FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Data-Driven Modeling of Aircraft Engine Performance

Yashovardhan S. Chati, Hamsa Balakrishnan

4 November 2015

Berkeley UNIVERSITY OF CALIFORNIA

Massachusetts Institute of Technology

NSF

MICHIGAN

VANDERBILT UNIVERSITY

Research Objectives

* ‘Hybrid’* models of aircraft engine performance
  * Physical principles + operational flight data (e.g. from FDR)
  * Aircraft engine performance + statistical data analysis/machine learning
* Intended intellectual outputs of the research
  * Methodology and techniques used to build the models
  * The models themselves

*Jaw, L. C., and Mattingly, J. D., *Aircraft Engine Controls: Design, System Analysis, and Health Monitoring*, AIAA, Inc., Reston, Virginia, 2009, Chap. 8.

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Current Practice

* Gas turbine performance simulation software
  * Require knowledge of engine parameters – not easy to access
* Data-driven models of engine performance
  * Non-operational data from flight manuals, ground tests
  * Inability to quantify variability in performance for the same engine type (pilot behavior, operational and maintenance procedures, etc.)
* ICAO Aircraft Engine Emissions Databank
* Base of Aircraft Data (BADA)

Regression Methodology

* Predictor variables: pressure altitude (in m) $[h]$, normalized ground $[v]$ and vertical speeds $[\dot{h}]$, normalized aircraft mass $[W]$
* Response variable: normalized fuel flow rate per engine $[\dot{m}_f]$
* Methods
  * Classification And Regression Trees (CART) model
  * Ensemble (boosted CART) model
  * Bayesian Multiple Linear Regression (BMLR) model

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Performance of CART Models v/s BADA

* Sample mean prediction error:

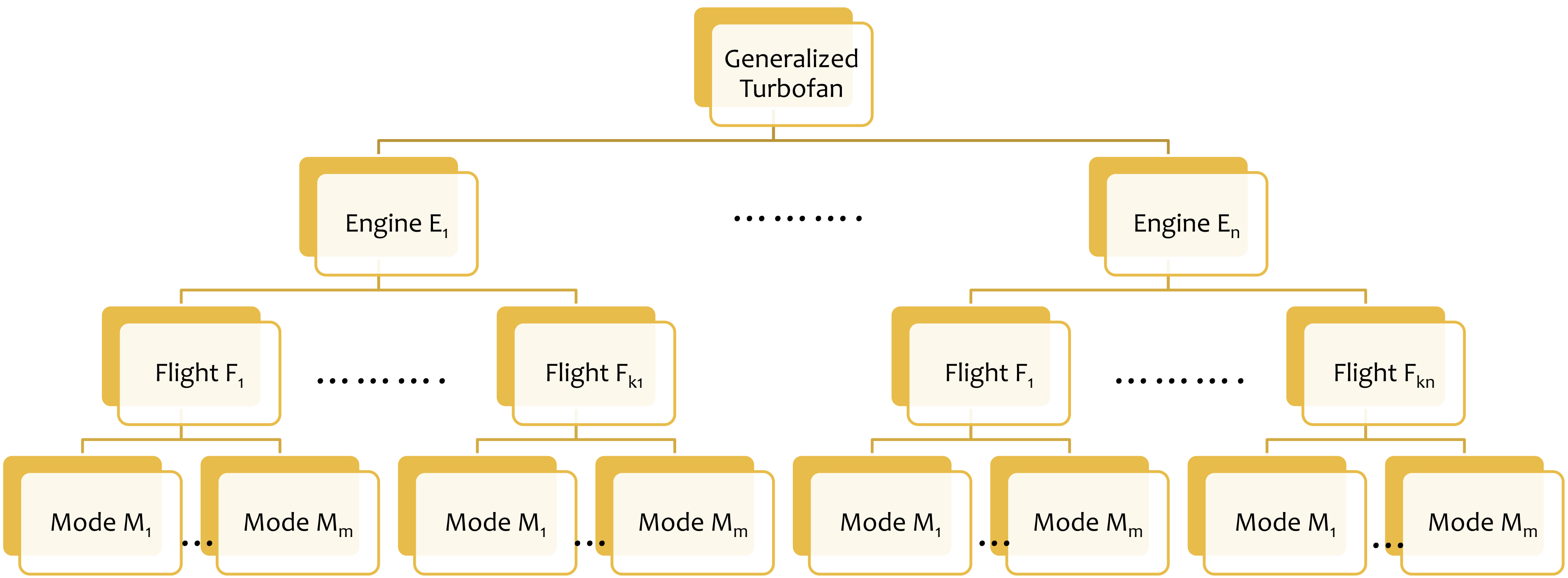
Phase/Method	CART	Boosted CART	BADA
Ascent	1.4 – 2.5%	0.7 – 1.4%	5.9 – 22.4%
Cruise	2.8 – 8.2%	2.0 -- 6.3%	12.2 -- 108.1%
Descent	15.0 – 18.6%	9.9 – 12.5%	31.4 – 59.8%

