weather pdf and from the mean weather (i.e., assuming that the weather duration is equal to its mean value), using the simple recursions of the jump-linear approach. I see that there is a large offset between the two dynamics, as reflected by the measures such as the maximum mean backlog and the duration of excessive delay. This insight is indeed informative since it is typical in practice to use mean weather duration for traffic system performance evaluation, due to the difficulty in performance evaluation under uncertain dynamical weather. The use of mean weather condition as a deterministic condition for performance evaluation avoids dealing with the stochastic weather. However, this example shows that mean weather is not sufficient for a good prediction, and the availability of richer weather information (such as a pdf) allows a

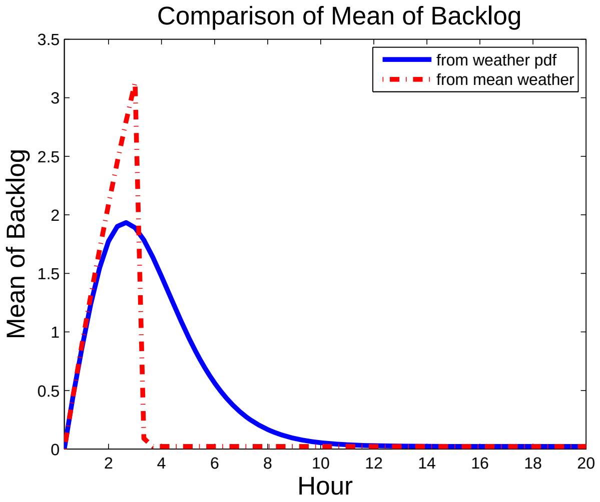


FIGURE 4.10. Comparison of backlog predicted from weather pdf and mean weather.

more precise prediction.

Here let us summarize the features of the jump-linear approach in air traffic system performance evaluation under weather uncertainty:

- **Efficiency.** As shown in the previous section, the jump-linear approach provides an effective way to evaluate traffic backlog (and other performance metrics) under uncertainty. There is no need to carry out intensive Monte Carlo simulations to understand the performance under uncertainty. The approach is also much more effective than the integrated Markov chain approach. The dimension of the recursion to find means in the jump-linear approach is $2n$, where n is number of states in the Markov chain model for weather. Meanwhile, the integrated Markov chain requires a recursion with the order mn, where m is the number of states in the truncated Markov chain that tracks the queue length. As the truncated Markov chain is used to approximate the infinite-state Markov chain, m is large for a good approximation (i.e., $m >> 2$).

- **Precision.** The jump-linear approach allows the prediction of backlog statistics with weather modeled as a stochastic automaton. As shown by Figure 4.9, the use of a precise stochastic model for uncertain weather provides an evaluation of backlog with nice precision. In fact, the only offset comes from the use of linear restriction to capture the saturation constraint. Moreover, the prediction is much more precise compared to the use of mean weather information as shown in the example in this section.
- **Scalability** The approach is promising to be generalized to evaluate performance at a broad spatial scale, with the whole system modeled as a big jump-linear system. The significant characteristic of the approach is that the dimension of computation grows linearly with the increase of the number of regions in consideration. As such the jump-linear approach has significant potential for the evaluation of NAS performance for the NextGen.
- **Designability** The design of optimal management actions using the jump-linear approach is concerned with choosing parameters of the linear restriction for best statistical performance under stochastic weather and flow. This task is in essence related to the control of jump-linear systems [40, 109]. The tractability of the jump-linear model makes the design problem tractable. I leave the design problem to future work.

4.5. Concluding Remarks and Future Work

Weather uncertainty plays a critical role in the performance of air traffic systems, especially in the strategic time-frame. Systematic and effective evaluation of system performance under dynamical weather uncertainty is a crucial step toward the design of strategic management actions. In this chapter, I have made two contributions to the study of strategic traffic management under weather uncertainty:

(1) I have formulated integrated models of stochastic weather and traffic flow. Specifically, I have introduced models in which traffic flows and flow constrictions have parameters that are modulated by an underlying transient or long-duration weather process. I have argued that such models permit useful analysis of traffic performance metrics under weather uncertainty.

(2) I have developed two methods, namely the integrated Markov approach and the jump-linear approach, that allow prediction of performance statistics like backlog under uncertain weather. Of particular note, the jump-linear approach models flow restriction using linear relationships, with parameters modulated by a Markov chain describing weather uncertainty. The tractability and scalability of the approach makes it promising for the evaluation and design of strategic management actions under uncertain weather at a broad spacial scale.

CHAPTER 5

PERFORMANCE EVALUATION AND OPTIMAL DECISION-MAKING FOR STRATEGIC AIR TRAFFIC MANAGEMENT UNDER WEATHER UNCERTAINTY

5.1. Introduction

Strategic air traffic flow management is concerned with planning air traffic 2-15 hours in advance. The decision-making process at this long look-ahead time is complicated by a variety of uncertainties, the most prominent of which include weather and demand uncertainties [84, 85, 112, 118, 140]. In particular, traffic demand is subject to a range of uncertain variabilities, caused by delay events, management initiatives, pop-up flights, etc [140]. With respect to weather, a precise prediction of the intensities of further weather events is currently unavailable due to the limitation of meteorological techniques. Because of the uniqueness of weather condition in the United States, convective weather events (such as storms) account for the primary reason of traffic delays, according to the Federal Aviation Administration (FAA) [4].

Effectively finding the best management strategies is nontrivial considering the weather and demand uncertainties, and nonlinear impact of management strategies and convective weather on flows [130, 135]. As such, the most intuitive way to assess weather impact and optimal management strategy design is to use Monte Carlo simulations as shown in Figure 5.1. In particular, this process involves finding the performance of each management strategy through simulating it against each possible pair of weather and demand ensembles. As both weather and demand are stochastic processes, enumerating pairs of weather and demand ensembles typically results in a very large ensemble space and thus intensive simulation runs. In this chapter, I seek effective simulation and analysis methods to reduce the computation load for weather impact and optimal management studies.

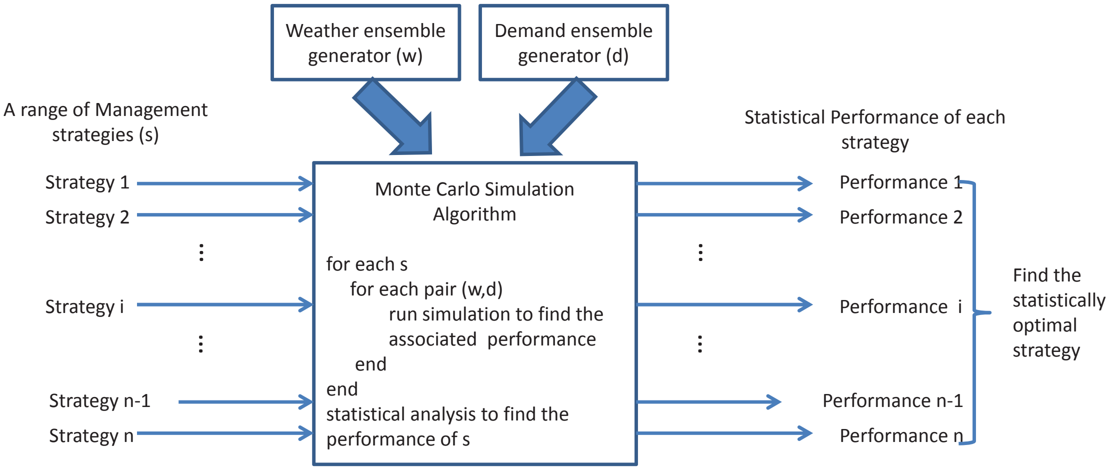


FIGURE 5.1. Illustration of the Monte Carlo decision making process.

In order to permit effective evaluation and design methodologies, I introduce an integrated weather, demand, and management modeling framework. In particular, I model demand as a stochastic process. Examples of demand models include Poisson processes and more realistic models that capture fixed flight plans and uncertain pop-up flights (obtained from historical data) [130, 140]. I also model uncertain dynamic weather impact as Markov chains [112, 151]. Weather impact is captured by capacity reduction, due to the increased in-trail distance requirements and controller workload. Moreover, I model traffic management plans and capacity constraints as queuing service rates. Their effects are to shape downstream flows at the cost of accumulated backlog (i.e., the number of aircraft being delayed due to the service rate) [130, 135]. The integration of the above three models permits an analytical framework for weather impact evaluation and optimal management design under uncertainty.

As a step toward the optimal management design at a National Airspace System (NAS) level, I focus on a simple scenario in this chapter: a stream of flow enters an uncertain weather zone. I then study: 1) the predicted statistics of delayed aircraft due to uncertain weather, and 2) the design of optimal management restrictions subject to weather and demand uncertainties. The contribution of this chapter is summarized in the following.

(1) *Evaluation of uncertain weather impact on the statistics of delayed aircraft.*

Transient backlog is an important metric to evaluate the impact of uncertain weather on flows. In [151], we studied a master-Markov approach and a jump-linear approach to predict transient mean backlog. In this chapter, I extend the previous results to higher moments, and show that these statistics can also be easily obtained through a simple recursion. These higher moments provide rich information about uncertain weather impact, such as the range of variability around the mean backlog.

(2) *Performance metric capturing both management and non-management induced delays.*

Interestingly, I find that the mean total backlog when no management is applied represents the minimum achievable one in the presence of uncertain weather. As flow management is always designed in advance, it is impossible for any specific pre-planned management to be optimal for each of the possible weather ensembles. This observation suggests to me that total mean backlog cannot be used as the sole cost function for optimal management design. In this chapter, I suggest the use of a combination of management and non-management induced backlogs as the optimization cost functions.

(3) *Design of optimal flow management subject to weather and demand uncertainties.*

I provide a systematic approach to find the best management plan under weather and demand uncertainties. The method is based upon a mesh of an effective simulation method called Probabilistic Collocation Method (PCM) and a Markov chain (or jump-linear) analysis to quickly identify a low-order mapping between management plan and total cost, from which the best management solution is obtained.

The remainder of the chapter is organized as follows. In Section 5.2, I present the modeling framework and preliminary results required for the development of this chapter. In Section 5.3, I present a closed-form analytical approach to predict weather impact. In Section 5.4, I introduce the PCM approach for optimal management design. Section 5.5 includes a brief conclusion.

5.2. The Modeling Framework and Problem Formulation

In this section, I describe the modeling framework, and then formulate the weather impact evaluation and optimal management design problems to be investigated in this chapter.

5.2.1. Modeling Framework

The modeling framework integrates the demand model, weather model, and queuing model as shown in Figure 5.2. Let me describe the details of each model.

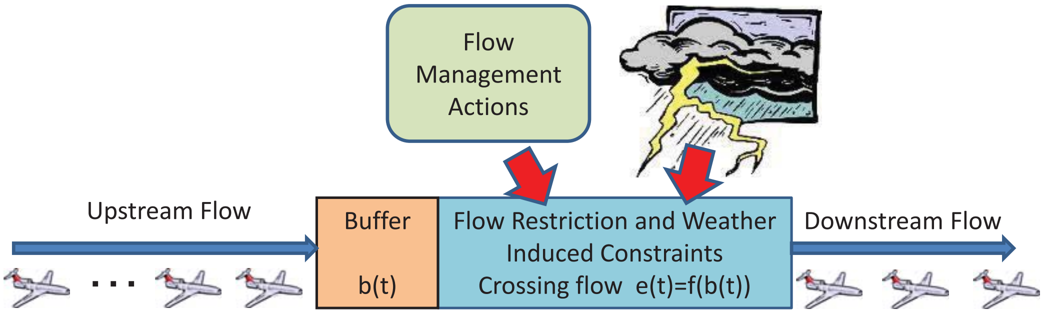


FIGURE 5.2. Illustration of the modeling framework.

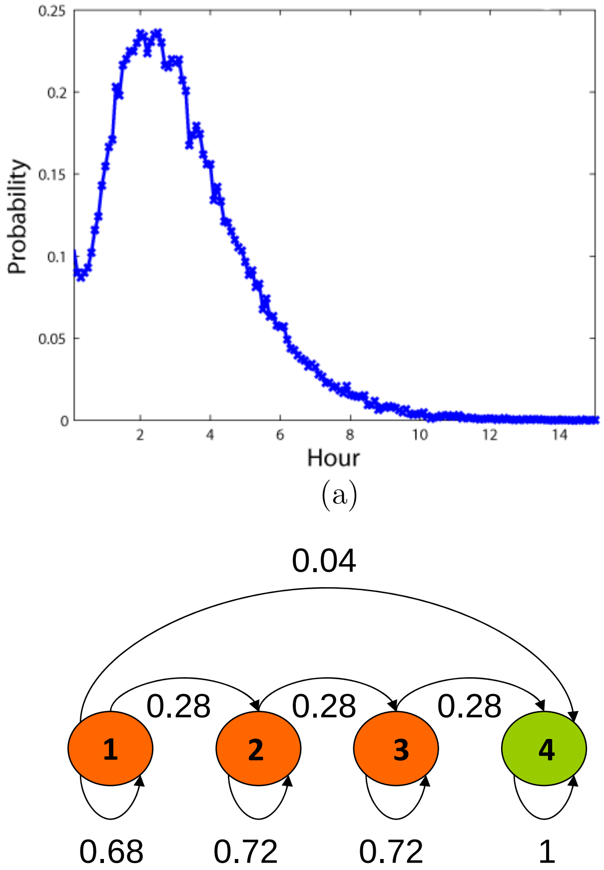
5.2.1.1. Demand Model

Demand, capturing the number of aircraft coming to a region, can be modeled as a stochastic process $x[k]$. The most commonly used stochastic demand model is the Poisson process [97]. Poisson process is an independent stochastic process, in which $x[k]$ at each time k is a Poisson random variable with mean $\lambda\Delta T$ and variance $\lambda\Delta T$, where λ is the average number of coming aircraft per unit time and ΔT is the time interval between adjacent time steps. In [140], more complicated demand models that utilize both deterministic flight plan information and uncertain historical pop-up flights information are introduced. In this study, I model the demand as a general independent stochastic process $x[k]$ with mean $u[k]$ and variance $v[k]$.

5.2.1.2. Weather Model

Stochastic weather impact is modeled as a finite state Markov chain. In the Markov chain, each state s_i represents a particular stage of weather development, and the transition probability $P_{w_{ij}}$ represents the chance for the weather to progress from state s_i to s_j during a time interval ΔT. I denote the transition probability matrix as P_w, in which the ith row, jth column entry is $P_{w_{ij}}$. Each state $s_i \in \mathcal{S}$ is associated with a flow restriction constraint $w(s_i) \in \mathcal{W} = [w_1, w_2, ...]$ that is related to weather intensity, reflecting capacity reduction. For the convenience of my presentation, we introduce a vector $q[k]$ to indicate the state of the Markov chain at time k. In particular, $q[k]$ is of length n, where n is the total number of states. All the entries in $q[k]$ are 0, except the ith entry $q_i[k]$ being 1, indicating the state of Markov chain at time k to be s_i.

Let me use an example to illustrate the weather model. If the weather-induced capacity reduction has two levels, no reduction (with flow restriction rate w_1) and partial reduction (with flow restriction rate w_2), and if the duration of weather-induced reduction is captured by the pdf shown in Figure 5.3a, the stochastic weather impact can be modeled as 4-state Markov chain shown in Figure 5.3b. Also, $w([1\ 0\ 0\ 0]') = w([0\ 1\ 0\ 0]') = w([0\ 0\ 1\ 0]') = w_2$ and $w([0\ 0\ 0\ 1]') = w_1$. The symbol $'$ denotes the transpose of a vector or a matrix.



(a)

(b)

Figure 5.3. (a) Probability distribution function (pdf) of the duration of capacity reduction at a single region, (b) Discrete-time 4-state Markov model to generate the pdf in the left ($\Delta T = 20$ min) [151].

5.2.1.3. Queuing Model

As shown in Figure 5.2, a queuing model is a natural framework for capturing the impact of weather and management on traffic flows [135]. A queuing model is composed of three parts: buffer, flow restriction, and crossing flow. A discrete-time queue model works as follows. All coming aircraft to a region enters the buffer automatically. I denote the buffer length as $b[k]$. At any time k, a portion of aircraft in the buffer (denoted by $f(b[k-1])[k]$) can pass the flow restriction and enter the downstream region, where $f()$ is the restriction function dependent on the type of queue. This crossing flow is denoted as $e[k]$. In this chapter, I consider a discrete-time version of the G/D/1 (general inflow, deterministic service rate, and single server) queue. In this queuing model, $f()[k]$ is a saturation function, i.e. if $b[k-1]$ is larger than the maximum service rate $N[k]$, $f(b[k-1])[k] = N[k]$; otherwise $f(b[k-1])[k] = b[k-1]$. The dynamics of such a queue are captured by the following equations [130]. In this equation, the backlog $B[k]$, which captures the number of delayed aircraft at time k, is a natural performance metric. $B[k]$ can be found as the number of aircraft in the buffer at time $k-1$ subtracting the crossing flow at time k.

$$\begin{aligned} e[k] &= \begin{cases} b[k-1], & (b[k-1] \leq N[k]) \\ N[k], & (b[k-1] > N[k]) \end{cases} \\ b[k] &= b[k-1] + x[k] - e[k] \\ B[k] &= b[k-1] - e[k] \end{aligned} \tag{77}$$

At the limit of small ΔT, the saturation model approaches the G/D/1 model [130, 135].

5.2.2. Problem Formulation

Convective weather events reduce region capacity, inducing accumulated backlog. As discussed in [151], the statistics of transient backlog is a natural metric to evaluate weather impact. This is the first problem that I investigate in this chapter. Besides weather impact evaluation, I also study optimal strategic traffic management before the occurrence of convective weather. Strategic management can increase the efficiency of traffic systems for several reasons: 1) it enhances safety as pilots and controllers can prepare ahead-of-time, rather than having to deal with urgent events; 2) it can reduce costs as advanced strategic management strategies such as Ground Delay Programs (GDP) can reduce fuel consumption;

and 3) it also permits NAS-wise planning to better allocate limited resources. In this chapter, I investigate how to design statistically optimal management strategies under the weather and demand uncertainties. Let us formulate these two problem mathematically below.

PROBLEM 1: EVALUATION OF TRANSIENT WEATHER IMPACT

As I discussed above, weather impact is captured by the statistics of backlog $B[k]$. In this chapter I evaluate the first two moments of $B[k]$, namely $E[B[k]]$ and $\sigma^2[B[k]]$, given the stochastic demand model $x[k]$ and the Markov chain stochastic weather model. These moments provide rich information about the severity and variability of weather impact on flows.

Since convective weather reduces region capacity (or flow restriction rate), I can capture weather impact by expressing the maximum service rate $N[k]$ in Equation 77 as the weather-induced flow restriction rate w_i at time k. As weather is modulated by a Markov chain, $N[k]$ is a function of the Markov state $q[k]$

$$N[k] = w(q(k)). \tag{78}$$

For instance, in the example of weather model shown in Figure 5.2, we have $N[k] = w_1$ when $q[k] \in \{[1\ 0\ 0\ 0]', [0\ 1\ 0\ 0]', [0\ 0\ 1\ 0]'\}$, and $N[k] = w_2$ when $q[k] = [0\ 0\ 0\ 1]'$.

PROBLEM 2: OPTIMAL MANAGEMENT DESIGN UNDER WEATHER AND DEMAND UNCERTAINTIES

Management strategy is captured by a series of deterministic flow restrictions $N_c[k]$. When both flow management and convective weather are in place, $N[k]$ in Equation 77 is captured by the minimum of the management restriction and the weather-induced flow restriction, i.e.,

$$N[k] = min(N_c[k], w(q[k])). \tag{79}$$

The optimal management is a series of $N_c[k]$ that minimizes a cost function (typically connected to delay statistics), given the stochastic demand and convective weather models.

In this chapter, I consider a simple scenario: the convective weather starts at the initial time $k = 0$ with uncertain ending time. When the convective weather is present, the restriction rate is w_2; and when the weather event is over, the restriction rate goes back to the normal w_1. Designing the optimal management plan is concerned with finding the start time, end time, and intensity of $N_c[k]$, given all the uncertainties. As convective weather starts from the initial time with a fixed intensity w_2, it is a reasonable practice to start the flow management $N_c[k]$ at $k = 0$ with intensity w_2. As such, I am concerned with finding the optimal end time of the management $N_c[k]$.

5.2.3. Overview of Our Approaches

In this chapter, I introduce two effective approaches to address the above weather impact evaluation and optimal flow management design problems. In particular, I take an analytical approach to characterizing transient impact statistics, specifically one based upon a Markov jump-linear approximation. This approach provides simple assessment of backlog and its variability caused by uncertain bad weather in a particular day. I also suggest using a smart simulation technique known as PCM for designing the flow management capability. As a purely analytical method is difficult to obtain while Monte Carlo simulations are time-consuming (for a problem of realistic scale), PCM permits an approximated suboptimal but much faster approximation, which also incidentally gives the mapping between design parameters and cost. The results in this chapter are presented in a very concise form, as they are heavily built upon our previous modeling and analysis efforts [112, 130, 132, 135, 151]. Please refer to these earlier papers for the background.

5.3. Using Jump-Linear Approach to Evaluate Uncertain Weather Impact

The Monte Carlo approach (as illustrated in Figure 5.1) is a straightforward approach to evaluate uncertain weather impact. However, as both weather and demand are subject to uncertainties, the Monte Carlo approach results in a large number of simulations and thus is time consuming for real-time management. As such, I seek a systematic analytical approach to effectively evaluate uncertain weather impact. This approach is based on a stochastic linearization of the original dynamics (Equation 77), transforming it into a Markov jump-linear system (see Equation 80) [151]. Markov jump-linear system is the type of linear systems with parameters modulated by a Markov chain [110].

$$\begin{aligned} e[k] &= a(q[k])b[k-1] + c(q[k]) \\ b[k] &= b[k-1] + x[k] - e[k] \\ B[k] &= b[k-1] - e[k] \end{aligned} \tag{80}$$

Equation 80 leads to the following dynamics of the backlog $B[k]$ through simple algebra.

$$B[k+1] = (1 - a(q[k]))(B[k] + x[k]) - c(q[k]) \tag{81}$$

I note that in the jump-linear representation, the cross flow $e[k]$ is a *stochastic linear* approximation of the original nonlinear saturation function shown in Equation 77. The parameters a and c are associated with different values for different $w(q(k))$. As such, a and c are also functions of the Markov state $q[k]$. The selection of these two parameters is critical for the performance of the jump-linear approximation. Please refer to [151] for the procedure to choose a and c. A complete treatment of selecting these two parameters is left to future work.

Let me demonstrate the use of the jump-linear formulation to analyze transient uncertain weather impact. In particular, I show how the first and second moments of backlog can be analyzed using simple recursions.

5.3.1. Transient Analysis of Mean Backlog

In [151], I introduced a procedure to find transient mean backlog when the demand is a Poisson flow. This result can be easily generalized to the case when demands are modeled as general independent stochastic processes with mean $u[k]$ and variance $v[k]$. Here I only summarize the result. To obtain the transient mean backlog, I introduce a vector $\sigma[k] = q[k] \otimes \begin{bmatrix} B[k] \\ 1 \end{bmatrix}$. The mean of $\sigma[k]$ can be found through a linear recursion as shown below.

$$E[\sigma[k+1]] = {P_w}' E[q[k]] \otimes \begin{bmatrix} u[k] \\ 0 \end{bmatrix} + P'_w \otimes \begin{bmatrix} 1 - a[q[k]] & c[q[k]] \\ 0 & 1 \end{bmatrix} E[\sigma[k]] \tag{82}$$

The mean backlog $E[B[k]]$ at each time k is found to be $\mathbf{1}_{1\times l_1} E[\sigma[k]] - 1$, where l_1 is the length of the vector $\sigma[k]$.

5.3.2. Transient Analysis of the Variance of Backlog

Higher moments can be obtained using a similar approach. In this section, I show how the variance of transient backlog $\sigma^2[B[k]]$ is obtained through a simple recursion. Variance of backlog describes performance variability in a range of weather possibilities, which helps in estimating the robustness of management initiatives.

To permit the analysis, I introduce a vector $\sigma_{\mathbf{2}}[k]$ which contains both $B[k]$ and $B^2[k]$.

$$\sigma_{\mathbf{2}}[k] = q[k] \otimes \begin{bmatrix} B^2[k] \\ B[k] \\ 1 \end{bmatrix}. \tag{83}$$

Squaring both ends of Equation 81 leads to

$$\begin{aligned} B^2[k+1] &= (1-a(q[k]))^2 B^2[k] + \left(2(1-a(q[k]))^2 x[k] - 2(1-a(q[k]))c(q[k])\right) B(k) \\ &\quad +(1-a(q[k]))^2 x^2[k] + c^2(q[k]) - 2(1-a(q[k]))c(q[k])x[k], \end{aligned} \tag{84}$$

from which the conditional mean of $E[B^2[k+1]]$ given $B[k]$ can be found as

$$\begin{aligned} & E[B^2[k+1]|B[k]] \\ = \;& (1-a[q[k]])^2 B^2[k] + \left(2(1-a[q[k]])^2 u[k] - 2(1-a[q[k]])c[q[k]]\right) B(k) \\ + \;& (1-a(q[k]))^2(v[k] + u^2[k]) + c^2[q[k]] - 2(1-a[q[k]])c[q[k]]u[k]. \end{aligned} \tag{85}$$

For the ease of presentation, I denote $A_2 = 2(1-a[q[k]])^2 u[k] - 2(1-a[q[k]])c[q[k]]$ and $B_2 = (1-a(q[k]))^2(v[k]+u^2[k])+c^2[q[k]]-2(1-a[q[k]])c[q[k]]u[k]$. The conditional expectation $E[\sigma_2[k+1]|\sigma_2[k]]$ can then be written as

$$\begin{aligned} E[\sigma_{\mathbf{2}}[k+1]|\sigma_{\mathbf{2}}[k]] &= E\left[q[k+1] \otimes \begin{bmatrix} B^2[k+1] \\ B[k+1] \\ 1 \end{bmatrix} |B[k], q[k] \right] \\ &= E[q[k+1]|q[k]] \otimes E\left[\begin{bmatrix} B^2[k+1] \\ B[k+1] \\ 1 \end{bmatrix} |B[k], q[k] \right] \end{aligned} \tag{86}$$

$$= P_w{}'q[k] \otimes \begin{bmatrix} 1-a^2[q[k]] & A_2 & B_2 \\ 0 & 1-a[q[k]] & (1-a[q[k]])u[k]-c[q[k]] \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} B^2[k] \\ B[k] \\ 1 \end{bmatrix}$$

$$= P_w{}' \otimes \begin{bmatrix} 1-a^2[q[k]] & A_2 & B_2 \\ 0 & 1-a[q[k]] & (1-a[q[k]])u[k]-c[q[k]] \\ 0 & 0 & 1 \end{bmatrix} \sigma_2[k].$$

Taking the expectation of both sides of this equation with respect to $\sigma_2[k]$, the mean of $\sigma_2[k+1]$ can be obtained as

$$E[\sigma_2[k+1]] = P_w{}' \otimes \begin{bmatrix} 1-a^2[q[k]] & A_2 & B_2 \\ 0 & 1-a[q[k]] & (1-a[q[k]])u[k]-c[q[k]] \\ 0 & 0 & 1 \end{bmatrix} E[\sigma_2[k]]. \tag{87}$$

The above recursion can be used to obtain $E[\sigma_2[k]]$ at any time from the the initial condition $E[\sigma_2[0]]$. $E[B^2[k]]$ can then be derived from $E[\sigma_2[k]]$ by realizing that

$$E[B[k]] = \mathbf{1}_{1\times n} \otimes [0\ 1\ 0]' E[\boldsymbol{\sigma_2}[k]] \tag{88}$$

$$E[B^2[k]] = \mathbf{1}_{1\times l_2} E[\boldsymbol{\sigma_2}[k]] - 1 - E[B[k]], \tag{89}$$

where n is the length of $q[k]$ and l_2 is the length of the vector $\sigma_2[k]$. The variance of $B[k]$ can be expressed as

$$\sigma^2[B[k]] = E[B^2[k]] - (E[B[k]])^2. \tag{90}$$

The result suggests that the first and second moments of demand and weather Markov chain are sufficient to infer the mean and variance of backlog at any time using the jump-linear approach. This fast approach to second moment analysis is of significant value, as it could potentially be used to evaluate the local impacts of TMIs on performance in flow contingency management (FCM) [123], without requiring simulation of the full NAS; it also gives an indication of the variability of performance over the range of weather outcomes.

5.3.3. An Illustrative Example

Let me use a simple example to illustrate the jump-linear approach to obtain the variance of transient backlog. In the example, demand is modeled as Poisson process $x[k]$ with mean u[k]=4.9/20$mins$. Moreover, weather is represented by a 4-state Markov chain with two intensity levels ($w_1 = 10/20min$ and $w_2 = 5/20min$) as shown in Figure 5.3b. In particular, the transition probability matrix of the Markov chain is $P_w = \begin{bmatrix} 0.68 & 0.28 & 0 & 0.04 \\ 0 & 0.72 & 0.28 & 0 \\ 0 & 0 & 0.72 & 0.28 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

Under good weather conditions (when $q[k] = [0\ 0\ 0\ 1]'$), I have $a[q[k]] = a_1 = 0.98$ and $c[q[k] = c_1 = 0.09$. Under bad weather conditions (when $q[k] = [1\ 0\ 0\ 0]'$, $[0\ 1\ 0\ 0]'$, or $[0\ 0\ 1\ 0]'$), I have $a[q[k]] = a_2 = 0.04$ and $c[q[k]] = c_2 = 4.22$.

To facilitate the presentation, I denote the matrix $\begin{bmatrix} 1 - a^2[q[k]] & A_2 & B_2 \\ 0 & 1 - a[q[k]] & (1 - a[q[k]])u[k] - c[q[k]] \\ 0 & 0 & 1 \end{bmatrix}$ in Equation 87 as k_1 when $q[k] = [0\ 0\ 0\ 1]'$, and as k_2 when $q[k] = [1\ 0\ 0\ 0]', [0\ 1\ 0\ 0]'$, or $[0\ 0\ 1\ 0]'$. It is easy to obtain that $k_1 = \begin{bmatrix} 4.84 \times 10^{-4} & 9.59 \times 10^{-4} & 0.0028 \\ 0 & 0.022 & 0.0218 \\ 0 & 0 & 1 \end{bmatrix}$ and $k_2 = \begin{bmatrix} 0.91 & 0.87 & 4.68 \\ 0 & 0.95 & 0.46 \\ 0 & 0 & 1 \end{bmatrix}$ by substituting the values of a_1, a_2, c_1, and c_2. According to Equation 87, the recursion of $E[\sigma_{\mathbf{2}}[k]]$ can thus be expressed as:

$$E[\boldsymbol{\sigma_2}[k+1]] = \begin{bmatrix} 0.68k_2 & 0k_2 & 0k_2 & 0k_1 \\ 0.28k_2 & 0.72k_2 & 0k_2 & 0k_1 \\ 0k_2 & 0.28k_2 & 0.72k_2 & 0k_1 \\ 0.04k_2 & 0k_2 & 0.28k_2 & 1k_1 \end{bmatrix} E[\boldsymbol{\sigma_2}[k]] \tag{91}$$

Assuming that there is no backlog at the initial time, and weather starts from the state s_1, the initial condition can be expressed as $E[\boldsymbol{\sigma_2}[0]] = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}' \otimes \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}'$. The variance $\sigma^2[B[k]]$ at each time step can thus be obtained using Equation 91.

5.4. Optimal Management Design under Both Weather and Demand Uncertainties

In the previous section, I provide an effective approach to predict the statistics of weather-induced transient backlog. In practice, it is unrealistic to directly use stochastically-varying weather-induced capacity reduction as flow management restrictions. This is because strategic management is planned in advance using weather prediction, and hence its update always lags the change of weather. As described in Problem 2, optimal management design is concerned with devising deterministic flow restrictions under the uncertainties of weather and demand. In this section, I first discuss the performance metrics to be used as optimization cost functions, and then present a novel method to design optimal management under the uncertainties.

5.4.1. Metrics to Evaluate the Performance of Flow Management

The usefulness of optimal management solutions is highly contingent upon the correctness of performance metrics (used as the optimization cost functions). As such, before I present the optimal design, I first introduce and evaluate three candidate cost functions. The first function is defined as the total backlog over a duration of interest. I show that this cost function cannot capture the advantage of flow management, and always results in "no management" as the optimal solution. To account for this problem, I then introduce a second cost function, by separating management and non-management induced backlogs. Finally, I present a third performance metric to penalize high transient backlogs.

I note that my focus here is not to determine the exact form of cost functions through real cost analysis. Instead, I try to understand the type of cost functions that allow the optimal solution to reflect the considerations in reality.

5.4.1.1. Total Backlog

The most intuitive performance measure is proportional to the total backlog

$$O_1 = \sum_{k=0}^{k_p} C_1 E(B[k]), \tag{92}$$

where $k_p > 0$ is the end time of consideration, and C_1 is a scaling factor. I prove that this performance metric always results in "no management" as the optimal solution. Or in another words, the cost O_1 (or total backlog) caused by pure uncertain weather with zero management is always the minimum among all deterministic management plans.

To prove this statement, it is sufficient to show that for any possible combinations of weather and demand ensembles at any time k, the backlog $B[k]$ caused by a deterministic management $N_c[k]$ is equal to or greater than that caused by just weather $w(q[k])$. This is straightforward by considering the following two cases. If $N_c[k] < w(q[k])$, I have $N[k] = min(N_c[k], w(q[k])) = N_c[k] < w(q[k])$, and clearly the backlog with management is larger than that with no management according to Equation 77. On the other hand, if $N_c[k] \geq w(q[k])$, I have $N[k] = min(N_c[k], w(q[k])) = w(q[k])$, and thus the backlog with management is equal to that with only weather. The proof is thus complete.

The above proof suggests that any deterministic management plan always results in higher total backlog. As such, the total backlog-based metric O_1 cannot capture the advantage of flow management, and thus I need to modify it.

5.4.1.2. Separation of Management and Non-Management induced Costs

To account for the advantage of flow management, I note that the unit cost associated with management-induced backlogs is typically less than that associated with weather-induced backlogs. For instance, if a severe storm is predicted, aircraft can be delayed taking off through the Ground Delay Program (GDP). If no management is adopted in this circumstance, aircraft has to be instead held in air before being allowed to enter a weather zone, incurring more costs, such as extra fuel, higher challenge to controllers, and endangered safety-levels. Therefore, even though flow management plans (such as GDPs) may result in higher total backlog (as suggested in Section 5.4.1.1), they are still desirable in practice. To account for the low costs associated with management-induced backlogs, I separate management and weather induced backlogs in the cost function, and assign different unit cost to each. In particular, the cost function is defined as

$$O_2 = \sum_{k=0}^{k_p} \left(C_2 E(B_m[k]) + C_3 E(B_{nm}[k]) \right), \tag{93}$$

where $B_m[k]$ and $B_{nm}[k]$ represent the management and non-management (such as convective weather and normal capacity constraints) induced backlogs at time k respectively, and C_2 and C_3 are scaling factors representing the unit costs associated with each type.

Typically, the management induced cost C_2 is much smaller than the non-management induced cost C_3. As such, I can phrase that the purpose of flow management is to *transform weather-induced backlogs into management-induced backlogs* so as to reduce the total costs. I also note that if multiple managements are applied concurrently, the cost associated with each management may need to be considered.

5.4.1.3. Penalization for High Transient Backlogs

High transient backlogs may not be allowable in reality, and as such need to be panelized. For instance, the maximum number of aircraft being held in air (caused by management strategies such as MIT/MINIT) cannot pass a limit. The cost function O_2 cannot reflect this panelization as the relationship between backlog and cost is linear (see Figure 5.4 for a comparison). I am thus motivated to modify O_2 and change the backlog-cost relationship to be nonlinear. A general representation of the cost function is:

$$O_3 = \sum_{k=0}^{k_p} \left(f_1(E(B_m[k])) + f_2(E(B_{nm}[k])) \right), \tag{94}$$

where f_1 are f_2 are nonlinear functions that capture the negative impact of high transient (management and non-management induced) backlogs. Possible candidates of $f_i(x)$ include:

$$f_i(x) = C_i x + H_i \mathbf{1}(x - \Gamma_i), \tag{95}$$

where Γ_i is a threshold and H_i is a scaling factor representing the extra unit cost of high transient backlog. $\mathbf{1}(x)$ is the function defined as $\mathbf{1}(x) = x$ if $x \geq 0$ and $\mathbf{1}(x) = 0$ if $x < 0$; or simply,

$$f_i(x) = C_i x + H_i x_{peak} \delta(k - k_{peak}), \tag{96}$$

where x_{peak} is the peak value of x, k_{peak} is the peak time, and $\delta(x)$ is an impulse function, i.e., $\delta(x) = 1$ if $x = 0$ and $\delta(x) = 0$ if x$\neq$ 0. The first candidate (in Equation 95) penalizes transient backlogs higher than a threshold, and the second candidate (in Equation 96) penalizes the peak transient backlog. Choosing the specific format needs to be determined based on real cost data.

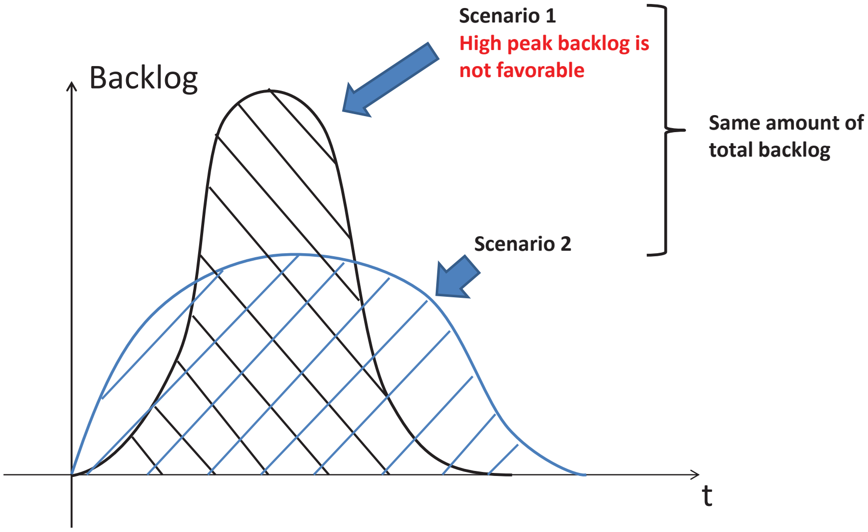


Figure 5.4. Comparison between two backlog scenarios. The cost functions O_1 and O_2 cannot distinguish between these two, as both have the same total backlog.

5.4.2. A PCM Approach to Find Optimal Flow Management

In this section, I introduce a PCM approach to find the optimal flow management plan. Compared to the Monte Carlo approach (see Figure 5.1), PCM consumes much less computational time, as it requires to evaluate only a few number of management solutions to find the optima. In this section, I first review the PCM approach, and then discuss how PCM is adapted for optimal management design.

5.4.2.1. Review of the PCM Approach

PCM, originally developed for power applications [58, 108], is an effective method to evaluate uncertainty. Consider a system with uncertain input and complicated input-output relationship. The Monte-Carlo approach can be used to find the output statistics through enumerating a large number of inputs. Unlike Monte-Carlo, PCM is concerned with *smartly* choosing only a few samples from the input distribution. Evaluating system outputs at these selected inputs produces a low-order input-output mapping, from which the output statistics can be obtained. It was shown in [58, 132] that for a higher-order input-output polynomial mapping of order $2m-1$, PCM can find the same output expectation with only m input samples. We have also shown that the PCM mapping has several other nice properties: 1) it is the minimum-mean-square-estimation (MMSE) mapping among polynomials of certain degree, and 2) it predicts the correct cross-correlation up to certain degree [132]. In paper [152], I extended the single-parameter PCM to the multivariate case.

Because of the good performance of PCM, I can imagine that it may be used as an optimization method. In particular, the low-order input-output PCM mapping can be directly used to identify the input that optimizes the output. In this section, I discuss how PCM can be used to find the best flow management that minimizes the cost functions discussed in Section 5.4.1.2. Before I present the PCM approach, let me first discuss a property of the optimal management solution for a fixed weather ensemble. In particular, I show that the optimal management duration and weather duration are the same. This property will be used later for optimal management planning.

