Data Analytics in Air Transportation Systems II

Dr. Emre Koyuncu (Istanbul Technical University)

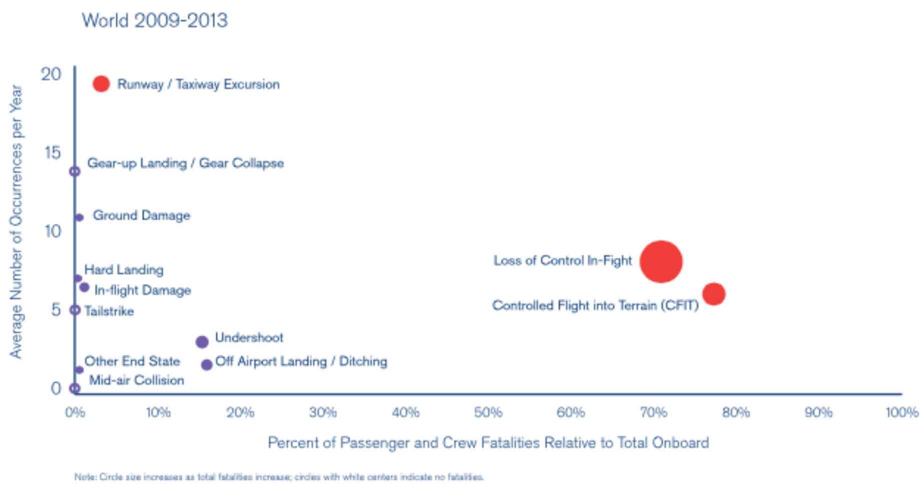
Istanbul Technical University
Air Transportation Management
M.Sc. Program

Advanced Information Systems
Module 5 : 2 June 2015

TURKISH AIRLINES İTÜ

- Data Analytics in AT:
  - TBO Flight Operation case
  - Flight Incidents case
  - FDM based flight performance analysis
  - Delay Propagation in ATM Network

BACKGROUND



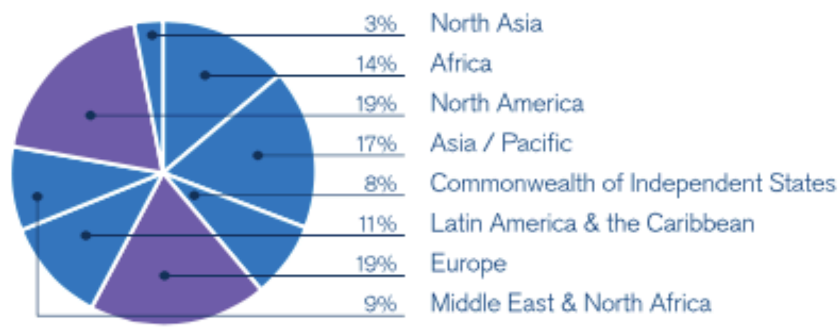
*Loss of control (L-CIF) usually occurs because the aircraft enters a flight regime which is outside its normal envelope, usually, but not always at a high rate, thereby introducing an element of surprise for the flight crew involved.

*Controlled flight into terrain (CFIT) describes an accident in which an airworthy aircraft, under pilot control, is unintentionally flown into the ground, a mountain, water, or an obstacle.

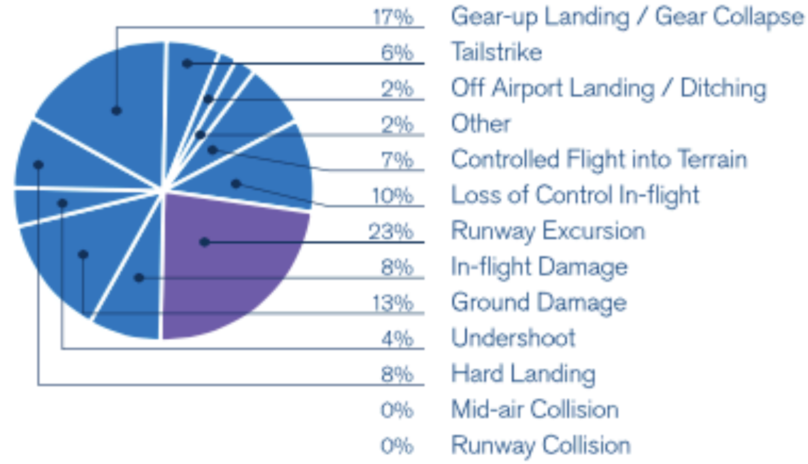
*Runway excursion is overrun off the runway surface

- 2009-2013 Aircraft Accidents

Breakdown per Operator Region

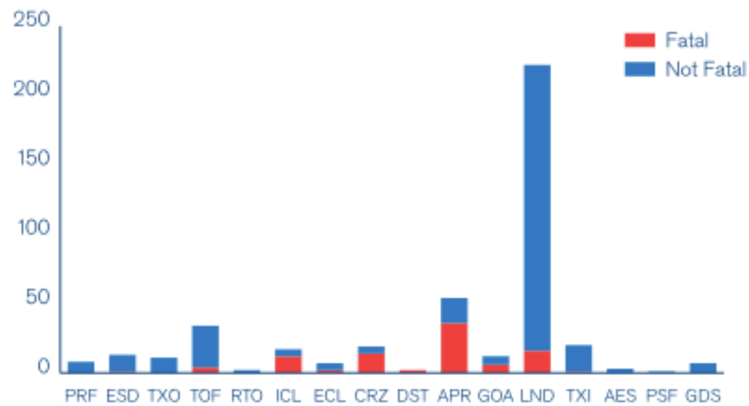


Breakdown per Accident Category



Accidents per Phase of Flight

See Annex 1 for detailed "Phase of Flight" definitions



CONTRIBUTING FACTORS

Top Contributing Factors, 2009-2013

See Annex 2 for "Contributing Factors" definitions

Latent Conditions
(deficiencies in...)

- **45%** Regulatory oversight
- **38%** Safety management
- **23%** Flight operations: training systems
- **20%** Technology and equipment

Threats

Environmental

- **45%** Meteorology:
  - Poor visibility/IMC (64% of these cases)
  - Wind/wind shear/gusty wind (22% of these cases)
- **26%** Ground-based navigation aids malfunctioning or not available
- **13%** Lack of visual reference

Airline

- **31%** Aircraft malfunction:
  - Contained engine failure/ powerplant malfunction (48% of all malfunctions)
  - Fire/smoke (cockpit/cabin/ cargo) (28% of all malfunctions)
- **9%** Maintenance events
- **8%** Operational pressure

Flight Crew Errors
(relating to...)

- **40%** SOP adherence/cross-verification
  - Intentional non-compliance (72% of these cases)
  - Unintentional non-compliance (28% of these cases)
- **33%** Manual handling/flight controls
- **15%** Callouts
- **14%** Pilot-to-pilot communication

Undesired Aircraft States

- **36%** Vertical, lateral or speed deviations
- **18%** Controlled flight towards terrain
- **16%** Operation outside of aircraft limitations
- **13%** Unstable approach
- **11%** Unnecessary weather penetration

Countermeasures

- **41%** Overall crew performance
- **31%** Monitor/cross-check
- **16%** Contingency management
- **16%** Leadership

Additional Classifications

- **10%** Insufficient data for contributing factors

Relationships of Interest, 2009-2013

38% of fatal accidents occurred during the approach phase of flight. Of these, **56%** were due to controlled flight into terrain and **63%** involved flight crew vertical, lateral or speed deviations.

Deficiencies in the operator's safety management were noted in **79%** of events where inadequate standard operating procedures for flight crew were noted as a factor.

Note: Nine accidents were not classified due to insufficient data; these accidents were removed from the count for the contributing factors and relationships of interest.

INCIDENT ANALYSIS

$$P_{Incident} = \frac{Frequency\ of\ Incident*}{Number\ of\ Flights}$$

Classical statistical approach

$$P_{Incident} = \frac{0}{400\,000} = 0$$

Runway overrun example

vs.