* 95% prediction interval coverage:

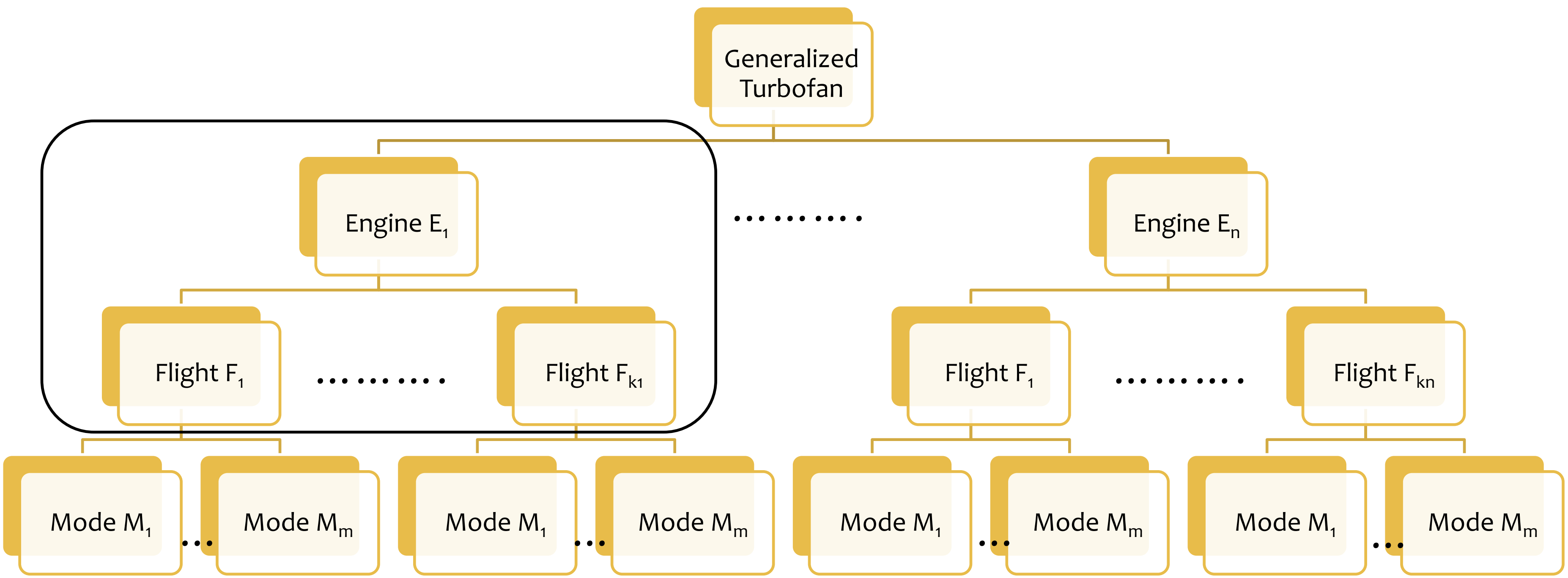
Phase/Method	CART	Boosted CART	BADA
Ascent	55.0 – 60.3%	70.0 -- 77.0%	0
Cruise	50.2 – 62.4%	59.7 -- 66.2%	0
Descent	52.6 -- 59.8%	63.6 -- 67.7%	0