5.4.2.2. Discussion on the Optimal Management Duration for Deterministic Weather

I assume that a deterministic weather ensemble lasts from time 0 to t_0 with restriction rate w_2, and the management restriction rate is also w_2 (see the problem formulation in Section 5.2.2). To find the optimal length of management (denoted as t_c), I compare the following four categories: 1) no management ($t_c = 0$), 2) appropriate management ($t_c = t_0$), 3) mild management ($t_c < t_0$), and 4) excessive management ($t_c > t_0$).

Figure 5.5 displays the transient mean backlog for each category of management duration. As weather and management have the same restriction rates, management induced backlogs (show in dashed lines) occur whenever management actions are applied. The non-management induced backlogs (shown in real lines) occur after the complication of management. In this case, the flow is either subject to weather restriction w_2 or normal capacity constraint w_1.

The cost associated with each category can be expressed in terms of the area under the curves (representing the total backlog in a time-span). Without loss of generality, let me consider the cost function O_2. According to Equation 93, the total cost for "no management" (Figure 5.5a) is $C_3(B_1+B_2)$ as all backlogs are caused by non-management flow restrictions. In the "appropriate control" case (Figure 5.5b), the total backlog is of the same amount, but the total cost is $C_2B_1+C_3B_2$ since the backlog B_1 is induced by flow management, and B_2 is induced by normal capacity constraint. In the "mild control" case (Figure 5.5c), despite the same total backlog as the previous two cases, the total cost is $C_2B_1' + C_3\hat{B}_1 + C_3B_2$, where $B_1' + \hat{B}_1=B_1$. Because the management induced unit cost C_2 is typically smaller than non-management induced unit cost C_3, it is easy to conclude that the cost relationship among the three cases is that "appropriate management" < "mild management" < "no management". The "excessive control" case (Figure 5.5d) results in a cost $C_2\tilde{B}_1+C_3\tilde{B}_2$ which is higher than that of the "appropriate management", as $\tilde{B}_1 > B_1$ and $\tilde{B}_2 > B_2$. Therefore, I conclude from the above comparison that the "appropriate management" represents the best management strategy for any fixed weather condition. As the analysis for the cost function O_3 is similar to that for O_2, I omit the proof here.

5.4.2.3. Using PCM to Find the Optimal Flow Management Plan

Using PCM to find the optimal management duration involves a process of 1) constructing a low-order PCM mapping between the management duration (as the input) and the total cost (as the output), and 2) finding the optimal solution from the mapping. Constructing the low-order PCM mapping requires selecting a few critical management dura-

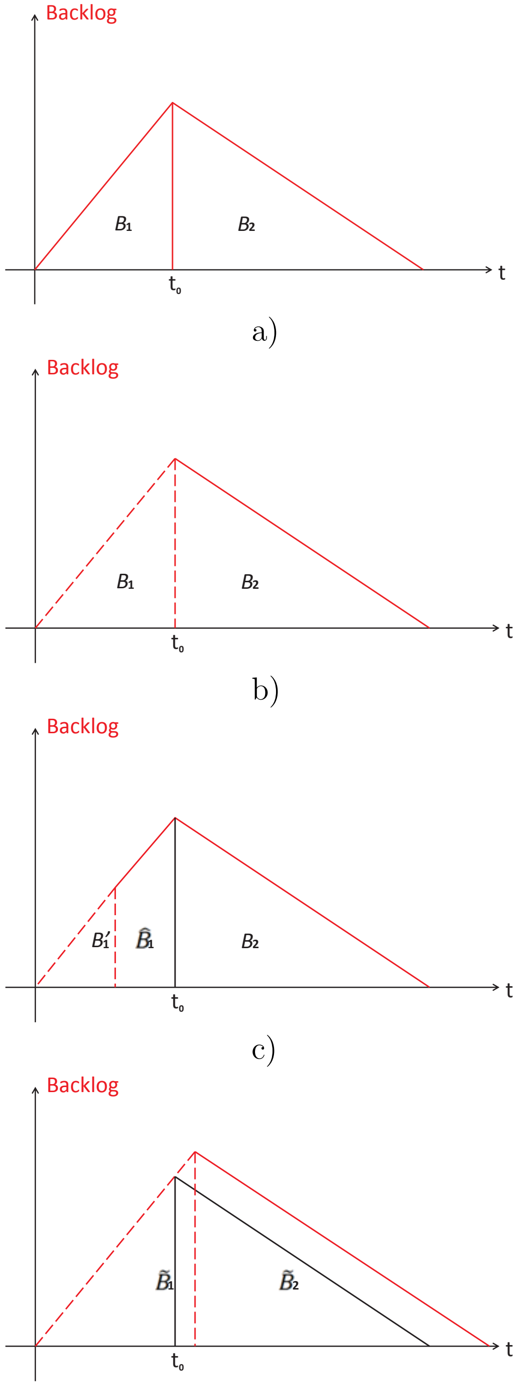


FIGURE 5.5. The relationship between management durations and backlogs: a) no management; b) appropriate management; c) mild management; d) excessive management. The dashed lines represent management induced backlogs, and the real lines represent the non-management induced backlogs.

tions, and obtaining the cost associated with each. Let me first discuss how these critical management durations are selected.

According to PCM, selecting the critical management durations requires the knowledge of their distribution. As the the distribution of weather duration is known, and for each fixed weather duration the optimal management duration is the same (as discussed in Section 5.4.2.2), the management duration can be assumed to follow the same distribution of the weather. This observation allows me to find the optimal management duration using the procedures summarized below.

Step 1: Selection of m simulation points

Based upon the distribution of weather duration, choose m durations as the PCM evaluation points. According to PCM, these m points are the roots of the mth-order orthonormal polynomial $h_i(x)$ that satisfy [34, 58]

$$\begin{aligned} (h_i(x), h_j(x)) &= \int h_i(x) h_j(x) f_X(x)\, dx \\ &= \begin{cases} 1, & if\ i = j \\ 0, & if\ i \neq j \end{cases} \\ h_0(x) &= 1, \end{aligned} \tag{97}$$

where $f_X(x)$ is the pdf of weather duration.

Step 2: Cost Evaluation of the m Selected Management Durations

This step is concerned with finding the cost associated with each selected management duration, under both weather and demand uncertainties. Two approaches can be adopted for an effective evaluation: 1) the Master Markov approach (as illustrated in [151]) and 2) the jump-linear approach as described in Section 5.3.

Step 3: Derivation of the Optimal Management Solution

The last step is concerned with deriving a $(m-1)$th order polynomial mapping that passes through the selected evaluation points [58, 132]. The optimal solution can be quickly solved through taking a derivative of the PCM polynomial mapping.

5.4.3. An Optimal Management Design Example

Let us present a simple example to illustrate the PCM-based optimal design approach. As suggested by this example, the PCM approach obtains near-optimal flow management solutions, with many fewer simulations than that would be needed by the Monte Carlo approach.

In this example, the convective weather model is shown in Figure 5.3, with reduced flow restriction rate $w_2 = 1/20mins$ and normal restriction rate $w_1 = 6/20mins$. The demand is modeled as a Poisson flow with inflow rate $u[k] = 3.5/20mins$. If no strategic management initiatives is enacted, aircraft will be held in air through tactical management strategies, wasting fuel and causing safety concerns. Instead, the strategic management initiative GDP can significantly reduce fuel cost and controller workload, as it delays aircraft on the ground. In this example, I aim to find the optimal duration of GDP to minimize a cost function O_2 (which weights the costs of management and non-management induced backlogs), and O_3 that further penalizes high transient backlogs. I first demonstrate how the optimal solutions are found using the PCM approach, and then discuss the impact of weighting factors in cost functions on the selection of optimal management solutions.

From the weather pdf, four weather duration instances (or equivalently management duration instances) are selected to serve as the PCM evaluation points: $12.48h$, $7.19h$, $3.49h$, $1.039h$. As the resolution of management duration is 20 minutes in this particular example, I approximate these four points as $12.33h$, $7.33h$, $3.33h$, $1h$ (see Figure 5.6). Let me show the design of optimal GDP solutions for the following two cost functions:

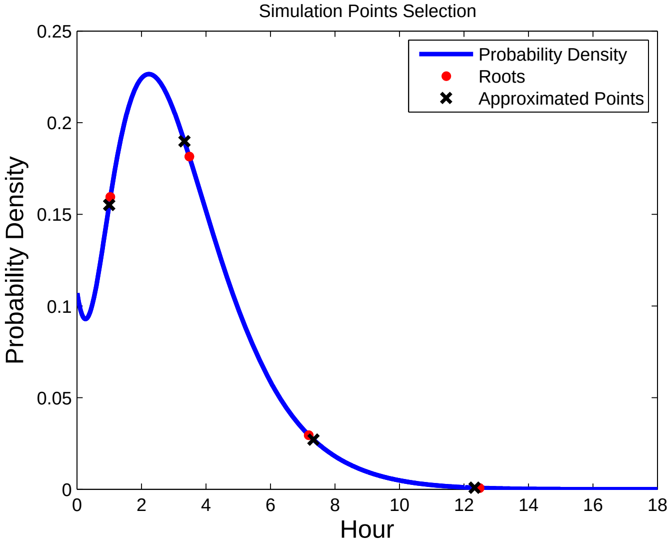


Figure 5.6. Weather pdf and the selection of four PCM points

Cost Function O_2: As GDP delays aircraft on the ground, the unit costs of GDP-induced backlog and weather-induced backlog should be significantly different. In this example we use C_2=1 to capture the unit cost of GDP-induced backlog whereas C_3=5 to capture the non-management-induced unit cost.

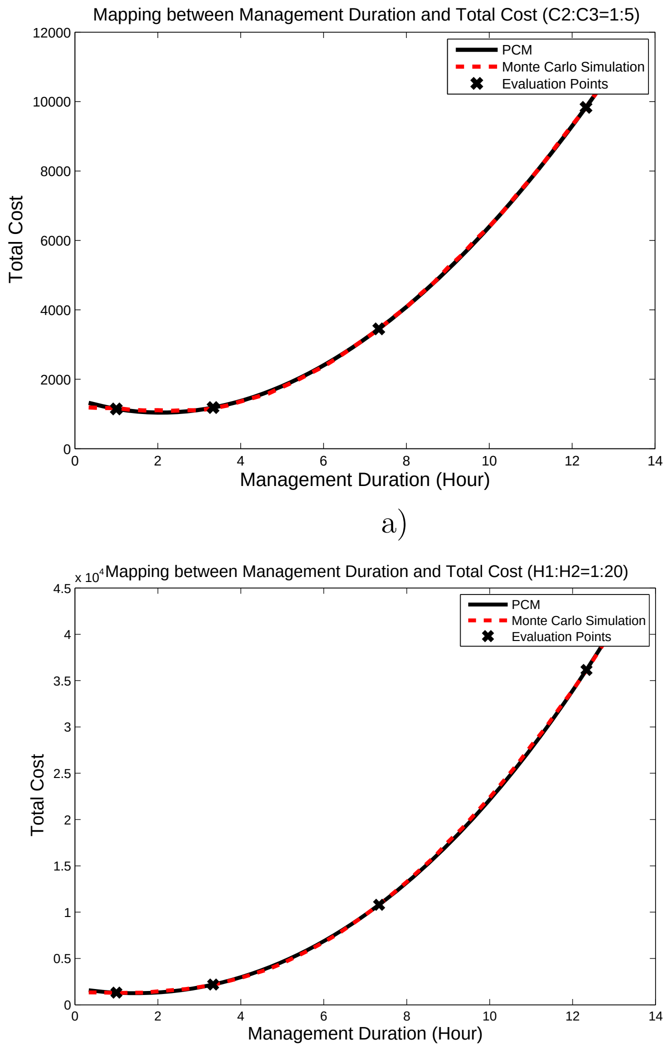
I then evaluate the total cost O_2 for each of the four selected management durations. Using the Markov approach, I find the management duration—total cost pairs to be $(12.33, 9835)$, $(7.33, 3453)$, $(3.33, 1184)$, $(1, 1147)$. These four points define a third-order PCM mapping $O_2(k) = -0.7265k^3 + 95.52k^2 - 386.78k + 1438.49$ between management duration and total cost. As shown in Figure 5.7a, The PCM mapping matches very well with the mapping obtained using the Monte Carlo approach, but with much less computational cost. The optimal management duration $2.07h$ is obtained through taking the derivative of the polynomial mapping $O_2(k)$. This solution is close to $2.33h$, the optimal solution of the Monte Carlo approach. The mismatch is caused by the resolution of management durations used in Monte Carlo. As the management is updated every 20 minutes, the optimal solution found by Monte Carlo is only available at a 20-minute resolution.

Cost Function O_3: As there is always a limit on the number of maximum allowed holding aircraft in air for safety concerns, it is important to penalize high weather-induced backlogs in the cost function. In this example, I assume that the number being held in air is not suggested to pass a threshold $\Gamma = 10$. To capture this realistic consideration, I penalize the number of non-management-induced backlog over the threshold Γ by an extra unit cost $H_2 = 20$. I also penalize the GDP-induced backlog over the same threshold by a small extra unit cost $H_1 = 1$.

Using the same Markov approach, I obtain the management duration—total cost pairs as $(12.33, 36152)$, $(7.33, 10799)$, $(3.33, 2183)$, and $(1, 1306)$. As shown in Figure 5.7b, these evaluation points lead to a PCM mapping $O_3(k) = 3.82k^3 + 236k^2 - 705.84k + 1771.6$. The optimal management $1.44h$ is close to $1.33h$ obtained from the Monte Carlo approach. Again, the small offset is mainly caused by the resolution of management durations.

Comparing the above two optimal management solutions suggests that the selection of cost functions plays a significant role in the optimal flow management planning. As such, I end this chapter with a comparison study to understand the impact of weighting factors in the cost functions on the optimal management solutions. Specifically, I keep the unit cost C_2 associated with management-induced backlog as 1 without the loss of generality, and vary the unit cost of non-management-induced backlogs C_3 and the penalties of high transient backlogs H_1 and H_2.

Figure 5.8a shows the effect of C_3 on the optimal management duration. In this comparison, the management and non-management penalization scaling factors H_1 and H_2 are set to be 1 and 1.5, respectively. As seen from the plot, the optimal management duration increases monotonically with the increase of C_3. This is reasonable as the higher



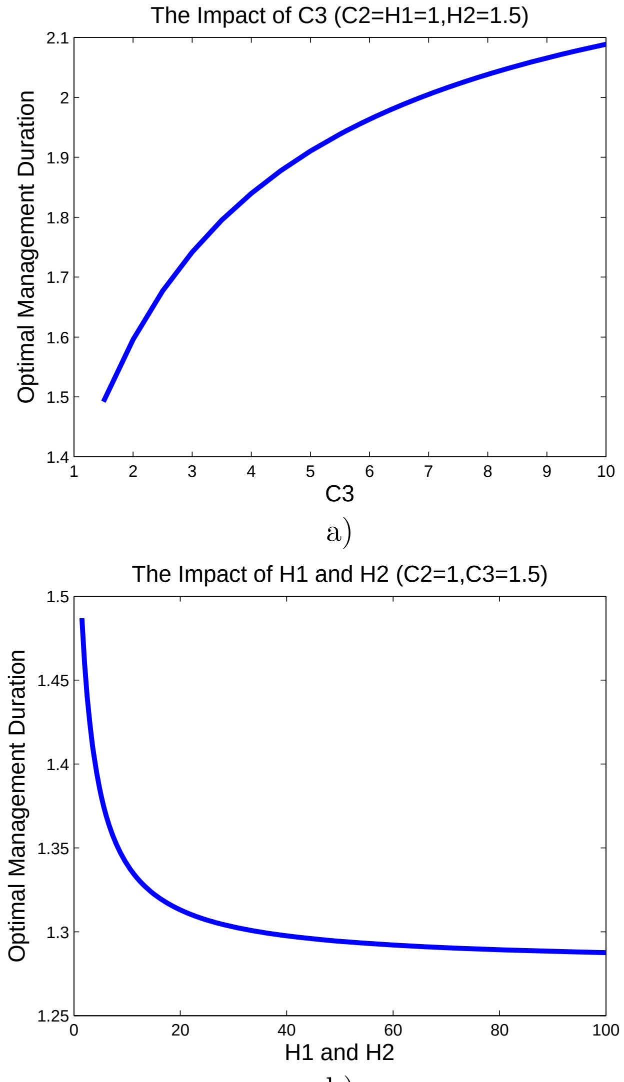
a)

b)

FIGURE 5.7. a) Performance comparison between PCM and Monte Carlo Simulation for the cost function O_2 ($C_2 = 1$ and $C_3 = 5$); b) Performance comparison between PCM and Monte Carlo Simulation for the cost function O_3 ($H_1 = 1$ and $H_2 = 20$).

cost of non-management-induced backlogs will naturally shift the optimal solution toward the transformation of non-management-induced backlogs to management-induced backlogs through prolonged management durations.

Next, I fix non-management-induced unit cost as $C_3 = 1.5$, and vary the penalties for management and non-management-induced backlogs H_1 and H_2 together. As shown in figure 5.8b, increasing H_1 and H_2 results in reduced length of management durations. The result is also understandable, as extended-length of flow management durations easily leads to accumulated large backlogs that pass the maximum allowable thresholds.



a)

b)

FIGURE 5.8. a)The relationship between C_3 and optimal management duration; b)The relationship between $H_1 = H_2$ and optimal management duration.

5.5. Concluding Remarks and Future Work

In this chapter, I introduce an integrated demand-weather-management modeling framework to evaluate weather impact and design optimal flow management under uncertainties. In particular, I model demand as a stochastic process, convective weather as a Markov chain, and flow management as queuing restrictions. The contributions of this chapter include:

(1) I develop a closed-form jump-linear approach to quickly evaluate the mean and variance of weather impact. The variance of weather impact is a critical measure that indicates the variability of weather severity.

(2) I provide a discussion on the candidate cost functions for optimal management. I note that the total backlog cannot capture the benefits of strategic flow management. As such, I induce a cost measure that separates management and non-management induced backlogs. I also introduce several nonlinear cost measures that penalize high transient backlogs. The effect of parameters in the cost functions on optimal management solutions is studied in the example at the end of this chapter.

(3) I introduce a PCM-based approach to quickly find the optimal management. The method is concerned with establishing a low-order mapping between flow management and total cost, and finding the optimal solution from the mapping. PCM is used to smartly identify a limited set of evaluation points to construct the low-order mapping. Markov or Jump-linear approach can be used to effectively evaluate the cost associated with each selected management plan.

CHAPTER 6

MULTIVARIATE PROBABILISTIC COLLOCATION METHOD FOR EFFECTIVE UNCERTAINTY EVALUATION WITH APPLICATION TO AIR TRAFFIC FLOW MANAGEMENT

6.1. Introduction

There is a growing need to effectively and strategically manage large-scale infrastructure systems, such as air traffic systems, power grids, and environmental systems. These systems are typically subject to a wide range of uncertainties, which significantly complicate the evaluation and management of system performance. To give some examples, flow contingency management solutions are being developed for air traffic systems, which seek to automatically generate management plans over a 2-15 hour lookahead time that are robust to weather uncertainties [123, 124]. In analogy, strategic resource scheduling and real-time surveillance/control algorithms are sought for the power grid, that are flexible to uncertainties in renewable generation and load, and robust to complex and uncertain fault events [58, 133]. As a step toward real-time management, techniques are needed for accurate yet computationally efficient evaluation/prediction of system performance over a range of parametric uncertainties. To address this need in broad infrastructure system applications, this chapter develops a systematic method to effectively evaluate output statistics for systems with multiple uncertain input parameters.

System uncertainty evaluation problems are typically addressed from two angles: analytical solutions and simulations. However because of the large scale and complicated nature of large-scale infrastructure systems, analytical solutions for system dynamics are typically unavailable; therefore, simulations using complicated computerized models are the primary approaches for uncertainty evaluation. Monte Carlo methods have been widely used to evaluate system performance under uncertainty by practitioners in many application domains [144]. Broadly, Monte Carlo methods have three main steps: 1) generation of a large number of samples covering the range of parameter uncertainties, 2) simulation of system performance for each parameter sample, and 3) summary of simulation outputs to obtain the system-performance statistics. Often, Monte Carlo methods may not be suitable for large-scale infrastructure applications, because of their inherent computational cost. Specifically, Monte Carlo methods typically require evaluation of a large number of simulations of the mapping of interest; since infrastructure-network simulations are rather computationally intensive, such exhaustive simulations are often impossible (especially when real-time decision-making is needed).

Driven by this limitation, our group (as well as others) have sought for alternative simulation methods to effectively evaluate parametric uncertainties. Our philosophy is that a limited number of simulations, if appropriately chosen, can provide adequate approximations of the mapping between uncertain input parameters and system performance over the range of likely parameter values. In particular, an adequate low-order mapping allows us to obtain statistical characterizations of system performance, and to evaluate system performance at any particular parameter value of interest. How to *smartly choose the values of input parameters as simulation points* and to construct the low-order mapping that allows best characterization of system-performance statistics is the key.

The probabilistic collocation method (PCM) is a method to evaluate the uncertainty of computationally expensive models at a low computational cost. It suggests a smart way of selecting simulation points to construct a low-order polynomial mapping between uncertain parameters and output or performance variable, that performs well over the likely range of parameter values [121, 141]. The idea behind PCM is as follows: although system parameters are uncertain, we typically have some statistical knowledge about the parametric uncertainties; smartly utilizing this information allows us to find representative simulation points, from which a reduced-order mapping of high fidelity can be constructed. In particular, it was proved in [58] that the low-degree polynomial mapping generated by PCM is able to predict the mean output (performance) correctly, even if the actual mapping is a much higher degree polynomial. In [133], we further explored the statistical performance of PCM, and the practical use of it when data or empirical low-order statistics is available instead of the probabilistic density function (pdf).

Several studies have detailed applications of PCM in transportation- and power- network management [58, 78, 79, 108]. While these efforts are promising, a limitation is that only one uncertain input parameter is considered. In practice, large-scale infrastructure systems typically involve multiple (sometimes a large number of) spatially/temporally distributed uncertain parameters. In these cases, more than one uncertain parameter (which may be correlated) may exert significant impact on system performance or other output variables. Thus, we are motivated to develop smart simulation techniques analogous to PCM, and to understand their effectiveness and cost, when multiple correlated, uncertain parameters are present. Specially in this chapter, we extend the formal analysis of the single-variable PCM to the multivariate case, where the uncertain parameters may or may not be independent.

The essence of our multivariate PCM approach lies in the smart selection of simulation points. Parsimonious selection of sampling points is needed for a range of applications, and has been widely studied (including in the specific context of uncertainty evaluation and mapping identification for complex systems). Here we provide a brief and incomplete review of these methods, focusing primarily on differentiating our approach from related ones. For the purpose of polynomial interpolation, Chebyshev nodes are widely used to overcome the Runge's phenomenon observed in using equally-spaced sampling [9]. For the purpose of uncertainty evaluation, besides Monte Carlo methods (also called random samping), several other techniques have been proposed. For instance, the stratified sampling and Latin hypercube Design (LHD) and their variants were developed to improve the coverage and projective properties [61, 66, 81]. Multiple steps are involved in these methods: first the space is subdivided to ensure full coverage, and then random sampling is used for each portion. These above methods are primarily designed to ensure appropriate coverage of the sampling space and to select sampling densities in the space, rather than to guarantee accurate estimation of output statistics. Gaussian Quadrature techniques also have been widely studied for uncertainty evaluation, both under the heading of PCM and using other terminology [65, 142, 147]. When multiple *independent* input parameters are involved, [65, 141, 142] suggest different procedures to select simulation points, among which [65] involves the most number of points. Gaussian Quadrature techniques have also been used to solve ODE/PDEs with uncertain parameters (see e.g. [10, 42, 78, 96]). These works mesh finite element decomposition in space and collocation methods on random variables, to approximate continuous high-dimensional solutions. All these works are relevant to our study, as they also consider the utilization of Gaussian Quadrature techniques when multiple uncertain variables are present. However, the purpose of our investigation in this chapter is different from these studies in two aspects. First, these studies consider the approximation of system mapping using expansions of orthogonal polynomials (e.g., Hermite for Gaussian distribution, Laguerre for Gamma distribution, Jocobi for Beta distribution, and Legendra for Uniform distribution) with small error, instead of *general* polynomial system mapping and its low-order approximation that predicts *precisely* the same statistics as we do. Second, their studies are not focused on understanding how *correlations* in parametric uncertainties may affect expected performance while our study does consider correlated uncertainty.

Our major contribution here is a formal investigation of the properties of PCM for systems with multiple uncertain input parameters (see also [154] for the brief conference version). Specifically, we 1) identify precise conditions on mapping functions and distributions to permit zero-error mean prediction, in both the independent and correlated cases; 2) in turn develop algorithms that obtain the best statistical performance; and 3) provide additional performance analyses, such as cross-statistics prediction, relation to minimum mean square estimation (MMSE), and computational feasibility analysis for large dimensional data. The multivariate PCM method that we develop overcomes the inadequacies of analytical uncertainty evaluation methods that do not work for complex infrastructure systems and Monte Carlo simulation methods that are computationally costly. In addition, our theoretical development provides a comprehensive understanding of the precise performance of PCM in multivariate settings and informs its practical use in performance evaluation and real-time decision making in large-scale infrastructure-type applications.

The remainder of this chapter is organized as follows. In Section 6.2, we describe and evaluate multivariate PCM when input parameters are independent. In Section 6.3, we discuss further properties of the independent multivariate PCM, and computational issues. In Section 6.4, we provide results of the multivariate PCM when parameters are correlated. In Section 6.5, we discuss the practical use of PCM when data or empirical low-order moments on low-order parameters, rather than explicit distributions of these parameters, are available. In Section 6.6, we present two examples to demonstrate the use of multivariate PCM to evaluate air traffic system performance under weather uncertainties. Section 6.7 concludes the chapter.

6.2. Independent Multivariate PCM

In this section, we consider a costly-to-simulate system with multiple input parameters subject to independent uncertainties. To effectively evaluate the dependence of system output on these input parameters and to obtain the statistics of system output under uncertainty, we study how to smartly select a limited number of simulations to construct a low-order polynomial mapping between the inputs and the output with good statistical performance.

6.2.1. A Simple Two-Variable Case

To simplify the development of multivariate PCM, we first focus on the two-variable case. Specifically, let us consider a system of interest, for which we wish to identify the mapping between two variable input parameters x and y and an output of interest that is functionally dependent on the two parameters, say $g(x, y)$. Unfortunately, it is costly to compute the output for an input pair (e.g., because it requires a time-consuming simulation or a costly experiment), and hence we can evaluate $g(x, y)$ for only a limited number of pairs (x, y). With this limited ability to probe the system, we seek to accurately identify the mapping over a useful range of parameter values. More precisely, we model the two input parameters, for the evaluation task of interest, as independent random variables with known probability distribution $f_{X,Y}(x, y) = f_X(x)f_Y(y)$.

We propose a technique for selecting input pairs for simulation and in turn approximating the mapping of interest, which we call the two-variable PCM. The conceptual basis for the two-variable PCM (which is analogous to the one-variable case) is the following. A sparse number of points (pairs) are selected for evaluation, and these evaluations are used to obtain a low-degree polynomial mapping between the input parameters and the output; the points are specially selected, in such a way that the obtained low-order polynomial mapping can predict the mean output correctly even if the mapping in reality is of a much higher degree (see Figure 6.1 for an illustration).

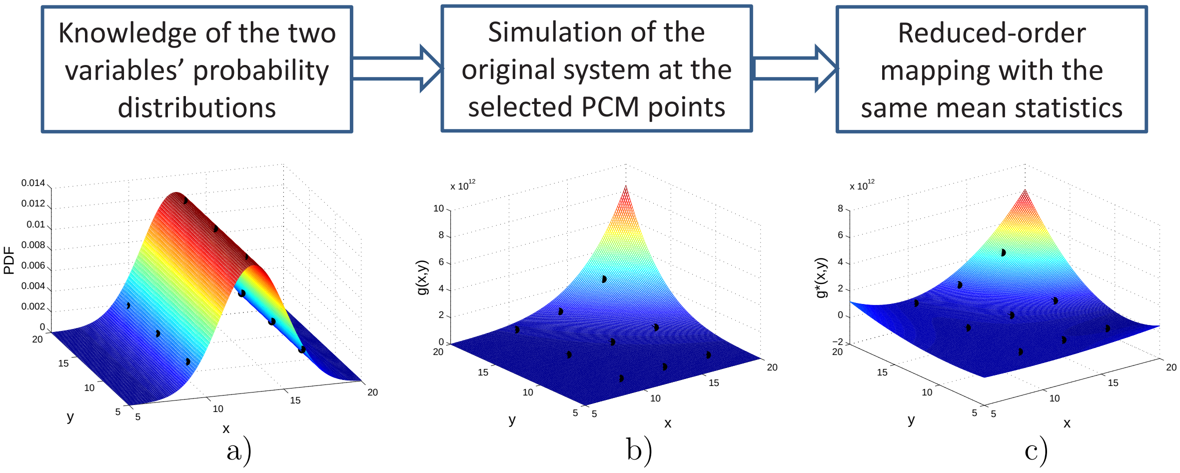


FIGURE 6.1. a) The joint probability density function $f_{X,Y}(x, y)$, b) The original mapping $g(x, y)$, c) The reduced-order multivariate PCM mapping $g^*(x, y)$. The black dots represent PCM points.

6.2.1.1. Main Results

In this subsection, we first illustrate the algorithm of the two-variable PCM in finding the PCM simulation points, the coefficients of the low-order PCM mapping, and then the predicted output mean. We then in Theorem 6.1 formalize that the low-degree approximation achieves the same mean as a higher-degree mapping of a general form, if simulation points are selected according to the algorithm. A simple example then follows to illustrate the use of two-variable PCM. The rest of the chapter extends this theorem in multiple directions toward 1) more than two input variables, 2) dependency among the input variables, 3) different general forms of original mappings, 4) further statistical performance analysis, and 5) input variables whose distributions are not explicitly known but specified by sample data.

Algorithm:

The following algorithm shown in Table 6.1, based on the distributions of random variables X and Y, $f_X(x)$ and $f_Y(y)$, constructs a low-order mapping of the form

$$g^*(x,y) = \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} B_{i,j} x^i y^j, \tag{98}$$

with the statistical property suggested by Theorem 6.1, where the coefficients $B_{i,j} \in R$.

In Theorem 6.1, we prove that a two-variable mapping $g(x,y)$ with the degrees of x up to $2n-1$ and y up to $2m-1$ can be approximated by a low-order mapping $g^*(x,y)$ with the degrees of x up to $n-1$ and y up to $m-1$, with the same expected values. We note that identifying the low-order PCM mapping $g^*(x,y)$ requires evaluating $g(x,y)$ at the simulation points (x_i, y_j), where x_i and y_j are the roots of the orthonormal polynomials $h_n(x)$ and $h'_m(y)$, respectively. Please refer to Appendix.A for the proof.

THEOREM 6.1. *Consider a two-variable mapping $g(x,y)$ of the form*

$$g(x,y) = \sum_{i=0}^{2n-1} \sum_{j=0}^{2m-1} A_{i,j} x^i y^j, \tag{99}$$

where the coefficients $A_{i,j} \in R$, and n and m are integers greater than 1. With the assumption that the two variables x and y follow independent distributions $f_X(x)$ and $f_Y(y)$ respectively, the mapping $g(x,y)$ can be approximated by a low-order mapping $g^*(x,y)$ of the form shown in Equation 98, such that $E[g(x,y)] = E[g^*(x,y)]$.

6.2.1.2. Discussion and Example

Now let us discuss the efficiency of the independent two-variable PCM. As the original mapping $g(x,y)$ is a polynomial with x up to the degree of $2n-1$ and y up to $2m-1$, a total of $(2n)(2m)$ simulations are required to uniquely identify the mapping. However, identifying

$g^*(x, y)$ requires only nm simulations, as there are n roots for $h_n(x)$ and m roots for $h'_m(y)$. As such, the two-variable PCM can reduce the number of simulations by $3nm$.

Finally we use a simple example to illustrate the procedure and performance of the independent two-variable PCM as shown in Figure 6.1. We consider a two-variable mapping $g(x, y) = x^5y^5 - 2x^5y^4 + 3x^5y^3 - 4x^5y^2 + 5x^5y - 6x^5 - 2x^4y^5 + 4x^4y^4 - 6x^4y^3 + 8x^4y^2 - 10x^4y + 12x^4 + 3x^3y^5 - 6x^3y^4 + 9x^3y^3 - 12x^3y^2 + 15x^3y - 18x^3 - 4x^2y^5 + 8x^2y^4 - 12x^2y^3 + 16x^2y^2 - 20x^2y + 24x^2 + 5xy^5 - 10xy^4 + 15xy^3 - 20xy^2 + 25xy - 30x - 6y^5 + 12y^4 - 18y^3 + 24y^2 - 30y + 36$ as shown in Figure 6.1b. Assuming that x is normally distributed with mean 13 and variance 4.6225 and y is uniformly distributed between 5 and 20, we show that the mapping can be well approximated by a polynomial with the degrees of x and y each up to 2.

To do that, we choose 9 simulation points based on the pdf of x and y, according to Step 1 in the algorithm. The locations of the simulation points are $(9.2761, 6.6905)$, $(9.2761, 12.5000)$, $(9.2761, 18.3095)$, $(13, 6.6905)$, $(13, 12.5000)$, $(13, 18.3095)$, $(16.7239, 6.6905)$, $(16.7239, 12.5000)$, and $(16.7239, 18.3095)$ as shown in Figure 6.1a. We then evaluate $g(x, y)$ at these 9 locations according to Step 2, and obtain the coefficients of the lower-oder PCM mapping according to the Step 3. The resulting PCM mapping up to the order of x^2y^2 is shown in Figure 6.1c.

From Figure 6.1a, we see that the 9 simulation points are located at the range of likely parameter values, in response to the probability distributions of x and y. Comparing Figures 6.1b and 6.1c shows that the PCM mapping $g^*(x, y)$ approximates $g(x, y)$ very well over the likely domain of parameter values, despite the significantly reduced mapping orders and number of simulations required for the construction.

6.2.2. General Theorem and Procedure on the Independent Multivariate PCM Mapping

Theorem 6.1 can be easily generalized to the multivariate PCM with more than two variables, as shown in Theorem 6.2. Please refer to Appendix.A for the proof.

THEOREM 6.2. *Consider a multivariate mapping* $g(x_1, x_2, ..., x_m)$ *of the form*

$$g(x_1, x_2, ..., x_m) = \sum_{k_1=0}^{2n_1-1} \sum_{k_2=0}^{2n_2-1} ... \sum_{k_m=0}^{2n_m-1} A_{k_1,...,k_m} \prod_{i=1}^{m} x_i^{k_i}, \tag{100}$$

where the coefficients $A_{k_1,...,k_m} \in R$, *and* n_1, n_2, $...n_m$ *are integers larger than* 1. *Assume that the variables* x_1, x_2, ..., x_m *follow independent distributions* $f_{X_1}(x_1)$, $f_{X_2}(x_2)$, ... , *and* $f_{X_m}(x_m)$ *respectively. Then the mapping can be approximated by a low-order mapping* $g^*(x_1, x_2, ..., x_m)$ *of the form*

$$g^*(x_1, x_2, ..., x_m) = \sum_{k_1=0}^{n_1-1} \sum_{k_2=0}^{n_2-1} ... \sum_{k_m=0}^{n_m-1} B_{k_1,...,k_m} \prod_{i=1}^{m} x_i^{k_i}, \tag{101}$$

such that $E[g(x_1, x_2, ..., x_m)] = E[g^*(x_1, x_2, ..., x_m)]$, *where the coefficients* $B_{k_1,...,k_m} \in R$.

Theorem 6.2 shows that the independent multivariate PCM method can significantly reduce the number of simulations needed to construct a low-order mapping of the same mean. Similar to what we have argued in the two-variable case, constructing $g(x_1, x_2, ..., x_m)$ as shown in Equation 100 requires at least $2^m \Pi_{i=1}^m n_i$ simulations. This is because $2n_i$ simulations are required to uniquely identify each variable up to the degree of $2n_i - 1$. However, the reduced-order mapping $g^*(x_1, x_2, ..., x_m)$ as shown in Equation 101 requires only $\Pi_{i=1}^m n_i$ simulations evaluated at the roots of each $h_{n_i}^i(x_i)$. The independent multivariate PCM can thus save $(2^m - 1)\Pi_{i=1}^m n_i$ simulations without introducing any error to the mean prediction.

To complete the presentation and facilitate the use of the independent multivariate PCM, we briefly summarize the procedure to find the PCM mapping in its general form (as shown in Equation 101).

Step 1: Simulation point selection. As a generalization to the two-variable case, $\Pi_{i=1}^{m} n_i$ simulation points are required to identify a m-variable PCM mapping shown in Equation 101. Because of the independence of variables, we only need to identify n_i points along the dimension of each x_i, for $i \in \{1, ..., m\}$. Combinations of the n_i points along each dimension constitute the set of $n_1 ... n_m$ simulation points. The n_i points associated with random variable x_i are the roots of the orthonormal polynomial $h^i_{n_i}(x_i)$, which can be found through the recursion illustrated in Step 1 in the algorithm in Section 6.2.1.

Step 2: Evaluation of system outputs at selected simulation points. For each simulation point identified in Step 1, we find the associated output through simulation. In total, $\Pi_{i=1}^{m} n_i$ simulations are needed. We stress that in many applications, this is the most time-consuming step. As the multivariate PCM can significantly reduce the number of simulations required to produce a mapping with correct mean prediction, this method is of significance to find high-dimensional mappings that are common to large-scale infrastructure applications.

Step 3: Identification of Mapping Coefficients. Using the simulations in Step 2, we can explicitly express the low order mapping $g^*(x_1, ..., x_m)$ by identifying the coefficients in its orthonormal polynomial form. This result is a simple generalization of Step 3 of the algorithm in Section 6.2.1 for the two-variable case. In particular, the coefficients $a_{k_1, ..., k_m}$ can be calculated using the matrix operation in Equation 102. The vector a is arranged in a descending order of the subscripts starting from the last bit to the first bit, and the vector g is arranged in an increasing order of the root indices (i), again from the rightmost entry to the leftmost entry, where $1 \leq i \leq n_m$.

$$L \begin{bmatrix} a_{n_1-1,\ldots,n_m-1} \\ \vdots \\ a_{n_1-1,\ldots,0} \\ a_{n_1-1,\ldots n_{m-1}-2,n_m-1} \\ \vdots \\ a_{n_1-1,\ldots n_{m-1}-2,0} \\ \vdots \\ a_{0,\ldots,0} \end{bmatrix} = \begin{bmatrix} g(x_{1(1)}, x_{2(1)}, \ldots, x_{m(1)}) \\ \vdots \\ g(x_{1(1)}, x_{2(1)}, \ldots, x_{m(n_m)}) \\ g(x_{1(1)}, x_{2(1)}, \ldots, x_{m-1(2)}, x_{m(1)}) \\ \vdots \\ g(x_{1(1)}, x_{2(1)}, \ldots, x_{m-1(2)}, x_{m(n_m)}) \\ \vdots \\ g(x_{1(n_1)}, x_{2(n_2)}, \ldots, x_{m(n_m)}) \end{bmatrix} \tag{102}$$

Here $L \in R^{n_1 n_2 \ldots n_m \times n_1 n_2 \ldots n_m}$ takes the following form

$$\begin{bmatrix} h^1_{n_1-1}(x_{1(1)})\ldots h^m_{n_m-1}(x_{m(1)}) & \cdots & h^1_0(x_{1(1)})\ldots h^m_0(x_{m(1)}) \\ h^1_{n_1-1}(x_{1(1)})\ldots h^m_{n_m-1}(x_{m(2)}) & \cdots & h^1_0(x_{1(1)})\ldots h^m_0(x_{m(2)}) \\ \vdots & \ddots & \vdots \\ h^1_{n_1-1}(x_{1(n_1)})\ldots h^m_{n_m-1}(x_{m(n_m)}) & \cdots & h^1_0(x_{1(n_1)})\ldots h^m_0(x_{m(n_m)}) \end{bmatrix}$$

In each row, the subscripts of h are arranged in a descending order from $n_1-1, n_2-1, \ldots, n_m-1$ to $0, 0, \ldots 0$. In each column, the indices of roots are in an increasing order from $1, 1, \ldots, 1$ to $n_1, n_2, \ldots n_m$. The coefficients $B_{k_1,\ldots,k_m}$ shown in Equation 101 can then be easily derived from $a_{k_1,\ldots,k_m}$.