Simple statistical approach is inappropriate and **unsuitable for rare events**

*__Serious incidents__ as defined in ICAO Annex 1

Predictive Analysis:

Making quantitative statements about the future state based on:

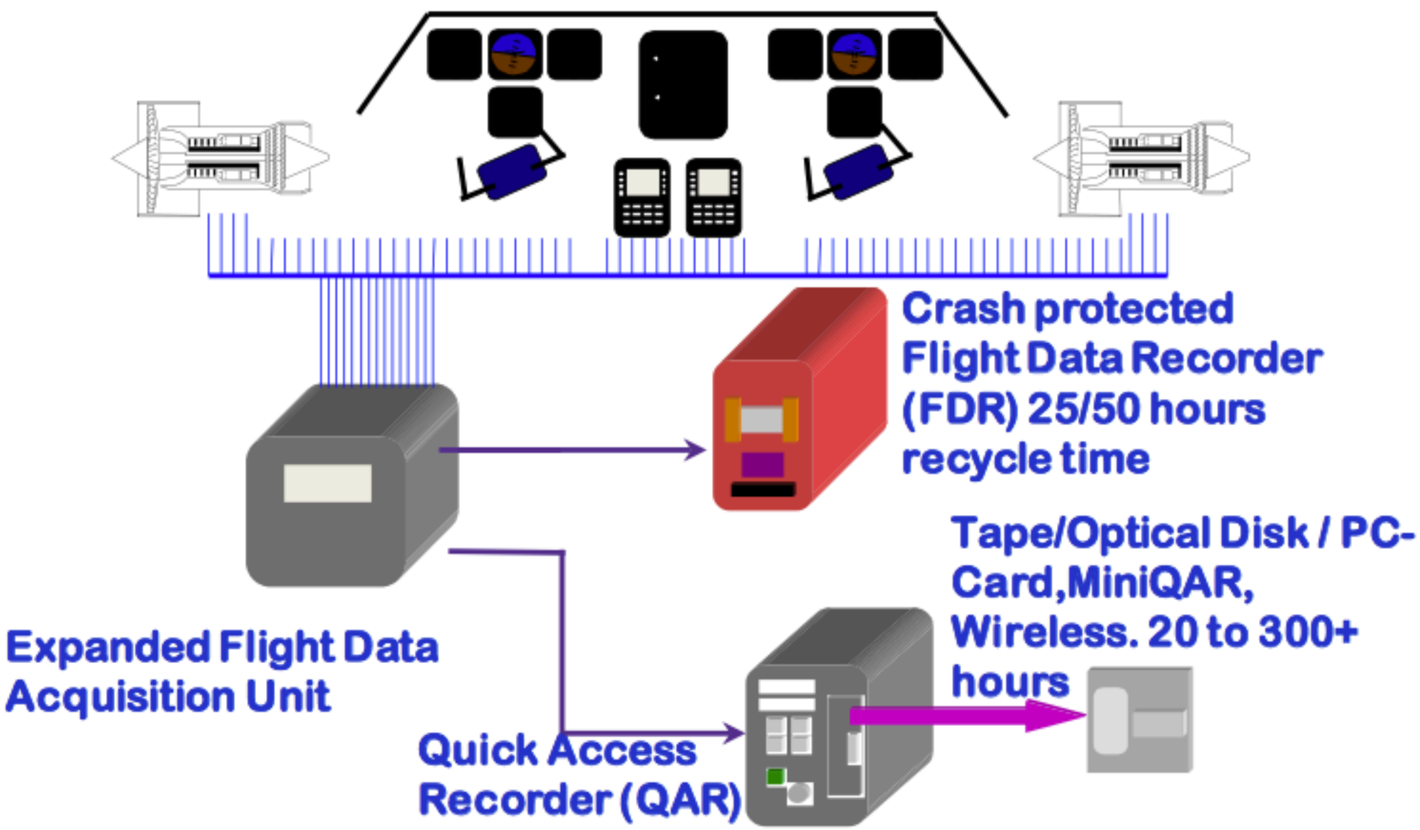
- previous experience
- knowledge

previous experience = data/evidence driven	• recorded data • known accident types and their causes
knowledge	• physical relation between contributing factors and accident • known cause-consequence-chains

Basic Hypothesis:

1. Accidents cannot be directly observed in daily operation, however, the contributing factors still occur at high frequency so they can be measured or observed with statistical significance.
2. The relation between the contributing factors and the accident can be described by the laws of physics and cause-consequence-chains based on operational and procedural knowledge.