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Hierarchical Model

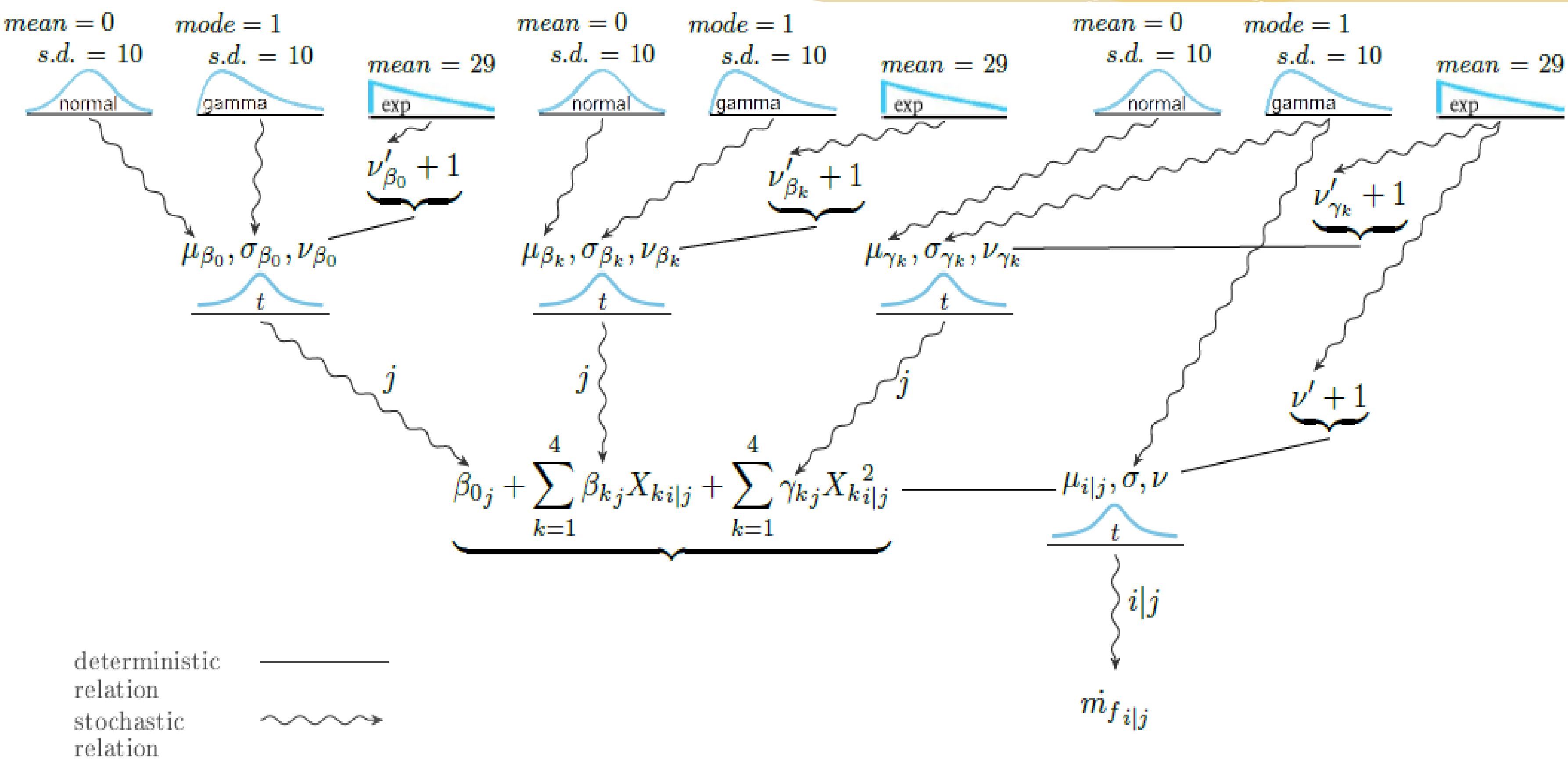


Hierarchical Model



FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Hierarchical BMLR for A320-200 in Cruise



$j = 1, 2, 3, \ldots, 169$

$k = 1, 2, 3, 4$

$X_{k=1} : h,\ X_{k=2} : v,\ X_{k=3} : \dot{h},\ X_{k=4} : W$

BMLR: Bayesian Multiple Linear Regression

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Hierarchical BMLR Model Training and Testing

* Gibbs sampling
  * 3 Markov Chain Monte Carlo (MCMC) chains
  * 10000 samples
  * 15000 burn in samples
  * Thinning = 110
* Good mixing among chains => convergence to target posterior distribution
* Posterior distributions of aircraft type-level regression coefficients used to develop posterior predictive distributions on test data

Predictive Performance Comparison across Models

Model	Mean Error (%)	Mean Length of Prediction Interval	Percent Coverage (%)
Hierarchical BMLR	6.7	0.0204	37.6
Non hierarchical BMLR	6.2	0.0067	14.9
Robust Least Squares	6.3	0.0088	18.1
CART	6.4	0.0503	58.4
Boosted CART	4.0	0.0318	65.6

BMLR: Bayesian Multiple Linear Regression
CART: Classification And Regression Trees

Discussion

* Bayesian models yield the complete posterior distributions
  * Making predictions on new data computationally simpler
  * Larger training time
* Hierarchical models: model relationships among different flights and aircraft types
* Nonparametric methods give better predictive performance
  * Bayesian trees?
* Extension to other aircraft/engine types and flight modes
  * Time series analysis using Bayesian nonparametric methods to identify the modes