6.3. Properties of the Independent Multivariate PCM

As proved in Section 6.2, the independent multivariate PCM can find a low-order polynomial mapping that correctly predicts the mean of the original higher-order mapping. In this section, we further characterize the performance of the independent multivariate PCM mapping. In particular, we show in Section 6.3.1 that the low-order mapping also predicts the correct cross-statistics. We then compare its performance with that of the MMSE. In

Section 6.3.2, we briefly discuss the numerical capability of the independent multivariate PCM in processing high-dimensional data.

6.3.1. Performance of the Independent Multivariate PCM

In Theorem 6.3, we show that the low-order independent multivariate PCM can precisely predict the cross-statistics up to certain degree. Please refer to Appendix.A for the proof.

THEOREM 6.3. *Consider the use of a multi-variable PCM mapping* $g^*(x_1, x_2, ..., x_m)$ *of the form* $\sum_{k_1=0}^{n_1-1} \sum_{k_2=0}^{n_2-1} ... \sum_{k_m=0}^{n_m-1} B_{k_1,k_2,...,k_m} \prod_{i=1}^{m} {x_i}^{k_i}$ *to approximate an original polynomial mapping* $g(x_1, x_2, ..., x_m)$ *of the form* $\sum_{k_1=0}^{n_1+p_1} \sum_{k_2=0}^{n_2+p_2} ... \sum_{k_m=0}^{n_m+p_m} A_{k_1,k_2,...,k_m} \prod_{i=1}^{m} {x_i}^{k_i}$*, for all* $p_i \in \{0, ..., n_i - 1\}$*. Assuming that all variables are independent, the low-order PCM can correctly predict the cross-statistics, i.e.,*

$$E\left[\left(\prod_{i=1}^{m} {x_i}^{l_i}\right) g^*(x_1, x_2, ..., x_m)\right] = E\left[\left(\prod_{i=1}^{m} {x_i}^{l_i}\right) g(x_1, x_2, ..., x_m)\right] \tag{103}$$

for all $l_i \in \{0, ..., n_i - 1 - p_i\}$*.*

In Theorem 6.4, we discuss the relationship between the independent multivariate PCM mapping and the MMSE estimator. In general, a low-order PCM mapping may not be the MMSE estimator among all polynomials of the same degree. However, the performance of PCM is attractive as reflected by the following two additional results: 1) the mean square error (MSE) performance of a PCM mapping may not be improved by adding any polynomial up to certain degree, and 2) a PCM mapping up to certain degree is the MMSE estimator among all polynomials of the same degree. Please refer to the Appendix.A for the proof.

THEOREM 6.4. *Consider a multivariate mapping* $\sum_{k_1=0}^{n_1+p_1} \sum_{k_2=0}^{n_2+p_2} ... \sum_{k_m=0}^{n_m+p_m} A_{k_1,k_2,...,k_m} \prod_{i=1}^{m} {x_i}^{k_i}$

for some $p_i \in \{0, ..., n_i - 1\}$. *If a PCM mapping of the form*

$$g^*(x_1, x_2, ..., x_m) = \sum_{k_1=0}^{n_1-1} \sum_{k_2=0}^{n_2-1} ... \sum_{k_m=0}^{n_m-1} a_{k_1,...,k_m} \prod_{i=1}^{m} h^i_{k_i}(x_i)$$

is used to fit the original mapping, the mean square error of the PCM fit cannot be improved by adding any polynomial with the degree of x_i *up to* $n_i - 1 - p_i$. *Moreover, the lower order PCM mapping*

$$g^*_r(x_1, x_2, ..., x_m) = \sum_{k_1=0}^{n_1-1-p_1} ... \sum_{k_m=0}^{n_m-1-p_m} a'_{k_1,...,k_m} \prod_{i=1}^{m} h^i_{k_i}(x_i)$$

is the MMSE mapping, among all polynomials with the degree of each x_i *up to* $n_i - 1 - p_i$.

6.3.2. Discussion on Numerical Issues

We note that the performance of PCM may be affected by numerical issues. As seen from Equation 102, finding the coefficients $a_{n_1-1,...,n_m-1}$, and then $B_{n_1-1,...,n_m-1}$ in Equation 101 (according to the procedure discussed in Section 6.2.2) involves solving a large equation array. Numerical issues (such as the loss of precision and even failure of getting a solution) may appear especially when the dimension of the array is large. Instead of providing a formal proof, we discuss the feasibility of this procedure through comparing it with an alternative procedure. We first describe the alternative procedure and then use the example presented in Section 6.2.1 to compare the performance of these two.

We note that once PCM points are selected and outputs at those points are simulated, we can find the mapping coefficients $B_{n_1-1,...,n_m-1}$ directly, instead of using the orthonormal bases. In particular,

$$L' \begin{bmatrix} B_{n_1-1,\ldots,n_m-1} \\ \vdots \\ B_{n_1-1,\ldots,0} \\ B_{n_1-1,\ldots n_{m-1}-2,n_m-1} \\ \vdots \\ B_{n_1-1,\ldots n_{m-1}-2,0} \\ \vdots \\ B_{0,\ldots,0} \end{bmatrix} = \begin{bmatrix} g(x_{1(1)}, x_{2(1)}, \ldots, x_{m(1)}) \\ \vdots \\ g(x_{1(1)}, x_{2(1)}, \ldots, x_{m(n_m)}) \\ g(x_{1(1)}, x_{2(1)}, \ldots, x_{m-1(2)}, x_{m(1)}) \\ \vdots \\ g(x_{1(1)}, x_{2(1)}, \ldots, x_{m-1(2)}, x_{m(n_m)}) \\ \vdots \\ g(x_{1(n_1)}, x_{2(n_2)}, \ldots, x_{m(n_m)}) \end{bmatrix} \tag{104}$$

where $L' \in R^{n_1 n_2 \ldots n_m \times n_1 n_2 \ldots n_m}$ takes the following form

$$\begin{bmatrix} x_1^{n_1-1}(x_{1(1)})\ldots x_m^{n_m-1}(x_{m(1)}) & \cdots & x_1^0(x_{1(1)})\ldots x_m^0(x_{m(1)}) \\ x_1^{n_1-1}(x_{1(1)})\ldots x_m^{n_m-1}(x_{m(2)}) & \cdots & x_1^0(x_{1(1)})\ldots x_m^0(x_{m(2)}) \\ \vdots & \ddots & \vdots \\ x_1^{n_1-1}(x_{1(n_1)})\ldots x_m^{n_m-1}(x_{m(n_m)}) & \cdots & x_1^0(x_{1(n_1)})\ldots x_m^0(x_{m(n_m)}) \end{bmatrix}$$

In the matrix L', $x_i^k(x_{i(j)})$ represents the k-th power of the variable x_i evaluated at the j-th simulation point of the variable x_i. The L' matrix is arranged as follows. In each row, the superscripts of x are arranged in a descending order from $n_1 - 1, n_2 - 1, \ldots, n_m - 1$ to $0, 0, \ldots 0$. In each column, the indices of roots are in an increasing order from $1, 1, \ldots, 1$ to $n_1, n_2, \ldots n_m$.

The feasibility of finding the coefficients B in Equation 104 or a in Equation 102 is reflected by *condition number*, which describes the accuracy of numerical solutions to linear equation arrays [26]. Condition number is defined as the ratio between the largest and the smallest eigenvalues. We often consider the logarithm of condition number to be the loss of accuracy, although this estimation is not always precise. At the extreme, infinite condition number means that the matrix is singular and the equation array is unsolvable. We note that ill-conditioning typically leads to a drastic failure in solving equations rather than small numerical issues; hence devising means for increasing solution precision rather than possible impacts of numerical errors may be the key concern. The possibility for ill-conditioning is dependent on a number of factors, including the number of variables, number of PCM points, and probability distribution; the possibility for numerical error arising from ill-conditioning may also depend on the nature of the system mapping, and of course the precision of the computing device. We here calculate and compare the condition numbers of the two matrices L' in Equation 104 and L in Equation 102 respectively, for the example in Section 6.2.1. The condition number associated with L is 2.53, and associated with L' is 9.54×10^6 which is significantly larger. This simple comparison clearly shows the advantage of using orthogonal bases (Equation 102) over the x and y bases (Equation 104) to calculate PCM coefficients, especially for high-dimensional mappings.

6.4. Correlated Multivariate PCM

The development in the previous sections is based on the assumption that uncertain variables are independent. In realistic modern infrastructure applications, system variables are often correlated. In general correlated settings, the PCM mapping obtained using the procedure described in Section 6.2.2 does not predict the correct mean. In this section, we explore conditions on the forms of original mapping and joint distribution to maintain the correct mean statistics.

6.4.1. Main Results

We first discuss the case when the original mapping does not contain any cross-terms among variables, and then the general case when cross-terms are present. In the first case, a low-order multivariate PCM mapping can be obtained in a way very similar to the independent case. The proof is omitted as it is a simple variation of the independent multivariate PCM [58].

Theorem 6.5. *Consider a multivariate mapping* $g_n(x_1, x_2, ..., x_m)$ *of the form*

$$g_n(x_1, x_2, ..., x_m) = \sum_{i=1}^{m} \sum_{k_i=0}^{2n_i-1} A'_{i,k_i} x_i^{k_i}, \tag{105}$$

where the coefficients $A'_{i,k_i} \in R$ *and* n_i *is an integer larger than* 1*, where* $i \in 1, 2, ..., m$*. Assume that the variables* x_1*,* x_2*, ...,* x_m *follow a joint distribution* $f_{X_1X_2...X_m}(x_1, x_2, ...x_m)$*. The original mapping can be approximated by a lower-order mapping* $g_n^*(x_1, x_2, ..., x_m)$ *of the form*

$$g_n^*(x_1, x_2, ..., x_m) = \sum_{i=1}^{m} \sum_{k_i=0}^{n_i-1} B'_{i,k_i} x_i^{k_i}, \tag{106}$$

such that $E[g_n(x_1, x_2, ..., x_m)] = E[g_n^*(x_1, x_2, ..., x_m)]$*, where the coefficients* $B'_{i,k_i} \in R$ *for* $i \in 1, 2, ..., m$*.*

We now discuss the case when cross-terms are present. We show that under certain assumptions on conditional moments, a low-order PCM mapping can correctly predict the mean of an original mapping up to certain degree. Please refer to Appendix.A for the proof.

Theorem 6.6. *Consider a multivariate mapping* $g_c(x_1, x_2, ..., x_m)$ *of the form*

$$g_c(x_1, x_2, ..., x_m) \;=\; \sum_{k_1=0}^{2n-1} \sum_{k_2=0}^{2n-1-k_1} \sum_{k_3=0}^{2n-1-(k_1+k_2)} \cdots \sum_{k_m=0}^{2n-1-(k_1+k_2+...+k_{m-1})} A'_{k_1,...,k_m} \prod_{i=1}^{m} x_i^{k_i},$$

where the coefficients $A'_{k_1,...,k_m} \in R$ *and* n *is an integer larger than* 1*. Assume that the variables* x_1*,* x_2*, ...,* x_m *follow a joint distribution* $f_{X_1,X_2,...,X_m}(x_1, x_2, ..., x_m)$*, and the* r*-th order conditional moment of* x_i *on* x_{i+1}*, ...,* x_m *is a polynomial of* x_{i+1}*, ...,* x_m *with the total degree not greater than* r*. The original mapping can be approximated by a lower-order mapping* $g_c^*(x_1, x_2, ..., x_m)$ *with the same mean. The degree of each variable in* $g_c^*(x_1, x_2, ..., x_m)$ *is not greater than* $n - 1$*.*

Theorem 6.6 shows that in the two-variable case, if the r-th conditional moment of one variable given the other is a r-th order polynomial, the low-order PCM mapping with each variable up to the degree of $n-1$ predicts the correct mean when the total degree of the original mapping does not exceed $2n-1$. The procedure to pick PCM points here is different from that in the independent case, in the sense that the point locations are selected based on the marginal probability of one variable, and the conditional probability of the other. In particular, PCM points are selected through: 1) determining the y coordinates (denoted as $y_1, ..., y_n$) of the PCM points according to the marginal probability of y, 2) finding the PCM points' x coordinates according to the conditional probability $f_{X|Y}(x|y_i)$ for each selected y_i.

6.4.2. Discussion on the Reduction of Computation Load

Here, let us also discuss the reduction of computational load. In the correlated case, the effectiveness of the multivariate PCM is generally not as good as that in the independent case. The original high-order mapping requires C_{2n-1+m}^{m} simulations, where C_p^q denotes the combination of p elements taken q of them at a time without repetition. However the number of computations that can be reduced is generally not easy to write in the closed-form, due to the complicated conditional relationship. We here only discuss the two-variable and three-variable cases.

In the two-variable case, clearly, identifying the low-order PCM mapping $g_c^*(x, y)$ requires n^2 simulations to correctly predict the mean output. However, as the total degree of the original mapping is $2n-1$, $2n^2+n$ simulations are needed to identify all the coefficients in the original mapping. As such, more than half of the simulation time can be saved. Finally, we also note that to identify the low-order mapping, we need to directly find $B'_{i,j}$, instead of $e'_{l,k}$ at the orthogonal coordinates, as $e'_{l,k}$ is a varying function of y. In the three-variable case, the total number of simulations to identify $g_c(x_1, x_2, x_3)$ equals to $\frac{4}{3}n^3+n^2+\frac{2}{3}n$, because

$g_c(x_1, x_2, x_3)$ has this number of coefficients. Similarly, we can check that the low-order PCM mapping requires $n^3 - C_n^3$ simulations. Hence, the number of simulations that can be saved is $\frac{1}{3}n^3 + n^2 + \frac{2}{3}n + C_n^3 = \frac{1}{2}n^3 + \frac{1}{2}n^2 + n$.

6.4.3. Discussion on the Assumption

We note that the assumption on conditional moments is the key to ensure correct mean prediction in the correlated case (see Lemma 6.10 and Theorem 6.6). Here we use the two-variable case to understand the generality of this condition, and then show examples to illustrate this condition.

In the two-variable case, the r-th conditional moment of x can be written as:

$$\int_{\alpha(y)}^{\beta(y)} x^r f_{X|Y} f(x|y)dx = \int_{\alpha(y)}^{\beta(y)} \frac{f_{X,Y}(x,y)}{\int_{\alpha(y)}^{\beta(y)} f_{X,Y}(x,y)dx} x^r dx = \frac{\int_{\alpha(y)}^{\beta(y)} f_{X,Y}(x,y)x^r dx}{\int_{\alpha(y)}^{\beta(y)} f_{X,Y}(x,y)dx}. \tag{107}$$

If $f_{X,Y}(x,y)$ does not depend on x (i.e. $f_{X,Y}(x,y)$ is a constant or a function of y), Equation 107 can be further simplified to:

$$\begin{aligned} \int_{\alpha(y)}^{\beta(y)} x^r f_{X|Y}(x|y)dx &= \frac{\frac{1}{r+1}[\beta^{r+1}(y) - \alpha^{r+1}(y)]}{\beta(y) - \alpha(y)} \\ &= \frac{1}{r+1}\left(\sum_{i=0}^{r} \beta^i(y)\alpha^{r-i}(y)\right) \end{aligned} \tag{108}$$

for $0 \leq i \leq r$. Clearly, if $\beta(y)$ and $\alpha(y)$ are linear functions of y, the conditional moment is a r-th order polynomial of y.

Some examples are shown in Figure 6.2. Specifically, in Figure 6.2a, random variables x and y are uniformly distributed in the parallelogram defined by $0 \leq y \leq 2$ and $y \leq x \leq y+2$. As $\alpha(y) = y$ and $\beta(y) = y+2$ are linear functions of y, and moreover $f_{X,Y}(x,y) = 0.25$, the conditional moment requirement is clearly satisfied, and hence the low-order multivariate PCM predicts the correct mean. Similarly, in Figure 6.2b, a uniform distribution in the trapezium (e.g., specified by the boundaries $\alpha(y) = y$ and $\beta(y) = 0.25y + 2$) also satisfies

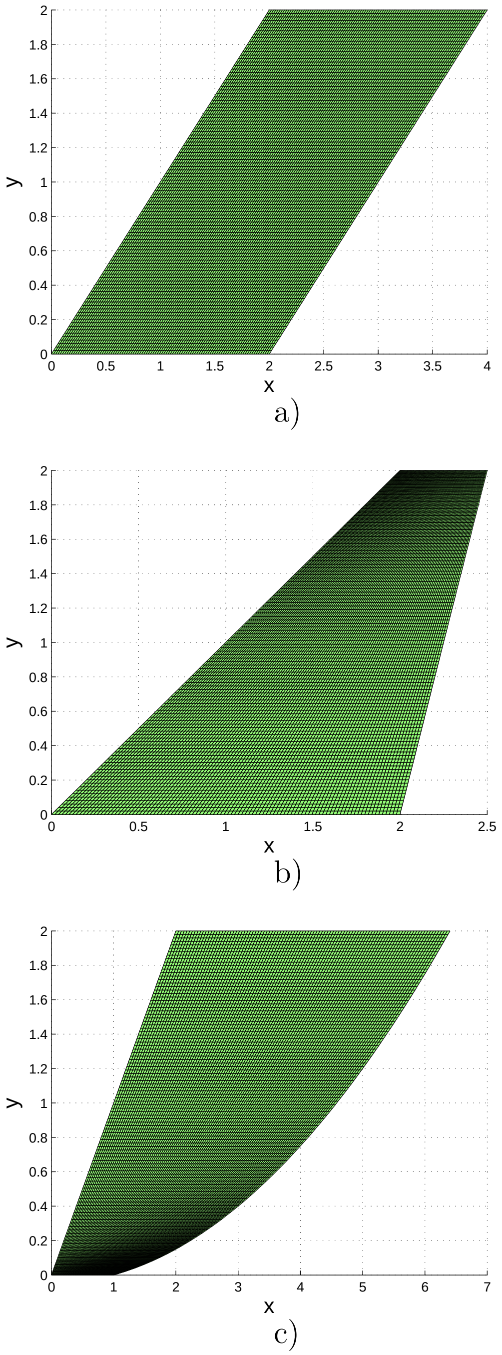


FIGURE 6.2. Two random variables x and y are uniformly distributed in a) a parallelogram, b) a trapezium, and c) an irregular area. The conditional moment condition is satisfied for cases a) and b), but not for c).

the conditional moment requirement. In Figure 6.2c, the variables x and y are subject to a uniform distribution in the region $0 \leq y \leq 2$, $y \leq x \leq \sqrt{20y+1}$. As $\beta(y) = \sqrt{20y+1}$ is nonlinear, the conditional moment condition fails. Furthermore, we note that the joint distribution does not have to be uniformly distributed. As shown in Figure 6.3a, $f_{X,Y}(x,y)$ is $\frac{72y}{25(8-3y)}$ in the area defined by $0 \leq y \leq \frac{5}{3}$ and $y \leq x \leq 0.25y+2$. As $f_{X,Y}(x,y)$ is only related to y, the conditional moment condition holds. Figure 6.3b is another example. In this case, $f_{X,Y}(x,y)$ is $\frac{1}{2}\frac{1}{\sigma\sqrt{2\pi}}e^{-(y-\mu)^2/2\sigma^2}$ in an infinite area defined by two parallel lines (e.g., $y \leq x \leq y+2, -\infty \leq y \leq \infty$). The last example is concerned with three variables x, y, z. When these three variables are uniformly distributed in a volume (e.g., specified in Figure 6.3c), the conditional moment condition also holds.

6.4.4. A Comparative Example

We use a simple example to illustrate the advantage of the correlated multivariate PCM approach. In particular, we consider a two-variable mapping with the original form of $g(x,y) = x^3 + y^3 + xy^2 + x^2y + xy + y + x^2 + y^2 + 1$ as shown in Figure 6.4a. The two random variables x and y are uniformly distributed in the parallelogram area bounded by $0 \leq y \leq 2$ and $y \leq x \leq y+2$, which is marked by the black dash lines in the Figure 6.4a with the joint pdf $f_{X,Y}(x,y) = 0.25$. The independent PCM method does not work, as it results in two simulation points outside the probability range (see the red spots in 6.4a). The correlated multivariate PCM method following Theorem 6.6 results in the four PCM points and the reduced-order mapping $1.33x - 29.67y + 23xy + 7$ (shown in Figure 6.4b) which predicts the correct mean of the original mapping 33.67. For illustration, we have also used the LHD method [147] to generate 4 simulation points. The produced low-order mapping $34xy - 47y - 11.96x + 27$ and the 4 simulation points are shown in Figure 6.4c, with a predicted mean of 35.7, different from the original mean.

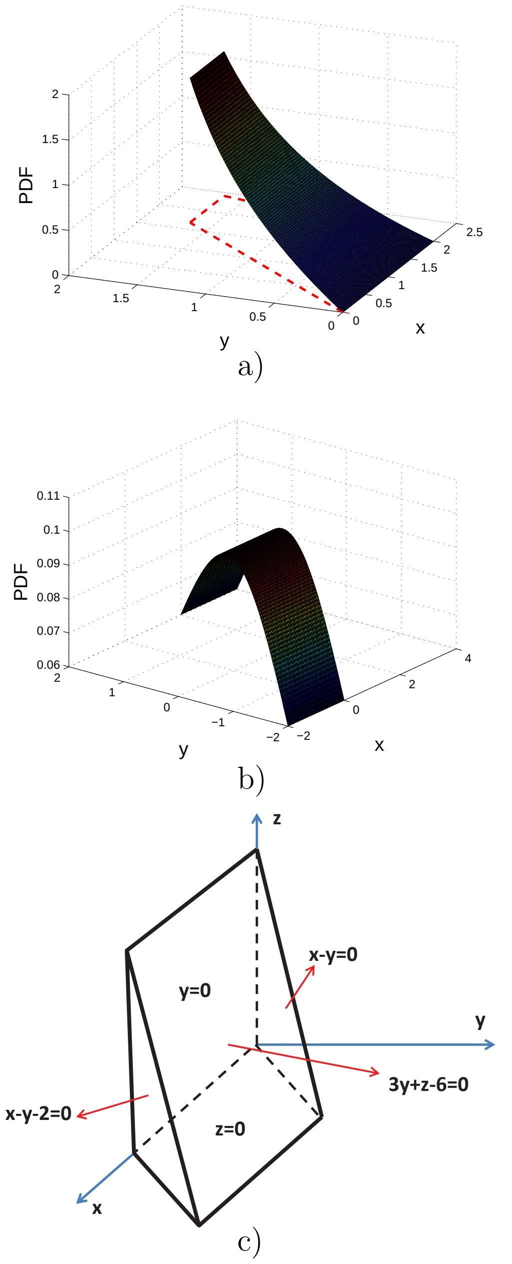


FIGURE 6.3. The first two are examples when the joint distribution is not uniform. Region a) is a trapezium, and region b) is an infinite area bounded by parallel lines. In both cases, the conditional moment requirements are satisfied. In the third example, x, y and z are uniformly distributed in a three-dimensional volume c). Every surface of this volume is described by a linear function. The conditional moment requirement is satisfied in this case.

6.5. Empirical Data Based Multivariate PCM

The full pdf of input parameters may not always be available in real applications. Instead, it is more common in practice that the distribution of input parameters is described in one of the following two forms: 1) sample data of input parameters based on historical records, and 2) low-order moment description of the parameters. For the completeness of our presentation, we extend the results in [133] and discuss the use of sample-data-based and empirical low-order moment-based multivariate PCM.

6.5.1. Large Data Set

The procedure to use the sample-data-based multivariate PCM is summarized as follows: 1) finding from data the moments of x_i (and conditional moments in the correlated case), 2) calculating the (conditional) orthogonal polynomials according to Equation 1 in [133], 3) finding the roots of the orthogonal polynomials as PCM points, and 4) following Steps 2 and 3 in Section 6.2.2 to obtain the low-order PCM mapping in the independent case, or the brief discussion after Theorem 6.6 in the correlated case.

Theorem 6.7. *Consider that each tuple in the data sets* $\{(x_{11}, x_{21}, ..., x_{m1}), (x_{12}, x_{22}, ..., x_{m2}), ..., (x_{1p}, x_{2p}, ..., x_{mp})\}$ *is drawn from a joint distribution* $f_{X_1,X_2,..,X_m}(x_1, x_2, ..., x_m)$*. With the assumption that* x_1*,* x_2*, ... and* x_m *have finite moments, the sample-data-based PCM mapping approaches to the pdf-based PCM mapping with probability* 1 *as* $p \to \infty$*.*

Proof. To prove that the sample-data-based PCM matches with the pdf-based PCM with probability 1, we only need to show that the PCM points obtained directly using sample data converge to those obtained using the pdf, as $p \to \infty$. As PCM points are the roots of orthogonal polynomials or conditional orthogonal polynomials as shown in the proofs of Theorems 6.1-6.6, it suffices to show that the coefficients of the (conditional) orthogonal polynomials obtained using sample data converge to those obtained using the pdf. This problem is thus reduced to the convergence of sample (conditional) moments to real (conditional) moments with increasing size of data. The result follows naturally. Please refer to Theorem 4 in [133] for the details of the proof in a single-variable case. ☐

6.5.2. Small Data Set and Low-Order Moments

In the case that neither the full pdf nor large sample data sets is available, we may find PCM mapping based on very limited information such as small data sets, or low moments such as the mean and variance. Please refer to [133] for a detailed treatment of PCM with sparse data or low moments. We summarize the key procedure for the complete presentation here.

The best approach to find PCM mappings in these cases is to fit these low moments/small sample data sets with typical probability distributions (i,e., those are mathematically determined by very small number of parameters, such as the Gaussian, uniform, Gamma, and Beta distributions). Then, we can follow the same procedure as that of the pdf-based PCM to find low-order PCM mappings. Please refer to standard estimation approaches such as the maximum-likelihood estimation [97].

The probability distribution to choose is generally not easy due to the sparsity of information. Here we briefly discuss the guidelines to select distribution functions. Firstly, the selection should be application-dependent (see [97]). For instance, noise is typically captured by the Gaussian distribution, whereas waiting time is captured by the Gamma distribution. Secondly, hypothesis testing techniques can help to determine the best distribution. It also helps to understand the impact of choosing different distribution functions. One example suggests that choosing the Gaussian distribution instead of the uniform distribution will result in more-spread-out PCM points that better account for parameter values further away from the mean, but at a cost of worse estimation close to the mean. Rigorous analysis on the impact of higher moments on the movement of PCM points have been studied in [133].

6.6. Applications to Air Traffic Flow Management

The development of multivariate PCM in this chapter was motivated by practical needs in the field of air traffic flow management (ATFM), which is typically concerned with managing traffic flows at a long lookahead time frame (e.g., 2-15 hours). As traffic flow management plans are decided hours in advance, a wide range of weather possibilities exist. Therefore, it is important to effectively evaluate the performance of the air traffic system under uncertainty, so as to design the best management strategies that are robust to such uncertainty in real time. Here we first provide a brief review of the literature on the evaluation and design of ATFM strategies at the long lookahead time, so as to make clear the contribution of our proposed approach.

Simple models of uncertainty as scenario trees and disturbances have been used to model uncertainty in ATFM; stochastic programming and model predictive control techniques have been applied on these models for decision-making [5–7, 28, 29]. More complicated stochastic models such as Bayesian networks and Markov chains have also been used (e.g., [98, 151]). As pointed out in these papers, these methods become less effective when the scale and complexity of the traffic system increase. Because air traffic systems have complex nonlinear and intertwined dynamics [135], predicting their performance may require system-wide simulations. The Monte Carlo method has thus been studied [30, 103, 139]. The Monte Carlo approach does not enable real-time analysis under weather uncertainty, as the evaluation of system performance under multiple (and even spatiotemporally-correlated) weather uncertainties requires time-consuming simulations for a large number of weather ensembles to obtain converging performance statistics. The multivariate PCM approach provides an alternative effective way for such simulations.

In this section, we demonstrate the use of multivariate PCM to evaluate the performance of air traffic systems under weather uncertainties using two examples. As seen from these examples, an important *preliminary step* for applying the multivariate PCM is to capture the range of weather uncertainties using a number of random variables. Next, the PCM points are selected along the dimension of each variable/parameter and then these selected points are simulated to provide a fast evaluation of traffic system performance. In the first simple example, we assume that the basic properties of an uncertain weather zone are captured by two independent variables (intensity and duration) with simple distributions, and then the total number of delayed aircraft (called backlog) for the flow entering this single zone is evaluated. The purpose of this example is not to precisely model uncertain weather, but to 1) provide some background on air traffic flow management, including the performance metric and the simulation model, and 2) show the practical use of PCM in the independent case. The second more realistic example is concerned with the evaluation of backlogs in two spatiotemporally-correlated weather zones.

6.6.1. A Simple Single Region Example

In this simple example, we consider a single region (e.g., a sector) in the airspace subject to severe weather prediction. We aim to evaluate the impact of uncertain weather on the congestion in the region. Let us first describe the simple queuing model for our study. We note that realistic air traffic system is much more complicated. However, the abstracted model is justified based on the uncertainty present in the data at this time frame and the computational requirements of real-time systems [130, 135, 151].

In particular, severe weather events reduce the capacity $N[k]$ (i.e., the maximum number of aircraft allowed to enter at time k) of the region. The aircraft that are not allowed to pass due to the reduced capacity can be modeled as waiting in the queue at the boundary of the region [130, 135, 151]. The length of the queue, named backlog $B[k]$, is a natural performance metric to capture the congestion of the region. The dynamics of the queuing simulation model is shown below (see more details at [151]):

$$B[k] = max(B[k-1] + x[k] - N[k], 0) \tag{109}$$

where $x[k]$ is the incoming flow. The equation suggests that at each time step, a maximum number of $N[k]$ aircraft can enter the region; the remaining flow will wait to enter at the next time step, together with the new incoming flow. We note that the capacity $N[k]$ varies with time k; in particular, the presence of convective weather at time k reduces the capacity $N[k]$.

As weather duration and intensity may not be precisely predicted at a long lookahead time, weather-induced capacity reduction and the accumulated transient backlog become uncertain. It is of practical value to predict the backlog statistics under uncertain weather duration and intensity.

Assuming that no information about the two uncertain random variables is available other than the ranges, we model both variables as uniformly distributed for simplicity. In particular, we assume that the severe weather starts at the current time, but with uncertain intensity and end time. When the severe weather is present, it reduces the number of aircraft to cross the region in a unit time (20 minutes) $N[k]$ from 6 to a, where a is captured by a random variable uniformly distributed between 1 and 2.67. Moreover, we assume that the duration of severe weather is a random variable uniformly distributed between 2 and 6 hours. When the severe weather completes, the capacity raises back to 6. For simplicity, we assume that the incoming flow (demand) $x[k]$ to the region is a deterministic sequence sampled from a Poisson process with the mean arrival rate of 3 per 20 minutes.

It is worthwhile to notice that the actual relationship between the output (total backlog for a span of 30 hours from the current time) and inputs (weather intensity and duration) is *nonlinear* [130, 151]. We show here that based on the distributions of weather intensity and duration, we can smartly choose only a limited number of simulations to construct a low-order multivariate PCM mapping that matches well with the original nonlinear mapping obtained using Monte Carlo simulations. Figure 6.5a and 6.5b show the original mapping and the 2×2-order PCM mapping using the algorithm in Section 6.2.1.1. The two mappings match very well with a mean square error of 2.4 within the probability range. The mean total backlog obtained using the two mappings are 163.91 and 163.42, respectively. It is also interesting to notice that the performance of mean predictor improves significantly when the order of the PCM mapping increases from 1×1 (with the mean backlog of 165.52) to 2×2, but not much when the order is further increased to 3×3 (with the mean backlog of 163.64). Finally, Figure 6.5c demonstrates the computational load that can be saved using the low-order 2×2 PCM. Monte Carlo simulation typically requires a large number of runs to converge (in this case 676 runs to estimate with an error less than 2%), while the 2×2-order PCM mapping only needs 9 simulations to achieve a similar performance.

6.6.2. A Spatiotemporally Correlated Two-Region Example

In the second example, we apply empirical data-based multivariate PCM to effectively evaluate the impact of spatiotemporally correlated uncertain weather events on transient traffic backlogs. In particular, we consider two scheduled streams of flows entering two neighboring regions A and B. Furthermore, a cold front (indicating of weather front associated with convective weather) passes Region A and then Region B, producing capacity reductions (see Figure 6.6a and 6.6b). Predicting transient backlog statistics caused by the uncertain progress of cold front is critical to design effective flow-management initiatives to reduce traffic delay.

Here we used six random parameters to capture the stochastic weather propagation: S_A, N_A, D_A, S_B, N_B, and D_B. In particular, the cold front hits Region A from time S_A with a reduced capacity N_A. We assume that the capacity reduction remains constant for a span of D_A before the weather leaves the region. After that, its capacity returns to its regular value N_{RA}. Similarly, Region B undergoes weather-induced capacity reduction from time S_B, with capacity reduced from the normal value N_{RB} to N_B, for a span of D_B.

Different from the previous example, the full pdf of the aforementioned weather parameters is not directly available. We utilized the weather simulator [112] to generate a large set of weather ensembles covering the range of weather uncertainty. In particular, the simulated airspace is decomposed into small grids, with white color representing normal capacity and black color denoting capacity reduction caused by convective weather (see Figure 6.6). The stochastic propagation of convective weather is governed by the influence model [8], with parameters estimated from hourly probabilistic weather forecasts[1]. The estimated weather simulator runs at a finer resolution of 15 minutes[112].

For each generated weather ensemble, we then found the six weather parameters (S_A, N_A, D_A, S_B, N_B, and D_B). Here, weather start times S_A and S_B are marked by the first time the region has at least two black grids. Durations D_A and D_B are similarly defined by the differences between the first time the region has at most two black grids afterwards and S_A and S_B. Average capacities N_A and N_B during the span of weather are calculated as the normal capacity scaled by the average fraction of white grids during this span. Because of spatiotemporal correlation of weather propagation, these parameters are subject to interdependency. In particular, as the weather front passes Region A and then

---

[1] A majority of weather forecasts are deterministic. The Very Short Range Ensemble Forecast System (VSREF) provides hourly probabilistic weather forecasts [148].

region B, due to the spatial weather propagation delay, we expect the start time S_B is closely related to the end time of region A (which is expressed as S_A+D_A), and therefore both S_A and D_A. Similarly, due to the correlation of propagation speeds for weather in these two regions, we expect that the duration D_B is dependent on D_A. Moreover, due to the correlation of weather intensities across time, we assume that the capacity N_B is dependent upon N_A. In this example, the normal capacity N_{RA} and N_{RB} of the two regions are both 10 per unit time (defined as $\Delta t = 15mins$). The simulator generates a large number of ensembles (one million in this case to guarantee the convergence of probability distributions).

We then applied the empirical data-based multivariate PCM approach to find the PCM points. To do that, we first found the sample moments of D_A, S_A, and N_A, as they serve as the parameters that D_B, S_B, N_B are conditioned upon. From these sample moments, the PCM points along each of the dimensions D_A, S_A, and N_A were obtained. We note that S_A plays a more important role in the uncertainty evaluation and needs to be sampled with more number of points for the following reasons: 1) due to the growth of uncertainty over time (which can be seen from the distributions plotted in Figure 6.7), the information closest to the initial simulation time is the most trustworthy, 2) this trustworthy early information is important to be estimated correctly as its error will affect the estimation performance of other parameters capturing later time characteristics. We therefore chose 5 points along S_A as $0.75hour$, $1hour$, $1.25hour$, $1.75hour$, and $2.5hour$. We also selected 2 PCM points along the dimension of D_A as $1hour$ and $1.75hour$, and 2 points for N_A as 4.0602, and 5.6358 respectively. Next, we identified PCM points along the dimensions of D_B, S_B and N_B. Because of the spatiotemporal correlation of weather parameters in Region A and Region B, each 3-tuple PCM point (D_A, S_A, and N_A) resulted in different conditional distributions of weather parameters for Region B (see examples shown in Figure 6.8), and thus different set of PCM points. Specifically, the PCM points along the dimensions of D_B, S_B were then selected based on the conditional sample moments of D_B and S_B for each combination of the PCM points for D_A and S_A. Similarly, the PCM points along the dimension of N_B were selected from the conditional sample moments of N_B on each of the PCM points for N_A. A total of $2^5 \times 5 = 160$ PCM points were selected.

We next evaluated the total backlog $\sum B[k]$ during a span of 24 hours at each of the 6-tuple PCM points. We assumed that two streams of deterministic flows sampled from a Poisson process (with mean 5 per 15 minutes) enter the two regions. Using Equation 109, the total backlogs at all PCM points were then used to obtain the low-order polynomial PCM mapping $f(D_A, S_A, N_A, D_B, S_B, N_B)$ (with 160 terms) between weather parameters and the total backlog, from which the mean backlog can be obtained. Furthermore, we noted that as the total backlog of the two regions is the summation of backlogs at each region, the mapping $f(D_A, S_A, N_A, D_B, S_B, N_B)$ can be expressed as $f_A(D_A, S_A, N_A) + f_B(D_B, S_B, N_B)$, where f_A and f_B are functions of weather parameters in individual regions. This expression significantly reduced the coefficients/terms in the polynomial mapping function from 160 to $5 \times 2^2 + 2^3 = 28$, by ignoring the cross terms that involve weather parameters of both regions. Least Square Mean Estimation was then used to identify the 28 coefficients in the reduced order mapping.

We compared the performance of the PCM mappings with that of the Monte Carlo simulation. As shown in Table 6.2, both PCM mappings predicted the total mean backlogs well with errors less than 2%. We also evaluated the efficiency of the PCM approaches. As seen in Figure 6.9, the Monte Carlo method requires 11716 simulations (marked as the black spot) for the mean prediction to fall within a threshold marked by the black dashed lines. However, the PCM method only needs 160 simulations (marked as the red spot) to reach the same threshold. We also note that selecting PCM points based on conditional distributions improves the performance of the mean prediction. To display the results intuitively, we demonstrate the mappings generated by the aforementioned empirical data-based multivariate PCM approaches, when four out of the six weather parameters are fixed: $S_A = 1.25hour$, $D_A = 1hour$, $D_B = 1.75hour$ and $N_A = 5.6358$. As shown in Figure 6.10, the 160-term PCM mapping and the 28-term PCM mapping generated by selecting 4 PCM points along the dimensions of S_B and N_B match well with the mapping generated by the Monte Carlo method.

6.7. Concluding Remarks and Future Work

Motivated by effective uncertainty evaluation needs in large-scale infrastructure systems, we develop in this chapter the multivariate PCM approach which allows using a few smartly selected simulation points to construct a low-order mapping between multiple uncertain input parameters and system output, which predicts the correct mean output of the original system of a higher-order. Besides describing the algorithm of the multivariate PCM, we develop mathematical conditions to permit correct mean output prediction in terms of the probability distributions of the input parameters and forms of original system mappings. Both independent and correlated cases are discussed. We also provide additional performance analysis of the multivariate PCM in terms of predicting other important statistics, and the practical use of the method when data or low-order moments are available instead of probability distribution. In terms of numerical issues, we note that in the independent PCM case, using orthonormal bases to calculate PCM mapping coefficients reduces the ill-conditioning of the calculation; this advantage does not exist for the correlated PCM. The final example at the end of this chapter demonstrates the use of this method to evaluate the impact of multiple dependent weather uncertain parameters on the statistics of air traffic system performance. Future works include applying this method for large-scale traffic examples, exploring capabilities of the reduced-order mapping such as optimization and sensitivity analysis, and selecting a subset of the PCM points when further knowledge of the original system mapping is available.

6.8. Appendix

Proof of Theorem 6.1:

Proof. In order to prove $E[g(x,y)] = E[g^*(x,y)]$, we start with computing $E[g(x,y)]$, then construct $g^*(x,y)$ along this process, and finally show that the means of both $g(x,y)$ and $g^*(x,y)$ can be reduced to the same value.