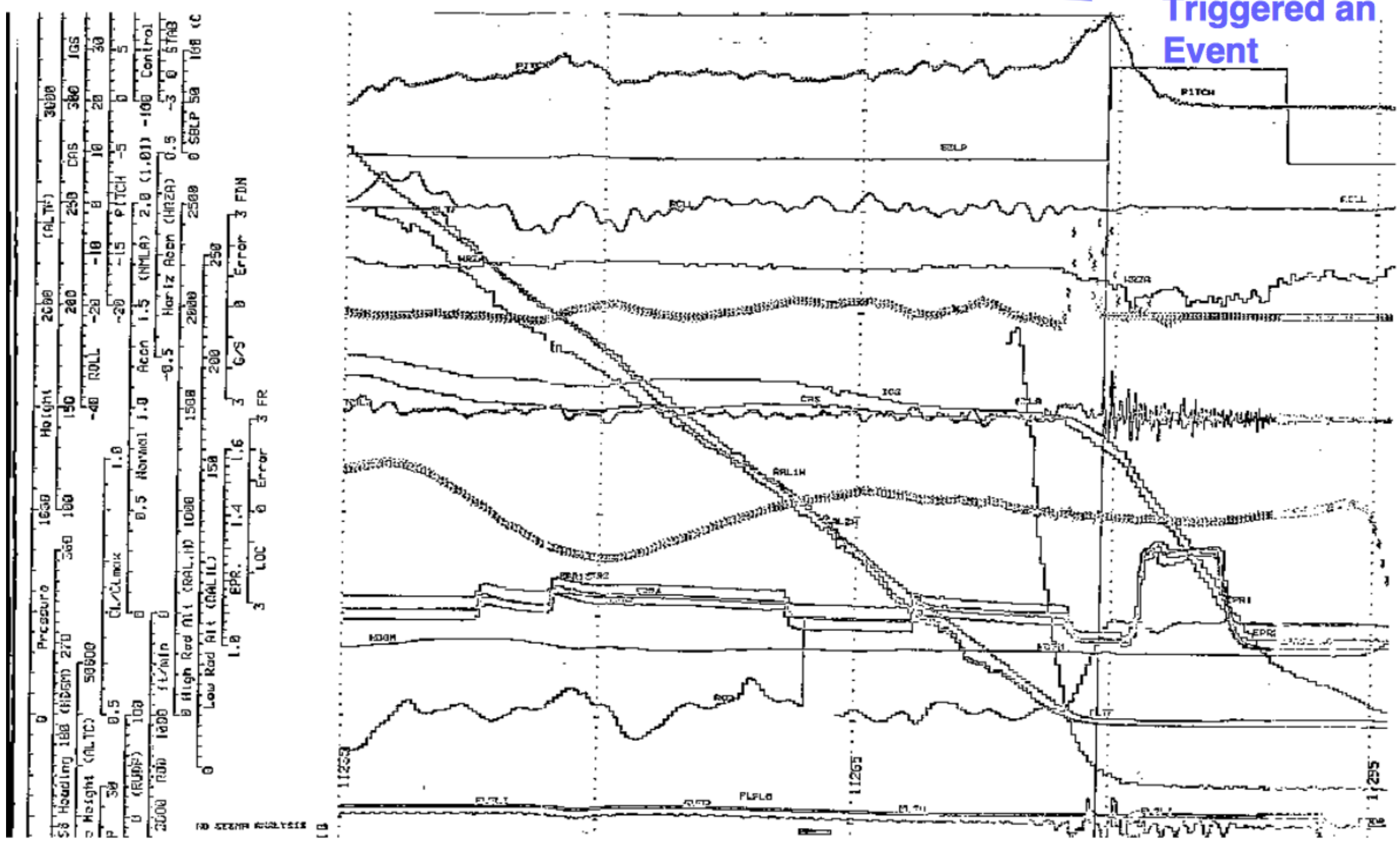
A Flight Data Recording System

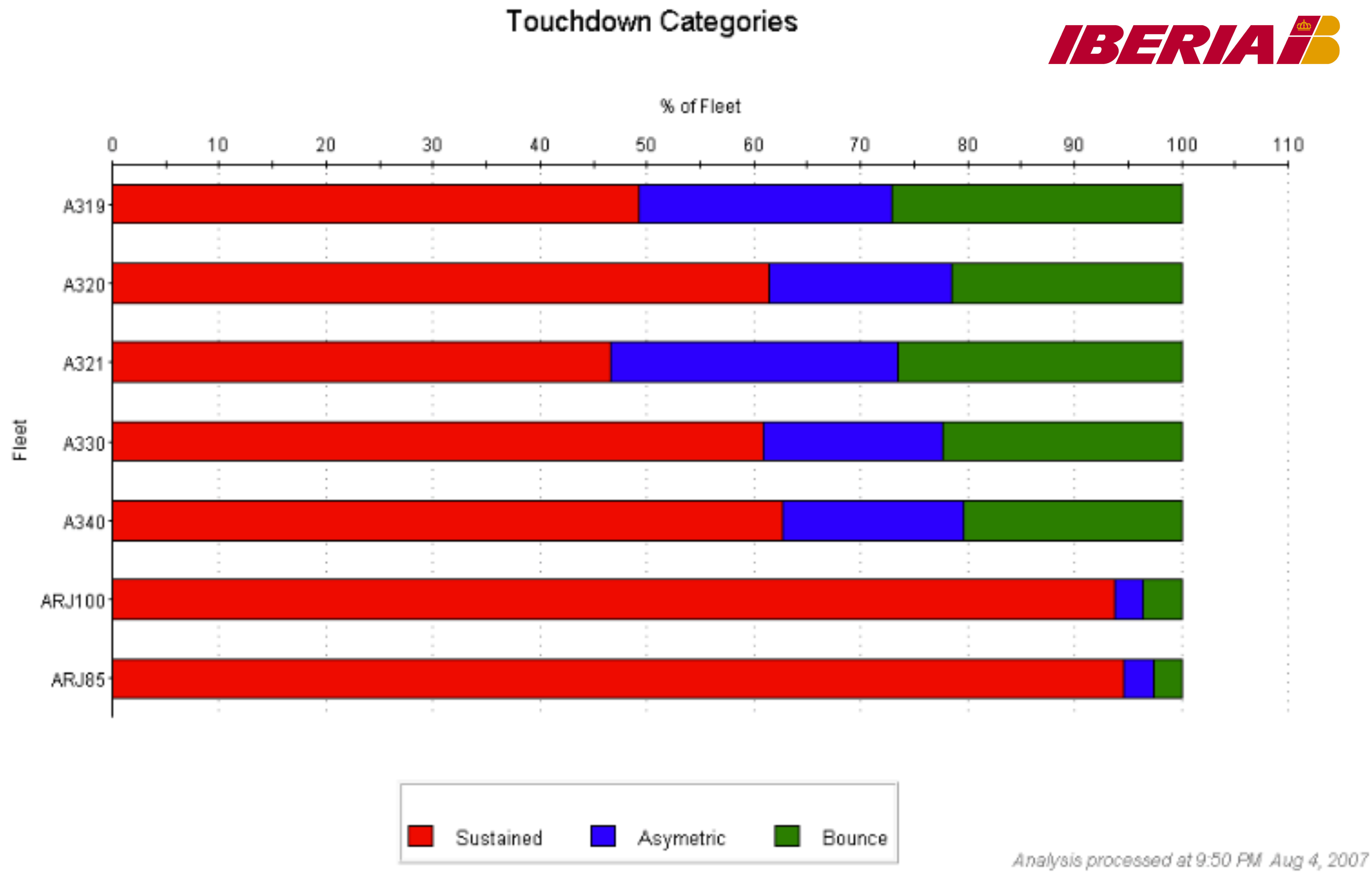


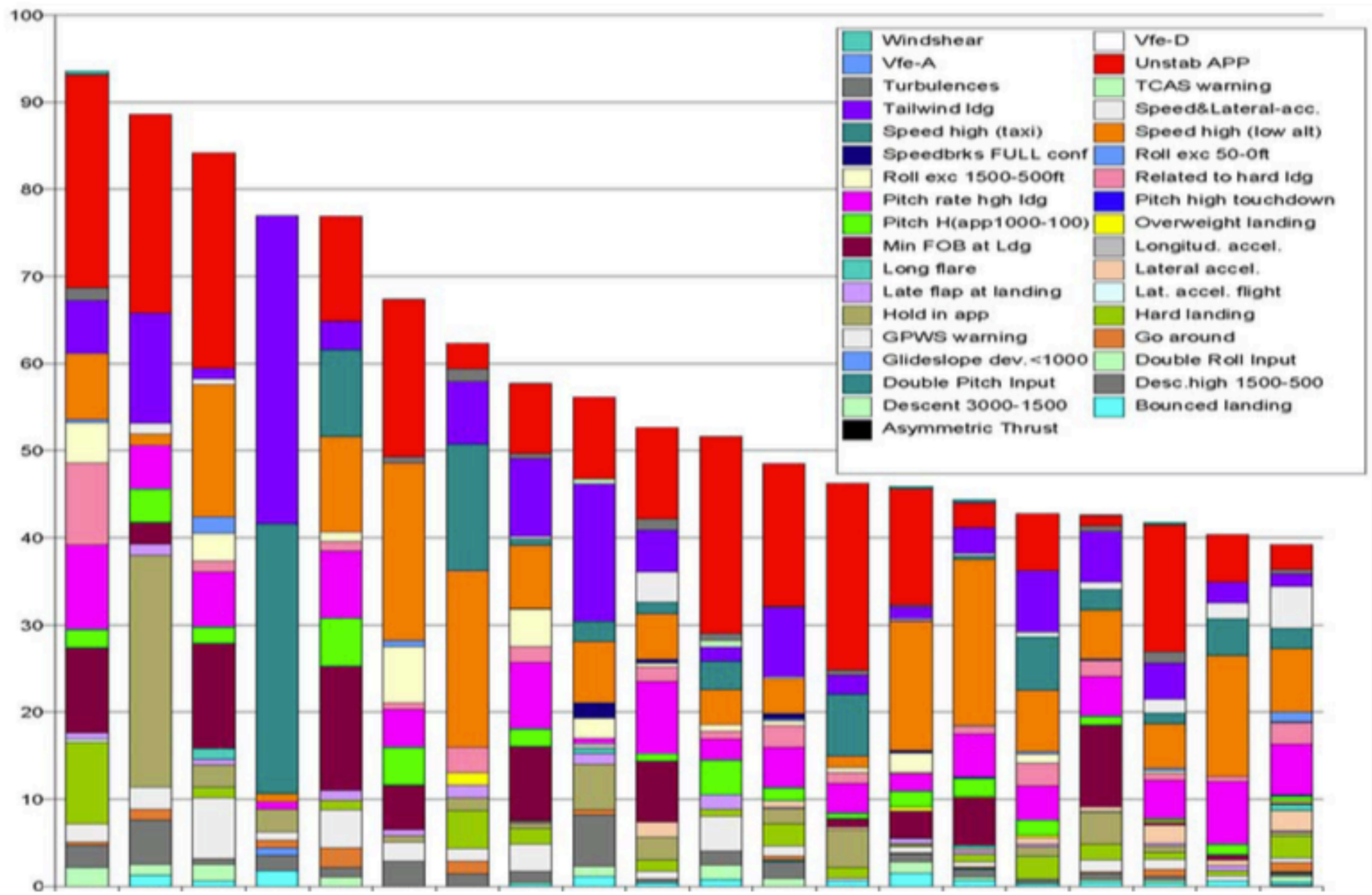
Flight Data Recording

Event Detection

High Pitch on Landing Triggered an Event

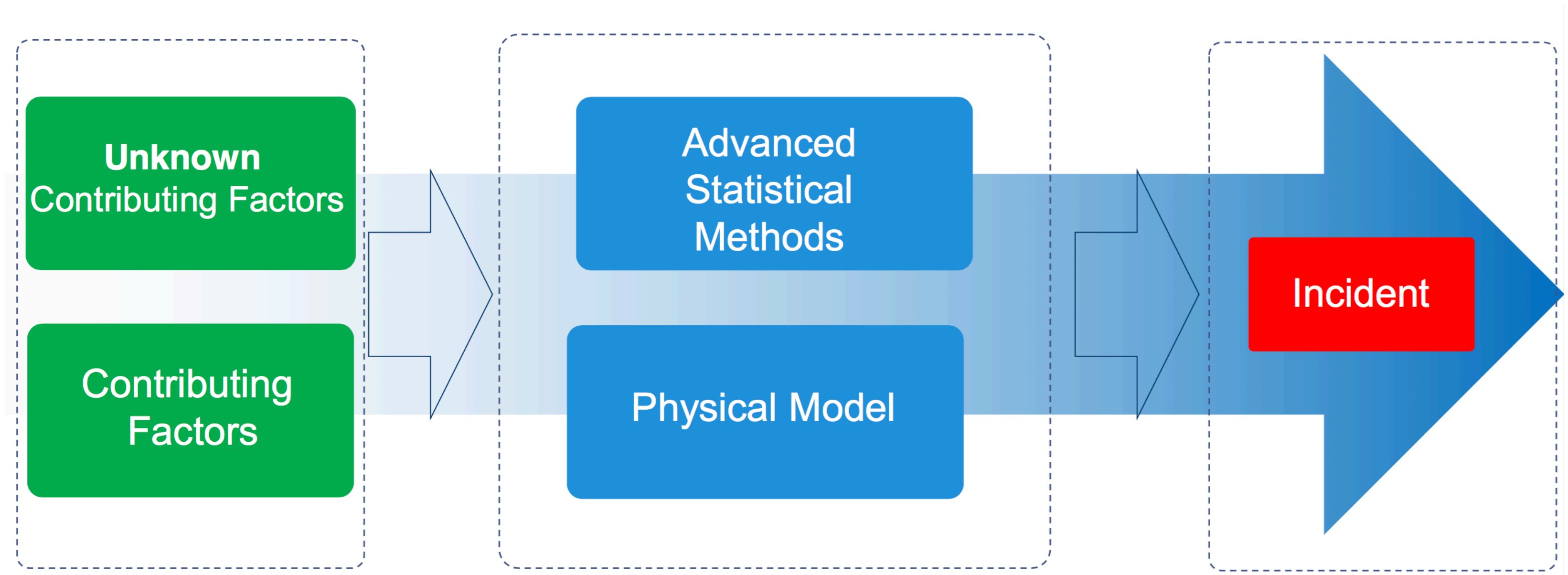



Touchdown Categories
IBERIA
% of Fleet
0
10
20
30
40
50
60
70
80
90
100
110
Fleet
A319
A320
A321
A330
A340
ARJ100
ARJ85
Sustained
Asymetric
Bounce
Analysis processed at 9:50 PM Aug 4, 2007