THANK YOU

BACKUP SLIDES

Foreseen Uses of Our Research

* Generation of fuel burn and emissions inventories
* Development of flight paths optimal on fuel burn
* Methodology behind model building can give insights into the application of data analysis techniques to aeronautical datasets
* Models built on operational flight data are abstractions of such data
  * Can be used by researchers as tools in the absence of raw operational data

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Novelty in Our Approach

- Use of operational flight data from FDR to build models
  - Bypass need to know internal engine parameters
  - Capture variability in performance of the same engine type
- Combination of physical insights and data analysis techniques
  - Ensure data-based models conform to physical principles governing engine performance

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Risks and Payoffs

* Risks
  * Performance of models outside range of training data?
  * Scale of model applicability (aircraft types, O-D pairs)?
* Payoffs
  * Potential 'proof of concept'
  * Merits of using operational data to model engine performance
  * Methods can be used to expand the models as more data are available

Performance of CART Models v/s ICAO Databank

* Sample mean prediction error:

Phase/Method	CART	Boosted CART	ICAO Databank
Climb out (<= 3000' AGL)	1.0 – 4.8%	0.3 – 2.5%	6.2 – 33.8%
Approach (<= 3000' AGL)	13.7 – 20.5%	6.5 – 12.8%	35.0 – 96.3%

* 95% prediction interval coverage:

Phase/Method	CART	Boosted CART	ICAO Databank
Climb out (<= 3000' AGL)	55.3 – 63.8%	69.8 – 79.1%	0
Approach (<= 3000' AGL)	55.0 – 61.2%	63.4 – 70.9%	0