Because x and y are independent random variables, we find

$$\begin{aligned} E[g(x,y)] &= \iint \sum_{i=0}^{2n-1} \sum_{j=0}^{2m-1} A_{i,j} x^i y^j f_X(x) f_Y(y) dxdy \\ &= \int \sum_{j=0}^{2m-1} y^j f_Y(y) \int \sum_{i=0}^{2n-1} A_{i,j} x^i f_X(x) dxdy. \end{aligned} \tag{110}$$

The terms inside the internal integral $\sum_{i=0}^{2n-1} A_{i,j} x^i$ (for any j) can be expressed in terms of the series of orthonormal polynomials $h_n(x), ..., h_0(x)$ of degrees $n, ..., 0$ [58]:

$$\sum_{i=0}^{2n-1} A_{i,j} x^i = h_n(x) \left(\sum_{i=0}^{n-1} a_{i+n,j} h_i(x) \right) + \sum_{i=0}^{n-1} a_{i,j} h_i(x), \tag{111}$$

where the coefficients $a \in R$, the first subscript of a represents the degree of x, and the second subscript stands for the degree of y. The orthonormal polynomials $h_i(x)$ satisfy

$$\left\langle h_i(x), h_j(x) \right\rangle = \int h_i(x) h_j(x) f_X(x)\, dx = \begin{cases} 1, & if\ i = j \\ 0, & if\ i \neq j \end{cases} \tag{112}$$

$$h_0(x) = 1.$$

A particular note is that the above definition leads to

$$\int h_i(x)h_0(x)f_X(x)dx = \int h_i(x)f_X(x)dx = 0 \tag{113}$$

for all $i \geq 1$, which we will frequently use later. We denote the roots of $h_n(x)$ as $x_1, x_2, ..., x_n$. In a single-variable PCM, these roots are the PCM points selected for computationally intensive simulations [133].

Due to the orthonormal properties of the variable x (Equation 112), $\int \sum_{i=0}^{2n-1} A_{i,j}x^i f_X(x)dx$ shown in the form of Equation 111 can be reduced to $\int \sum_{i=0}^{n-1} a_{i,j}h_i(x)f_X(x)dx$. As such, Equation 110 becomes

$$E[g(x,y)] = \int \sum_{j=0}^{2m-1} y^j f_Y(y) \int \left(\sum_{i=0}^{n-1} a_{i,j}h_i(x) \right) f_X(x)dxdy. \tag{114}$$

We now follow the same procedure to reduce the order of y. By rearranging the terms in Equation 114 in a descending degree of $h_i(x)$ for all $i \in \{0, ..., n-1\}$, we obtain

$$E[g(x,y)] = \int \sum_{i=0}^{n-1} h_i(x)f_X(x) \int \left(\sum_{j=0}^{2m-1} a_{i,j}y^j \right) f_Y(y)dydx. \tag{115}$$

Introducing the jth-degree orthonormal polynomials $h'_j(y)$ with respect to the probability distribution $f_Y(y)$, and denoting the roots of $h'_m(y)$ as $y_1, y_2, ..., y_m$, we can express any polynomial of y up to the order of $2m-1$ in terms of $h'_0(y), ..., h'_m(y)$. In particular,

$$\sum_{j=0}^{2m-1} a_{i,j}y^j = h'_m(y) \left(\sum_{j=0}^{m-1} b_{i,j+m}h'_j(y) \right) + \sum_{j=0}^{m-1} b_{i,j}h'_j(y), \tag{116}$$

where $b \in R$ are the coefficients. Again, the first subscript of b represents the degree of x, and the second stands for the total degree of y. Applying orthonormal properties again, Equation 115 is further reduced to

$$E[g(x,y)] = \int\int \sum_{i=0}^{n-1}\sum_{j=0}^{m-1} b_{i,j} h_i(x) h'_j(y) f_X(x) f_Y(y) dx dy. \tag{117}$$

Note that the expression inside the above double integrals is $g^*(x,y)$. As $h_i(x)$ is an ith-order polynomial and $h'_j(y)$ is a jth-order polynomial, we can easily find the parameters $B_{i,j}$ from $b_{i,j}$ and express Equation 117 in terms of x and y as

$$g^*(x,y) = \sum_{i=0}^{n-1}\sum_{j=0}^{m-1} b_{i,j} h_i(x) h'_j(y) = \sum_{i=0}^{n-1}\sum_{j=0}^{m-1} B_{i,j} x^i y^j. \tag{118}$$

Observation of Equations 111 and 116 clearly suggests that $g(x,y)$ and $g^*(x,y)$ pass through the same set of points defined by (x_i, y_j) for all $i \in \{1, ..., n\}$ and $j \in \{1, ..., m\}$, as $h_n(x_i) = 0$ and $h'_m(y_i) = 0$. Finally, we also notice that both $E[g(x,y)]$ and $E[g^*(x,y)]$ can be reduced to $b_{0,0}$ through applying Equation 113 to variable x first and then variable y. □

Proof of Theorem 6.2:

Proof. As this theorem is a straightforward generalization of the two-variable case, we only sketch the outline of the proof.

Through recursively expressing x_1, x_2,... x_m in terms of orthonormal polynomials, and applying the orthonormal properties (Equation 112), we obtain the reduced-order mapping $g^*(x_1, x_2, ..., x_m)$ of the form

$$g^*(x_1, x_2 ..., x_m) = \sum_{k_1=0}^{n_1-1}\sum_{k_2=0}^{n_2-1} ... \sum_{k_m=0}^{n_m-1} a_{k_1,...,k_m} \prod_{i=1}^{m} h^i_{k_i}(x_i), \tag{119}$$

where the function $h^i_{k_i}(x_i)$ denotes the k_ith degree orthonormal polynomial with respect to the parameter x_i, and $a_{k_1,...,k_m}$ are the coefficients. Furthermore,

$$E\left[g(x_1, x_2, ..., x_m)\right] = E\left[g^*(x_1, x_2, ..., x_m)\right] = a_{0,...,0}. \tag{120}$$

Rearranging the terms in Equation 119 in terms of $x_1, x_2, ..., x_m$, we can easily find the coefficients $B_{k_1,...,k_m}$ such that $g^*(x_1, x_2, ..., x_m)$ is in the form of Equation 101. □

Proof of Theorem 6.3:

We prove the two-variable case for the clarity of presentation. In this case, the theorem is reduced to the following Lemma. The general case follows naturally.

LEMMA 6.8. *Consider the use of a two-variable PCM mapping $g^*(x, y)$ of the form $\sum_{i=0}^{n-1}\sum_{j=0}^{m-1} B_{i,j}x^iy^j$ to approximate an original polynomial mapping $g(x, y)$ of the form $\sum_{i=0}^{n+n_1}\sum_{j=0}^{m+m_1} A_{i,j}x^iy^j$, for some $n_1 \in \{0, ..., n-1\}$ and $m_1 \in \{0, ..., m-1\}$. Assuming that the two variables are independent, the low-order PCM can correctly predict the cross-statistics up to certain degree. In particular,*

$$E\left[x^ly^kg^*(x, y)\right] = E\left[x^ly^kg(x, y)\right] \tag{121}$$

for all $l \in \{0, ..., n-1-n_1\}$ and $k \in \{0, ..., m-1-m_1\}$.

PROOF. The cross-statistics can be expressed in the following due to the independence of variables:

$$\begin{aligned} E\Big[x^ly^kg(x, y)\Big] &= \iint x^ly^k \textstyle\sum_{i=0}^{n+n_1}\sum_{j=0}^{m+m_1} A_{i,j}x^iy^jf_X(x)f_Y(y)dxdy \\ &= \textstyle\int y^k \sum_{j=0}^{m+m_1} y^jf_Y(y) \int x^l \sum_{i=0}^{n+n_1} A_{i,j}x^if_X(x)dxdy. \end{aligned} \tag{122}$$

For any fixed j, $\sum_{i=0}^{n+n_1} A_{i,j}x^i$ can be represented in terms of orthonormal polynomials (similar to Equation 111)

$$\sum_{i=0}^{n+n_1} A_{i,j}x^i = h_n(x)\left(\sum_{i_1=0}^{n_1} a_{n+i_1,j}h_{i_1}(x)\right) + \sum_{i=0}^{n-1} a_{i,j}h_i(x), \tag{123}$$

where the first subscript in the coefficient a represents the total degree of this term with respect to x, and the second represents the degree with respect to y inside the summation operator in Equation 122. As $x^l\left(\sum_{i_1=0}^{n_1} a_{n+i_1,j}h_{i_1}(x)\right)$ is a polynomial with the degree of x less than or equal to $n-1$, where $j \in \{0, ..., m+m_1\}$, the orthogonality naturally leads to

$$\begin{aligned} E\left[x^l y^k g(x,y)\right] &= \int y^k \textstyle\sum_{j=0}^{m+m_1} y^j f_Y(y) \int x^l \sum_{i=0}^{n-1} a_{i,j}h_i(x)f_X(x)dxdy \\ &= \int x^l \textstyle\sum_{i=0}^{n-1} h_i(x)f_X(x) \int y^k a_{i,j} \sum_{j=0}^{m+m_1} y^j f_Y(y)dydx. \end{aligned} \tag{124}$$

Next, we reduce the order of y to $m-1$ similar to the above process. It is then not difficult to obtain

$$\begin{aligned} E\left[x^l y^k g(x,y)\right] &= \int x^l \sum_{i=0}^{n-1} h_i(x)f_X(x) \int y^k b_{i,j} \sum_{j=0}^{m-1} h'_j(y)f_Y(y)dydx \\ &= \iint x^l y^k \sum_{i=0}^{n-1}\sum_{j=0}^{m-1} b_{i,j}h_i(x)h'_j(y)f_X(x)f_Y(y)dxdy \\ &= E\left[x^l y^k g^*(x,y)\right]. \end{aligned} \tag{125}$$

□

Proof of Theorem 6.4:

Again, we prove the two-variable case shown in the following lemma. The proof of the general case follows naturally.

Lemma 6.9. *Consider a mapping* $g(x,y)=\sum_{i=0}^{n+n_1}\sum_{j=0}^{m+m_1}A_{i,j}x^iy^j$ *for some* $n_1\in\{0,...,n-1\}$ *and* $m_1\in\{0,...,m-1\}$*. If a PCM mapping of the form* $g^*(x,y)=\sum_{i=0}^{n-1}\sum_{j=0}^{m-1}b_{i,j}h_i(x)h'_j(y)$ *is used to fit the original mapping, the MSE of the PCM fit cannot be improved by adding any polynomial with the degree of* x *up to* $n-1-n_1$*, and the degree of* y *up to* $m-1-m_1$*. Moreover, the lower order mapping* $g_r^*(x,y)=\sum_{i=0}^{n-1-n_1}\sum_{j=0}^{m-1-m_1}b'_{i,j}h_i(x)h'_j(y)$ *is the MMSE mapping, among all polynomials with the degree of* x *up to* $n-1-n_1$ *and the degree of* y *up to* $m-1-m_1$*.*

Proof. To prove the first part of the theorem, we construct $\bar{g}(x,y)=g^*(x,y)+\sum_{i=0}^{n-1-n_1}\sum_{j=0}^{m-1-m_1}C_{i,j}x^iy^j$, and show that $E\left[\left(g(x,y)-g^*(x,y)\right)^2\right]\leq E\left[\left(g(x,y)-\bar{g}(x,y)\right)^2\right]$. Here, the coefficients $C_{i,j}\in R$. The mean square error between $g(x,y)$ and $\bar{g}(x,y)$ can be expressed as:

$$\begin{aligned}E\left[\left(g(x,y)-\bar{g}(x,y)\right)^2\right]&=E\left[\left(g(x,y)-g^*(x,y)\right)^2\right]\\&+2E\Big[\left(g(x,y)-g^*(x,y)\right)\left(g^*(x,y)-\bar{g}(x,y)\right)\Big]\\&+E\left[\left(g^*(x,y)-\bar{g}(x,y)\right)^2\right].\end{aligned}\tag{126}$$

We note that $E\left[\left(g^*(x,y)-\bar{g}(x,y)\right)^2\right]$ is always nonnegative. Therefore, it is sufficient to show
$E\Big[\left(g(x,y)-g^*(x,y)\right)\left(g^*(x,y)-\bar{g}(x,y)\right)\Big]=0$. To do that, we first notice

$$g^*(x,y)-\bar{g}(x,y)=-\left(\sum_{i=0}^{n-1-n_1}\sum_{j=0}^{m-1-m_1}C_{i,j}x^iy^j\right).\tag{127}$$

Moreover, $g(x,y)-g^*(x,y)$ is the sum of three terms: $\sum_{i=0}^{n_1}\sum_{j=0}^{m_1}b_{n+i,m+j}h_n(x)h_i(x)h'_m(y)h'_j(y)$, $\sum_{i=0}^{n_1}\sum_{j=0}^{m-1}b_{n+i,j}h_n(x)h_i(x)h'_j(y)$, and $\sum_{i=0}^{n-1}\sum_{j=0}^{m_1}b_{i,m+j}h_i(x)h'_m(y)h'_j(y)$. As each $h_i(x)$ in the first two terms is a polynomial of degree at most n_1, $-h_i(x)\left(\sum_{i=0}^{n-1-n_1}\sum_{j=0}^{m-1-m_1} C_{i,j}x^i y^j\right)$ is a polynomial of degree at most $n-1$ of variable x. Applying orthogonality with respect to the variable x, it is straightforward to find $E\Big[\big(g(x,y)-g^*(x,y)\big)\big(g^*(x,y)-\bar{g}(x,y)\big)\Big]$ equals to zero for the first two terms of $g(x,y)-g^*(x,y)$. Using the same argument for the third term and apply orthogonality to the variable y, we can find that $E\left[\big(g(x,y)-g^*(x,y)\big)\big(g^*(x,y)-\bar{g}(x,y)\big)\right]$ also equals to 0 for the third term.

We next prove the second part of the theorem. This is sufficient to show that $E\left[x^l y^k(g(x,y)-g_r^*(x,y))\right]$ equals to zero, for $l\in\{0,...,n-1-n_1\}$ and $k\in\{0,...,m-1-m_1\}$. Notice that

$$
\begin{aligned}
E\left[x^l y^k\big(g(x,y)-g_r^*(x,y)\big)\right] &= E\left[x^l y^k\big(g(x,y)-g^*(x,y)\big)\right] \\
&\quad +E\left[x^l y^k\big(g^*(x,y)-g_r^*(x,y)\big)\right].
\end{aligned} \tag{128}
$$

The first term is zero according to Theorem 6.3. The second term is also zero according to a proof similar to that of the first part of this theorem. In particular, each term in $g^*(x,y)-g_r^*(x,y)$ contains $h_i(x)h'_j(y)$ with degrees either $i\geq n-n_1$ or $j\geq m-m_1$. Invoking orthogonality, the second part is proved as well. □

Proof of Theorem 6.6:

To ease understanding, we first investigate the correlated two-variable PCM in Lemma 6.10.

LEMMA 6.10. *Consider a two-variable mapping $g_c(x,y)$ of the form*

$$
g_c(x,y)=\sum_{j=0}^{2n-1}\sum_{i=0}^{2n-1-j} A'_{i,j}x^i y^j, \tag{129}
$$

where the coefficients $A'_{i,j} \in R$, *and* n *is an integer greater than* 1. *Assume that the two variables* x *and* y *follow a joint distribution* $f_{X,Y}(x, y)$ *and the* r*-th conditional moment of one variable given the other is at most a* r*-th degree polynomial of the other variable. Then the mapping* $g_c(x, y)$ *can be approximated by a lower-order PCM mapping* $g_c^*(x, y)$ *of the form*

$$g_c^*(x, y) = \sum_{j=0}^{n-1} \sum_{i=0}^{n-1} B'_{i,j} x^i y^j, \tag{130}$$

such that $E[g_c(x, y)] = E[g_c^*(x, y)]$, *where the coefficients* $B'_{i,j} \in R$.

Proof. Without loss of generality, we assume that $\int x^r f_{X|Y}(x|y)dx$ is a r-th degree polynomial of y. We first construct the expression of $g_c^*(x, y)$ from $g_c(x, y)$ without changing its mean. We then verify that $g_c(x, y)$ and the constructed $g_c^*(x, y)$ have the same values at the set of PCM points selected from the marginal and conditional probabilities. We can thus use these PCM points to uniquely identify $g_c^*(x, y)$.

To calculate the mean of the original mapping $g_c(x, y)$, we note that because $E[g_c(x, y)]$ can be expressed as $\sum_{j=0}^{2n-1} \sum_{i=0}^{2n-1-j} E\left[A'_{i,j} x^i y^j\right]$, we can focus on the calculation of each term $E\left[A'_{i,j} x^i y^j\right]$ first. To simplify $E\left[A'_{i,j} x^i y^j\right]$, we consider three cases with different ranges of i and j: Case 1 ($0 \le i < n$, $n-1 < j \le 2n-1$), Case 2 ($n-1 < i \le 2n-1$, $0 \le j < n$), and Case 3 ($0 \le i < n$, $0 \le j < n$). As in Case 3, the polynomial is already in the form of $g_c^*(x, y)$, we only focus on the first two cases. We show the calculation of $E\left[A'_{i,j} x^i y^j\right]$ for Case 1 only, as Case 2 follows a similar procedure.

In Case 1, when $n-1 < j \le 2n-1$, we can express y^j using the orthogonal polynomials of y. In particular, $y^j = h_n(y)\big(\sum_{k=0}^{j-n} b'_{j,k+n} h_k(y)\big) + \sum_{k=0}^{n-1} b'_{j,k} h_k(y)$, where the coefficients $b' \in R$, the first subscript of b' stands for the total degree of y, and the second represents the degree of y in each term. Moreover, as $0 \le i < n$, for each particular y, we can write x^i

in terms of the conditional orthogonal polynomials of x given y, denoted as $h'_k(x|y)$, where $k \in \{0, ..., i\}$ represents the degree of the orthogonal polynomial. In particular, $A'_{i,j}x^i = \sum_{l=0}^{i}(a'_{i,l}|y)h'_l(x|y)$, where $(a'_{i,l}|y)$ is a parameter dependent upon y, and the format of the subscripts is the same as that of b', but for the variable x. The conditional orthogonal polynomial $h'_l(x|y)$ is associated with the conditional pdf $f_{X|Y}(x|y)$. As such, at different y, $h'_l(x|y)$ may have different expressions. Now, we are ready to calculate $E\left[A'_{i,j}x^i y^j\right]$ as

$$
\begin{aligned}
&E\Big[A'_{i,j}x^i y^j\Big] \\
&= \iint A'_{i,j}x^i y^j f_{X|Y}(x|y)f_Y(y)dxdy = \iint \sum_{l=0}^{i}(a'_{i,l}|y)h'_l(x|y) \\
&\left(h_n(y)\big(\sum_{k=0}^{j-n} b'_{j,k+n}h_k(y)\big) + \sum_{k=0}^{n-1} b'_{j,k}h_k(y)\right) f_{X|Y}(x|y)f_Y(y)dxdy \\
&= \iint \sum_{l=0}^{i}(a'_{i,l}|y)h'_l(x|y)\left(h_n(y)\big(\sum_{k=0}^{j-n} b'_{j,k+n}h_k(y)\big)\right) \\
&f_{X|Y}(x|y)f_Y(y)dxdy + \iint \sum_{l=0}^{i}(a'_{i,l}|y)h'_l(x|y)\left(\sum_{k=0}^{n-1} b'_{j,k}h_k(y)\right) \\
&f_{X|Y}(x|y)f_Y(y)dxdy = \int h_n(y)\left(\sum_{k=0}^{j-n} b'_{j,k+n}h_k(y)\right) \\
&\left(\int \sum_{l=0}^{i}(a'_{i,l}|y)h'_l(x|y)f_{X|Y}(x|y)dx\right) f_Y(y)dy \\
&+ \iint \sum_{l=0}^{i}(a'_{i,l}|y)h'_l(x|y)\left(\sum_{k=0}^{n-1} b'_{j,k}h_k(y)\right) f_{X|Y}(x|y)f_Y(y)dxdy.
\end{aligned} \tag{131}
$$

Now let us show that the first double integration in the last equation equals 0. Note that for each y, $\sum_{l=0}^{i}(a'_{i,l}|y)h'_l(x|y)$ is a polynomial of x with the degree up to i. Furthermore, according to the assumption that $\int x^i f_{X|Y}(x|y)dx$ is at most an i-th degree polynomial of y, we find that $\left(\sum_{k=0}^{j-n} b'_{j,k+n}h_k(y)\right)$

$\left(\int \sum_{l=0}^{i}(a'_{i,l}|y)h'_l(x|y)f_{X|Y}(x|y)dx\right)$ is a polynomial of y with the degree up to $i+j-n \leq 2n-1-n = n-1$. Applying orthogonality with respect to variable y, we can simplify Equation 131 to

$$
\begin{aligned}
E[A'_{i,j}x^iy^j] &= \iint \sum_{l=0}^{i}(a'_{i,l}|y)h'_l(x|y)\left(\sum_{k=0}^{n-1} b'_{j,k}h_k(y)\right) f_{X|Y}(x|y)f_Y(y)dxdy \\
&= \iint \sum_{l=0}^{i}\sum_{k=0}^{n-1}(c'_{l,k}|y)h'_l(x|y)h_k(y)f_{X,Y}(x,y)dxdy
\end{aligned} \tag{132}
$$

for $0 \leq i < n$, $n-1 < j \leq 2n-1$. Here $(c'_{l,k}|y)$ is some parameter dependent upon y.

Following a similar procedure, we can show that in Case 2,

$$
\begin{aligned}
E[A'_{i,j}x^iy^j] &= \iint \Big(h'_n(x|y)\big(\sum_{l=0}^{i-n}(a'_{i,l+n}|y)h'_l(x|y)\big) + \sum_{l=0}^{n-1}(a'_{i,l}|y)h'_l(x|y)\Big) \\
&\quad \Big(\sum_{k=0}^{j} b'_{j,k}h_k(y)\Big)f_{X,Y}(x,y)dxdy \\
&= \iint \sum_{l=0}^{n-1}\sum_{k=0}^{j}(d'_{l,k}|y)h'_l(x|y)h_k(y)f_{X,Y}(x,y)dxdy,
\end{aligned} \tag{133}
$$

where $(d'_{l,k}|y)$ is some parameter dependent upon y. As for both cases, $E[A'_{i,j}x^iy^j]$ can be expressed using orthogonal and conditional orthogonal polynomials with the degree up to $n-1$, we can write

$$
\begin{aligned}
E[g_c(x,y)] &= \sum_{j=0}^{2n-1}\sum_{i=0}^{2n-1-j} E[A'_{i,j}x^iy^j] \\
&= \iint \sum_{l=0}^{n-1}\sum_{k=0}^{n-1}(e'_{l,k}|y)h'_l(x|y)h_k(y)f_{X,Y}(x,y)dxdy,
\end{aligned} \tag{134}
$$

where $(e'_{l,k}|y)$ is some parameter dependent upon y. Noticing that $\sum_{l=0}^{n-1}\sum_{k=0}^{n-1}(e'_{l,k}|y)h'_l(x|y)$ $h_k(y)$ is a polynomial with the degree of each variable up to $n-1$, we can write

$$E[g_c(x,y)] = \iint \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} B'_{i,j} x^i y^j f_{X,Y}(x,y) dxdy = E[g_c^*(x,y)]. \tag{135}$$

It is clear that $g_c^*(x,y)$ and $g_c(x,y)$ have the same mean.

Finally, let us verify that both $g_c^*(x,y)$ and $g_c(x,y)$ pass through the same set of selected PCM points. Let us denote the roots of $h_n(y)$ as y_1, y_2 , ... , y_n and the roots of $h'_n(x|y_j)$ as x_{ij} for $1 \leq i \leq n$, $1 \leq j \leq n$. The n^2 pairs of inputs (x_{ij}, y_j) are the selected PCM points. By expressing $g_c(x,y)$ and $g_c^*(x,y)$ in orthogonal forms, and from observing Equations 131, 132 and 134, it is clear to see that $g_c(x,y)$ and $g_c^*(x,y)$ are identical at each pair of the PCM points. The proof is now complete. □

We now generalize to more than two variables, and sketch the key steps to prove Theorem 6.6.

PROOF. First of all, note that there is only one variable in the original mapping with degree greater than $n-1$, as the total degree of $A'_{k_1,k_2,...,k_m} \prod_{i=1}^{m} x_i^{k_i}$ is at most $2n-1$. Without loss of generality, let us denote this variable as x_t, where $1 \leq t \leq m$. As such, we only need to find a low-order mapping with the degree of this variable x_t reduced to $n-1$.

To do that, we express x_t in terms of orthogonal polynomials defined upon the conditional pdf $f_{X_t|X_{t+1},X_{t+2},...,X_m}(x_t|x_{t+1}, x_{t+2}, ..., x_m)$. Let us denote these orthonormal polynomials as

$h_i^t(x_t|x_{t+1}, ..., x_m)$, where i is the degree of this polynomial. It is then easy to see that any term in $g_c(x_1, x_2, ..., x_m)$ that includes x_t, denoted as $A'_{k_1,k_2,...,k_t,...,k_m} \prod_{i=1}^{m} x_i^{k_i}$, can be expressed as

$$A'_{k_1,k_2,...,k_t,...,k_m} \prod_{i=1}^{m} x_i^{k_i} = \left(\prod_{i=1}^{t-1} x_i^{k_i} \right) \Bigg(h_n^t(x_t|x_{t+1}, ..., x_m) \Big(\sum_{l=0}^{k_t - n} \left(a'^t_{k_t, l+n} | x_{t+1}, ..., x_m \right)$$

$$h_l^t(x_t|x_{t+1},...,x_m)\Big) + \sum_{l=0}^{n-1} \left(a'^t_{k_t,l}|x_{t+1},...,x_m\right) h_l^t(x_t|x_{t+1},...,x_m)\Bigg) \left(\prod_{i=t+1}^{m} x_i^{k_i}\right), \tag{136}$$

where $a'^t \in R$ are the corresponding coefficients. Again, the first subscript of a'^t stands for the total degree of x_t, and the second represents the degree of x_t in each term.

In order to show that the degree of x_t can be reduced to a value smaller than n, we only need to show that in $E[g_c(x_1, x_2, ..., x_m)]$, any term involved with $h_l^t(x_t|x_{t+1}, ..., x_m)$ and with total degree greater than n equals 0. It is sufficient to show that

$$\begin{gathered} \iint ... \int \left(\prod_{i=1}^{t-1} x_i^{k_i}\right) h_n^t(x_t|x_{t+1},...,x_m) \\ \left(\sum_{l=0}^{k_t-n} \left(a'^t_{k_t,l+n}|x_{t+1},...,x_m\right) h_l^t(x_t|x_{t+1},...,x_m)\right) \\ \left(\prod_{i=1}^{t} f_{X_i|X_{i+1},...,X_m}(x_i|x_{i+1},...,x_m)\right) dx_1 dx_2 ... dx_t = 0. \end{gathered} \tag{137}$$

Recall the assumption that each conditional moment of x_i (i.e. $\int x_i^{k_i} f_{X_i|X_{i+1},X_{i+2},...,X_m}$ $\big(x_i|x_{i+1}, x_{i+2}, ..., x_m\big)dx_i)$ is a polynomial of $x_{i+1}, x_{i+2}, ..., x_m$ with total degree k_i, $1 \le i \le t-1$. We could then find that the conditional moment of $\prod_{i=1}^{t-1} x_i^{k_i}$ expressed as

$$\iint ... \int \left(\prod_{i=1}^{t-1} x_i^{k_i}\right) \left(\prod_{i=1}^{t-1} f_{X_i|X_{i+1},...,X_m}(x_i|x_{i+1},...,x_m)\right) dx_1 ... dx_{t-1}$$

is a polynomial of $x_t, x_{t+1}, ..., x_m$ with the total degree not exceeding $\sum_{i=1}^{t-1} k_i$. When multiplying this polynomial with $\sum_{l=0}^{k_t-n} \left(a'^t_{k_t,l+n}|x_{t+1},...,x_m\right) h_l^t(x_t|x_{t+1},...,x_m)$, we find that the maximum degree of x_t is $\sum_{i=1}^{t-1} k_i + k_t - n = \sum_{i=1}^{t} k_i - n \le n-1$. Equation 137 is proved according to the orthogonality.

The rest of the proof on selecting PCM points and constructing $g_c^*(x_1, x_2, ..., x_m)$ follows directly from the proof of Theorem 6.6, and is thus omitted here. □

TABLE 6.1. Algorithm for the Independent Two-Variable PCM

Step 1: *Simulation point selection*

1.1 Compute the orthonormal polynomials of degree i for the random variable X, $h_i(x)$, for $i = 1, ..., n$ according to the following:

a) Initialize $H_{-1}(x) = h_{-1}(x) = 0$ and $H_0(x) = h_0(x) = 1$

b) **For** $i = 1$ to n

$$H_i(x) = xh_{i-1}(x) - \langle xh_{i-1}(x), h_{i-1}(x)\rangle\, h_{i-1}(x) - \langle H_{i-1}(x), H_{i-1}(x)\rangle^{\frac{1}{2}}\, h_{i-2}(x),$$

$$h_i(x) = H_i(x) / \langle H_i(x), H_i(x)\rangle^{\frac{1}{2}}.$$

End

Note that $H_i(x)$ represents the orthogonal polynomial of degree i for the random variable X. The second equation in the FOR loop normalizes $H_i(x)$. $\langle p(x), q(x)\rangle$ denotes the integration operation $\int p(x)q(x)f_X(x)dx$.

1.2 Find the roots of $h_n(x) = 0$ as the n PCM simulation points for X, denoted as x_1, x_2,, ..., x_n.

1.3 Repeat 1.1 and 1.2 for the random variable Y and obtain the orthonormal polynomials $h_j(y)$, for $j = 1, ...m$, and the m PCM simulation points for Y, denoted as y_1, y_2, ..., y_m.

Step 2: *Evaluation of system outputs at selected simulation points*

2.1 For each simulation points (x_i, y_j) where $i \in \{1, ...n\}$ and $j \in \{1...m\}$, run simulation and find the associated output $g(x_i, y_j)$.

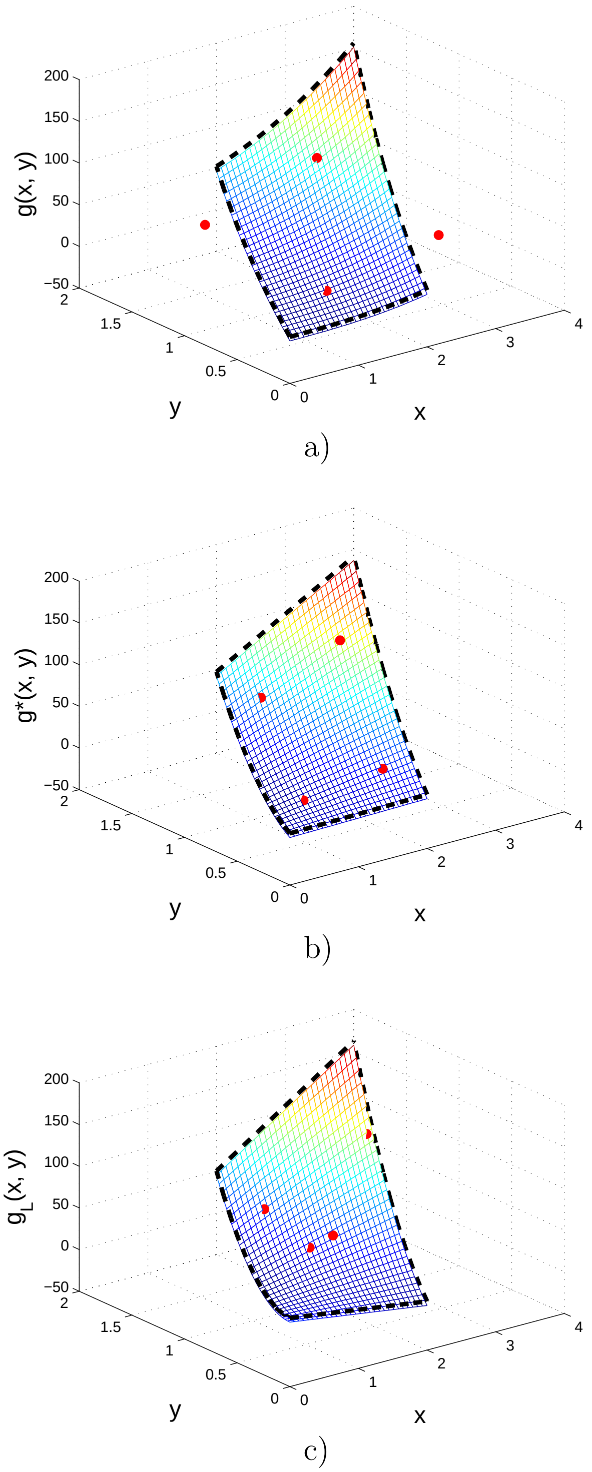
Step 3: *Identification of Mapping Coefficients*

3.1 Find the coefficients $b_{i,j}$ in the low-order PCM mapping: $g^*(x, y) = \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} b_{i,j} h_i(x) h'_j(y)$ following:

$$\begin{bmatrix} b_{n-1,m-1} \\ \vdots \\ b_{n-1,0} \\ \vdots \\ b_{0,m-1} \\ \vdots \\ b_{0,0} \end{bmatrix} = K^{-1} \begin{bmatrix} g(x_1, y_1) \\ \vdots \\ g(x_1, y_m) \\ \vdots \\ g(x_n, y_1) \\ \vdots \\ g(x_n, y_m) \end{bmatrix}, \text{ where } K = \begin{bmatrix} h_{n-1}(x_1)h'_{m-1}(y_1) & \cdots & h_{n-1}(x_1)h'_0(y_1) & \cdots & h_0(x_1)h'_0(y_1) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ h_{n-1}(x_1)h'_{m-1}(y_m) & \cdots & h_{n-1}(x_1)h'_0(y_m) & \cdots & h_0(x_1)h'_0(y_m) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ h_{n-1}(x_n)h'_{m-1}(y_1) & \cdots & h_{n-1}(x_n)h'_0(y_1) & \cdots & h_0(x_n)h'_0(y_1) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ h_{n-1}(x_n)h'_{m-1}(y_m) & \cdots & h_{n-1}(x_n)h'_0(y_m) & \cdots & h_0(x_n)h'_0(y_m) \end{bmatrix}.$$

3.2 The predicted output mean is $b_{0,0}$.

3.3 Reorganize terms in the low-order PCM mapping, and obtain coefficients $B_{i,j}$ in the low-order PCM mapping: $g^*(x, y) = \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} B_{i,j} x^i y^j$.



a)

b)

c)

FIGURE 6.4. a) The original mapping and the four PCM points generated using the independent PCM method; b) the four PCM points and low-order mapping generated using the correlated PCM approach; and c) the four simulation points and mapping generated by the LHD approach.

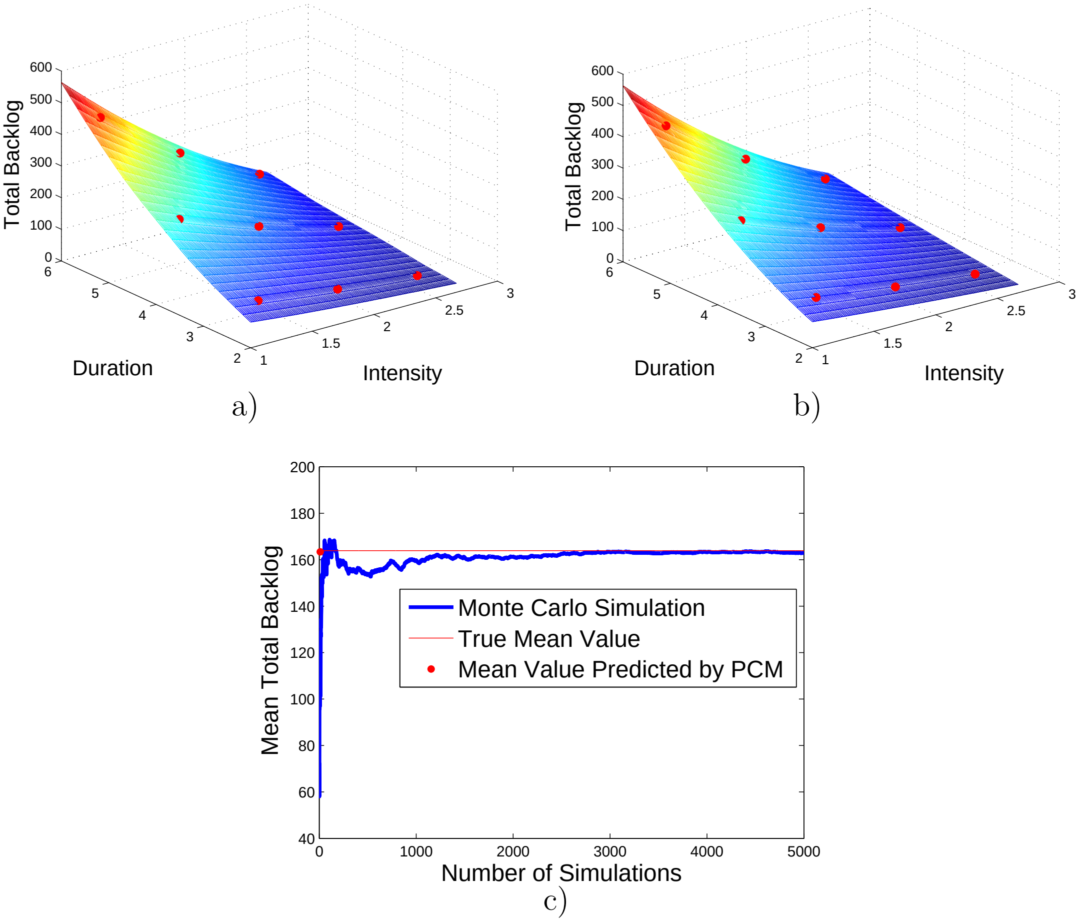
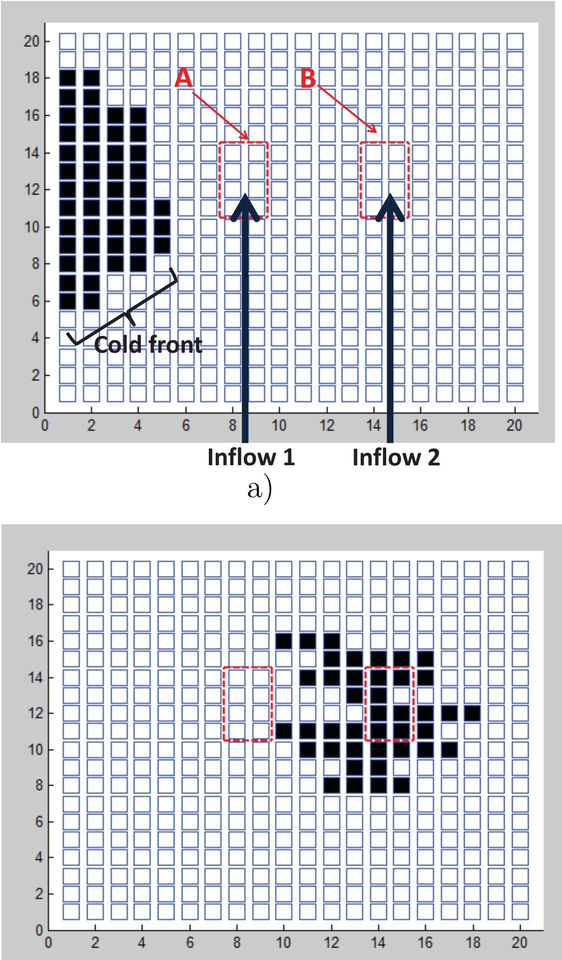


FIGURE 6.5. a) The original mapping between the input set including weather duration and weather intensity, and the output (30-hour total backlog) obtained using the Monte Carlo simulation. b) The reduced-order multivariate PCM mapping is based on the 9 sample points marked as red spots on the plot. c) Comparison of the number of simulation runs needed to predict the correct mean total backlog over a 30-hour span. The blue curve shows the means predicted by the accumulative Monte Carlo runs with the number specified on the x axis. The red thin line shows the true mean estimated by the value that the Monte Carlo simulation (of two random variables) converges to. The red spot corresponds to the 9 simulation runs required for the PCM method to predict the true mean.



a)

b)

FIGURE 6.6. a) Illustration of a spatiotemporal correlated two-region example; b) the development of cold font causes capacity reduction.