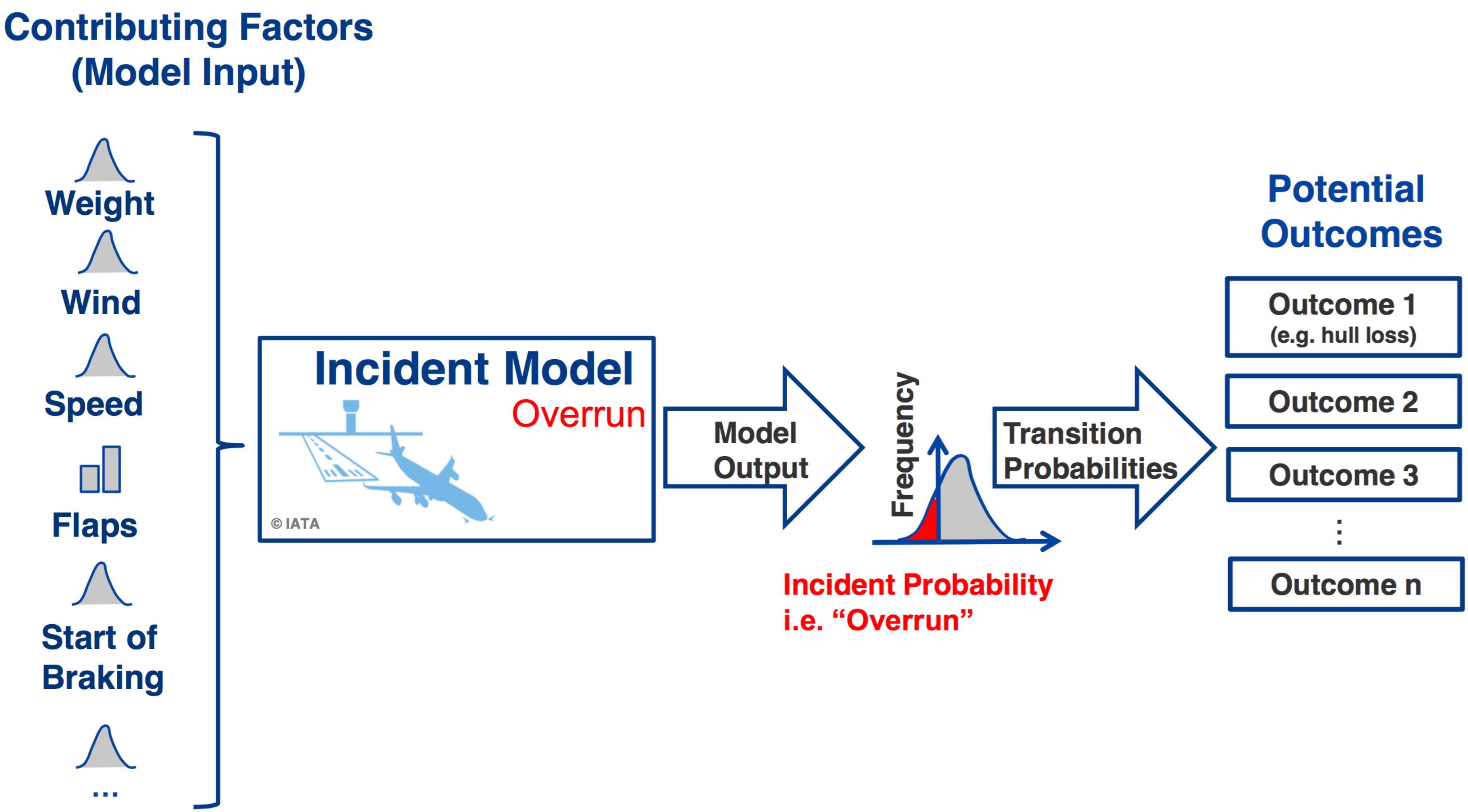


Ranking of European Airports with FDM events (*names erased purposedly*)

Predictive Incident analysis



PREDICTIVE INCIDENT ANALYSIS



HARD LANDING

Hard Landing

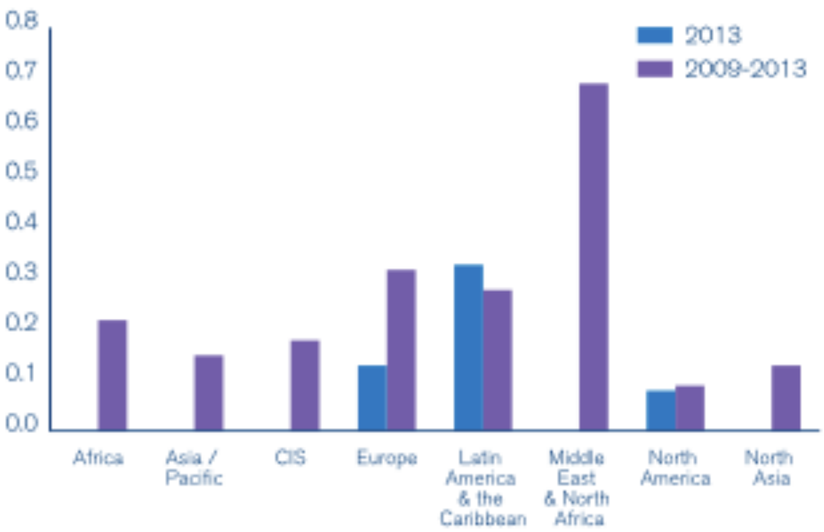
2013	3 Accidents
2009-2013	35 Accidents

	2013	'09-'13
IATA Members	0%	34%
Hull Losses	100%	31%
Fatal	0%	3%
Accident Rate	0.08	0.20

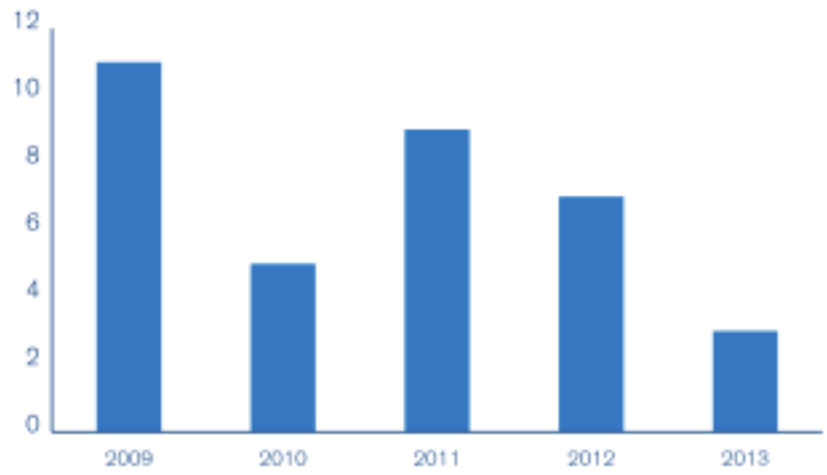
	Passenger	Cargo	Ferry	Jet	Turboprop
2013	67%	0%	33%	33%	67%
2009-2013	77%	20%	3%	71%	29%

Accident Rates per Operator Region

Accidents per million sectors flown for all aircraft types



Accidents per Year



Top Contributing Factors, 2009-2013

See Annex 2 for "Contributing Factors" definitions

Latent Conditions
(deficiencies in...)

31% Flight operations:
Training systems
(100% of these cases)
SOPs & checking
(40% of these cases)

16% Safety management

Threats

Environmental

47% Meteorology:
Wind/wind shear/gusty wind
(80% of these cases)
Poor visibility /IMC
(20% of these cases)

Airline

None noted.

Flight Crew Errors
(relating to...)

63% Manual handling/flight controls

28% Failure to go around after destabilized approach

22% SOP adherence/SOP cross-verification:
Unintentional non-compliance
(86% of these cases)

9% Automation

Undesired Aircraft States

75% Long/floated/bounced/firm/off-center/crabbed landing

22% Unstable approach

19% Vertical, lateral or speed deviations

13% Abrupt aircraft control

13% Continued landing after unstable approach

Countermeasures

25% Monitor/cross-check

25% Overall crew performance

16% Contingency management

13% Automation management

Additional Classifications

9% Insufficient data for contributing factors

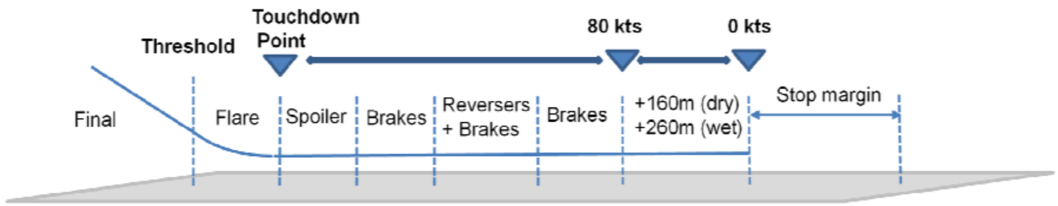
Relationships of Interest, 2009-2013

Manual handling of the aircraft was a factor in **71%** of hard landings where the crews landed long, floated, bounced, off-center or crabbed.

Deficiencies in flight operations training were noted in **56%** of hard landing accidents where the crew decision not to go around after destabilization also contributed to the accident.

Note: Three accidents were not classified due to insufficient data; these accidents were removed from the count for the contributing factors and relationships of interest.

Step 1 Incident metric



Runway overrun:
Stop margin < 0

Step 2 Functional relationships between contributing factors:

- **Physical relationships**

Aerodynamics

+

Propulsion

+

Brakes

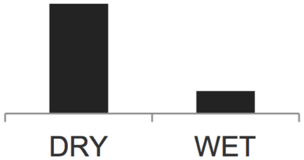
+

Gravitation

$$(\dot{u}_K^G)_B^{EB} = \frac{1}{m}\cdot\left[-m\cdot g\cdot sin\theta + \frac{\rho}{2}\cdot(V_A^A)^2\cdot S\cdot(-cos\beta_A\cdot C_D - cos\beta_A\cdot sin\beta_A\cdot C_Q) + (X_P^G)_B + \mu\cdot\left(-m\cdot g\cdot cos\Theta\cdot cos\Phi - \frac{\rho}{2}\cdot(V_A^A)^2\cdot S\cdot(-C_L)\right)\right]$$

- **Operational relationships**

Runway Condition



Procedures

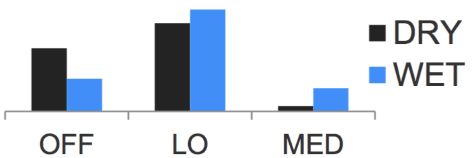
AUTO/BRK AS RQRD B

A330-300 / A340-300

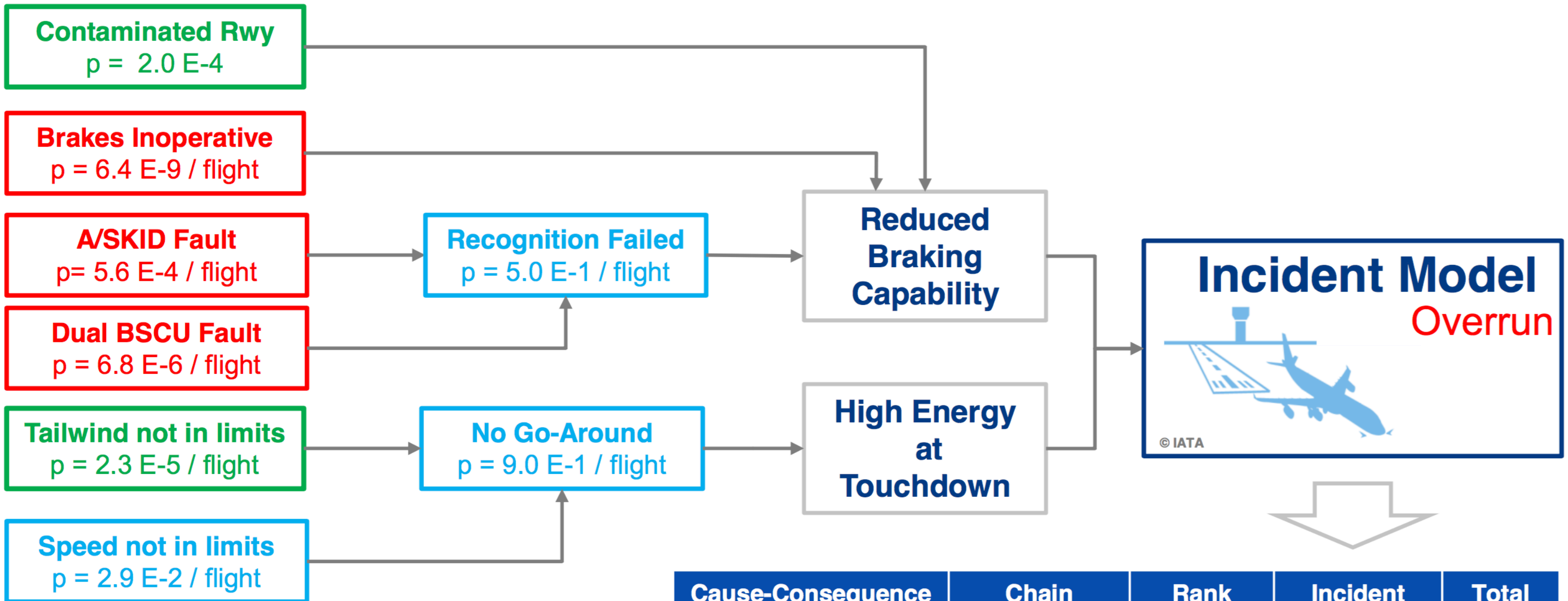
Use of autobrake is recommended.
Use of MAX mode is not recommended for landing.
On short or contaminated runways use MED mode.
On long runways, LO mode is recommended.

Note: If, on very long runways, it is anticipated that braking will not be needed, use of autobrake is

A/BRK Selection



CAUSE-CONSEQUENCE CHAINS



Cause-Consequence Chain	Chain Probability	Rank Chain	Incident Probability	Total Rank
Chain 1	2.6*10E-2	1	1.1 *10E-6	1
Chain 2	2.8*10E-4	2	2.4 *10E-9	3
Chain 3	2.1*10E-5	3	4.0 *10E-4	2
Chain 4	3.4*10E-9	4	2.3 *10E-11	4
Chain …	…	…	…	…