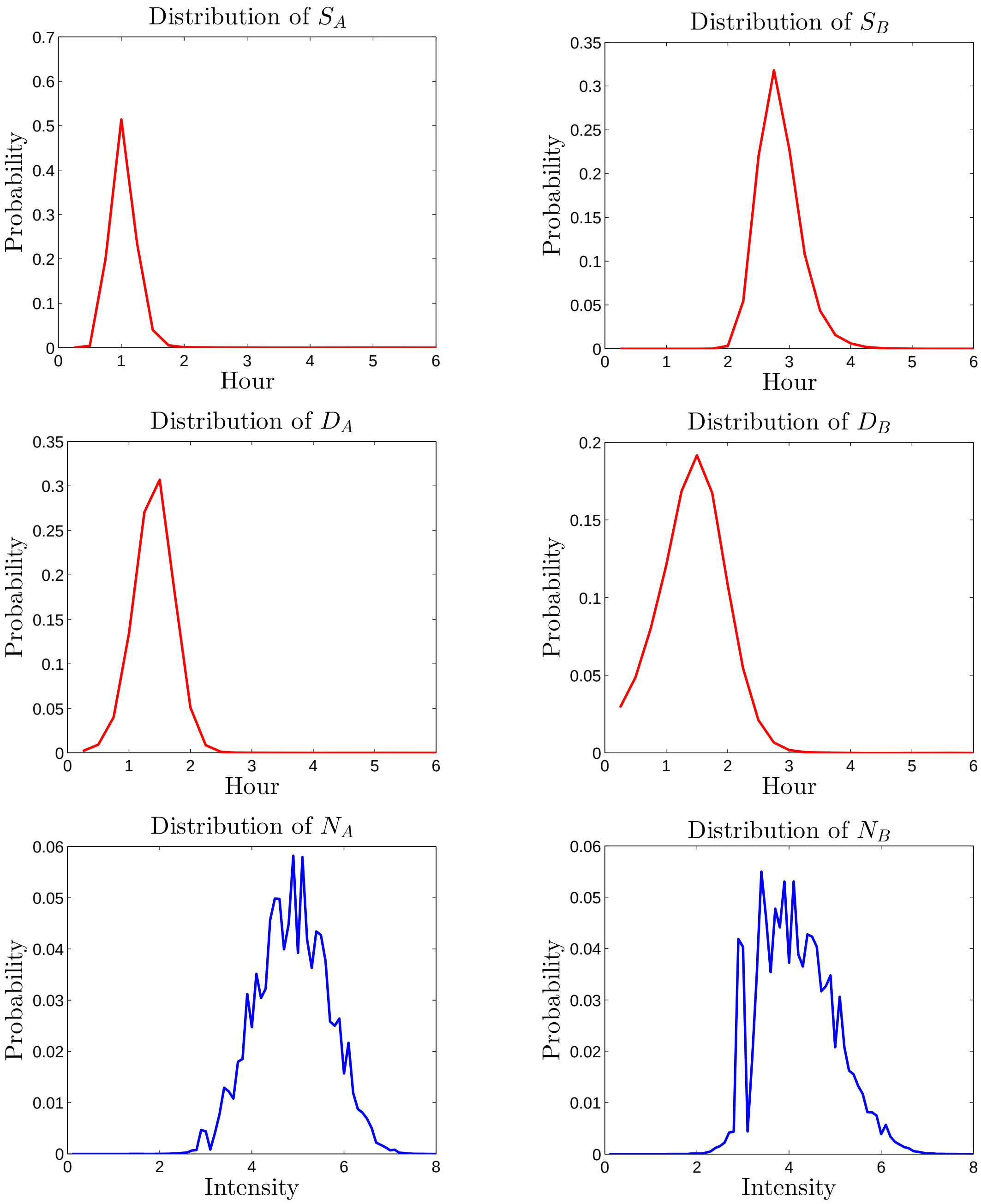
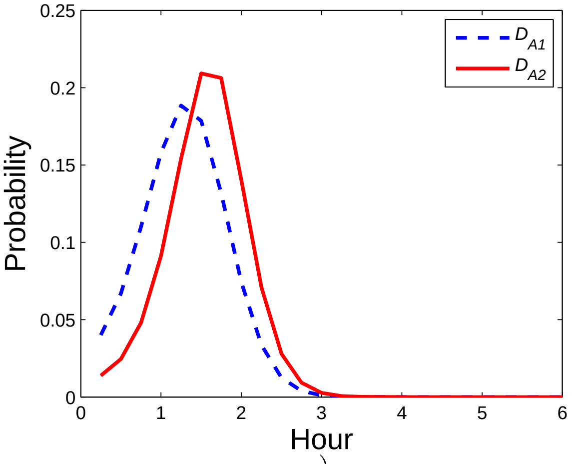
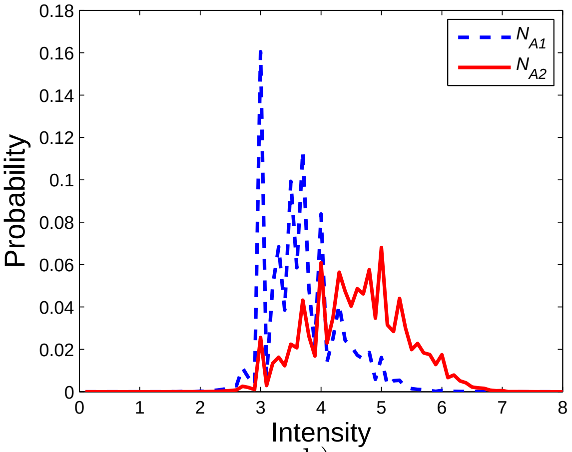


FIGURE 6.7. Probability density functions of six weather parameters.



a)



b)

FIGURE 6.8. a) Distribution of D_B conditioned upon 2 selected PCM coordinates of D_A. b) Distribution of N_B conditioned upon 2 selected PCM coordinates of N_A.

Table 6.2. Performance Comparison among the Monte Carlo and PCM Approaches with Two Different Expressions

	Backlog of Region A	Backlog of Region B	Total Backlog
Monte Carlo Simulation	40.7	72.09	112.8
PCM with 160 terms	40.5	74.2	114.7
PCM with 28 terms	40.5	74	114.5

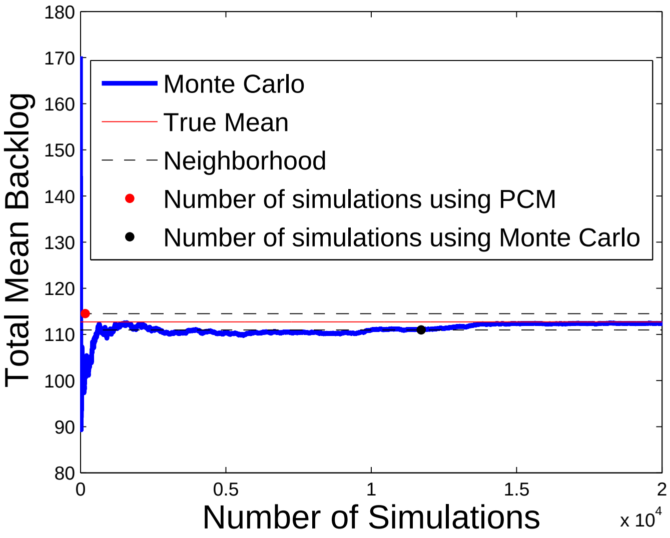


Figure 6.9. Comparison of simulation time between the Monte Carlo and PCM approaches. The mean backlog estimated by the Monte Carlo method is shown in blue. The mean backlog estimated using the PCM approach is marked as the red spot. The dashed lines show a neighborhood around the true mean defined by the estimated mean using the PCM approach, marked by the red spot. The black spot denotes the number of simulations for the Monte Carlo method to reach and stay within the neighborhood.

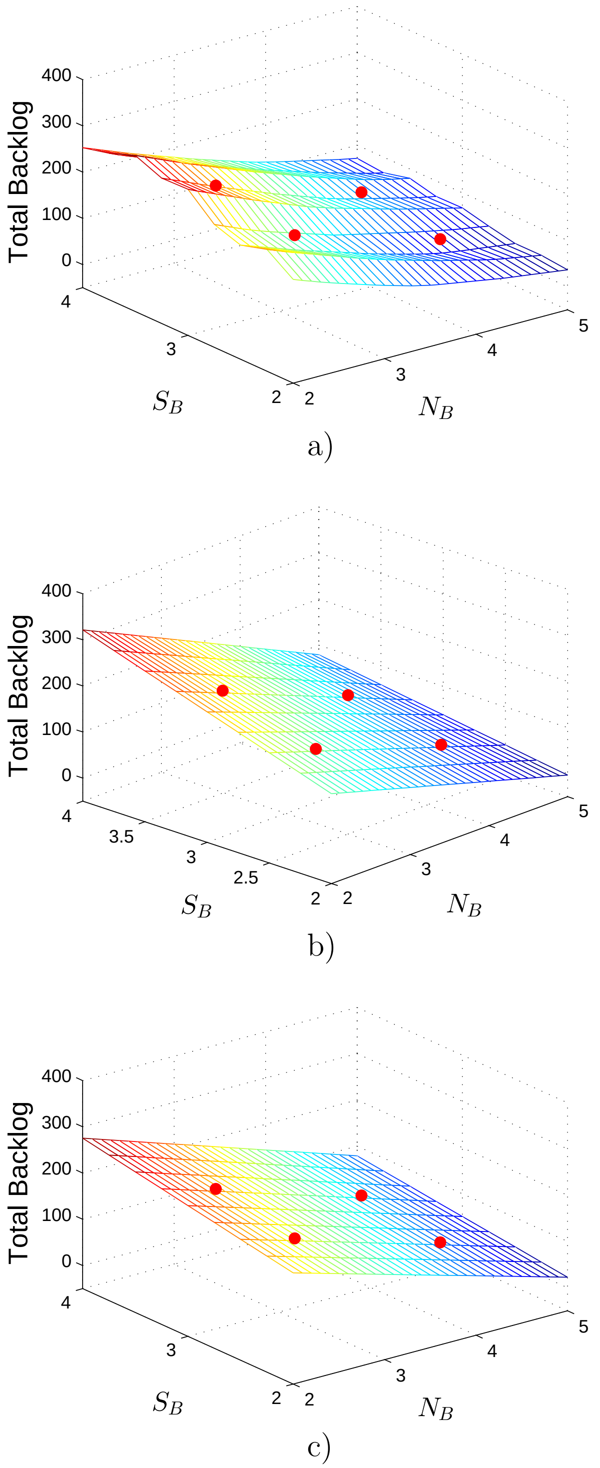


FIGURE 6.10. Mappings with fixed S_A, D_A, D_B and N_A generated by a) Monte Carlo Simulation, b) PCM mapping with 160 terms, and c) PCM mapping with 28 terms.

CHAPTER 7

A PROBABILISTIC COLLOCATION METHOD-BASED APPROACH FOR OPTIMAL STRATEGIC AIR TRAFFIC FLOW MANAGEMENT UNDER WEATHER UNCERTAINTIES

7.1. Introduction

Convective weather events reduce region capacities in the National Airspace System (NAS), and account for the majority of traffic delays in the United States. Strategic Air Traffic Flow Management (TFM) is concerned with re-allocating limited resources 2-15 hours in advance to resolve the discrepancy between demand and reduced capacity. A major challenge facing strategic TFM is the existence of weather uncertainties at this long look-ahead time and the need to design traffic management plans that work well under a range of weather possibilities.

In order to design effective approaches for optimal management strategies under weather uncertainties, we have put forth efforts along two parallel directions. First is the modeling of stochastic weather propagation dynamics. A stochastic automation-based weather simulator has been developed to track the spatiotemporal propagation of convective weather [112, 149, 150]. Second is the modeling of traffic management strategies, followed by the evaluation and parameter optimization of these strategies under uncertainty [136, 151, 154, 155]. In [136], five typical flow management strategies, including routing, Miles-in-trail/minutes-in-trail (MIT/MINIT), Time-based metering (TBM), Airspace Flow Program (AFP) and Ground Delay Program (GDP) are captured as queuing-type flow restrictions. In [151, 155], effective methods are developed to evaluate the impact of these management strategies on the performance of traffic systems. Along the direction of optimal strategy design, we studied in [155] a simple optimal management strategy design problem:

the design of MINIT restriction on a flow entering a single stochastic weather zone. The optimal restriction design is based upon an effective simulation method, called Probabilistic Collocation Method (PCM) [58, 133].In particular, PCM permits the use of a few smartly selected weather samples to define a reduced-order polynomial representation of the design space under uncertainty, from which the optimal restriction can be easily derived.

Despite these advances, we note the above analysis does not easily generalize to the design of multiple flow restrictions for multiple regions (or even NAS-wide). This is because multiple-region design problems cannot be simply decomposed into single-region design problems, due to the correlation of weather propagation at these regions. To achieve optimal flow restriction design for multiple correlated regions, we conducted fundamental studies on the PCM method when multiple correlated parameters are involved. In particular, we developed the multivariate PCM approach, which permits the selection of simulation points on correlated multi-dimensional uncertain parameter space to produce reduced-order mappings with zero-error mean prediction [154].

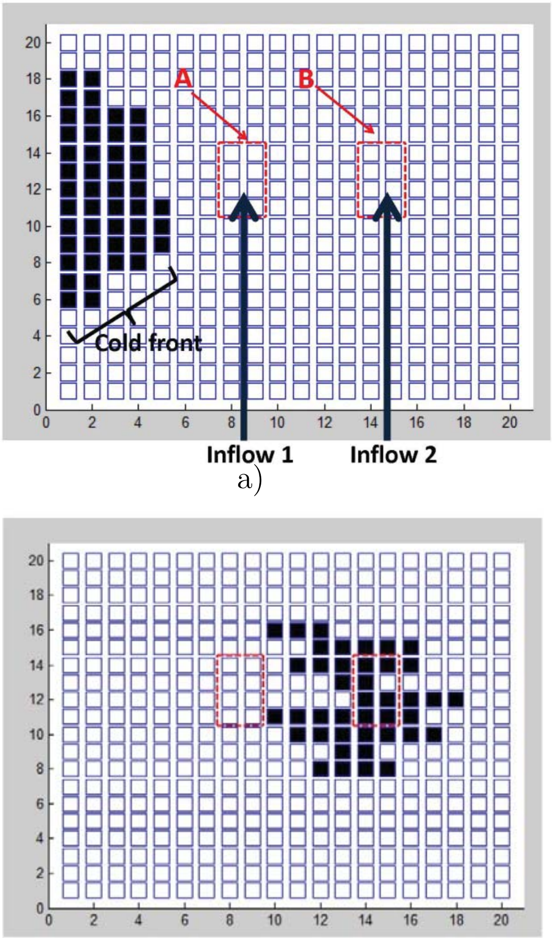
In this chapter, we demonstrate the capability of the multivariate PCM in designing optimal management strategies for multiple regions subject to spatiotemporal correlated weather events. In particular, we consider the scenario of multiple flows entering neighboring weather zones, and provide solutions to address the following two typical management design problems: 1) the design of optimal MINIT management initiatives for these regions, and 2) the design of rerouting management initiative (flow fraction assignment) for these multiple flows. To permit the PCM approach for optimal management design, we also study how a few key statistical parameters (such as weather starting time, duration, and intensity) can be used to capture the key characteristics of correlated weather events. The organization of the chapter is arranged as follows. Section 7.2 includes the modeling framework and the problem formulation. Section 7.3 contains a brief review of the PCM method. In Section 7.4, the empirical data-based multivariate PCM approach is applied to design multiple MINIT restrictions in correlated weather zones. In Section 7.5, a similar approach is applied to optimal rerouting design. The difference resides in the additional effort to find the distribution of routing parameters. Discussions and conclusions are provided in Section 7.6.

7.2. Model Description and Problem Formulation

In this section, we first describe the integrated weather and flow restriction models that permit our analysis, and then mathematically formulate the two optimal air traffic flow management design problems: MINIT restriction design and rerouting design. In order to better illustrate our modeling and optimal design approaches, we use a specific two-region example throughout this chapter as shown in Figure 7.1. Despite the simplicity of the example, the same design procedure applies to larger-scale problems. We also note that both the weather and flow restriction models are highly abstract models, ignoring a variety of operational issues. However, such abstractions are reasonable considering the following two facts: 1) there is large uncertainty in the strategic time-frame, and 2) the goal is strategic management of traffic flows instead of tactical control of flight trajectories.

7.2.1. Weather Model

The weather model that we use in this chapter adopts the stochastic weather simulator developed in [112]. In particular, the simulated airspace is decomposed into small grids, with white color representing normal capacity and black color denoting capacity reduction caused by convective weather (see Figure 7.1. The stochastic propagation of convective weather is governed by the influence model [8], with parameters estimated from hourly probabilistic weather forecasts. Running the simulator can produce a rich number of weather ensembles covering the range of weather possibilities. From analyzing the weather ensembles or directly the stochastic weather model, we can obtain the statistics of weather propagation, which is



a)

b)

FIGURE 7.1. The development of a cold font causes correlated capacity reduction in the two neighboring regions A and B: a) at the beginning of simulation; b): after the cold front leaves region A.

the core information needed for our optimal strategy design under uncertainty.

In this chapter, we consider two scheduled streams of flows entering two neighboring regions A and B with the same capacities. A cold front (indicating a weather front associated with convective weather) passes through these two regions sequentially, producing capacity reductions as shown in Figure 7.1b. The percentage of capacity reduction in each region is defined as the fraction of black grids in the region at each moment. For instance, the capacity reduction in region B is 7/8 according to Figure 7.1b.

Here we use six random parameters to capture the stochastic weather propagation: S_1, N_1, D_1, S_2, N_2, and D_2. In particular, the cold front hits region A from time S_1 with a reduced capacity N_1. We assume that the capacity remains constant (i.e. average capacity reduction) for a span of D_1 before the weather leaves the region. After that, its capacity returns to its original value. Similarly, region B undergoes weather-induced capacity reduction from time S_2, with capacity reduced from the normal value $N_{rb}(= N_{ra})$ to N_2, for a span of D_2. As the same cold font passes through the neighboring two regions, there is interdependency between the sets of random weather parameters associated with the two regions. For instance, the start time S_2 is dependent upon the duration D_1 and start time S_1 of region A due to spatial weather propagation delay; the duration D_2 is dependent upon the duration D_1 due to the correlation of propagation speed; and the capacity N_2 is dependent upon N_1 due to the correlation of weather intensity across time.

7.2.2. Flow Restriction Model

For each region, we use the discrete-time queuing model (or saturation model) developed in [130, 136, 151]to capture the impact of weather and management restrictions on flows (as shown in Figure 7.2). Assume that at each time step k, a total of $x[k]$ aircraft come into the region. Because of the reduced capacity $N[k]$ (caused by both the management restriction and the weather constraint), not all aircraft can enter the region. In particular, the crossing flow $e[k]$ (number of aircraft passing the region) cannot exceed the capacity constraint. The delayed aircraft are assumed to be accumulated in the buffer at the boundary of the region, waiting to pass in the next time step $k + 1$. The number of aircraft in the buffer $B[k]$ is called the backlog at time k. Backlog is an intuitive measurement to abstract region congestion.

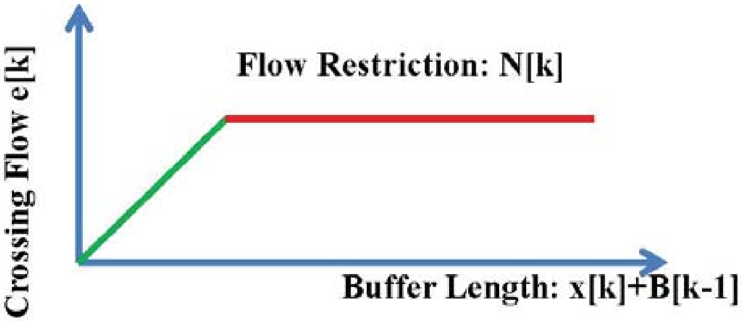


Figure 7.2. Illustration of saturation model

Management-induced backlog is usually associated with lower cost compared with the backlog directly caused by unexpected weather. This is because well-planned management reduces the cost for pilots and human controllers to deal with the emergent encountering of bad weather. In order to capture this realistic consideration, we differentiate weather-induced backlog $B_w[k]$ and management-induced backlog $B_m[k]$. In our model, when both weather-induced capacity constraint $N_w[k]$ and management restriction $N_m[k]$ are acting on a flow, we assume that the flow first passes the management restriction and then the weather-induced capacity constraint. The dynamic of the saturation model is captured by Equations 138–141 in [151]:

$$B_w[k] = max(B_m[k-1] + B_w[k-1] + x[k] - N_m[k], 0) \tag{138}$$

$$e_m[k] = min(B_m[k-1] + B_w[k-1] + x[k], N_m[k]) \tag{139}$$

$$B_w[k] = max(e_m[k] - N_m[k], 0) \tag{140}$$

$$e_w[k] = min(e_m[k], N_m[k]) \tag{141}$$

When there is no management restriction, backlog is only caused by the weather-induced capacity constraint $N_w[k]$. In this case, the dynamic of the model can be simplified to:

$$e_w[k] = min(B_w[k-1] + x[k], N_w[k]) \tag{142}$$

$$B_w[k] = max(e_w[k] - N_w[k], 0) \tag{143}$$

We integrate the above stochastic weather model and the flow restriction model to simulate the impact of stochastic weather and management restriction on flows. In particular, each weather ensemble generated by the stochastic weather simulator produces a reduced capacity sample $N_w[k]$ for each k. The capacity sample and the sequence of deterministic management restrictions (which are subject to design), feed to the above dynamics to obtain the associated total cost over a time span.

7.2.3. Problem Formulation

In this section, we mathematically formulate the design of two types of flow management initiatives: 1) MINIT restriction design for these two regions, and 2) rerouting of flows to these two regions. As convective weather events are random at the strategic time-frame, the resulting backlogs are also uncertain. As such, we define the performance metric using backlog statistics, and in particular the mean backlog under a range of weather possibilities. Let us detail the formulation of each optimal management design problem.

Problem 1: Optimal MINIT restriction design under weather uncertainty.

The aim of this management is to transfer weather-induced backlogs (associated with high unit cost) to management-induced backlogs (associated with low unit cost), through the design of start, duration, and strength of MINIT initiatives.

This problem is formulated as designing a set of management design variables to minimize a cost function as shown below:

- Management design variables: the start time S_A, duration D_A, and strength N_{mA} of the MINIT restriction for region A. A similar set of design variables are defined for region B, denoted as S_B, D_B, and N_{mB}.
- Cost function: total mean cost of weather- and management- induced backlogs over a time span $[0, k_p]$, defined as $\sum_{k=0}^{k_p}(C_1 E(B_m[k]) + C_2 E(B_w[k]))$. C_1 and C_2 are scaling factors representing the unit costs associated with each type of backlogs.

Problem 2: Optimal Rerouting design under weather uncertainty.

The aim of this management action is to route flows to regions less affected by convective weather. In particular, when bad weather is not present in region A or region B, total inflows are equally distributed to the two regions. When the cold front comes to region A, the region has less capacity than region B, and as such more flow should be routed to region B. At this stage, we use P_{A1} to denote fraction of the total flow assigned to region A, and in consequence $(1 - P_{A1})$ of the total flow is assigned to region B. Later when the cold front propagates to region B, more fraction of flow should be assigned to region A instead. We then denote a fraction P_{A2} of the total flow distributed to region A, and $(1 - P_{A2})$ to region B.

This problem is then formulated as designing a set of management design variables to minimize a cost function as shown below:

- Management design variables: the flow redistribution start time S_A and flow fraction P_{A1} when region A is affected by convective weather; flow redistribution switching time $S_w$and the corresponding flow fraction P_{A2} when weather is affecting regions B; and time $S_E$when both flows restore their original fractions 0.5.
- Cost function: total mean cost of weather-induced backlogs over a time span $[0, k_p]$,, defined as $C_2 \sum_{k=0}^{k_p} E(B_w[k])$. C_2 is the scaling factor representing the unit cost

associated with weather-induced backlog. Since no management restriction acts on flows, the cost is determined only by weather-induced backlog.

7.3. Review of the Probabilistic Collocation Method

The probability collocation method is an effective method to evaluate the uncertainty of computationally expensive systems. For a system with an uncertain input parameter, the typical method to evaluate the output performance of the system is to enumerate a large number of samples for this input parameter, and then simulate the system output at each of these input samples. The above so-called Monte Carlo method is time-consuming due to the large number of computationally expensive simulations. The PCM approach can significantly save computation, through smartly choosing a few parameter samples and simulating the system output at these selected parameter samples. PCM works because the limited parameter samples (PCM points) are selected based upon the probabilistic distribution of the parameter, and thus cover the likely range of parameter values. It was shown that an nth-order polynomial PCM mapping constructed from the selected PCM points and their corresponding simulated outputs can predict the same mean of the original system mapping of order not greater than 2n-1 [132]. We have shown in [132] that the PCM mapping does not only produce the correct mean statistics but also other interesting statistics such as cross-correlation, and is the minimum mean-square estimator among polynomials of certain order. This statistical analysis fundamentally suggests that the PCM mapping is a good approximation of the original mapping over the likely range of parameter values.

Another relevant work we have developed on PCM theory is the use of PCM when multiple uncertain parameters are involved. A two-variable example is shown in Figure 7.3. In general, the single-variable PCM approach cannot be directly generalized to the multivariate case. We studied the general form of original mappings and conditions on parameter distributions that permit the multivariate PCM mapping to produce the same mean of the original mapping in [154]. Here is a brief summary of the main results. If the multiple uncertain parameters are independent, or the original mapping does not include cross-terms (e.g., terms with the multiplication of multiple uncertain parameters), we can select PCM points based upon the marginal distribution of each parameter independently. The mapping constructed from the evaluation of these PCM points estimates the correct mean. If any of the above two conditions are violated, choosing PCM points based upon marginal distributions in general cannot produce the correct mean. In this case, we show that a PCM point selection procedure based upon marginal distributions can lead to a mapping with the correct mean under certain conditions. The multivariate PCM approach is widely applicable to applications that involve computationally-expensive simulation of multiple uncertain parameters.

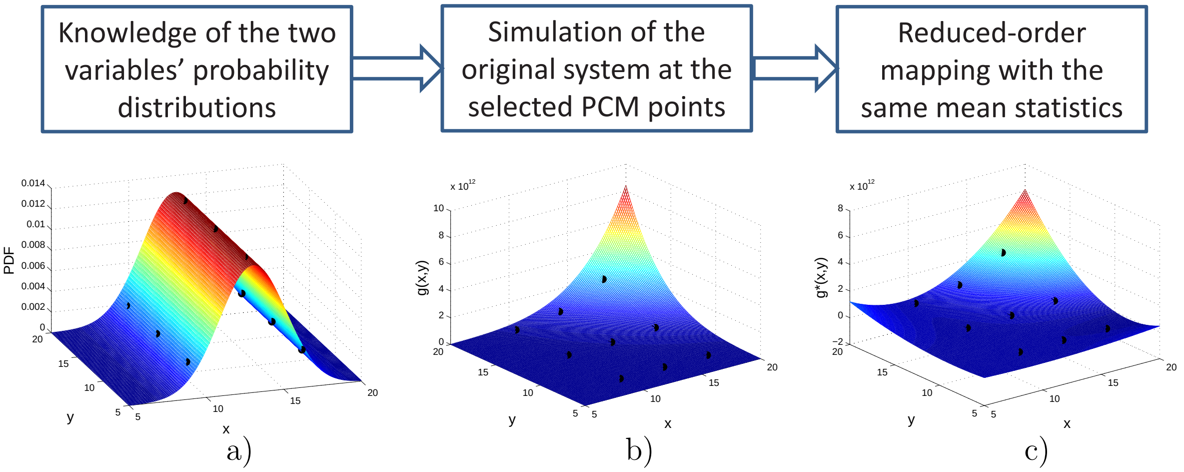


FIGURE 7.3. a) The joint probability distribution; b) The original mapping; and c) The reduced-order multivariate PCM mapping. The black dots represent PCM points.

As in many applications parameter distributions are not directly available, we developed an empirical data-based PCM approach to obtain PCM points directly from parameter data set [132]. The procedures of the empirical-date-based PCM points selection are briefly summarized here. We first find the moments or conditional moments of each input parameter from the sample data set. Then, orthogonal polynomial or conditional orthogonal polynomial [34] for each parameter can be constructed accordingly. Their roots are the PCM points used for simulation and uncertainty evaluation. In this chapter, we take one step further in the sense of using the PCM mapping to find optimal decision-making solutions. We will use the empirical data-based multivariate PCM approach to design optimal air traffic management initiatives robust to a range of weather possibilities.

7.4. Optimal MINIT Design under Weather Uncertainty

In this section, we first elaborate upon the procedures to use the multivariate PCM-based approach for optimal MINIT restriction design. We then show the optimal design results and the advantages of the proposed approach.

7.4.1. Procedures of the Multivariate PCM-based Approach for MINIT Design

Step 1: Weather data generation.

We utilize the weather simulator [112] to generate a large set of weather ensembles, and then find the six weather parameter values (S_1, N_1, D_1, S_2, N_2, and D_2) associated with each weather ensemble. Because of spatiotemporal correlation of weather propagation, these parameters are subject to interdependency. In particular, the start time S_2 is dependent upon D_1 and S_1; the duration D_2 is dependent upon the duration D_1; and the reduced capacity N_2 is dependent upon N_1. Here, weather start time S is marked by the first time the region has at least two black grids. The difference between the last time the region has at least two black grids and S is defined as duration D. The average reduced capacity during the span of weather is calculated as the normal capacity scaled by the average fraction of white grids during this span.

Step 2: Selection of multi-dimensional weather PCM points.

Based upon the set of weather parameters obtained in Step 1, the empirical data-based multivariate PCM is used to find the critical PCM points [58]. To do that, we first find the sample moments of D_1, S_1 and N_1, as they serve as the parameters that D_2, S_2, N_2 are conditioned upon. From these sample moments, the PCM points along each of the dimensions D_1, S_1 and N_1 are obtained. The PCM points along the dimension of D_2 and S_2 are then selected based upon the conditional sample moments of D_2 and S_2 for each combination of the PCM points for D_1 and S_1. Similarly, the PCM points along the dimension of N_2 are selected from the conditional sample moment of N_2 on each of the PCM points for N_1.

Step 3: Identification of PCM management mapping.

Our previous work [155] suggests that for each deterministic weather data ensemble, the best MINIT design concurs with the set of deterministic weather parameters, i.e. $S_A = S_1, S_B = S_2, D_A = D_1, D_B = D_2, N_{mA} = N_1$, and $N_{mB} = N_2$. As such, the weather PCM points selected from Step 2 can directly be used as the PCM points for management evaluation. For each 6-tuple management PCM point denoting a specific design, all weather PCM points are used to find the associated total mean cost defined as $\sum_{k=0}^{k_p}(C_1 E(B_m[k]) + C_2 E(B_w[k]))$ in section 7.2.3. The low-order PCM mapping $f(S_A, D_A, N_{mA}, S_B, D_B, N_{mB})$ is then uniquely defined based upon the 6-tuple management design points and their associated total mean costs.

Step 4: Derivation of optimal management solution.

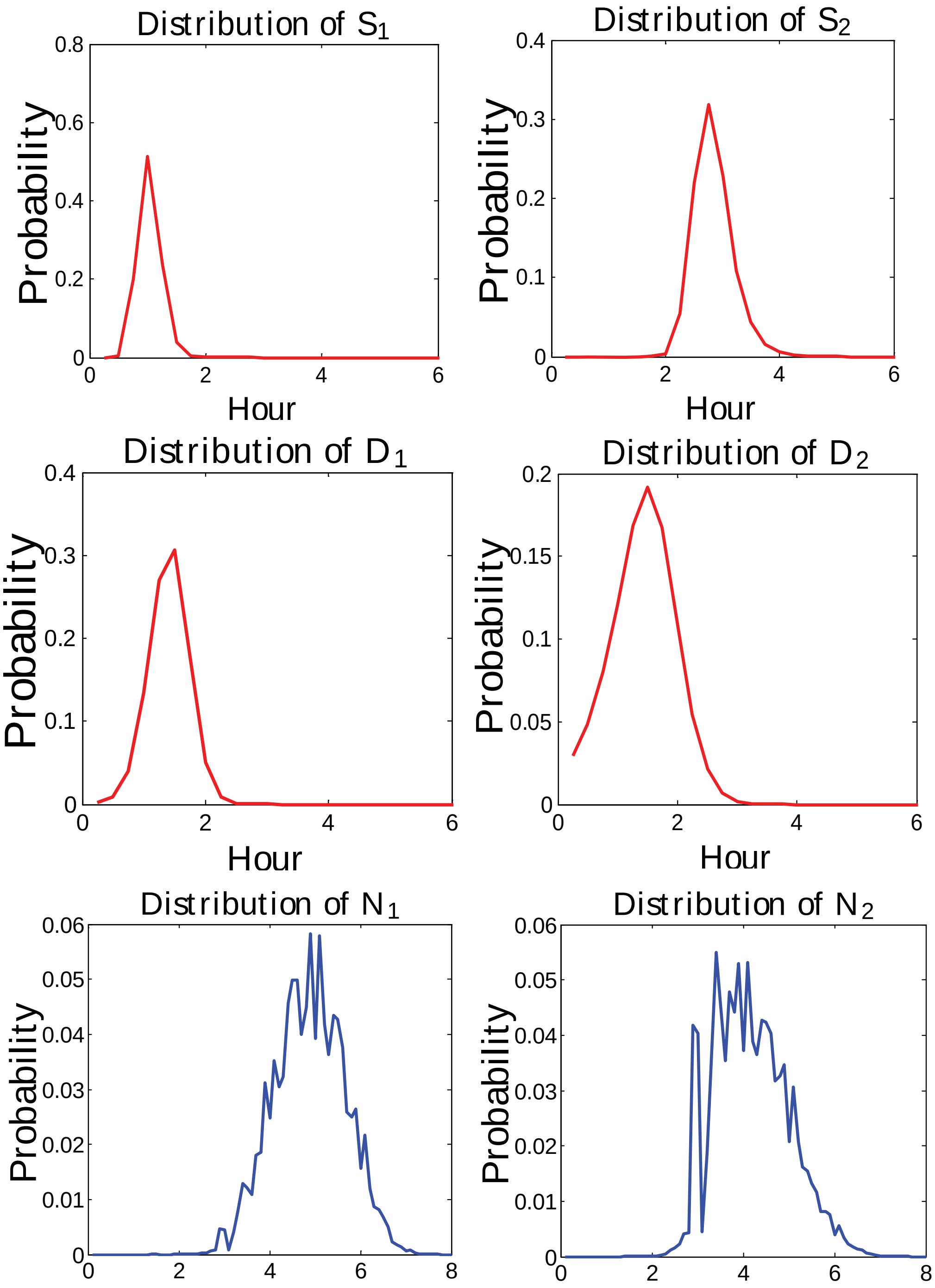
As the management mapping is a multi-dimensional polynomial function defined in a feasible space of the six management design variables, we can find the optimal solution, though, for example, making the partial derivatives of the total mean cost with respect to each of the variables zero when the function is convex in the feasible space. The solution associated with the minimum total mean cost in the feasible management space represents the optimal management design.

7.4.2. Optimal Management Design Results

In this two-region example, the normal capacity and of the two regions are both 10 per unit time (defined as $\Delta t = 15min$). The simulator generates 138611 weather ensembles to cover the range of weather possibilities. The distribution of each of the six weather parameters is shown in Figure 7.4.

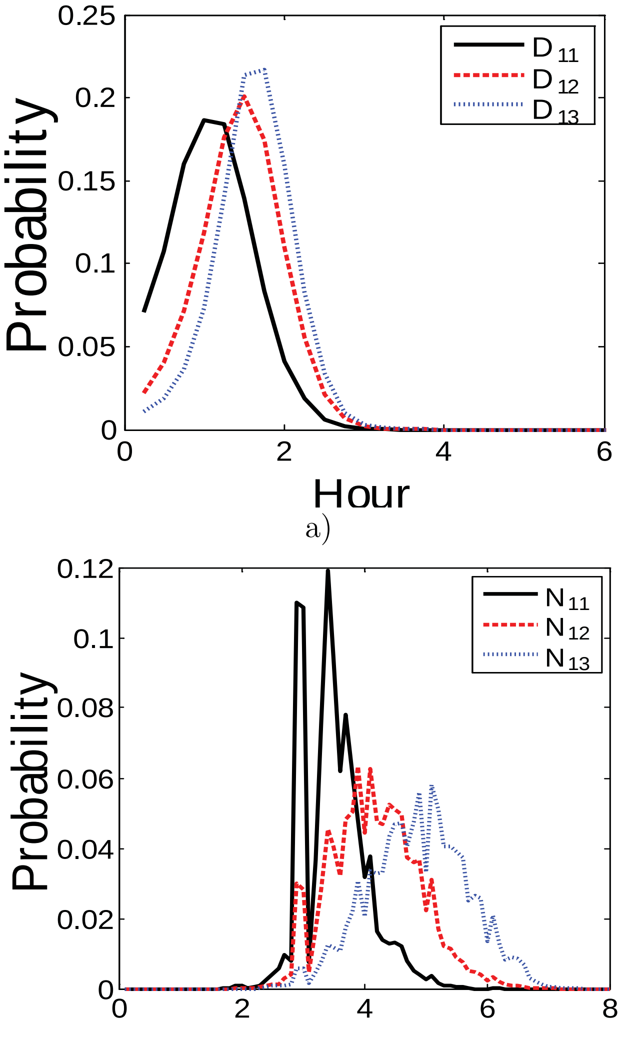
We construct a PCM mapping $f(S_A, D_A, N_{mA}, S_B, D_B, N_{mB})$ with the degree of each variable not greater than 2 to approximate the true mapping between management design variables and the total mean cost. To come up with this mapping, we apply empirical data-based multivariate PCM approach to identify 3 PCM points along each dimension of weather parameters according to the explanation in Step 3. In particular, we first choose 3 PCM points along D_1 as $0.75hour, 1.25hour, 1.75hour$, along S_1 as $0.75hour, 1.5hour, 2hour$, and along N_1 as $3.5561, 4.8602, 6.1310$ based on their sample moments respectively. Next, we identify PCM points along the dimensions of D_2, S_2 and N_2. Because of spatiotemporal correlation of weather parameters in region A and region B, each 3-tuple PCM point (D_1, S_1, N_1) results in different conditional distribution of weather parameters associated with region B (see examples shown in Figure 7.5), and thus different set of PCM points. Specifically, as D_2 and N_2 are conditional upon D_1 and N_1 respectively, 3 selected PCM points for each of D_1 and N_1 will result in 9 selected PCM points each for D_2 and N_2; Similarly, since S_2 is dependent upon both S_1 and D_1, $3 \times 3 = 9$ combinations of the PCM points of S_1 and D_1 identify $9 \times 3 = 27$ PCM points for S_2. Therefore, a total of $3^6 = 729$ weather PCM points are selected.

We assume that two streams of deterministic flows sampled from a Poisson process with mean 5 per 15 minutes enter the two regions. As weather PCM points con-



a) b)

FIGURE 7.4. a) Distribution of weather parameters in region A. b) Distribution of weather parameters in region B.



a)

b)

FIGURE 7.5. a) Distribution of D_B conditioned upon 3 selected PCM coordinates of D_A. b) Distribution of N_B conditioned upon 3 selected PCM coordinates of N_A.

cur with management PCM points, we utilize this property to identify the management mapping. In particular, for each management PCM point, we simulate and evaluate the total cost at each of the selected 729 weather PCM point based on the cost function $\sum_{k=0}^{k_p}(C_1 E(B_m[k]) + C_2 E(B_w[k]))$, where $k_p = 24$ hours, $C_1 = 1$ and $C_2 = 4$. Then, all of the 729 management PCM points and associated total mean costs are used to identify the coefficients in the reduced-order PCM mapping between management variables and total mean cost. In the derived polynomial mapping, the degree of each variable is less than 3. Furthermore, we notice that the total mean cost of the two regions is the summation of that of each region. In particular, the mapping $f(S_A, D_A, N_{mA}, S_B, D_B, N_{mB})$ can be decomposed as $f_A(S_A, D_A, N_{mA}) + f_B(S_B, D_B, N_{mB})$, where f_A and f_B are the functions subject to management variables in each region. This expression can significantly reduce the terms/coefficients of the polynomial mapping function from 729 to $3^3 + 3^3 = 54$, by eliminating the cross terms involving management variables of both regions. Least Square Estimation (LSE) is then applied to identify the 54 coefficients of this further reduced-order mapping. Optimal management solution can be found as $S_A = 1.23hour$,$S_B = 2.86hour$,$N_{mA} = 3.24$, $N_{mB} = 4.59$,$D_A = 1.36hour$, and $D_B = 1.245hour$, through taking partial derivatives of this PCM mapping function with respect to each of the design variables. We expect that a good MINIT management planning should be able to reduce the total mean cost of air traffic congestion under weather uncertainty, because of its objective to transform high unit cost (C_2) weather-induced backlog into low unit cost (C_1) management-induced backlog. In our example, the total mean cost under optimal MINIT management for region A is 61.47, and region B is 152.39. If without management strategies, the total mean cost of two regions is 413.96, with 152.312 in region A and 259.65 in region B.

We also apply the Monte Carlo simulation to find the optimal solution: $S_A = 1hour$,$S_B = 2.5hour$, $N_{mA} = 3.3$ and $N_{mB} = 5$,$D_A = 1.5hour$, $D_B = 1.5hour$. The results match well with those obtained using the PCM approach. The small difference is mainly due to the 15-minute simulation resolution. Comparison suggests that the PCM approach is capable of identifying the near-optimal solution with significantly fewer simulations. Specifically, PCM approach needs only 729×729 simulations. This first 729 stands for the number of selected management PCM points and the second 729 denotes the number of selected weather PCM points to simulate. However, the Monte Carlo approach takes a tremendously larger number of simulations ($10^5 \times 138611$) to cover the 6-variable design space. This study suggests the main advantage of PCM-based approach is in terms of computational efficiency, making this approach promising for real time management.

As it is difficult to visualize our 7-dimentional management design mapping (including 6 input management variables and one output), we fix four parameters S_A, S_B, D_A, and N_{mA}at their optimal values, and only vary D_B and N_{mB}.The three-dimensional mapping is demonstrated in Figure 7.6a. We also apply the Monte Carlo simulation to construct a mapping as shown in Figure 7.6b. The optimal solutions suggested by the 3-dimentional mappings are marked as red spots in those two figures. We notice that the shapes of the two mappings are slightly different because we use a polynomial function to approximate a complex non-linear air traffic system. However, the optimal solutions generated by those two methods are close to each other, which verify the capability of the PCM-based approach to identify a good mapping and to obtain the best management design.

7.5. Optimal Rerouting Design under Weather Uncertainty

Re-routing is another effective method to reduce region congestion. In the two-region example, because of the sequential influence of the cold front to the two regions, it is unlikely that the two regions will be subject to capacity reduction at the same time. As such, redistributing flows into the two regions at different times can also lead to significant cost reduction. In this section, we first illustrate the procedures of multivariate PCM-based optimal rerouting design. We then present the results using the two-region example, and discuss the attractive performance of the PCM-based approach, in terms of accuracy and efficiency.

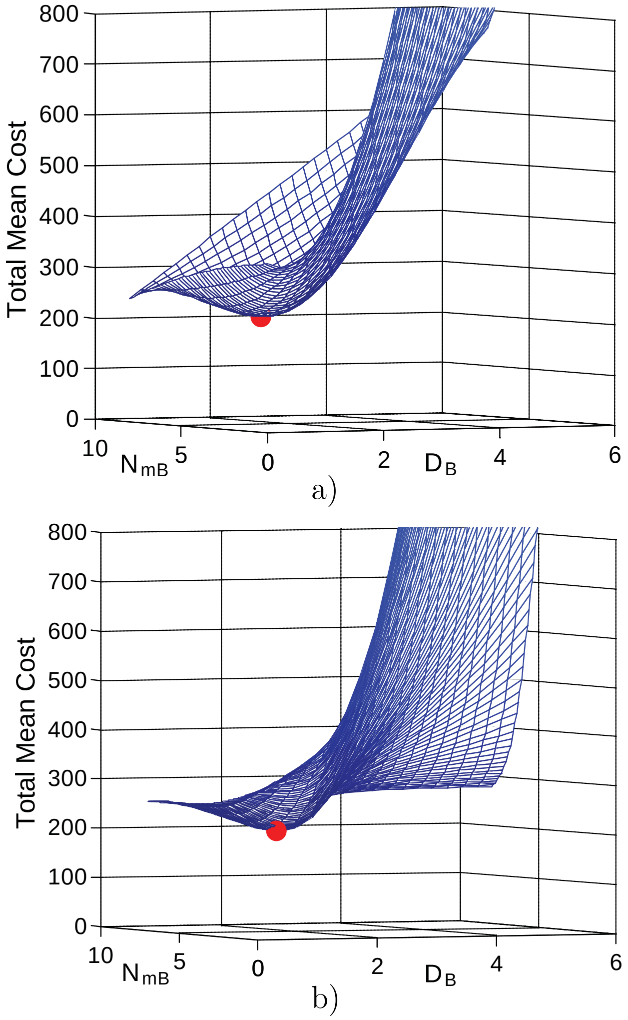


FIGURE 7.6. Mapping between two design variables and the total mean cost, generated by a) the PCM-based approach and b) the Monte Carlo simulation approach.

7.5.1. Procedures of the Multivariate PCM-based Approach for Optimal Rerouting Design

The first two steps regarding weather data generation and weather PCM point selection are the same as those in the MINIT design, and thus are skipped here. Please refer to Section 7.4.1 for the details.

Step 3: Identification of the rerouting management mapping.

In order to achieve the mapping between rerouting management parameters and the associated mean total costs, we first need to identify the distributions of routing design variables. This step is significantly different from that in the MINIT restriction design. In particular, in the MINIT restriction design, the distributions of design parameters are precisely same as those of the weather parameters; however in rerouting there does not exist such equivalence. In order to obtain the distributions of routing variables, we use the following approach.

For each selected 6-tuple weather PCM point, we find the associated optimal rerouting design $(S_A, P_{A1}, S_w, P_{A2}, S_E)$ using a simple search algorithm. This step does not need large computational time, as the initial optimal management parameter values to be used as the seeds in the search algorithm are very easy to guess, and in fact these initial guesses are likely to be very close to the optimal values. For instance, a good guess of the switching time S_w is the start time S_2 of weather in region B. Searching around S_2 can quickly lead to the optimal S_w. The distributions for these optimal solutions are the same as the distributions of the associated weather. Next, based upon the distributions of optimal solutions, we apply the multivariate PCM approach again to find the 5-tuple rerouting PCM points. To find the associated cost for each of these rerouting PCM points, the 6-tuple weather PCM points selected in step 2 are used to obtain the total mean cost defined as$C_2 \sum_{k=0}^{k_p} E(B_w[k])$ in section 7.2.3 A low-order mapping $f(S_A, P_{A1}, S_w, P_{A2}, S_E)$ between the rerouting PCM points and the associated costs can then be easily constructed.

Step 4: Derivation of the optimal management solution. The optimal management solution can then be found from the mapping $f(S_A, P_{A1}, S_w, P_{A2}, S_E)$. If the function is not convex, we can use search algorithms such as the trust-region-reflective algorithm to obtain the optimal routing solution in the feasible management space.

7.5.2. Optimal Rerouting Design Results

To illustrate the procedure and understand the performance of the multivariate PCM-based rerouting design approach, we consider the same weather ensemble data used in the MINIT design. In particular, the same 729 6-tuple representative weather samples are selected as the weather PCM points. In good weather conditions, the two streams of inflows entering the two regions are the same as those in MINIT example (i.e. each flow is a sampled Poisson process with mean 5 per 15 minutes). The rerouting plan works as follows: during the time span $[S_A, S_W)$, less flow (P_{A1} of the sum of the two inflows) is assigned to region A because of its associated weather-induced capacity reduction. Similarly, more flow (P_{A2} of the total inflows) is distributed to region A from $S_W$to $S_E$as the weather has passed region A and is affecting region B.

To obtain the multivariate PCM mapping associated with design variables S_A, P_{A1}, S_w, P_{A2}, andS_E, we must identify critical management PCM points first. However, the 6-tuple weather PCM points cannot directly related to the 5-tuple management PCM points. Additional effort is needed to derive the distributions of the management variables, from which the management PCM points can be found. Our idea is to use the 729 representative weather PCM samples (with their associated probabilities) to obtain the distribution of management variables. To do that, for each deterministic PCM weather sample, we use a simple search algorithm to find the set of best rerouting parameter values associated with the minimum cost $C_2 \sum_{k=0}^{k_p} E(B_w[k])$, where $k_p = 24$ hour and $C_2 = 4$, to be consistent with the settings of the MINIT design. This process leads to 729 typical optimal rerouting management of the form of $S_A, P_{A1}, S_w, P_{A2}, S_E$. Each of the 729 5-tuple optimal management designs is associated with the same probability of the corresponding weather scenario.

Then, we assume that the mapping between routing parameters and total mean cost can be well captured by a polynomial with degree of each variable less than 3. This is the minimum-degree mapping to find potential optimal solutions. Using the empirical data-based PCM approach, 3 management points for each routing variable are selected as S_A : $0.75hour, 1.25hour, 1.75hour$, P_{A1} : $0.1575, 0.2390, 0.3077$, S_w : $1.75hour, 2.5hour, 3hour$, P_{A2} : $0.8982, 0.7353, 0.5956$ and S_E : $3hour, 4.25hour, 5.5hour$, from their distributions respectively. We mentioned that our selection of PCM management points is essentially based upon the 729 typical weather scenarios instead of the entire 138611 weather ensembles to save computation. To justify the correctness of our selected management PCM points based upon the reduced weather ensemble set, we compare them with the selections obtained through Monte Carlo methods (see Figure 7.7). The PCM points for each rerouting management parameter obtained using the two methods are very close. This step results in a total $3^5 = 243$ management PCM points.

Next, we evaluate the total mean costs at these selected management PCM points to obtain the management PCM mapping. Again, for each management PCM point, we estimate the total mean cost based on the 729 weather PCM points instead of the large set of original Monte Carlo weather samples. To verify the accuracy of this estimation, we also compare the total mean backlogs associated with each of the management PCM points, obtained using the two methods. As shown in Figure 7.8, it is clear that the use of weather PCM points can estimate the total mean backlog very well despite the significant reduced simulations. Now, we are able to construct a 6-dimentional polynomial management mapping function $f(S_A, P_{A1}, S_w, P_{A2}, S_E)$ with degree of each management variable not greater than two, based upon the 243 PCM management points and the total mean cost associated with each of them.

We notice that this routing management mapping is not convex within the feasible management domain by checking its second partial derivative. Instead, we apply trust-region-reflective searching algorithm to find the best rerouting management as $S_A = 0.75hour, P_{A1} = 0.21, S_w = 2.59hour, P_{A2} = 0.75$ and $S_E = 5hour$. The corresponding total mean cost is 37.1420, which is much better than the one without management 413.96, and in fact is also better than that of the MINIT design. The results show that the delay can be effectively reduced by the rerouting initiative.

We also verify our results through a comparison with the Monte Carlo method. In particular, the Monte Carlo method finds the optimal solution to be: $S_A = 1hour, P_{A1} = 0.18, S_w = 2.5hour, P_{A2} = 0.72$ and $S_E = 5.25hour$. The total mean cost is 58. The comparison suggests that the PCM approach is able to find the near-optimal for rerouting design under weather uncertainty. The differences are again mainly caused by the simulation resolution (15mins). The PCM-based re-routing design needs much fewer simulation times (243 × 729) than that of Monte Carlo approach (50000 × 138611) due to the use of representative management PCM points and weather PCM points instead of the huge Monte Carlo ensemble space. The significant improvement of computational efficiency makes PCM a promising approach for real time management design.

As five management variables are involved in the rerouting design, it is hard to visualize the shape of 5-dimentional management design space. To demonstrate the mapping and the optimal solution in an intuitive way, we fix S_A, S_w,andS_E at their optimal values and only vary P_{A1} and P_{A2}. The resulting three dimensional mapping and the optimal solution are shown in Figure 7.9a. We similarly show the 3-dimentioanl mapping using the Monte Carlo approach in Figure 7.9b. As we use a polynomial function to approximate the real design mapping, it is generally different in shape from the true mapping; however the two mappings are similar in the likely range of parameter values. Moreover, the PCM-based approach can find the near optimal solution well.

7.6. Concluding Remarks and Future Work

This chapter introduces a novel multivariate-PCM-based approach to effectively design optimal strategic management strategies under spatiotemporally-correlated weather uncertainty. The key idea of this approach is to select representative weather scenarios based upon the understanding of spatiotemporally-correlated weather distributions; and then select representative management points to construct a low-order approximated multi-dimensional mapping between the management parameters and system performance, from which the near-optimal solutions can be achieved. In order to better illustrate this approach, we use an example with two correlated weather zones and develop in detail the design procedures for two widely-used management initiatives: MINIT and Rerouting. The performance of the PCM-based optimal management design approach is verified through a comparison with that of the Monte Carlo method. The advantages of the PCM approach are summarized below. First, compared to the Monte Carlo method to obtain optimal solutions, the multivariate PCM approach is more computationally effective, and thus more suitable for real-time applications. Second, the multivariate PCM approach produces an explicit polynomial approximation of the original system. Besides producing the optimal solution, the mapping can produce other useful insights, such as the sensitivities of design variables, the visualization of design space, etc. Despite the advantage of this approach for optimal decision-making, potential problems may arise especially when the number of uncertain parameters and design variables are large. We envision that the PCM approach can be combined with Model Predictive Control for a better design of optimal decision-making strategies under uncertainty. We will leave this direction to the future work.

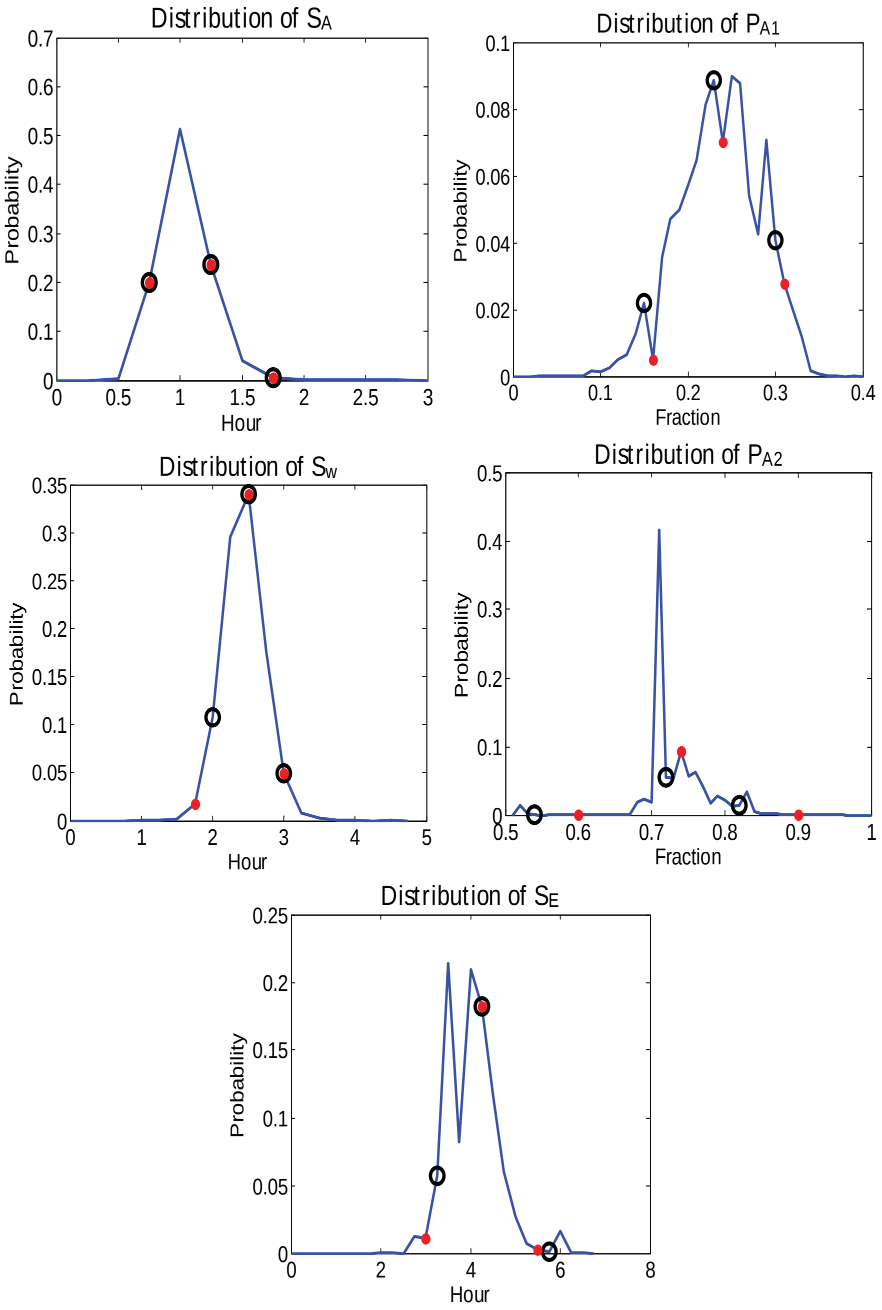


FIGURE 7.7. Comparison of 3 selected PCM points for each routing management parameter. The distribution of each management parameter is shown in blue. The red spots stand for the PCM management points obtained using 729 representative weather scenarios. The management PCM points derived using the Monte Carlo weather scenarios are marked in black circles.

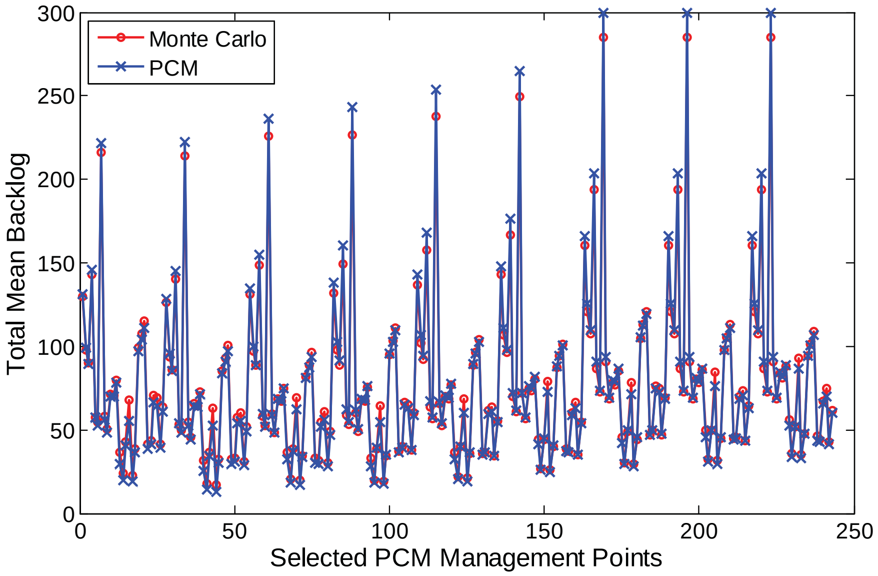
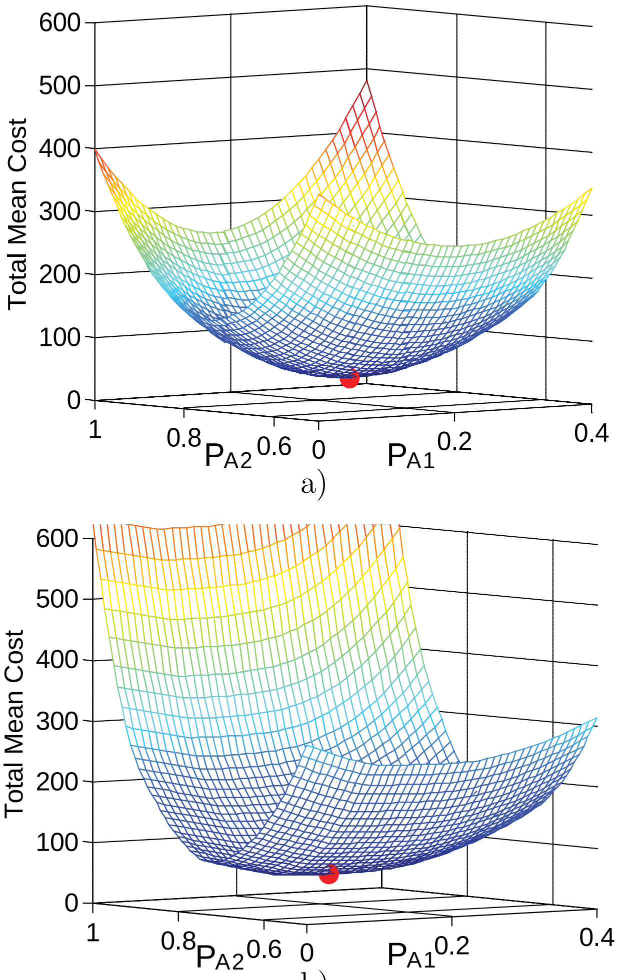


FIGURE 7.8. The comparison of total mean backlog at each selected management PCM point based upon Monte Carlo weather ensembles and the weather PCM points. We use red circles to represent the total mean backlog generated by the Monte Carlo simulation approach for each of the selected 243 PCM management points. Meanwhile, blue crosses represent total mean backlogs obtained using the PCM approach.



a)

b)

FIGURE 7.9. Mapping between two design variables and the total mean cost, generated by a) the PCM approach and b) the Monte Carlo simulation approach. The red spots are the best routing fractions whenS_A, S_w, and S_E are fixed at their optima.

CHAPTER 8

EFFECTIVE AND SCALABLE UNCERTAINTY EVALUATION FOR LARGE-SCALE COMPLEX SYSTEM APPLICATIONS

8.1. Introduction

Modern large-scale complex systems typically involve a large number of uncertain parameters, which modulate the systems' dynamics and pose significant challenges for real-time system evaluation and decision-support. For instance, the management of complex information systems requires methodologies to achieve high throughput and low latency under demand uncertainties. Similarly, strategic air traffic flow management is concerned with designing management initiatives that are robust to a wide range of weather uncertainties at a long look-ahead time. As a step toward real-time management, it is critical to develop a systematic procedure to evaluate statistical system performance in the presence of uncertain parameters. This problem can be formulated as the prediction of *output* statistics subject to a set of uncertain *input* parameters. The problem has been typically addressed using the Monte Carlo simulation method, which simulates at a *very large* set of randomly selected simulation points, and then calculates the output statistics using the simulated outputs. In large-scale complex system applications, each simulation consumes considerable computational time; as the Monte Carlo simulation method requires a very *large* number of simulations to *converge*, the method does not meet the requirement for real-time management. The uncertainty evaluation procedure needs to be efficient in time and also scalable with the number of uncertain parameters.

The Multivariate Probabilistic Collocation Method (M-PCM) was developed to effectively evaluate the output statistics of a system subject to multiple uncertain input parameters, which may or may not be correlated ([152]). Compared to the Monte Carlo method, the

M-PCM permits using a significantly reduced number of simulations to predict the *correct* mean of the original system output. Specifically, the method suggests a procedure to *smartly* select a few values for each uncertain input parameter as simulation points. Simulations evaluated at these points can identify a reduced-order mapping between input parameters and the output, from which the output statistics are readily obtained. The selection procedure is based on statistical knowledge of uncertain input parameters, such as joint probabilistic distribution functions (pdfs), historical data sets, or as simple as low-order moments, e.g., the mean and the variance. The correct predictions of mean output as well as several other important statistics suggest that the reduced-order polynomial mapping approximates the original mapping well over likely ranges of parameter values. The reduced-order mapping then facilitates further studies, including parameter sensitivity analysis, optimal decision-support under uncertainties, and the application to strategic air traffic management [79, 155, 156].

Although the M-PCM significantly reduces the number of simulations required to predict the correct mean output, it is not scalable to large-scale system applications that typically involve a large number of uncertain parameters. In particular, the number of simulations increases exponentially with the increase of the number of uncertain parameters, and thus leads to potential computational load issues for real-time applications. In this chapter, we study further reducing the number of simulations from the M-PCM designs.

Conceptually, this further reduction is possible, as the M-PCM assumes the existence of all cross-multiplication terms in system mappings. As many of these cross-terms do not exist in realistic applications, the number of mapping coefficients can be significantly reduced and thus requires a lower number of simulations to estimate. In this chapter, we investigate the selection of a subset of M-PCM points to predict the correct mean output. In addition, we note a practical numerical issue on the success of M-PCM. In particular, many system simulations have constraints on the *resolutions* of input parameters, and thus require numerical truncation of selected points for simulation. Such truncation may unfortunately fail the mean output prediction. As such, we also require the selected subset to be robust to such numerical errors.

In this chapter, we explore the use of an experimental design method, called the Orthogonal Fractional Factorial Design (OFFD) [36, 46, 102, 105] to further reduce the number of simulations from the M-PCM. Main contributions of this chapter are summarized in the following.

- *An integrated design to enhance the scalability and applicability of the M-PCM for uncertainty evaluation.* By integrating M-PCM with OFFDs, the number of simulations is significantly further reduced. We focus on the case that each parameter in the original system mapping has a degree of up to 3. We show that for an m-parameter system, the integration of M-PCM and OFFDs is able to reduce the number of simulations from 2^{2m} to the range of $\left[2^{\lceil log_2(m+1)\rceil}, 2^{m-1}\right]$, where $\lceil x \rceil$ denotes the nearest integer above the number x. We prove that the integrated M-PCM-OFFD predicts the correct mean of the original system mapping, and is the most robust to numerical errors among all designs of same the size. This study enhances M-PCM for practical uncertainty evaluation for large-scale systems.
- *A novel statistical measure and the optimality study of OFFDs.* We explore the performance of OFFDs in terms of the robustness to numerical errors for output statistics prediction, which has never been studied in the literature per knowledge of the authors. We adopt a quantitative robustness metric in the matrix theory [33], and show that the subset of simulations selected by OFFDs is optimal under this metric. This study provides new quantitative insights into the attributes of OFFDs, and broadens their application domains.

The remainder of this chapter is organized as follows. In Section 8.2, fundamentals of the M-PCM and OFFDs are introduced. In Section 8.3, the algorithm of the integrated M-PCM-OFFD is presented, and the main results on its performance are presented from two aspects: mean output prediction and robustness to numerical errors. Section 8.4 includes simulation studies on an illustrative example. Finally, a brief conclusion and a discussion of future works are included in Section 8.5.

8.2. Preliminaries

We first review fundamentals of the M-PCM and motivate our approach to further reduce the number of simulations. We then review basics of the OFFDs and discuss the feasibility of integrating the two methods.

8.2.1. M-PCM

The M-PCM was developed to effectively evaluate uncertainty for systems with multiple uncertain input parameters. In this chapter, we consider the case when these input parameters are independent; however we note that the correlated M-PCM was developed in [152]. As Theorem 1 in [152] generally shows, for a system mapping (called response surface in the experiment design literature) of m uncertain parameters with the degree k_i of each parameter x_i up to $2n_i - 1$, a total number of $2^m \prod_{i=1}^{m} n_i$ simulations are needed to uniquely determine its mapping $g(x_1, x_2, ..., x_m) = \sum_{k_1=0}^{2n_1-1} \sum_{k_2=0}^{2n_2-1} ... \sum_{k_m=0}^{2n_m-1} \Psi_{k_1,...,k_m} \prod_{i=1}^{m} x_i^{k_i}$, where $\Psi_{k_1,...,k_m} \in R$ are the coefficients. The M-PCM suggests a procedure to choose n_i simulation points for each parameter, and produces a reduced mapping $g^*(x_1, x_2, ..., x_m) = \sum_{k_1=0}^{n_1-1} \sum_{k_2=0}^{n_2-1} ... \sum_{k_m=0}^{n_m-1} \Omega_{k_1,...,k_m} \prod_{i=1}^{m} x_i^{k_i}$ with the degree of each parameter up to $n_i - 1$, where $\Omega_{k_1,...,k_m} \in R$ are the coefficients. The reduced-order mapping predicts the *correct mean output* of the original mapping.

8.2.1.1. Design Procedures

The three major steps of the M-PCM method are briefly summarized below. Please refer to [152] for the details.

Step 1: Choose simulation points. For each input parameter $x_i, i = 1, 2, \cdots, m$, find its orthonormal polynomial $h_i^{n_i}(x_i)$ of degree n_i based on the statistics of x_i, such as the pdf, historical data, or low-order moments. The roots of $h_i^{n_i}(x_i)$ are the n_i M-PCM simulation points for x_i, denoted as $x_{i(1)}, ..., x_{i(n_i)}$.

Step 2: Run simulations at selected simulation points. For each simulation point identified in Step 1, run simulation and find the associated output.

Step 3: Produce the low-order mapping. Calculate the coefficients $\Omega_{k_1,...,k_m}$ in the low-order mapping $g^*(x_1, x_2, ..., x_m)$ by

$$\begin{bmatrix} \Omega_{0,...,0} \\ \vdots \\ \Omega_{0,...,n_m-1} \\ \vdots \\ \Omega_{n_1-1,...,n_m-1} \end{bmatrix} = L^{-1} \begin{bmatrix} g(x_{1(1)}, ..., x_{m(1)}) \\ g(x_{1(1)}, ..., x_{m(2)}) \\ \vdots \\ g(x_{1(n_1)}, ..., x_{m(n_m)}) \end{bmatrix}, \text{ where}$$

$$(144) L = \begin{bmatrix} x_1^0(x_{1(1)})...x_m^0(x_{m(1)}) & \cdots & x_1^0(x_{1(1)})...x_m^{n_m-1}(x_{m(1)}) & \cdots & x_1^{n_1-1}(x_{1(1)})...x_m^{n_m-1}(x_{m(1)}) \\ x_1^0(x_{1(1)})...x_m^0(x_{m(2)}) & \cdots & x_1^0(x_{1(1)})...x_m^{n_m-1}(x_{m(2)}) & \cdots & x_1^{n_1-1}(x_{1(1)})...x_m^{n_m-1}(x_{m(2)}) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_1^0(x_{1(n_1)})...x_m^0(x_{m(n_m)}) & \cdots & x_1^0(x_{1(n_1)})...x_m^{n_m-1}(x_{m(n_m)}) & \cdots & x_1^{n_1-1}(x_{1(n_1)})...x_m^{n_m-1}(x_{m(n_m)}) \end{bmatrix}.$$

and $x_i^{k_i}(x_{i(j)})$ represents the k_i-th power of x_i evaluated at the simulation point $x_{i(j)}$.

Despite the significant computational load reduction enabled by the M-PCM, the method does not scale with the number of uncertain parameters m. In particular, $\prod_{i=1}^m n_i$

runs are still needed to predict the correct mean output. We note that if all coefficients in the low-order mapping $g^*(x_1, x_2, ..., x_m)$ are nonzero, the full set of M-PCM points is required to uniquely determine the mapping, and in turn the mean output. In realistic applications, however, many cross-terms in the mapping do not exist or have negligible effects on the output [46]. Under such assumptions, only a subset of the M-PCM points is required. In this chapter, we study using only a subset of the $\prod_{i=1}^{m} n_i$ simulations to predict the mean output. Next, we review the OFFDs, which we will use to achieve this further reduction.

8.2.2. OFFDs

Orthogonal fractional factorial designs (OFFDs) provide an approach to select a subset of experimental combinations that best estimate the main effects of single factors (or parameters) and low-order interaction effects of multiple factors on the output. Please refer to [36, 41, 46, 86, 102, 105] for the details of OFFDs.

8.2.2.1. Design Procedures

Consider an m-factor experiment, with each factor evaluated at P levels (or values). A specific OFFD is described by $P_R^{m-\gamma}$. The **fractionation constant**, $\gamma \in Z+$, indicates that a fraction of $P^{-\gamma}$ runs is selected from the full set of P^m runs [52, 102]. γ is in the range of $1 \leq \gamma \leq m - \lceil log_P(m+1) \rceil$ [46], with the upper bound determined by the minimum number of runs to estimate m main effects and the mean. γ also determines the minimum number of **generators**, which decides the effects (main effects or interactions) that are confounded together [52, 102]. The length of the shortest generator is defined as the **resolution** R, which is usually represented by Roman numerical subscript [102]. The procedures to generate the $P_R^{m-\gamma}$ OFFD are summarized in the following. We note that in statistical experiment design, the selections of γ and R need to balance the degree of tolerable confounding and OFFD sample sizes.

Step 1: Generate the $P^{m-\gamma}$ full factorial design for $m-\gamma$ factors. List all $P^{m-\gamma}$ combinations for $m-\gamma$ factors.

Step 2: Specify γ generators. The selection of generators is somewhat flexible. In principle, given γ and m, the highest resolution R is usually adopted to achieve the minimal aliasing for effect estimation [46]. We can also refer to standard designs [38] to select generators.

Step 3: Determine the levels of all other γ factors for each experimental run. The generators selected in Step 2 are used to generate the levels for all other factors.

If we view all PCM points selected from the M-PCM as a full factorial design, the OFFDs provide systematic procedures to select a subset of simulation points that breaks the curse of dimensionality. In the next section, we present the integrated algorithm, and show that it produces a low-order mapping that predicts the correct mean output of the original mapping, and is the most robust to numerical errors.

8.3. Integrated M-PCM and OFFDs

In this section, we investigate the integrated M-PCM and OFFDs that together break the curse of dimensionality for effective mean output prediction. For most of the analyses here, we assume that each of the parameters in the original system mapping $g(x_1, x_2, ..., x_m)$ is up to the degree of 3. This assumption is placed to facilitate the use of 2-level OFFDs, which have mature design procedures (and in particular the formulation of generators). In addition, we assume that uncertain input parameters are independent. We note that the correlation among parameters, if known, can be exploited to further reduce the number of simulations [72]. We leave these generalizations to the future work. We first present the integrated algorithm, and then analyze its optimality using two metrics: 1) mean output estimation, and 2) robustness to numerical errors. We note that the specific OFFD to choose is dependent on the knowledge that cross-terms have up to a certain number of parameters.

8.3.1. Algorithm Description

Consider an original system mapping of m input parameters, each with a degree up to 3. Mathematically,

$$g(x_1, x_2, ...x_m) = \sum_{k_1=0}^{3}\sum_{k_2=0}^{3}...\sum_{k_m=0}^{3}\Psi_{k_1,...,k_m}\prod_{i=1}^{m}x_i^{k_i}, \tag{145}$$

where the coefficients $\Psi_{k_1,...,k_m} \in R$. Assume that the random parameters x_1, x_2, ..., x_m follow independent distributions $f_{X_1}(x_1)$, $f_{X_2}(x_2)$, ... , and $f_{X_m}(x_m)$ respectively. In addition, assume that cross-terms involve at most τ parameters, where τ is an integer in the range of $1 \leq \tau \leq m$. In other words, $\Psi_{k_1,...,k_m} = 0$ if more than τ of $k_1, ..., k_m$ are non-zero.

The following integrated algorithm constructs a low-order mapping

$$g^*(x_1, x_2, ..., x_m) = \sum_{k_1=0}^{1}\sum_{k_2=0}^{1}...\sum_{k_m=0}^{1}\Omega_{k_1,...,k_m}\prod_{i=1}^{m}x_i^{k_i}, \tag{146}$$

where the coefficients $\Omega_{k_1,...,k_m} \in R$, and $\Omega_{k_1,...,k_m} = 0$ if more than τ of $k_1, ..., k_m$ are non-zero.

Algorithm:

Step 1: Choose 2^m M-PCM simulation points. Follow Step 1 of the M-PCM algorithm in Section 8.2.1.1 to select 2^m PCM points. Here $n_i = 2$ for all i. Check if $1 \leq \tau \leq \lceil \frac{m}{2} \rceil - 1$. If yes, move to Step 2; otherwise follow Steps 2 and 3 of the M-PCM algorithm as no simulations can be further reduced by OFFDs.

Step 2: Calculate γ_{max} to save $2^{-\gamma_{max}}$ simulations. Select $\gamma_{max} = m - \lceil log_2(\sum_{i=0}^{\tau}\binom{i}{m})) \rceil$ to save the maximum number of simulations. Here $l = \sum_{i=0}^{\tau}\binom{i}{m}$ is the number of coefficients in Equation 146.

Step 3: Select simulation subsets using the OFFD. Follow the three steps of the $2^{m-\gamma_{max}}$ OFFD algorithm in Section 8.2.2.1 to select $l_{offd} = 2^{m-\gamma_{max}}$ simulation points from the full set of 2^m simulations obtained in Step 1. These points constitute the M-PCM-OFFD simulation set.

Step 4: Run simulations. Run simulation at each of the $2^{m-\gamma_{max}}$ M-PCM-OFFD simulation points generated in Step 3.

Step 5: Produce the low-order mapping. If the number of coefficients, l, equals the number of simulation points selected using the OFFD, l_{offd}, find the coefficients in Equation 146 similar to Step 3 of the M-PCM algorithm, but with a reduced-size L matrix, denoted as the **input matrix** $L' \in R^{l\times l}$, which excludes those entries with rows representing points not selected in the reduced M-PCM-OFFD simulation set, and those columns with more than τ of $k_1, k_2, ...k_m$ being nonzero. If $l < l_{offd}$, the input matrix $L' \in R^{l_{offd}\times l}$ is not a square matrix. In this case, the coefficients can be instead found by replacing L'^{-1} with $(L'^T L')^{-1} L'^T$, according to the least square estimation [87].

We note that the ordering of entries in the L' matrix does not need to strictly follow that in Equation 144. They only need to match with the orderings of simulation points and the simulated outputs.

8.3.2. Performance of Algorithm on the Estimation of Mean Output

In this section, we show that the reduced M-PCM-OFFD simulation set in Section 8.3.1 estimates the correct mean output of the original system mapping with the degree of each parameter up to 3. We first present three lemmas. In Lemma 8.1, we show that the reduced-order mapping does not introduce additional cross-terms.

LEMMA 8.1. *Consider an original system mapping $g(x_1, x_2, ..., x_m)$ that contains cross-terms* *of at most τ parameters (Equation 145). The low-order mapping $g^*(x_1, x_2, ..., x_m)$ estimated using the M-PCM also contains cross-terms of at most τ parameters.*

Proof. According to the proofs for Theorems 1 and 2 in [152], the M-PCM recursively reduces the degree of each input parameter to produce a low-order mapping of the same mean output. As this procedure does not introduce new parameters to each cross-term, the numbers of parameters in all cross-terms in the low-order mapping do not increase. $\square$

In Lemma 8.2, we study the maximum number of simulations that can be further reduced using OFFDs, given the maximum number of parameters, τ, in cross-terms of $g(x_1, x_2, ..., x_m)$, or equivalently $g^*(x_1, x_2, ..., x_m)$ according to Lemma 8.1.

Lemma 8.2. *Consider the low-order mapping $g^*(x_1, x_2, ..., x_m)$ (Equation 146) estimated using the M-PCM, which contains cross-terms of at most τ parameters. An OFFD design can further reduce the number of simulations if $1 \leq \tau \leq \lceil \frac{m}{2} \rceil - 1$. The maximum fraction of simulations that can be reduced is $2^{-\gamma_{max}}$ using the $2^{m-\gamma_{max}}$ OFFD, where $\gamma_{max} = m - \lceil log_2(\sum_{i=0}^{\tau} \binom{i}{m}) \rceil$.*

Proof. The M-PCM produces 2^m simulation points. Let us first prove that an OFFD design can further reduce the number of simulations if $1 \leq \tau \leq \lceil \frac{m}{2} \rceil - 1$. As OFFDs reduce the number of simulations at least by half, the number of non-zero parameters $\sum_{i=0}^{\tau} \binom{i}{m}$ in $g^*(x_1, x_2, ..., x_m)$ must be less than or equal to 2^{m-1} to produce the same low-order mapping $g^*(x_1, x_2, ..., x_m)$. Note that $2^{m-1} = \sum_{i=0}^{\frac{m-1}{2}} \binom{i}{m}$ when m is odd and $\sum_{i=0}^{\frac{m}{2}-1} \binom{i}{m} < 2^{m-1} < \sum_{i=0}^{\frac{m}{2}} \binom{i}{m}$ when m is even. The maximum of τ thus satisfies

$$max(\tau) = \begin{cases} \frac{m-1}{2} & \text{if } m \mod 2 \equiv 1 \\ \frac{m}{2} - 1 & \text{if } m \mod 2 \equiv 0 \end{cases} = \lceil \frac{m}{2} \rceil - 1. \tag{147}$$

As τ is an integer greater than or equal to 1, it needs to be in the range of $1 \leq \tau \leq \lceil \frac{m}{2} \rceil - 1$, such that an OFFD can further reduce simulation points without altering the low-order mapping $g^*(x_1, x_2, ..., x_m)$.

Now we prove that the maximum fraction of simulations that can be reduced is achieved using the $2^{m-\gamma_{max}}$ OFFD, where $\gamma_{max} = m - \lceil log_2(\sum_{i=0}^{\tau} \binom{i}{m}) \rceil$. As the number of simulations must be larger than or equal to the number of parameters,

$$\gamma_{max} = max\{\gamma \mid 2^{m-\gamma} \geq \sum_{i=0}^{\tau} \binom{i}{m}\} = m - \lceil log_2(\sum_{i=0}^{\tau} \binom{i}{m}) \rceil. \tag{148}$$

□

In the next lemma, we prove that the matrix L' is full column rank. In this process, we show the general QR decomposition expression of L'. This lemma is central to the rest of the development in this chapter, as it establishes the direct relationship between the OFFD design table (captured by Q) and the input matrix L' which is used for our study of mapping construction and mean output prediction.