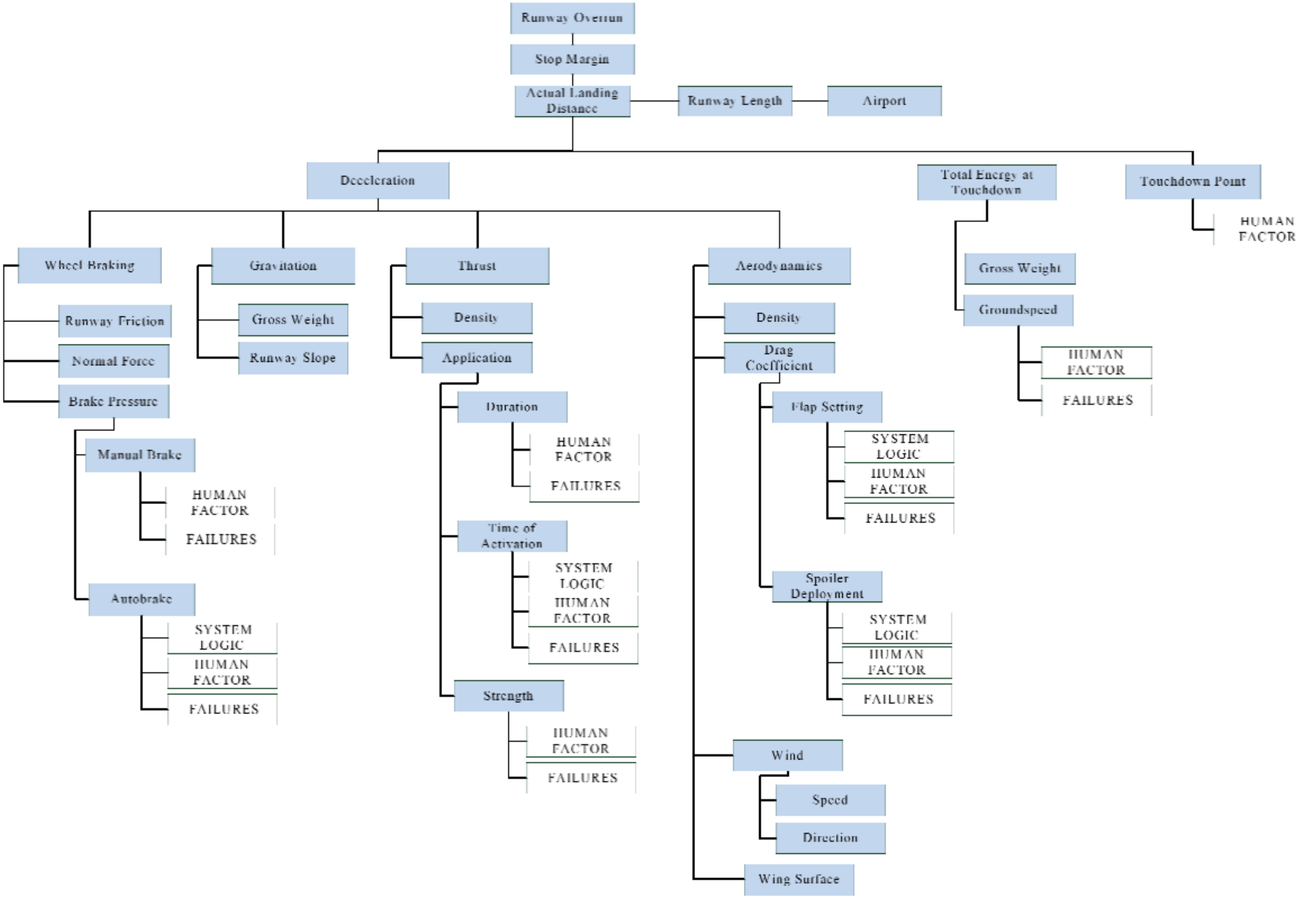
Human Performance

Environment

System Failures

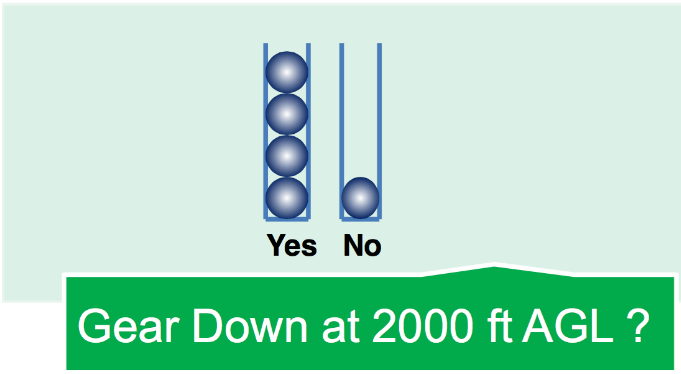
Numbers for illustration only!

CAUSE-CONSEQUENCE CHAINS

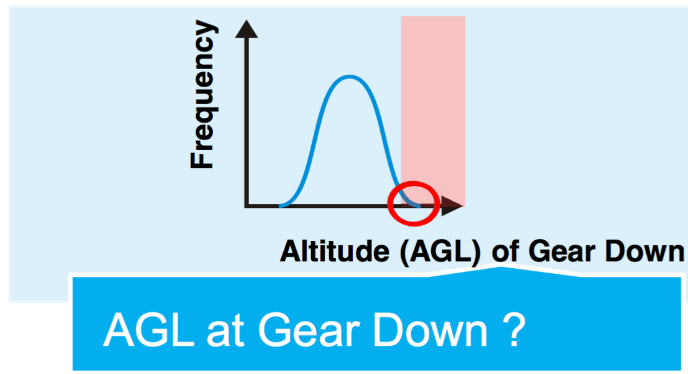


ASKING RIGHT QUESTION

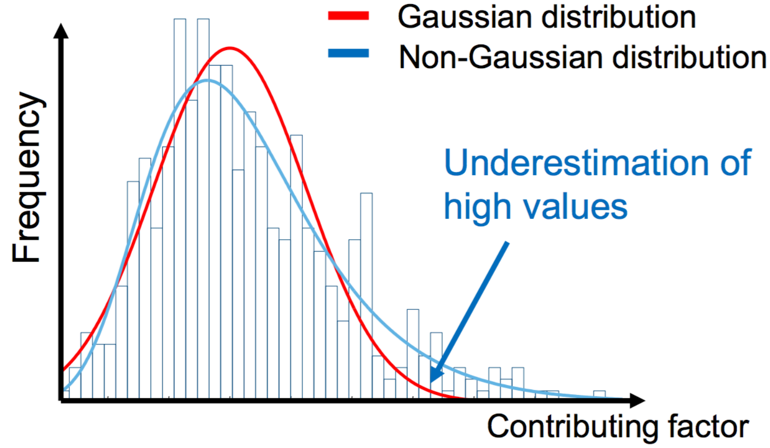
- Asking the right question can significantly increase the information we obtain.



vs.

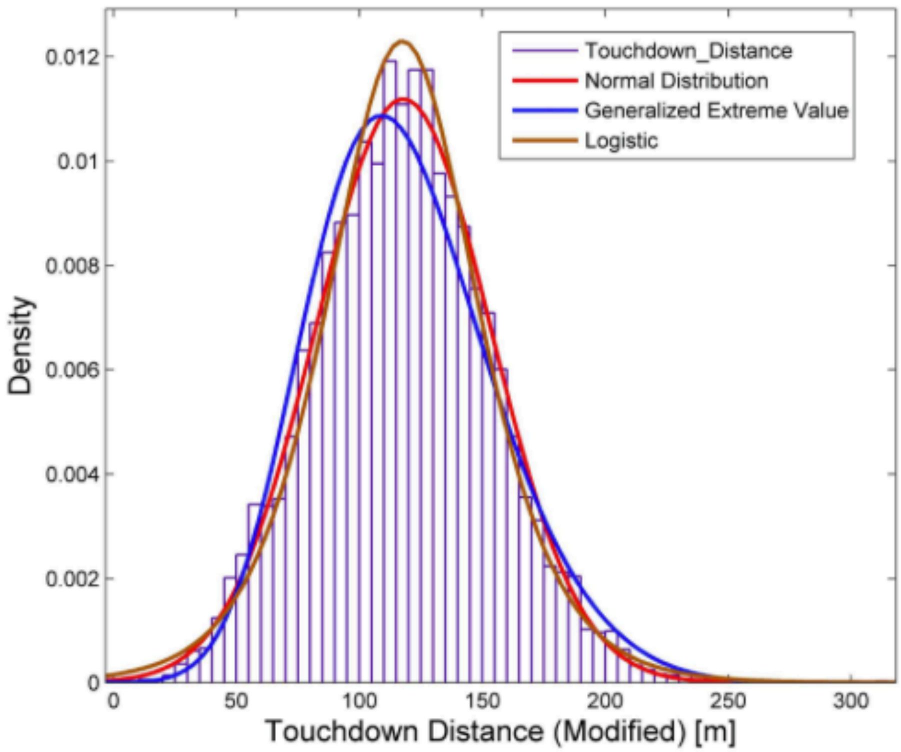


- Quality of statistical statements depend on how we look at the data.

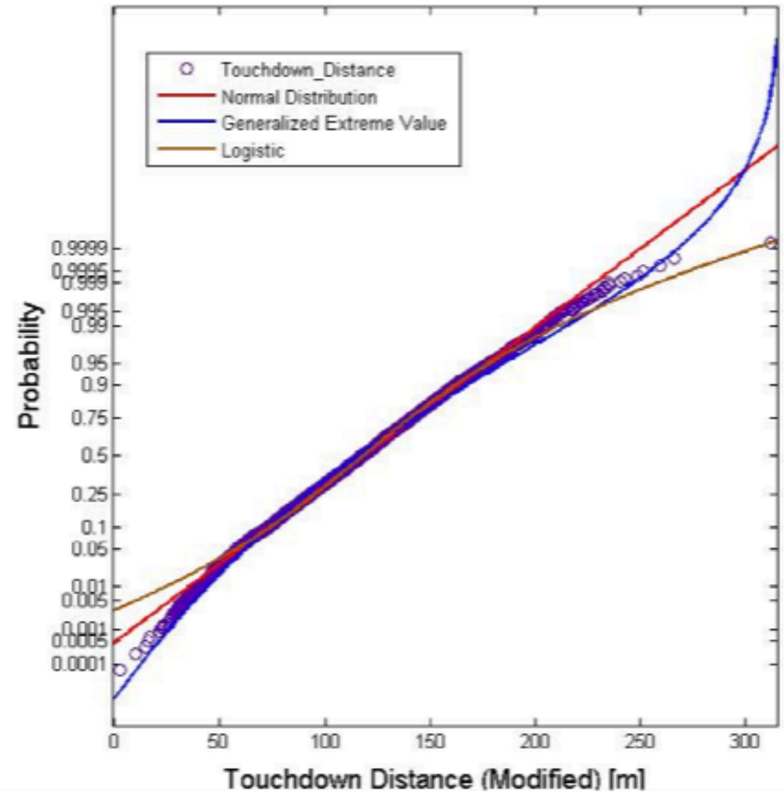


DISTRIBUTION FIT

Touchdown distances of 7263 landings in Frankfurt and Munich



a)