LEMMA 8.3. *The input matrix* $L' \in R^{l_{offd} \times l}, l_{offd} \geq l$ *constructed by the integrated M-PCM and OFFD is full column rank, and can be represented using the QR decomposition [56]*

$$L' = QU, \tag{149}$$

where $Q \in R^{l_{offd} \times l}$ *is an orthogonal matrix (i.e.* $Q^T Q = I$*) of the form:*

$$Q = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 & \cdots & \boldsymbol{q}_l \end{bmatrix} = \frac{1}{\sqrt{l_{offd}}} \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 & \cdots & \boldsymbol{v}_l \end{bmatrix}. \tag{150}$$

Here $\boldsymbol{q}_i \in R^{l_{offd} \times 1}$ *is the orthogonal basis with* $\boldsymbol{q}_i^T \boldsymbol{q}_j = 0$ *if* $i \neq j$ *and* $\|\boldsymbol{q}_i\|_2 = 1$*, where* $i, j \in \{1, 2, \cdots, l\}$*. Assume that* $x_{k(2)} > x_{k(1)}$ *WLOG.* $\boldsymbol{v}_i = \frac{1}{\sqrt{l_{offd}}} \boldsymbol{q}_i$ *contains entries of* ± 1*, and is obtained by replacing each entry* $x_{k(1)}$ *in the i-th column of* L' *with* -1 *and* $x_{k(2)}$ *with* 1*, where k is the index of input parameters. Alternatively, it is obtained by replacing '-' and*

'+' in the OFFD design table by '-1' and '+1' respectively, and adding an all '1' vector to the left. $U \in R^{l \times l}$ *is an upper triangular matrix, with the i-th diagonal entry* U_{ii} *expressed as*

$$U_{ii} = \begin{cases} \sqrt{l_{offd}} & if\ i = 1 \\ \frac{\sqrt{l_{offd}}}{2^{\xi_i}} \prod_{k \in S_i} \Delta x_k & if\ i \neq 1 \end{cases} \tag{151}$$

where $\Delta x_k = x_{k(2)} - x_{k(1)}$*,* $S_i \subseteq \{1, 2, \cdots, m\}$ *includes all the indices of input parameters in the i-th column of* L'*, and* ξ_i *is the size of* S_i*.*

Proof. The QR decomposition follows the recursive Gram-Schmidt procedure [56]. It is omitted due to the limited space. It is clear from the expressions of Q and R that L' is full column rank, as Q is invertible ($Q^{-1} = Q^T$) and the determinant of U is $\prod_i U_{ii} \neq 0$. □

Lemmas 8.1, 8.2 and 8.3 and Theorem 1 and 2 in [152] directly lead to the theorem on the performance of the integrated M-PCM-OFFD in terms of the correctness of predicting the mean output of the original mapping.

Theorem 8.4. *The low-order mapping* $g^*(x_1, x_2, ..., x_m)$ *(Equation 146) using the integrated M-PCM and* $2^{m-\gamma_{max}}$ *OFFD predicts the correct mean output of the original mapping, i.e.,*

$$E[g(x_1, x_2, ..., x_m)] = E[g^*(x_1, x_2, ..., x_m)]. \tag{152}$$

The number of simulations reduces from 2^{2m} *to* $2^{m-\gamma_{max}}$*, where* $\gamma_{max} = m - \lceil log_2(\sum_{i=0}^{\tau} \binom{i}{m}) \rceil$*. The maximum reduction is* $2^{2m} - 2^{\lceil log_2(m+1) \rceil}$*, and is achieved when* $\tau = 1$*.*

Proof. Theorem 1 and 2 in [152] suggests that the reduced-order mapping $g^*(x_1, x_2, ..., x_m)$ produced by the M-PCM predicts the correct mean output of $g(x_1, x_2, ..., x_m)$. Lemmas 8.1, 8.2 and 8.3 guarantee that the reduced M-PCM-OFFD simulation set produces the same mapping $g^*(x_1, x_2, ..., x_m)$. The result is then straightforward. □

8.3.3. Performance of the Algorithm on the Robustness to Numerical Errors

In this section, we study the robustness of the integrated design to numerical errors. We first introduce the robustness metric and formulate the problem in Section 8.3.3.1. We then show the optimality of the integrated design using this metric in Section 8.3.3.2.

8.3.3.1. Metric and Problem Formulation

Recall that the integrated algorithm involves the calculation of L'^{-1} or $(L'^T L')^{-1} L'^T$. This inversion is only feasible when L' is full column rank. In Lemma 8.3, we have shown that an OFFD guarantees that L' is full column rank. In this section, we further explore the computational feasibility by noticing that parameter resolutions of simulation software [152] and computational limitations of computing devices [64, 67] may unfortunately *fail* this calculation. In particular, when L' is close to losing column rank, a small disturbance introduced by the aforementioned numerical errors may easily push L' to lose rank. In addition, even if such L' under a disturbance does not directly lose rank, the correctness of L'^{-1} becomes sensitive to small perturbations [35]. In order to facilitate the inversion and minimize the impact of numerical errors, L' needs to have a *large margin to rank loss.*

Multiple metrics exist in the literature to measure the margin to invertibility, including the widely used *condition number* (the ratio between the largest eigenvalue to the smallest eigenvalue) [22]. Here we use a metric based on the perturbation theory ([67], [64]). Specially, the **full-column-rank margin** for the matrix L' to rank loss, $D(L')$, is measured by the norm of the smallest perturbation matrix to make L' lose rank. Here we use the Frobenius norm ("$|| \, ||_F$"), calculated by summing the squares of all its elements, and then taking a square root of the sum [69]:

$$D(L') = min\{\|e\|_F \mid rank(L' + e) < l\} \tag{153}$$

where $e \in R^{l_{offd} \times l}$ is a perturbation matrix.

Lemma 8.2 suggests that all simulation subsets of size $2^{m-\gamma_{max}}$ can predict the correct mean output, provided that the input matrix of the design, L', is full column rank. We show in the next section that when $\tau = 1$, the L' matrix of the OFFD design, denoted as L'_{offd} thereafter, has the largest margin to rank loss among all designs of the size $2^{m-\gamma_{max}}$. The results can be extended to the general case ($\tau \geq 1$) through a more complicated analysis and hence is ignored here for clarity.

8.3.3.2. Optimal Robustness of the Integrated M-PCM-OFFD

In this section, we study the robustness of the integrated M-PCM-OFFD to numerical errors for system mappings of m uncertain parameters and $\tau = 1$. Lemma 8.5 calculates the full-column-rank margin of the integrated M-PCM-OFFD. Theorem 8.6 shows that the OFFD produces the largest margin among all subsets of the same size.

LEMMA 8.5. *Consider an original system mapping $g(x_1, x_2, ...x_m)$ (Equation 145) with $\tau = 1$. The integrated M-PCM and $2^{m-\gamma_{max}}$ OFFD has the following full-column-rank margin:*

$$D(L'_{offd}) = \frac{\sqrt{l_{offd}}}{2} min\{\Delta x_1, \Delta x_2, \cdots, \Delta x_m\} \tag{154}$$

where $\gamma_{max} = m - \lceil log_2(\sum_{i=0}^{\tau} \binom{i}{m})) \rceil$.

PROOF. According to Lemma 8.3, L'_{offd} is full column rank and can be expressed as a multiplication of an orthogonal matrix Q and an upper triangular matrix U. As Q is full column rank, the rank of L'_{offd} is solely determined by U. Furthermore, as U is upper triangular, its determinant is the multiplication of all diagonal entries as shown below:

$$det(U) = \sqrt{l_{offd}} \cdot \prod_{i=2}^{l} (\frac{\sqrt{l_{offd}}}{2^{\xi_i}} \prod_{j \in S_i} \Delta x_j) \neq 0. \tag{155}$$

Now we find the minimum $\|e\|_F$ to make $L' + e$ lose rank, according to the definition of full-column-rank margin in Equation 153. We use $e_{x_{i(j)}}$ to represent the perturbation to $x_{i(j)}$ and $\hat{x}_{i(j)} = x_{i(j)} + e_{x_{i(j)}}$ to represent the corrupted parameter value. Similar to L'_{offd}, $L'_{offd} + e$ can also perform a QR decomposition, and the determinant of the upper triangular matrix is $det(\hat{U}) = \sqrt{l_{offd}} \cdot \prod_{i=2}^{l}(\frac{\sqrt{l_{offd}}}{2^{\xi_i}} \prod_{j \in S_i} \Delta\hat{x}_j)$, where $\Delta\hat{x}_j = \hat{x}_{j(2)} - \hat{x}_{j(1)}$. Clearly, $L'_{offd} + e$ loses rank if and only if at least one of $\Delta\hat{x}_i = 0, i \in \{1, 2, \cdots, m\}$. In the case of $\Delta\hat{x}_1 = 0$, we have $\Delta\hat{x}_1 = \hat{x}_{1(2)} - \hat{x}_{1(1)} = (x_{1(2)} + e_{x_{1(2)}}) - (x_{1(1)} + e_{x_{1(1)}}) = 0$ and therefore $e_{x_{1(1)}} = e_{x_{1(2)}} + x_{1(2)} - x_{1(1)} = e_{x_{1(2)}} + \Delta x_1$. As a consequence,

$$
\begin{aligned}
(156) \qquad \|e\|_F &= \sqrt{(\frac{l_{offd}}{2}e^2_{x_{1(1)}} + \frac{l_{offd}}{2}e^2_{x_{1(2)}}) + (\frac{l_{offd}}{2}e^2_{x_{2(1)}} + \frac{l_{offd}}{2}e^2_{x_{2(2)}}) + \cdots + (\frac{l_{offd}}{2}e^2_{x_{m(1)}} + \frac{l_{offd}}{2}e^2_{x_{m(2)}})} \\
\geq \quad & \sqrt{(\frac{l_{offd}}{2}e^2_{x_{1(1)}} + \frac{l_{offd}}{2}e^2_{x_{1(2)}})} = \sqrt{\frac{l_{offd}}{2}[(e_{x_{1(2)}} + \Delta x_1)^2 + e^2_{x_{1(2)}}]} = \sqrt{l_{offd}(e_{x_{1(2)}} + \frac{\Delta x_1}{2})^2 + \frac{l_{offd}}{4}\Delta x_1^2} \\
\geq \quad & \sqrt{\frac{l_{offd}}{4}\Delta x_1^2} = \frac{\sqrt{l_{offd}}}{2}\Delta x_1
\end{aligned}
$$

The equality holds when $e_{x_{1(1)}} = \frac{1}{2}\Delta x_1$, $e_{x_{1(2)}} = -\frac{1}{2}\Delta x_1$, and $e_{x_{j(1)}} = e_{x_{j(2)}} = 0$ for all $j \neq 1$. Similarly, we obtain $\|e\|_F \geq \frac{\sqrt{l_{offd}}}{2}\Delta x_2, \cdots, \frac{\sqrt{l_{offd}}}{2}\Delta x_m$. As such, $D(L'_{offd}) = \frac{\sqrt{l_{offd}}}{2} min\{\Delta x_1, \Delta x_2, \cdots, \Delta x_m\}$, and the minimum $\frac{\sqrt{l_{offd}}}{2}\Delta x_i$ is achieved when $\Delta x_i \leq \Delta x_j$ for all $j \neq i$, $e_{x_{i(1)}} = \frac{1}{2}\Delta x_i$, $e_{x_{i(2)}} = -\frac{1}{2}\Delta x_i$, and $e_{x_{j(1)}} = e_{x_{j(2)}} = 0$ for all $j \neq i$. □

THEOREM 8.6. *Consider an original system mapping $g(x_1, x_2, ...x_m)$ (Equation 145) with $\tau = 1$. From the 2^m M-PCM simulation points, the simulation subset selected by the M-PCM-OFFD has the largest full-column-rank margin among all subsets of $2^{m-\gamma_{max}}$ simulations points, where $\gamma_{max} = m - \lceil log_2(\sum_{i=0}^{\tau}\binom{i}{m})\rceil$. Mathematically,*

$$
(157) \qquad max(D(L')) = D(L'_{offd}).
$$

PROOF. We first construct the input matrix, L', from any subset of $2^{m-\gamma_{max}}$ simulation points selected from the 2^m M-PCM simulation points, and then show that the input matrix of the OFFD, L'_{offd}, has the largest full-column-rank margin.

Through simple row operations, the L' matrix constructed using any $2^{m-\gamma_{max}}$ M-PCM points can be transformed to an upper triangular matrix, where the first diagonal entry is 1, and the $(k+1)$-th diagonal entry is a multiple of Δx_k, i.e., $\lambda_{(k+1)(k+1)}\Delta x_k$, where $\lambda_{kk} \in Z$ and $k \in \{1, 2, \cdots, m\}$. As $\Delta x_k \neq 0$, any $\lambda_{kk} = 0$ will lead to $D(L') = 0$.

Now let us find the minimum $||e||_F$ to make $L' + e$ lose rank. The case that L' is not full rank is trivial, as e is the null matrix and $D(L') = 0$. When L' is full rank, the same procedures to calculate $D(L'_{offd})$ in the proof of Lemma 8.5 leads to

$$\begin{aligned} (158)\ D(L') &= min\{\sqrt{[c_1 e^2_{x_{1(1)}} + (l_{offd} - c_1)e^2_{x_{1(2)}}] + \cdots + [c_m e^2_{x_{m(1)}} + (l_{offd} - c_m)e^2_{x_{m(2)}}]}\} \\ &= min\{\sqrt{\frac{c_1(l_{offd} - c_1)}{l_{offd}}}\Delta x_1, \cdots, \sqrt{\frac{c_m(l_{offd} - c_m)}{l_{offd}}}\Delta x_m\} \end{aligned}$$

where c_i is the number of $x_{i(1)}$ in the $(i+1)$-th column of L'. The minimum at $\sqrt{\frac{c_i(l_{offd}-c_i)}{l_{offd}}}\Delta x_i$ is achieved, when $e_{x_{i(1)}} = \frac{(l_{offd}-c_i)\Delta x_i}{l_{offd}}$, $e_{x_{i(2)}} = -\frac{c_i \Delta x_i}{l_{offd}}$, and for all $j \neq i, j \in \{1, 2, \cdots, m\}$, we have $\sqrt{\frac{c_i(l_{offd}-c_i)}{l_{offd}}}\Delta x_i \leq \sqrt{\frac{c_j(l_{offd}-c_j)}{l_{offd}}}\Delta x_j$, and $e_{x_{j(1)}} = e_{x_{j(2)}} = 0$.

Since $\sqrt{\frac{c_i(l_{offd}-c_i)}{l_{offd}}} = \sqrt{-\frac{1}{l_{offd}}(c_i - \frac{l_{offd}}{2})^2 + \frac{l_{offd}}{4}} \leq \frac{\sqrt{l_{offd}}}{2}$, we have $\sqrt{\frac{c_i(l_{offd}-c_i)}{l_{offd}}}\Delta x_i \leq \frac{\sqrt{l_{offd}}}{2}\Delta x_i$. Equation 158 can then be further simplified to

$$D(L') \leq \frac{\sqrt{l_{offd}}}{2} min\{\Delta x_1, \Delta x_2, \cdots, \Delta x_m\} = D(L'_{offd}) \tag{159}$$

The equality is achieved by an OFFD. ☐

The robustness optimality of the integrated algorithm is brought by the balance and orthogonality of OFFDs. The orthogonality (i.e., the symbolic multiplication of each pair of columns in the design table sums up to 0) guarantees the full rank of Q (as shown in in Equation 150) and thus the invertability of L'_{offd}. Moreover, the balance property (i.e., each level is evaluated the same number of times for each factor) guarantees the maximal perturbation to spoil the invertability of L'_{offd} (as shown in Equation 159).

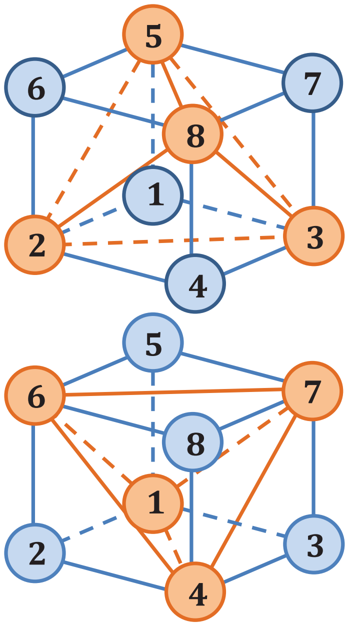
In this section, we use a simulation study to illustrate the procedures and properties of the integrated design presented in this chapter.

Consider an original mapping of the form $g(x_1, x_2, x_3) = x_1^3 + x_1^2 + x_1 + x_2^3 + x_2^2 + x_2 + x_3^3 + x_3^2 + x_3 + 1$, where x_1, x_2, x_3 are the three uncertain input parameters following independent distributions. In particular, x_1 follows an exponential distribution of $f_{X_1}(x_1) = 2e^{-2x_1}$; x_2 follows a uniform distribution of $f_{X_2}(x_2) = \frac{1}{15}, 5 \leq x_2 \leq 20$; and x_3 also follows a uniform distribution of $f_{X_3}(x_3) = \frac{1}{5}, 5 \leq x_3 \leq 10$. The output mean is $E[g(x_1, x_2, x_3)] = \iiint g(x_1, x_2, x_3) f_{X_1}(x_1) f_{X_2}(x_2) f_{X_3}(x_3) dx_1 dx_2 dx_3 = 3381.1$. Identifying all coefficients requires $4^3 = 64$ simulations.

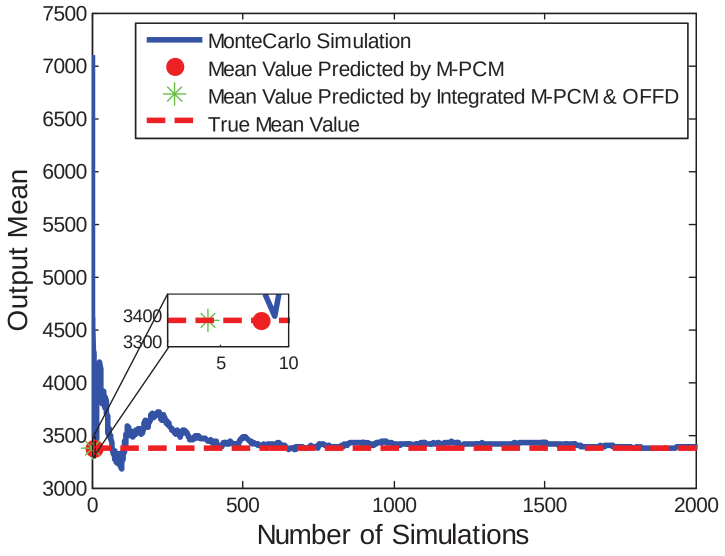
Now let us choose only 4 simulations, using the integrated M-PCM-OFFD. We first choose 8 M-PCM points based upon the pdf of each parameter. The 8 simulation points are $p_1 = (0.2929, 8.1699, 6.0566)$, $p_2 = (1.7071, 8.1699, 6.0566)$, $p_3 = (0.2929, 16.8301, 6.0566)$, $p_4 = (1.7071, 16.8301, 6.0566)$, $p_5 = (0.2929, 8.1699, 8.9434)$, $p_6 = (1.7071, 8.1699, 8.9434)$, $p_7 = (0.2929, 16.8301, 8.9434)$, and $p_8 = (1.7071, 16.8301, 8.9434)$. We then use 2^{3-1}_{III} OFFD (as the design table and 3-D cube show in Figure 8.1a,b) to select 4 M-PCM points, which are $\{p_2, p_3, p_5, p_8\}$ or $\{p_1, p_4, p_6, p_7\}$. The input matrix for the first design is $L'_{offd} = \begin{bmatrix} 1 & x_{1(1)} & x_{2(1)} & x_{3(2)} \\ 1 & x_{1(2)} & x_{2(1)} & x_{3(1)} \\ 1 & x_{1(1)} & x_{2(2)} & x_{3(1)} \\ 1 & x_{1(2)} & x_{2(2)} & x_{3(2)} \end{bmatrix}$. After that, we run simulations to evaluate $g(x_1, x_2, x_3)$ at these 4 simulation points and estimate the coefficients of the low-order mapping $g^*(x_1, x_2, x_3) = -4442.2 + 6.5x_1 + 513.5x_2 + 186.8x_3$. For illustration purpose,

	x_1	x_2	x_3	y
1	−	−	+	y_5
2	+	−	−	y_2
3	−	+	−	y_3
4	+	+	+	y_8

	x_1	x_2	x_3	y
1	−	−	−	y_1
2	+	+	−	y_4
3	−	+	+	y_7
4	+	−	+	y_6



(a) (b)



(c)

FIGURE 8.1. a) 2^{3-1}_{III} OFFD design table. b) 2^{3-1}_{III} OFFD with each point represented at the vertex of a 3-D cuboid. c) Comparison of the number of simulations needed to predict the correct output mean.

$U = \begin{bmatrix} 2 & * & * & * \\ 0 & 1.4142 & * & * \\ 0 & 0 & 8.6602 & * \\ 0 & 0 & 0 & 2.8868 \end{bmatrix}$ and $Q = \frac{1}{2}\begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$ which can be directly obtained from the OFFD design table.

The output mean of $g^*(x_1, x_2, x_3)$ is $E[g^*(x_1, x_2, x_3)] = \iiint g^*(x_1, x_2, x_3) f_{X_1}(x_1) f_{X_2}(x_2) f_{X_3}(x_3) dx_1 dx_2 dx_3 = 3381.1$, precisely the same as the original output mean. For comparison, we also use the Monte Carlo simulation to find the mean output. The number of simulations and associated mean output using these three methods are compared and shown in Figure 8.1c.

To check the robustness of the integrated design to numerical errors, we calculate the full-column-rank margin of the input matrix $L' \in R^{4\times4}$, and compare it with those of other designs. According to Lemma 8.5, we find $D(L'_{offd}) = min\{\Delta x_1, \Delta x_2, \Delta x_3\} = 1.4142$, where $\Delta x_1 = 1.4142$, $\Delta x_2 = 8.6602$ and $\Delta x_3 = 2.8868$.

For all simulation subsets of the same size, the margin $D(L')$ takes one of the following three values $\{0, 0.8660, 1.4142\}$. Therefore, the OFFD design is the most robust to numerical errors.

8.5. Concluding Remarks and Future Work

We developed an effective uncertainty evaluation method for large-scale complex systems with a large number of uncertain input parameters. The integrated M-PCM and OFFDs significantly reduces the number of simulations, whereas maintaining the statistical prediction performance of the M-PCM. Specially, for an original system mapping of m parameters with each parameter up to the degree of 3, the reduced-order mapping produced using the integrated method precisely predicts the mean output of the original system mapping, and reduces the number of mappings from 2^{2m} to at most $2^{\lceil log_2(m+1) \rceil}$. We also showed that the integrated design is the most robust to numerical errors, making it of practical use for simulations with constraints on parameter resolutions. The development in this chapter also provided new interpretations of the optimality of OFFDs, and gave rise to broad new usage of OFFDs for system mapping estimation and uncertainty evaluation. In the future work, we will generalize the degree of uncertain input parameters by exploring multiple-factor OFFDs and also exploit parameter dependency to further reduce the number of simulations required.

CHAPTER 9

A JUMP LINEAR APPROACH BASED SENSITIVITY STUDY FOR OPTIMAL AIR TRAFFIC FLOW MANAGEMENT UNDER WEATHER UNCERTAINTY

9.1. Introduction

Strategic air traffic management is concerned with coordinating traffic flow at strategic time frame. Management at this time scale is challenging due to the existence of uncertain convective weather. When convective weather is likely to occur, mild alteration of either traffic demand or airspace resource may have significant impact on the performance of the Nation Airspace System (NAS). Thus, it is necessary for us to study the sensitivity of air traffic congestion subject to those disturbances. From this study, not only can we understand the sources of disturbances that have significant impact on the traffic congestion, but also we are able to design appropriate management planning such as demand allocation or airspace reconfiguration.

There are several previous efforts relevant to our investigation in this chapter. A sensitivity study of the NAS performance subject to various disturbances shows that it can aid to design appropriate management planning [131]. However, this work does not take the weather impact on the air traffic into consideration. In our previous work, we develop a systematical jump linear approach to efficiently evaluate the number of aircraft delayed due to uncertain weather events [151]. In this chapter, we utilize this approach as a tool to identify the relationship between the sensitivity of total number of aircraft delayed during a time span of interest to the demand variations. Moreover, we want to show that well-designed inflows entering a region under weather uncertainty satisfy a special sensitivity structure. We envision that this study can be generalized to the NAS-level so that it can improve the airspace resource utility.

The main contributions of this chapter can be summarized as follows. 1) We use jump linear approach to derive total number of aircraft delayed in the weather-impacted region across time. The sensitivity of the performance with respect to its local disturbance is also discussed. 2) From the analysis, we are able to aid optimal flow distribution subject to unexpected weather events. In a weather-impacted region, we consider the total incoming flow with a fixed rate is apportioned into multiple flows to efficiently utilize the airspace. Our purpose is to determine the best flow rate separation such that the total number of aircraft delayed across time will be minimized. We show that the best distribution planning can be verified from the perspective of sensitivity analysis of the congestion associated with each of the separated incoming flows.

The content of this chapter is organized as follows. In section 9.2, we introduce the jump linear model and formulate our problem based on it. In section 9.3, we investigate the optimally distributed inflow rate design problem from sensitivity analysis perspective. Section 9.4 is the conclusion and work in progress.

9.2. The Jump Linear Model and Problem Formulation

In this section, we first introduce the air traffic system modeled as a jump linear system. We then continue with the derivation of the transient traffic congestion over a time span based upon our proposed modeling framework. The sensitivity of the total traffic congestion with respect to system parameter is also derived afterwards. Finally, we mathematically formulate our design problem: optimal flow distribution problem under weather uncertainties. We want to point out that our model is an abstract model but it is reasonable due to the huge uncertainty during strategic planning timeframe.

9.2.1. Jump Linear Model: A Merged Model of Air Traffic and Weather Impact

In order to introduce the jump linear representation of our air traffic system clearly, we first begin with the discrete queuing model (saturation model [130]) as the jump linear model is a further abstraction of that model. To reduce the dimensions of the complex air traffic system, we consider the traffic in the flow level instead of individual aircraft.

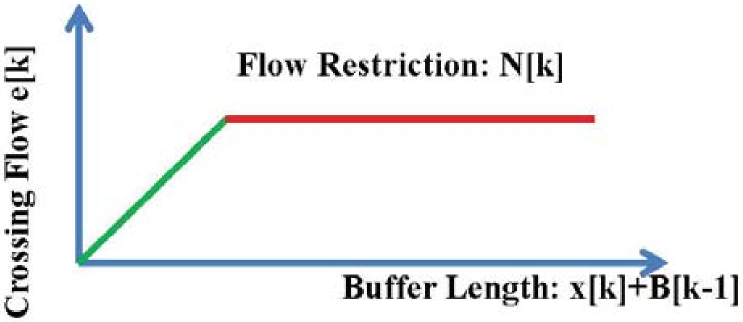


FIGURE 9.1. Illustration of saturation model

In particular, we assume a stream of flow x[k] entering a region subject to capacity constraint N[k] at each time step k. As the distance between each two aircraft cannot be less than a safety distance, we use N[k] to define the maximum number of aircrafts can be processed in that region at time k. The cross flow e[k] cannot exceed that constraint. The number of aircraft do not allow to pass the region are assumed to be accumulated at the boundary of the region. We define them as the backlog (B[k]) to measure the traffic congestion. The dynamic of the saturation model can be expressed in the following equations:

$$e[k] = \begin{cases} b[k-1], & (b[k-1] \leq N_c)) \\ N_c, & (b[k-1] \geq N_c)) \end{cases} \tag{160}$$

$$b[k] = b[k-1] + x[k] - e[k] \tag{161}$$

$$B[k] = b[k-1] - e[k] \tag{162}$$

Here, the buffer length b[k] is the number of aircraft waiting to cross the region, the capacity constraint N[k] is driven by weather condition. Because when bad weather present, the safety distance between two aircrafts will increase such that the capacity in that region will decrease accordingly. A simple illustration of the saturation model is shown in Figure 9.1. We notice that the performance of cross flow represented by a green solid line and a red solid line is piecewise linear as demonstrated in Figure 9.1. As such, we approximate e[k] as a linear function of b[k-1] as shown below:

$$e[k] = ab[k-1] + c \tag{163}$$

a and c is a pair of parameters relevant to x[k] and N[k].

In another aspect, we model the duration of capacity reduction caused by weather event as finite state Markov Chain [112, 149, 150]. Each state is associated with x[k]and N[k]. As such, the parameters a and c of Equation 163 are changing according to the states of the weather Markov Chain. Thus, the dynamic of the model can then be expressed as a jump linear system below:

$$e[k] = a(q[k])b[k-1] + c(q[k]) \tag{164}$$

$$b[k] = b[k-1] + x[k] - e[k] \tag{165}$$

$$B[k] = b[k-1] - e[k] \tag{166}$$

Here, we introduce an indicator vector $q[k] \in R^n$ has all the entries equal to 0 except one entry equals to 1, indicating the state of the weather Markov Chain at time step k. $a(q[k])$ and $c(q[k])$ represent the values of the parameters a and c associated with that state. For example, a jump linear model corresponding to the saturation model introduced in Figure 9.1 can be shown in Figure 9.2.

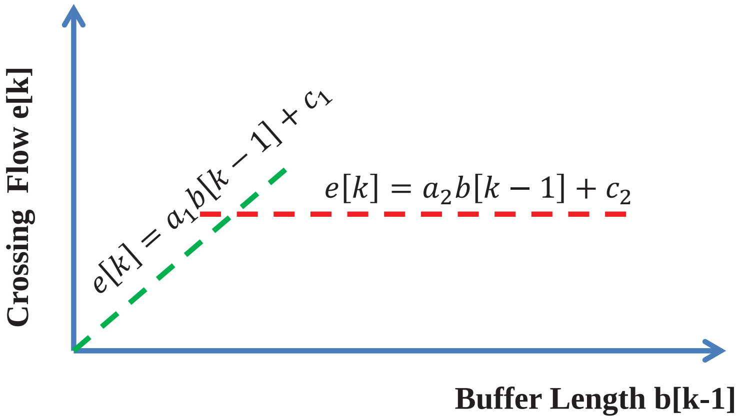


Figure 9.2. A jump linear approximation of the saturation model in Figure 1. The parameters a_1 and c_1 are associated with the states when the capacities are greater than the buffer length. The parameters a_2 and c_2 are associated with the states when the capacities are less or equal to the butter length.

9.2.2. Sensitivity of Total Mean Backlog

Based upon the jump linear representation of our model, the backlog at each time step kis easy to track. Our focus is on the sensitivity of the total mean backlog during a time span with respect to possible variation of the inflow rate of x[k]. Let us show the details of deriving the total mean backlog at first and then, the its sensitivity can naturally obtained.

First, the dynamic of the transient backlog can be derived in an explicit recursion form by applying a little bit algebra on Equations 164-166. In particular,

$$B[k+1] = (1 - a(q[k]))(B[k] + x[k]) - c(q[k]) \tag{167}$$

We then directly use the result in our previous work [130] to compute the total mean backlog over a time span. To be more specific, we first introduce a vector $\sigma[k] = q[0] \otimes \begin{bmatrix} B[k] \\ 1 \end{bmatrix}$ to capture the information of both the weather condition and backlog. The dynamic of $E[\sigma[k]]$ can be obtained as follows:

$$E[\sigma[k+1]] = {P_w}' \otimes \begin{bmatrix} (1-a[q[k]]) & (1-a[q[k]])u[k] - c[q[k]] \\ 0 & 1 \end{bmatrix} E[\sigma[k]] \tag{168}$$

Here, P_w is the transpose of transition matrix of weather Markov chain and $u[k]$ is the mean of $x[k]$. Thus, with the given initial condition $q[0]$ and $B[0]$, the mean backlog $E[B[k]]$ equals $1_{1\times l}E[\sigma[k]] - 1$. $1_{1\times l}$ is $a_1 \times l$ vector with all entries equal to one. l is the length of $\sigma[k]$. Now, we are ready to express our total mean backlog during a time span $[0, k_p]$. As Equation 168 can be rewritten into the following form:

$$E[\sigma[k]] = \left({P_w}' \otimes \begin{bmatrix} (1-a[q[k]]) & (1-a[q[k]])u[k] - c[q[k]] \\ 0 & 1 \end{bmatrix} \right)^k \left(q[0] \otimes \begin{bmatrix} B[0] \\ 1 \end{bmatrix} \right) \tag{169}$$

Then, the total mean backlog $T_{E(B)}$ from time 0 to k_p can be computed as:

$$T_{E(B)} = \sum_{i=0}^{k_p} 1_{1\times l} \left(\left({P_w}' \otimes \begin{bmatrix} (1-a[q[k]]) & (1-a[q[k]])u[k] - c[q[k]] \\ 0 & 1 \end{bmatrix} \right)^i \left(q[0] \otimes \begin{bmatrix} B[0] \\ 1 \end{bmatrix} \right) \right) - (K_p + 1) \tag{170}$$

To facilitate our analysis, in this chapter, we assume the inflow rate $u[k]$ is a time invariant constant u. Without loss of generality, we use a m_d-state Markov chain to represent the variation of weather condition. The subscript d is associated with weather intensity, which leads to capacity reduction. In the first m_1 states, the capacity reduction level 1 is determined by the system parameters a_1 and c_1. The dynamic of the transient backlog can be written as:

$$B[k+1] = (1-a_1)(B[k]+x[k]) - c_1 \tag{171}$$

Likewise, the transient backlog starts from state $m_{i-1}+1$ to m_i can be represented as:

$$B[k+1] = (1-a_i)(B[k]+x[k]) - c_i \tag{172}$$

where the capacity reduction level $i \in [2,d]$ are associated with the parameters a_i and c_i. As we use Markov Chain to model the weather evolution from bad condition to good condition, the d^{th} capacity reduction level is the normal capacity. a_d and c_d are the corresponding parameters. To facilitated our analysis, we define $\frac{(1-a_i)}{(1-a_d)} = f_i$ and $\frac{c_i}{c_d} = h_i$ for $i \in [1,d]$, where f_i and h_i are well-known ratios to scale different capacity reduction levels. The dynamic of the backlog can be rewritten as:

$$B[k+1] = f_i(1-a_d)(B[k]+x[k]) - h_i c_d, \tag{173}$$

Then, we apply Theorem 51 and Corollary 52 in [23] to simplify the first term of Equation 169. In particular,

$$\begin{aligned}
&\left({P_w}' \otimes \begin{bmatrix} (1-a[q[k]]) & (1-a[q[k]])u[k]-c[q[k]] \\ 0 & 1 \end{bmatrix} \right)^k \qquad (174)\\
= & \left(Per(2,m_d)^{-1} \left(\begin{bmatrix} (1-a[q[k]]) & (1-a[q[k]])u[k]-c[q[k]] \\ 0 & 1 \end{bmatrix} \otimes {P_w}' \right) Per(2,m_d) \right)^k \\
= & \; Per(2,m_d)^{-1} \left(\begin{bmatrix} (1-a[q[k]]) & (1-a[q[k]])u[k]-c[q[k]] \\ 0 & 1 \end{bmatrix} \otimes {P_w}' \right)^k Per(2,m_d)
\end{aligned}$$

$$= \quad Per(2,m_d)^{-1}\begin{bmatrix} A & C \\ 0 & {P_w}' \end{bmatrix}^k Per(2,m_d).$$

In this equation, $Per(2,m_d)$ is a $2m_d \times 2m_d$ permutation matrix calculated as $\sum_{i=1}^{2}\sum_{j=1}^{m_d} O_{ij} \otimes O_{ij}^T$, where O_{ij} is a $2 \times m_d$ matrix with only the ith row, jth column entry equals to one and the others are zero. A is defined as $(1-a_d)F$, where

$$F = \begin{bmatrix} f_1p_{1,1} & \cdots & f_1p_{m_1,1} & f_2p_{m_1+1,1} & \cdots & f_2p_{m_2,1} & \cdots & p_{m_{d-1}+1,1} & \cdots & p_{m_d,1} \\ f_1p_{1,2} & \cdots & f_1p_{m_1,2} & f_2p_{m_1+1,2} & \cdots & f_2p_{m_2,2} & \cdots & p_{m_{d-1}+1,2} & \cdots & p_{m_d,2} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ f_1p_{1,m_d} & \cdots & f_1p_{m_1,m_d} & f_2p_{m_1+1,m_d} & \cdots & f_2p_{m_2,m_d} & \cdots & p_{m_{d-1}+1,m_d} & \cdots & p_{m_d,m_d} \end{bmatrix}$$

and C is defined as $[(1-a_d)u]F - c_dH$, where

$$H = \begin{bmatrix} h_1p_{1,1} & \cdots & h_1p_{m_1,1} & h_2p_{m_1+1,1} & \cdots & h_2p_{m_2,1} & \cdots & p_{m_{d-1}+1,1} & \cdots & p_{m_d,1} \\ h_1p_{1,2} & \cdots & h_1p_{m_1,2} & h_2p_{m_1+1,2} & \cdots & h_2p_{m_2,2} & \cdots & p_{m_{d-1}+1,2} & \cdots & p_{m_d,2} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ h_1p_{1,m_d} & \cdots & h_1p_{m_1,m_d} & h_2p_{m_1+1,m_d} & \cdots & h_2p_{m_2,m_d} & \cdots & p_{m_{d-1}+1,m_d} & \cdots & p_{m_d,m_d} \end{bmatrix}$$

The element $p_{i,j}$ in matrices F and H is the transition probability from state i to state j, for $1 \le i,j \le m_d$. We let $M(k) := \begin{bmatrix} A & C \\ 0 & {P_w}' \end{bmatrix}^k = \begin{bmatrix} A^k & U(k) \\ 0 & {P_w}' \end{bmatrix}$, It is easy to derive $M(k+1) := \begin{bmatrix} A & C \\ 0 & {P_w}' \end{bmatrix}^{k+1} = \begin{bmatrix} A^{k+1} & U(k+1) \\ 0 & {P_w}'^{k+1} \end{bmatrix} = \begin{bmatrix} A & C \\ 0 & {P_w}' \end{bmatrix}\begin{bmatrix} A^k & U(k) \\ 0 & {P_w}'^{k} \end{bmatrix}$. Then, we naturally obtain $U(k+1) = AU(k) + CP_w^k$. Noticing that $U(0) = 0$, the general solution is $U(k) = \sum_{i=0}^{k-1} A^iCP_w^{k-1-i}$. From observing Equation 170 and applying the last equality of Equation

174, the total mean backlog $T_E(B)$ can be rewritten as:

$$\begin{aligned}
T_{E(B)} &= 1_{1\times 2m_d} \sum_{i=0}^{k_p} \left(\left({P_w}' \otimes \begin{bmatrix} (1-a[q[k]]) & (1-a[q[k]])u[k]-c[q[k]] \\ 0 & 1 \end{bmatrix} \right)^i \left(q[0] \otimes \begin{bmatrix} B[0] \\ 1 \end{bmatrix} \right) \right) - (K_p+1) \\
&= 1_{1\times 2m_d} \left(\sum_{i=0}^{k_p} Per(2,m_d)^{-1} \begin{bmatrix} A & C \\ 0 & {P_w}' \end{bmatrix}^i Per(2,m_d) \right) \left(q[0] \otimes \begin{bmatrix} B[0] \\ 1 \end{bmatrix} \right) - (K_p+1) \\
&= 1_{1\times 2m_d} Per(2,m_d)^{-1} \left(\sum_{i=0}^{k_p} \begin{bmatrix} A & C \\ 0 & {P_w}' \end{bmatrix}^i \right) Per(2,m_d) \left(q[0] \otimes \begin{bmatrix} B[0] \\ 1 \end{bmatrix} \right) - (K_p+1) \\
&= 1_{1\times 2m_d} Per(2,m_d)^{-1} \left(\begin{bmatrix} \sum_{i=0}^{k_p} A^i & \sum_{j=1}^{k_p} \sum_{i=0}^{j-1} A^i C {P_w}'^{j-1-i} \\ 0 & \sum_{i=0} k_p {P_w}'^i \end{bmatrix} \right) Per(2,m_d) \left(q[0] \otimes \begin{bmatrix} B[0] \\ 1 \end{bmatrix} \right) - (K_p+1)
\end{aligned}$$

We assume the initial condition is $q[0] = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{m_d\times 1}$, so the total mean backlog $T_{E(B)}$ can be further simplified as:

$$(175) \qquad T_{E(B)} = 1_{1\times 2m_d} \begin{bmatrix} \sum_{i=0}^{k_p} A^i & \sum_{j=1}^{k_p} \sum_{i=0}^{j-1} A^i C {P_w}'^{j-1-i} \\ 0 & \sum_{i=0} k_p {P_w}'^i \end{bmatrix} G - (K_p+1)$$

Where G is a $2m_d \times 1$ vector with the $(m_d+1)^{th}$ entry equals to one and the others equal to zero. We notice that $T_{E(B)}$ is a function with respect to a_d, c_d, and u. As such, the sensitivity S of the total mean backlog can be simply obtained by taking the derivative with respect to each of those parameters. Since our interest is to study the performance of the total mean backlog with respect to the variations of the incoming flow, the sensitivity of the flow rate $S(u)$ can be derived as follows.

$$(176) S(u) = \frac{d(T_{E(B)})}{du}$$

$$= \; 1_{1\times m_d} \frac{d(\sum_{j=1}^{k_p}\sum_{i=0}^{j-1} A^i C {P_w}'^{j-1-i})}{du} \begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix}_{m_d\times 1}$$

$$= \; 1_{1\times m_d}\big(\sum_{j=1}^{k_p}\sum_{i=0}^{j-1}[-(i+1)(1-a_d)^i \frac{d(a_d)}{du}u + (1-a_d)^{i+1}]F^{i+1}{P_w}'^{j-1-i} -$$

$$\sum_{j=2}^{k_p}\sum_{i=1}^{j-1}[i(1-a_d)^{i-1}\frac{d(a_d)}{du}c_d + (1-a_d)^i\frac{d(c_d)}{du}]F^iH{P_w}'^{j-1-i} - \sum_{j=1}^{k_p}\frac{d(c_d)}{du}H{P_w}'^{j-1}\big) \begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix}_{m_d\times 1}$$

We notice that for the jump linear system, the parameters a_d and c_d will vary with u, so the derivatives $\frac{d(a_d)}{du}$ and $\frac{d(c_d)}{du}$ are involved in Equation 176.