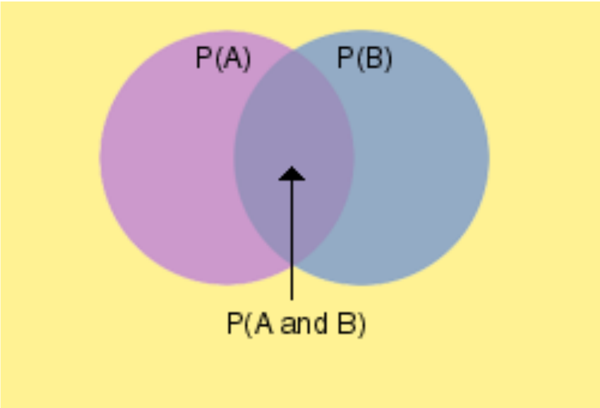
b)

$$P(Runway\_Overrun) = P(SM < 0) = \int_{-\infty}^{0} f(x)\,dx$$

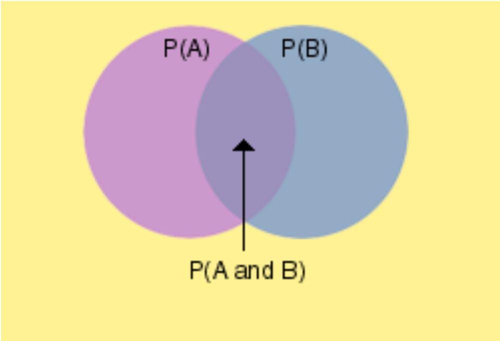
- to quantify the probability of these hazards
  - which happen quite often
  - use them to quantify the effect on the incident probability

CONDITIONAL PROBABILITY

- "Chance" of an event given that something is true
  - Notation:
  - $p(a|b)$
  - probability of event a, given b is true



- Diagnosis using a clinical test
  - Sample Space = all patients tested
    - Event A: Subject has disease
    - Event B: Test is positive



- Interpret:

$p(A \cap B)$ – Probability patient has disease and positive test (correct!)

$p(A \cap B')$ – Probability patient has disease BUT negative test (false negative)

$p(A' \cap B)$ – Probability patient has no disease BUT positive test (false positive)

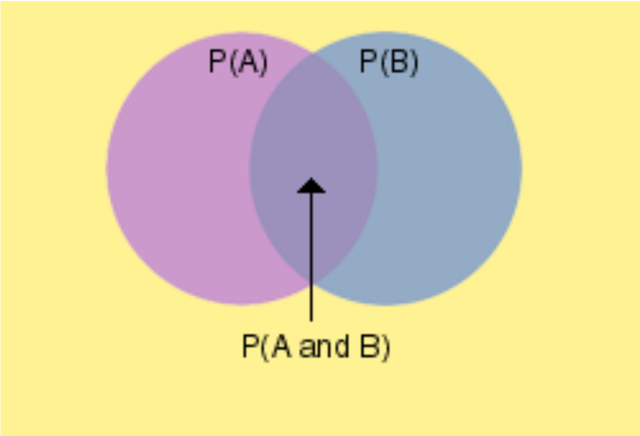
$p(A|B)$ – Probability patient has disease given a positive test

$p(A|B')$ – Probability patient has disease given a negative test

Conditional Probability Example

- If only data we have is B or not B, what can we say about A being true?
  - Not as simple as positive = disease, negative = healthy
  - Test is not infallible!
- Probability depends on intersection of A and B

$$p(A|B) = \frac{p(A \cap B)}{p(B)}$$



- Must Examine independence
  - Does $p(A)$ depend on $p(B)$?
  - Does $p(B)$ depend on $p(A)$?
  - Events are dependant

INDEPENDENCE

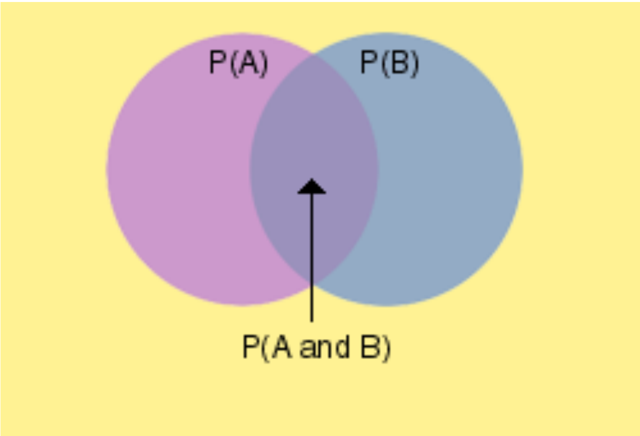
- Do A and B depend on one another?
  - Yes! B more likely to be true if A.
  - A should be more likely if B.
- If independent

$$p(A \cap B) = p(A) \cdot p(B)$$

$$p(A|B) = p(A) \quad p(B|A) = p(B)$$

- If dependent

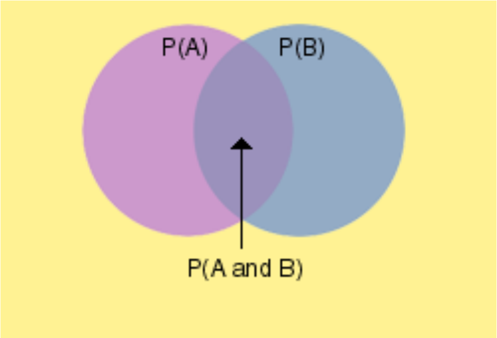
$$p(A \cap B) = p(B|A) \cdot p(A)$$



- Take events A_i for I = 1 to k to be:
  - Mutually exclusive: $A_i \cap A_j = 0$ for all i,j
  - Exhaustive: $A_1 \cup \cdots \cup A_k = S$
- For any event B on S

$$p(B) = p(B|A_1)p(A_1) + \cdots + p(B|A_k)p(A_k)$$

$$p(B) = \sum_{i=1}^{k} p(B|A_i)p(A_i)$$

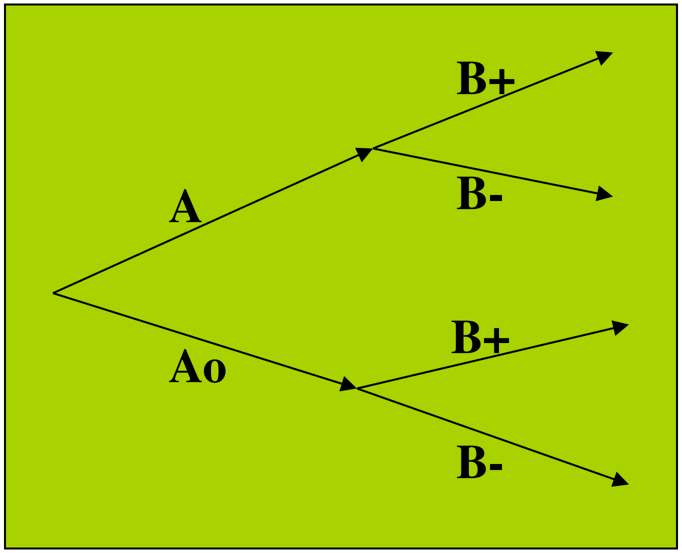


- Bayes theorem follows

$$p(A_j|B) = \frac{p(A_j \cap B)}{p(B)} = \frac{p(B|A_j) \cdot p(A)}{\sum_{i=1}^{k} p(B|A_i)p(A_i)}$$