9.2.3. Problem Formulation

In this section, we will mathematically formulate our flow apportion problem under weather uncertainty. As convective weather event is random during strategic timeframe, the backlog is also uncertain. We use the total mean backlog to measure the system performance. Now, let us describe the details of the design problem.

PROBLEM: OPTIMAL FLOW DISTRIBUTION UNDER WEATHER UNCERTAINTY

Consider n streams of flow entering a weather zone, the total rate of the those flows is u. The capacity assigned to each flow is well known. For example, the i^{th} flow is associated with d_i different levels of capacity reduction from N_{i1} to N_{id_i}.

We assign the i^{th} flow with rate u_i for $1 \leq i \leq n$, such that the total mean backlog (i.e. $\sum_{i=1}^{n} T_{E(Bi)}$) will be minimized, subject to the following constraints:

- $\sum_{i=1}^{n} u_i = u$, where the total inflow rate u is positive.

- $u_i \geq 0$

We denote the optimal rate for flow i as u_i^*. The illustration of our first design problem can be demonstrated in Figure 9.3.

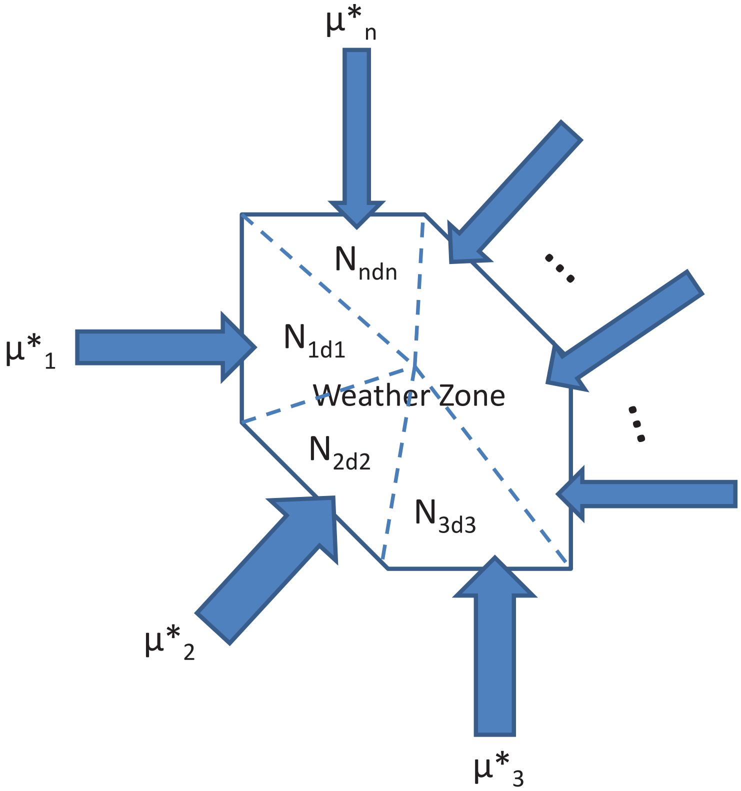


FIGURE 9.3. Illustration of optimal flow distribution problem. The capacity assignment N_{id_i} to each flow i is known. u_i^* is the optimal design parameter resulting in the minimum total mean backlog.

9.3. Optimal Flow Distribution Based upon Sensitivity Study

The goal of this work is to design an appropriate management planning built upon the sensitivity information of the total mean backlog mentioned in the previous section. Therefore, in this section, we will show the relationship between the result of the sensitivity study and the optimal inflow rate assignment. The backlog dynamic of the each flow can be captured by a jump linear model of its own as we discussed in section 9.2. For the flow distribution design problem, the capacity assignment N_{id_i} to each flow i is well known beforehand. We define the backlog of flow i at time step k as $B_i[k]$. Hence, the total mean backlog $T_{E(Bi)}$ during time span $[0, k_p]$ for flow i can be derived according to Equation 175 and the sensitivity $S(u_i)$ can be obtained similar as Equation 175. In the next theorem, we show that the sensitivity of each flows total mean backlog has a simple relationship with the optimal inflow rate distribution strategy. The proof of the theorem is similar to that of Theorem 1 in reference [131].

THEOREM 9.1. *Consider design problem mentioned in section 9.2.3. The optimal inflow rates* u_i^* *satisfy the following condition: there exist a constant* D *such that* $S(u_i) = D$ *for all* i.

PROOF. It is obvious that our design problem is an optimization problem with constraints. We directly apply Lagrange multiplier to find the solution. In particular, the Lagrangian associated with the objective function and constraints is $L = \sum_{i=1}^{n} T_{E(Bi)} + D(u - \sum_{i=1}^{n} u_i) + \alpha_i u_i$, where the constants D and α_i are non-negative. Taking the derivatives of the Lagrangian with respect u_i and D for all i, we obtain

$$S(u_i)D + \alpha_i = 0 \; \forall i \tag{177}$$

$$\sum_{i=1}^{n} u_i^* = u \tag{178}$$

$$\alpha_i u_i^* = 0 \; \forall i \tag{179}$$

As $u_i^* \neq 0$ for any i , $\alpha_i = 0$. We are able to obtain $S(u_i) = D$ for all i. □

The proof shows that for optimal inflow rate allocation, the sensitivity of each flow is equal. Thus, we can utilize this information to come up with the best design. The goal of this design is to minimize the total backlog of all flows during a time span. Noticing that $S(u_i)$ is associated with parameters a_{id_i}, c_{id_i} and u_i, and the values of a_{id_i}, c_{id_i} will change with u_i, therefore, in the practical operation, with the flow distribution rate u_i and capacity assignment N_{id_i}, we are able to identify the corresponding jump linear parameters $a_{id_i} c_{id_i}$, and their derivatives at given u_i. Then, the sensitivity $S(u_i)$ of each flow can be checked. As such, we can decide whether the performance is close to the optimal. In other words, if the sensitivities of several flows are too high or too low, then, it is necessary to adjust the inflows rates to improve the performance. Essentially, this theorem allows us to compute the optimal flow rate u_i^* when a_{id_i} and c_{id_i} are explicit functions relevant to u_i. Noticing that Equation 177 and Equation 178 involve n+1 equalities associated with $n+1$ unknown variables (i.e. u_i for $1 \leq i \leq n$ and D). As such, u_i^* can be directly obtained by solving those equalities.

9.4. Concluding Remarks and Future Work

This chapter discusses the sensitivity study of performance of the NAS with respect to inflow variations under weather uncertainties. We first utilize a jump linear approach to explicitly express one important performance metric (total mean backlog) during a time span. Then, from its sensitivity analysis, we gain some insight of designing optimal inflow rates with certain constraints. We have completed the theoretical part of this chapter, such as problem formulation, derivation of sensitivity regarding incoming flow and the usage of this information to aid optimal inflow distribution. In the future, a concrete real management situation will be taken into consideration and we will show how to distribute the incoming flow appropriately based on the theorem derived in Section 9.3. We will also test the performance of all possible flow apportion plans to verify that our approach ends up with the optimal planning.

BIBLIOGRAPHY

[1] http://ee.unt.edu/public/wan/SNDC.htm.

[2] http://en.wikipedia.org/wiki/Trianglewave.

[3] *Instrument flying handbook*, United States Department of Transportation, Federal Aviation Administration, Airman Testing Standards Branch, AFS-630, Oklahoma City, OK, 2008.

[4] Federal Aviation Administration (FAA), Aviation Capacity Enhancement Plan, FAA Administration Office of System Capacity, 216 pp., 2003.

[5] A. Agogino and J. Rois, *Robustness of two air traffic scheduling approaches to departure uncertainty*, Proceedings of 2011 IEEE/AIAA 30th Digital Avionics Systems Conference (Seattle, WA), October 2011.

[6] A. Agustin, A. Alonso-Ayuso, L. F. Escudero, and C. Pizarro, *On air traffic flow management with rerouting. part ii: stochastic case*, European Journal of Operational Research 219 (2012), no. 1, 167–177.

[7] G. Andreatta, P. Dell'Olmo, and G. Lulli, *An aggregate stochastic programming model for air traffic flow management*, European Journal of Operational Research 215 (2011), no. 3, 697–704.

[8] C. Asavathiratham, S. Roy, B. Lesieutre, and G. Verghese, *The influence model*, IEEE Control Systems Magazine 21 (2001), no. 6, 52–64.

[9] K. Atkinson, *An introduction to numerical analysis*, Wiley, Hoboken, NJ, 1988.

[10] I. Babuska, F. Nobile, and R. Tempone, *A stochastic collocation method for elliptic partial differential equations with random input data*, SIAM Journal on Numerical Analysis 45 (2007), no. 3, 1005–1034.

[11] F. Baccelli and P. Brémaud, *Elements of queuing theory: Palm martingale calculus and stochastic recurrence*, Spring Verlag, Germany, 1994.

[12] F. Bai and A. Helmy, *A survey of mobility modeling and analysis in wireless ad hoc networks*, Wireless Ad Hoc and Sensor Networks, Springer, New York, October 2006.

[13] A.M. Bayen, R.L. Raffard, and C.J. Tomlin, *Eulerian network model of air traffic flow in congested areas*, Proceedings of the 2004 American Control Conference (Boston, Massachusetts), June 2004.

[14] M. Baykal-Gursoy and Z. Duan, *M/m/c queues with markov modulated service processes*, Proceedings of Valuetools'06 (Pisa, Italy), October 2006.

[15] D. Bertsimas, G. Lulli, and A. Odoni, *The air traffic flow management problem: An integer optimization approach*, Integer Programming and Combinatorial Optimization (A. Lodi, A. Panconesi, and G. Rinaldi, eds.), Lecture Notes in Computer Sciences, no. 5035, Springer-Verlag Berlin Heidelberg, May 2008.

[16] C. Bettstetter, *Smooth is better than sharp: a random mobility model for simulation of wireless networks*, ACM International Workshop on Modeling, Analysis and Simulation of Wiresless and Mobile Systems (Rome, Italy), July 2001.

[17] ______, *On the connectivity of ad hoc networks*, The Computer Journal 47 (2004), no. 4, 432– 447.

[18] C. Bettstetter, H. Hartenstein, and X. Pérez-Costa, *Stochastic properties of the random waypoint mobility model*, Wirelss Networks 10 (2004), no. 5, 555–567.

[19] K. Biliia, B. Sridhar, G. Chatterji, K. Sheth, and S. Grabbe, *FACET: Future ATM concepts evaluation tool*, Air Traffic Control Quarterly 9 (2001), no. 1, 1–20.

[20] F. Borrelli, T. Keviczky, and G. J. Balas, *Collision-free uav formation flight using decentralized optimization and invariant sets*, in Proceedings of 43rd IEEE Conference on Decision and Control (2004).

[21] J. Boudec and M. Vojnovic, *Perfect simulation and stationarity of a class of mobility models*, Tech. report, EPFL/IC/2004/59, July 2004.

[22] Stephen P Boyd, Laurent El Ghaoui, Eric Feron, and Venkataramanan Balakrishnan, *Linear matrix inequalities in system and control theory*, vol. 15, SIAM, 1994.

[23] B. J. Broxson, *The kronecker product*, Master Thesis, Department of Mathmatics and Statistics, University of North Florida, 2006.

[24] S. E. Campbell, *Multiscale path optimization for the reduced environmental impact of air transportation*, IEEE Transactions on Intelligent Transportation Systems 13 (2012), no. 3, 1327–1337.

[25] Y. Cao and D. Sun, *A parallel computing framework for large-scale air traffic flow optimization*, IEEE Transactions on Intelligent Transportation Systems 13 (2012), no. 4, 1855–1864.

[26] E. Ward Cheney and David R. Kaincaid, *Numerical mathematics and computing*, Cengage Learning, Boston, MA, 2007.

[27] Z. Cheng and W. B. Heinzelman, *Exploring long lifetime routing (llr) in ad hoc networks*, Proceedings of the 7th ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems (New York, NY), 2004.

[28] G. Clare and A. Richards, *Disturbance feedback for handling uncertainty in air traffic flow management*, Proceedings of 2013 European Control Conference (Zurich, Switzerland), July 2013.

[29] G. Clare, A. Richards, J. Escartin, and D. Martinez, *Air traffic flow management under uncertainty: interactions between network manager and airline operations centre*, Proceedings of the second SESAR Innovation Days (Braunschweig, Geremany), November 2012.

[30] J.-P. B. Clarke and S. Solak, *Air traffic flow management in the presence of uncer-*

tainty, Proceedings of the Eighth USA/Europe Air Traffic Management Research and Development Seminar (Napa, CA, USA), June 2009.

[31] K. Conrad, *Probability distributions and maximum entropy*, http://www.math.uconn.edu/ kconrad/blurbs/entropy.pdf.

[32] O. L. V. Costa, *Linear minimum mean square error estimation for discrete-time markovian jump linear systems*, IEEE Transactions on Automatic Control 39, 1685–1689.

[33] Mohammed Dahleh, Munther A Dahleh, and George Verghese, *Lectures on dynamic systems and control*, A+ A 4 (2004), no. 100, 1–100.

[34] P. J. Davis and P. Rabinowitz, *Methods of numerical integration*, Harcourt Brace Jovanovich, Orlando, FL, 1984.

[35] James Weldon Demmel, *On condition numbers and the distance to the nearest ill-posed problem*, Numerische Mathematik 51 (1987), no. 3, 251–289.

[36] Aloke Dey, *Orthogonal fractional factorial designs*, Wiley New York, 1985.

[37] L. Ekroot and T. M. Cover, *The entropy of markov trajectories*, IEEE Transations on information theory 39 (1993), no. 4, 1418–1421.

[38] L.Trutna et al., *Summary tables of useful fractional factorial designs*, Last Modified Oct. 30, 2013. http://www.itl.nist.gov/div898/handbook/pri/section3/pri3347.htm, 2013.

[39] FAA, ATO NextGen and Operations Planning, NextGen Mid-Term Concept of Operations for the National Airspace System, June 2009.

[40] Y. Fang and K. Loparo, *Stabilization of continuous-time jump-linear systems*, IEEE Transactions on Automatic Control 47 (2002), no. 10, 1590–1603.

[41] JJ Filliben and E Simiu, *Tall building response parameters: sensitivity study based on orthogonal factorial experiment design technique*, Journal of structural engineering 136 (2010), no. 2, 160–164.

[42] Jasmine Foo, Xianliang Wan, and George Em Karniadakis, *The multi-element probabilistic collocation method (me-pcm): Error analysis and applications*, Journal of Computational Physics 46 (2008), no. 5, 2309–2345.

[43] N. M. Freris, H. Kowshik, and P. R. Kumar, *Fundamentals of large sensor networks: Connectivity, capacity, clocks and computation*, Proceedings of the IEEE 98 (2010), no. 11, 1828–1846.

[44] I .Frolow and J. Sinnott., *National airspace system demand and capacity modeling*, Proceedings of the IEEE 77 (1989), no. 11, 1618–1624.

[45] Bo Fu and Luiz A. DaSilva, *A mesh in the sky: A routing protocol for airborne networks*, Proceedings of IEEE Military Communications Conference (Orlando, FL), October 2007, pp. 1–7.

[46] W. G. Hunter G. E. Box, J. S. Hunter, *Statistics for experimenters*, Wiley New York, 2005.

[47] R. G. Gallager, *Discrete stochastic processes*, Kluwer, MBoston, MA, 1996.

[48] R. Ghanta and S. Suresh, *Influence of mobility models on the performance of routing protocols in ad-hoc wireless networks*, Proceedings of IEEE 59th Vehicular Technology Conference (Milan, Italy), vol. 4, May 2004, pp. 2185–2189.

[49] B. Gloss, M. Scharf, and D. Neubauer, *A more realistic random direction mobility model*, Proceedings of 4th Management Committee Meeting (Würzburg, Germany), October 2005.

[50] D. Gross, J. F. Shortle, J. M. Thompson, and C. M. Harris, *Fundamentals of queueing theory*, 4th ed., John Wiley and Sons, Inc., Hoboken, New Jersey, 2008.

[51] R. A. Guérin, *Channel occupancy time distribution in a cellular radio system*, IEEE Transactions on Vehicular Technology 35 (1987), no. 3, 89–99.

[52] Richard F Gunst and Robert L Mason, *Fractional factorial design*, Wiley Interdisci-

plinary Reviews: Computational Statistics 1 (2009), no. 2, 234–244.

[53] C. Gwigger and S. Nagaoka, *Recent models in the analysis of air traffic flow*, Proceedings: Aeronautics and Space Science 57/666 (2009), 210–216.

[54] M. Hansen, A. Odoni, D. Lovell, V. Manikonda, and A. Bayen, *Advanced stochastic network queueing models of the impact of 4d trajectory precision*, NEXTOR Research Symposium (FAA Headquarters, Washington D. C.), 2008.

[55] M. Hansen, A. R. Odoni, D. Lovell, T. Nikoleris, and K. Vlachou, *Use of queuing models to estimate delay savings from 4d trajectory precision*, presented at the 8th USA/Europe Air Traffic Manage. Res. Develop. Seminar (Napa, Valley, CA, USA), June 2009.

[56] James Hardy and Wilkinson, *The algebraic eigenvalue problem*, vol. 87, Oxford Univ Press, 1965.

[57] J. P. Helferty, *Improved tracking of maneuvering targets: the use of turn-rate distributions for acceleration modeling*, Proceedings of the 1994 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (Las Vegas, NV), October 1994, pp. 515–520.

[58] J. R. Hockenberry and B. C. Lesieutre, *Evaluation of uncertainty in dynamic simulations of power system models: The probabilistic collocation method*, IEEE Transactions on Power Systems 19 (2004), no. 3, 1483–1491.

[59] D. Hong and S. S. Rappaport, *Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures*, IEEE Transactions on Vehicular Technology 35 (1986), no. 3, 77–92.

[60] B.R. Horangic, *Some queuing models of airport delays*, M.S. thesis, Department of Electrical Engineering and Computer Science, Massachusetts Inst. Technol., Cambridge, February 1990.

[61] S. Hosder, R. W. Walters, and M. Balch, *Efficient sampling for non-intrusive polynomial chaos applications with multiple uncertain input variables*, Proceedings of AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference (Honolulu, Hawii), April 2007.

[62] E. Hyytiä, P. Lassila, and J. Virtamo, *Spatial node distribution of the random waypoint mobility model with applications*, IEEE Transactions on mobile computing 5 (2006), no. 6, 680–694.

[63] H. Idris, J.-P Clarke, R. Bhuva, and L. Kang, *Queuing model for taxi-out time estimation*, Air Traffic Control Quarterly 10 (2002), no. 1, 1–22.

[64] Ilse Ipsen and Shivkumar Chandrasekaran, *Perturbation theory for the solution of systems of linear equations*, YALE UNIV NEW HAVEN CT DEPT OF COMPUTER SCIENCE, 1991.

[65] S.S. Isukapalli, *Uncertainty analysis of transport-transformation models*, The State University of New Jersey, New Brunswick, N.J., 1999.

[66] R. Jin, W. Chen, and A. Sudjianto, *An efficient algorithm for constructing optimal design of computer experiments*, Journal of Statistical Planning and Inference 134 (2005), no. 1, 268–287.

[67] Tosio Kato, *Perturbation theory for linear operators*, vol. 132, springer, 1995.

[68] J. Kim, K. Palaniappan, and P. K. Menon, *Trajectory uncertainty modeling for queuing analysis of the national airspace system*, The Congress of International Council of the Aeronautical Science (ICAS) (Anchorage, Alaska), 14-19 September 2008.

[69] P. Knupp, *Matrix norms and the condition number*, Proceedings of 8th International Meshing Roundtable, Citeseer, 1999, pp. 13–22.

[70] E. Kuiper and S. Nadjm-Tehrani, *Mobility models for uav group reconnaissance applications*, Proceedings of International Conference on Wireless and Mobile Communica-

tions (Bucharest), July 2006, p. 33.

[71] ______, *Geographical routing with location service in intermittently connected manets*, IEEE Transactions on Vehicular Technology 60 (2011), no. 2, 592–604.

[72] Y. Lei, R. Kacker, D. Kuhn, V. Okun, and J. Lawrence, *Ipog/ipod: Efficient test generation for multi-way software testing*, Journal of Software Testing, Verification, and Reliability (2008), 125–148.

[73] X. R. Li and V. P. Jilkov, *A survey of maneuvering target tracking: dynamic models*, Proceedings of SPIE Conference on Signal and Data Processing of Small Targets (Orlando, FL), vol. AES-6, April 2000, pp. 212–235.

[74] B. Liang and Z. J. Haas, *Predictive distance-based mobility management for multidimensional pcs networks*, IEEE/ACM Transactions on Networking 11 (2003), no. 5, 718–732.

[75] G. Lim, K. Shin, S. Lee, H. Yoon, and J. Ma, *Link stability and route lifetime in ad-hoc wireless networks*, Proceedings of International Conference on Parallel Processing Workshops (Vancouver, B.C., Canada), August 2002, pp. 116–123.

[76] B. Liu, O. Dousse, P. Nain, and D. Towsley, *Dynamic coverage of mobile sensor networks*, Computing Research Repository abs/1101.0376 (2011).

[77] D. Long, D. Lee, J. Johnson, E. Gaier, and P. Kostiuk, *Modeling air traffic management technologies with a queuing network model of the national airspace system*, Tech. Report NASA/CR-1999-208988, Langley Research Center, VA, USA, July 1999.

[78] Q. Ma, Y. Wan, and D. Sun, *Probabilistic collocation method in solving euclidean model for air traffic flow management*, Proceedings of AIAA Infotech@aerospace Conference (Garden Grove, CA), June 2012.

[79] ______, *Stochastic analysis of air-traffic system and its corresponding application in parameters prediction*, Proceedings of American Control Conference (Montreal, Canada),

June 2012, pp. 1689–1694.

[80] K. M. Malone, *Dynamic queueing systems: behavior and approximations for individual queues and networks*, Ph.D. Thesis, Operations Research Center, Massachusetts Institute of Technology, Cambridge, June 1995.

[81] M. D. Mckay, R. J. Beckman, and W. J. Conover, *A comparison of three methods for selecting values of input variables in the analysis of output from a computer code*, Technometrics 21 (1979), no. 2, 239–245.

[82] P. K. Menon, G. D. Sweriduk, and K. Bilimoria, *New approach for modeling, analysis and control of air traffic flow*, AIAA Journal of Guidance, Control, and Dynamics 27 (2004), no. 5, 737–744.

[83] P. K. Menon, G. D. Sweriduk, T. Lam, and V. H. L Cheng, *Air traffic flow modeling, analysis and control*, AIAA Conference on Guidance, Navigation, and Control Conference and Exhibit (Austin, TX), August 2003, AIAA Paper 2003–5712.

[84] L. Meyn, *A new model to improve aggregate air traffic demand predictions*, AIAA Guidance, Navigation, and Control Conference and Exhibit (Hilton Head, SC), August 2007.

[85] S. P. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*, Springer Verlag, London, 1993.

[86] K. L. Mills, J. J. Filliben, D. Y. Cho, and E. J. Schwartz, *Predicting macroscopic dynamics in large distributed systems*, Proceedings of the ASME 2011 Pressure Vessels and Piping Division Conference (Baltimore, Maryland), July 17-22 2011.

[87] Douglas C. Montgomery, *Design and analysis of experiments*, vol. 7, Wiley New York, 1997.

[88] D. Moreau and S. Roy, *A stochastic characterization of en route traffic flow management strategies*, AIAA Conference on Guidance, Navigation, and Control Conference

and Exhibit (San Francisco, CA), July 2005.

[89] E.R. Mueller and G.B. Chatterji, *Analysis of aircraft arrival and departure delay characteristics*, Proceedings of the AIAA Aircraft Technology, Integration and Operations (ATIO) Conference (Los Angeles, CA), October 2002.

[90] T. Myers and D. Kierstead, *Network model to address capacity/demand imbalances in the national airspace system*, Proceedings of AIAA Guidance, Navigation and Control Conference and Exhibit (Honolulu, Hawaii), 16-18 August 2008.

[91] N. Nabaa and R. H. Bishop, *Validation and comparison of coordinated turn aircraft maneuver models*, IEEE Transactions on Aerospace and Electronic Systems 36 (2000), no. 1, 250–255.

[92] P. Nain, D. Towsley, B. Liu, and Z. Liu, *Properties of random direction models*, Proceedings of 24th Annual Joint Conference of the IEEE Computer and Communications Societies, March 2005, pp. 1897–1907.

[93] Hemanth Narra, Egemen K. Çetinkaya, and James P.G. Sterbenz, *Performance analysis of aerorp with ground station advertisements*, Proceedings of the first ACM MobiHoc workshop on Airborne Networks and Communications (New York, NY, USA), Airborne '12, ACM, 2012, pp. 43–47.

[94] A. Nilim, L. Ghaoui, and V. Duong, *Robust dynamic routing of aircraft under uncertainty*, Proceedings of the 21st Digital Avionics Systems Conference (Irvine, California), February 2002.

[95] ———, *Multi-aricraft routing and traffic flow management under uncertainty*, Recherche Informatique Vietnam-Francophone (Hanoi, Vietnam), February 2004.

[96] F. Nobile, R. Tempone, and C.Webster, *A sparse grid collocation method for elliptic partial differential equations with random input data*, SIAM Journal on Numerical Analysis 229 (2010), no. 5, 1536–1557.

[97] A. Papoulis and S. U. Pillai, *Probability, random variables and stochastic processes*, McGraw-Hill, New York, 2002.

[98] J. W. Pepper, K. R. Mills, and L. A. Wojoik, *Predictability and uncertainty in air traffic flow management*, PNAS 94 (1997), 814–819.

[99] Y. Peres, A. Sinclair, P. Sousi, and A. Stauffer, *Mobile geometric graphs: Detection, coverage and percolation*, January 2011, pp. 412–428.

[100] K. Peters, A. Jabbar, E.K. Cetinkaya, and J.P.G. Sterbenz, *A geographical routing protocol for highly-dynamic aeronautical networks*, Proceedings of IEEE Wireless Communications and Networking Conference (WCNC) (Cancun, Quintana Roo), March 2011, pp. 492–497.

[101] M. D. Peterson, D. J. Bertsimas, and A. R. Odoni, *Decomposition algorithms for analyzing transient phenomena in multiclass queuing networks in air transportation*, Operations Research 43 (1995), no. 6, 995.

[102] J. Prinz, P. Tobias, W.F.Guthrie, B.Hembree, M.C.Croarkin, J.J.Filliben, and N.A.Heckert, *Nist/sematech e-handbook of statistical methods*, (2013).

[103] J. Rios and A. Morando, *The value of reduced uncertainty in air traffic flow management*, Proceedings of the AIAA Guidance, Navigation, and Control Conference (Portland, Oregon), August 2011.

[104] R.Mori, *Aircraft ground-taxiing model for congested airport using cellular*, IEEE Transactions on Intelligent Transportation Systems 14 (2013), no. 1, 180–188.

[105] M. Robert, *A comprehensive guide to factorial two-level experimentation*, Springer, 2009.

[106] J. P. Rohrer, E. K. Cetinkaya, H. Narra, D. Broyles, K. Peters, and J. P. G. Sterbenz, *Aerorp performance in highly-dynamic airborne networks using 3d gauss-markov mobility model*, Proceedings of Military Communications Conference (Baltimore, MD),

November 2011.

[107] J.P. Rohrer, A. Jabbar, E.K. Cetinkaya, E. Perrins, and J.P.G. Sterbenz, *Highly-dynamic cross-layered aeronautical network architecture*, IEEE Transactions on Aerospace and Electronic Systems 47 (2011), no. 4, 2742–2765.

[108] S. Roy, D. Ramamurthy, and B. C. Lesieutre, *Studies on the probabilistic collocation method and its application to power system analysis*, 36th North America Power Symposium (Moscow, ID), August 2004.

[109] S. Roy and A. Saberi, *Static decentralized control of a single-integrator network with markovian sensing topology*, Automatica 41 (2005), no. 11, 1867–1877.

[110] S. Roy, G.C. Verghese, and B.C. Lesieutre, *Moment-linear stochastic systems*, Proceedings of First International Conference on Informatics in Control, Automation, and Robotics (Setubal, Portugal), August 2004.

[111] S. Roy and Y. Wan, *Quantifying the tradeoff between fairness and optimality in traffic flow mangement and planning: a queueing-theory approach*, AIAA Conference on Guidance, Navigation, and Control Conference and Exhibit (Chicago, IL), August 2009.

[112] S. Roy, Y. Wan, C. Taylor, and C. Wanke, *Stochastic network model for uncertain spatiotemporal weather impact at the strategic time horizon*, 10th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference (Fort worth, TX), April 2010.

[113] Sandip Roy, *Moment-linear stochastic systems and their applications*, Ph.D. thesis, Massachusetts Institute of Technology, June 2003.

[114] K. Sampigethaya, R. Poovendran, S. Shetty, T. Davis, and C. Royalty, *Future e-enabled aircraft communications and security: The next 20 years and beyond*, Proceedings of the IEEE 99 (2011), no. 11, 2040–2055.

[115] J. F. Shortle and B. L. Mark., *Efficient simulation of the national airspace system*,

Proceedings of 2003 Winter Simulation Conference, December 2003.

[116] R. A. Singer, *Estimating optimal tracking filter performance for manned maneuvering targets*, IEEE Transactions on Aerospace and Electronic Systems AES-6 (1970), no. 4, 473–383.

[117] B. Sridhar, S. Grabbe, and A. Mukherjee, *Modeling and optimization in traffic flow management*, Proceedings of the IEEE 96 (2008), no. 12, 2060–2080.

[118] M. Steiner and J. Krozel, *Translation of ensemble-based weather forecasts into probabilistic air traffic capacity impact*, Proceedings of the 28th Digital Avionics Systems Conference (Orlando, FL), October 2009.

[119] D. Sun and A.M. Bayen, *Multicommodity Eulerian-Lagrangian Large-capacity Cell Transmission Model for en route traffic*, AIAA Journal of Guidance, Control, and Dynamics 31 (2008), no. 3, 616–628.

[120] M. Tandale, P. Sengupta, P. K. Menon, and V. H. L Cheng, *Queueing network models of the national air space system*, The Congress of International Council of the Aeronautical Science (ICAS) (Anchorage, Alaska), 14-19 September 2008.

[121] Menner A. Tatang, *Direct incorporation of uncertainty in chemical and environmental engineering systems*, Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, 1995.

[122] C. Taylor, T. Masek, C. Wanke, Y. Wan, and S. Roy, *Defining multi-resolution networks for flow contingency management*, AIAA Aviation Technology, Integration, and Operations (ATIO) Conference (Virginia Beach, VA), September 2011.

[123] C. Taylor, C. Wanke, Y. Wan, and S. Roy, *A framework for flow contingency management*, Proceedings of AIAA Aviation Technology, Integration, and Operations Conference (Virginia Beach, VA), September 2011.

[124] ———, *A decision support tool for flow contingency management*, Proceedings of

AIAA Guidance Navigation and Control Conference (Minneapolis, Minnesota), August 2012.

[125] S. Tien, C. Taylor, Y. Zhou, Y. Wan, and C. Wanke, *A route-based queuing network model for air traffic flow contingency management*, Proceedings of AIAA Aviation Technology, Integration, and Operations (ATIO) Conference (Virginia Beach, VA), September 2011.

[126] A. Tiwari, A. Ganguli, A. Sampath, D.S. Anderson, B. h. Shen, N. Krishnamurthi, J. Yadegar, M. Gerla, and D. Krzysiak., *Mobility aware routing for the airborne network backbone*, Proceedings of Military Communications Conference (San Diego, CA), November 2008.

[127] C. Tyalor, C. Wanke, S. Zobell, R. Holland, M. Hokit, L. Wang, Y. Wan, S. Roy, and M. Xue, *Flow contingency management decision framework*, Tech. Report MP100309, MITRE, September 2010.

[128] Y. Wan, K. Namuduri, Y. Zhou, and S. Fu, *A smooth-turn mobility model for airborne networks*, IEEE Transactions on Vehicular Technology 62 (2013), no. 7, 3359–3370.

[129] Y. Wan, K. Namuduri, Y. Zhou, D. He, and S. Fu, *A smooth-turn mobility model for airborne networks*, Proceedings of the first ACM MobiHoc workshop on Airborne Networks and Communications (Hilton Head, SC), June 2012.

[130] Y. Wan and S. Roy, *A scalable methodology for evaluating and designing coordinated air traffic flow management strategies under uncertainty*, IEEE Transactions on Intelligent Transportation Systems 9 (2008), no. 4, 644–656.

[131] ______, *Sensitivity of national airspace system performance to disturbances: modeling, identification from data, and use in planning*, Proceedings of the AIAA Conference on Guidance, Navigation, and Control (Honolulu, HI), August 2008.

[132] ______, *Uncertainty evaluation through mapping identification in intensive dynamic*

simulations, AIAA Conference on Guidance, Navigation, and Control Conference and Exhibit (Honolulu, HI), August 2008.

[133] ______, *Uncertainty evaluation through mapping identification in intensive dynamic simulations*, IEEE Transaction on Systems, Man and Cybernetics, Part A: Systems and Humans 40 (2010), no. 5, 1094–1104.

[134] Y. Wan, S. Roy, and A. Saberi, *A new focus in the science of networks: towards methods for design*, Proceedings of the Royal Society A, vol. 464, March 2008, pp. 513–535.

[135] Y. Wan, C. Taylor, S. Roy, C. Wanke, and Y. Zhou, *Dynamic queuing network model for flow contingency management*, Proceedings of AIAA Guidance, Navigation and Control Conference and Exhibit (Portland, Oregon), August 2011.

[136] ______, *Dynamical queueing network model for flow contingency management*, IEEE Transactions on Intelligent Transportation Systems 14 (2013), no. 3, 1380–1392.

[137] L. Wang, C. Taylor, and C. Wanke, *An airport clustering method for air traffic flow contingency management*, Proceedings of the AIAA Aviation Technology, Integration, Operations Conference (Virginia Beach, VA), September 2011.

[138] W. Wang, X. Guan, B. Wang, and Y. Wang, *A novel mobility model based on semi-random circular movement in mobile ad hoc networks*, Information Science 180 (2010), no. 3, 399–413.

[139] C. Wanke, S. Mulgund, D. Greenbaum, and L. Song, *Modeling traffic prediction uncertainty for traffic management decision support*, AIAA Guidance, Navigation, and Control Conference and Exhibit (Providence, RI), 16-19 August 2004.

[140] C. Wanke, L. Wang, C. Taylor, S. Roy, and Y. Wan, *Modeling air traffic demand for a real-time queuing network model of the national airspace system*, AIAA MIST, August 2012.

[141] M. Webster, M. A. Tang, and G. J. McRae, *Application of the probabilistic collocation*

method for an uncertainty analysis of a simple ocean model, January 1996.

[142] D. L. Wei, Z. S. Cui, and J. Chen, *Uncertainty quantification using polynomial chaos expansion with points of monomial cubature rules*, Computers and Structures 86 (2008), no. 23-24, 2102–2108.

[143] L. A. Wojcik et al., *World regional air traffic modeling with DPAT*, MITRE Technical Report MTR-97W00000070 (1997).

[144] M. M. Woolfson and G. J. Pert, *An introduction to computer simulation*, Oxford University Press, New York, 1999.

[145] J. Xie, Y. Wan, and Y. Zhou, *Effective uncertainty evaluation in large-scale systems (book chapter)*, accepted for publication, April 2014.

[146] J. Xie, Y. Wan, Y. Zhou, K. Mills, J. Filliben, and Y. Lei, *Effective and scalable uncertainty evaluation for large-scale complex system applications*, Proceedings of the 2014 Winter Simulation Conference (Savannah, GA), December 2014.

[147] F. Xiong, W. Chen, Y. Xiong, and S. Yang, *A new weighted stochastic response surface method for uncertainty propagation*, Proceedings of AIAA/ISSMO Multidisciplinary Analysis Optimization Conference (Fort Worth), September 2010.

[148] F. Xue and P. R. Kumar, *The number of neighbors needed for connectivity of wireless networks*, The Computer Journal 10 (2004), no. 2, 169–181.

[149] M. Xue, S. Roy, S. Zobell, Y. Wan, C. Taylor, and C. Wanke, *A stochastic spatiotemporal weather-impact simulator: Representative scenario selection*, AIAA Aviation Technology, Integration, and Operations (ATIO) Conference (Virginia Beach, VA), 2011.

[150] M. Xue, S. Zobell, S. Roy, C. Taylor, Y. Wan, and C. Wanke, *Using stochastic, dynamic weather-impact models in strategic traffic flow management*, Proceedings of the 91st AMS Annual Meeting (Seattle, WA), January 2011.

[151] Y. Zhou, Y. Wan, S. Roy, C. Taylor, and C. Wanke, *A stochastic modeling and analysis*

approach to strategic traffic flow management under weather uncertainty, Proceedings of AIAA Guidance, Navigation and Control Conference and Exhibit (Portland, Oregon), August 2011.

[152] Y. Zhou, Y. Wan, S. Roy, C. Taylor, C. Wanke, D. Ramamurthy, and J. Xie, *Multivariate probabilistic collocation method for effective uncertainty evaluation with application to air traffic management*, IEEE Transactions on Systems, Man and Cybernetics: System (in press) (2014).

[153] Y. Zhou, Y. Wan, and J. Xie, *A jump linear approach based sensitivity study for optimal air traffic flow management under weather uncertainty*, submitted to the AIAA Science and Technology Forum 2015.

[154] Yi Zhou, Dinesh Ramamurthy, Yan Wan, Sandip Roy, Christine Taylor, and Craig Wanke, *Multivariate probabilistic collocation method for effective uncertainty evaluation with application to air traffic management*, American Control Conference (ACC), IEEE, 2013, pp. 6345–6350.

[155] Yi Zhou, Yan Wan, Christine Taylor, Sandip Roy, and Craig Wanke, *Performance evaluation and optimal decision-making for strategic air traffic management under weather uncertainty*, Proceedings of AIAA Infotech@ Aerospace Conf., Garden Grove, CA, USA, 2012.

[156] Yi Zhou, Yan Wan, Craig Wanke, Christine Taylor, and Sandip Roy, *A probabilistic collocation method-based approach for optimal strategic air traffic flow management under weather uncertainties*, Proceedings of AIAA ATIO Conference, August 2013.