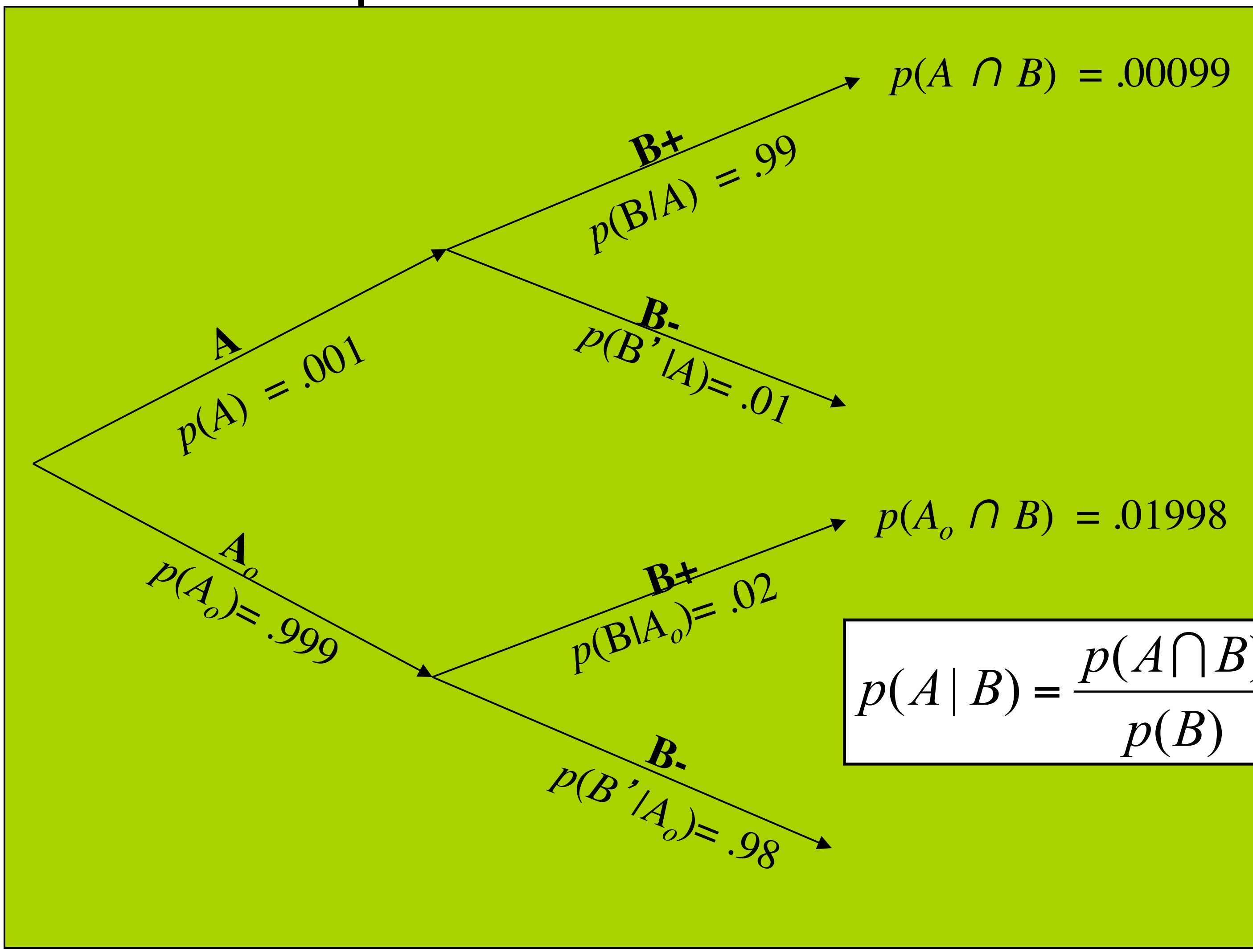
NUMERICAL EXAMPLE

- Only 1 in 1000 people have rare disease A
  - TP = .99 FP=.02
  - If one randomly tested individual is positive, what is the probability they have the disease
- Label events:
  - A = has disease A_o = no disease
  - B = Positive test result
- Examine probabilities
  - $p(A) = .001$
  - $p(A_o) = .999$
  - $p(B/A) = .99$
  - $p(B|A_o) = .02$



NUMERICAL EXAMPLE

- Examine probabilities



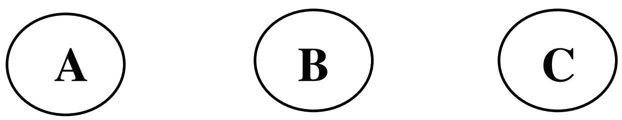
$$p(A \mid B) = \frac{p(A \cap B)}{p(B)} = \frac{.00099}{.02097} = .047$$

- Given a sequence of *n* outcomes $\{a_0, a_1, \ldots, a_n\}$
  - Where P(a_x) depends only on a_{x-1}

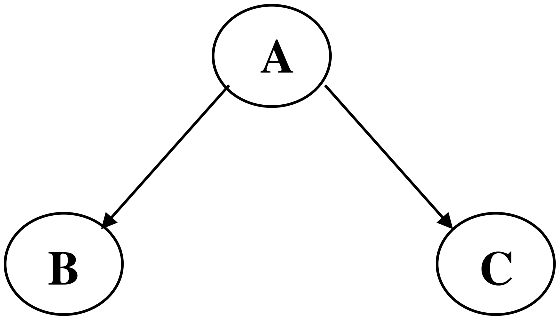
$$P(a_0, a_1, \ldots, a_n) = P(a_n \mid a_{n-`1}) \cdot P(a_{n-1} \mid a_{n-2}) \cdot \ldots \cdot P(a_1 \mid a_0) P(a_0)$$

- Probability of the sequence is given by the product of the probability of the first event with the probabilities of all subsequent occurrences
- Markov chains have been explored through simulation (Markov Chain Monte Carlo – MCMC)

EXAMPLES OF 3-WAY BAYESIAN NETWORKS



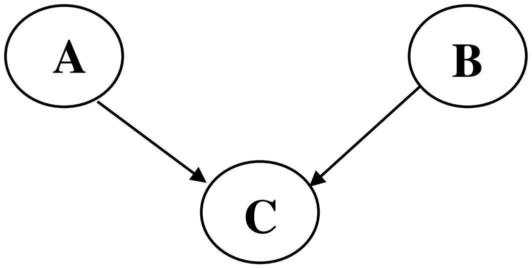
Marginal Independence:
p(A,B,C) = p(A) p(B) p(C)



Conditionally independent effects:
p(A,B,C) = p(B|A)p(C|A)p(A)

B and C are conditionally independent Given A

e.g., A is a disease, and we model B and C as conditionally independent symptoms given A

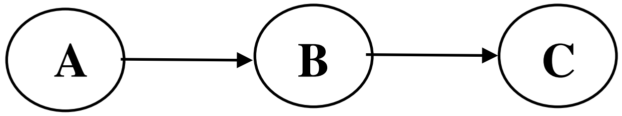


Independent Causes:
$p(A,B,C) = p(C|A,B)p(A)p(B)$

"Explaining away" effect:
Given C, observing A makes B less likely
e.g., earthquake/burglary/alarm example

A and B are (marginally) independent
but become dependent once C is known

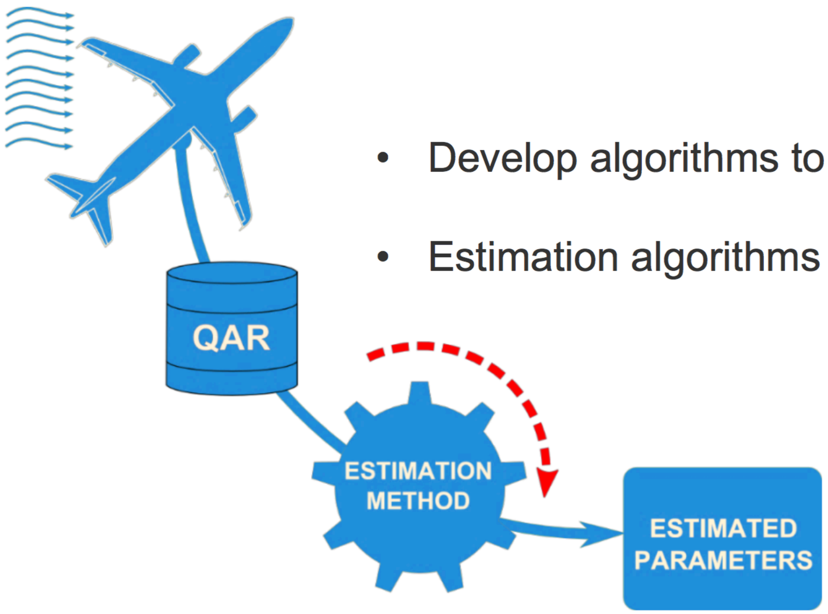
EXAMPLES OF 3-WAY BAYESIAN NETWORKS



Markov dependence:

$p(A,B,C) = p(C|B)\, p(B|A) p(A)$

PARAMETER ESTIMATION

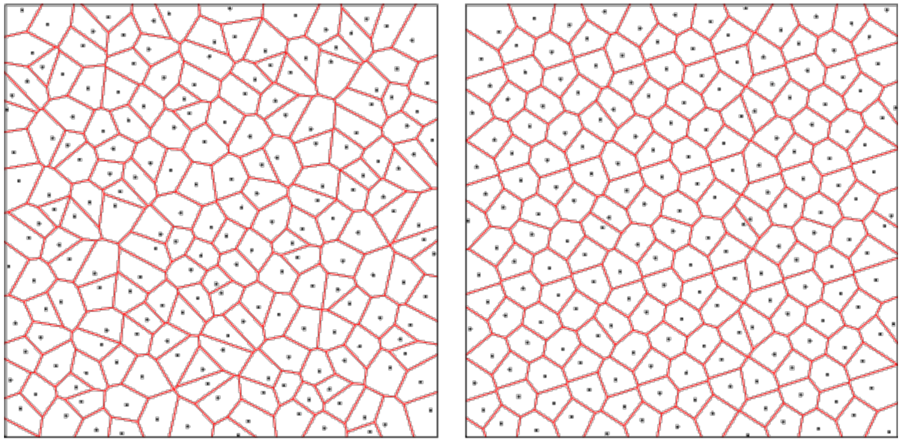


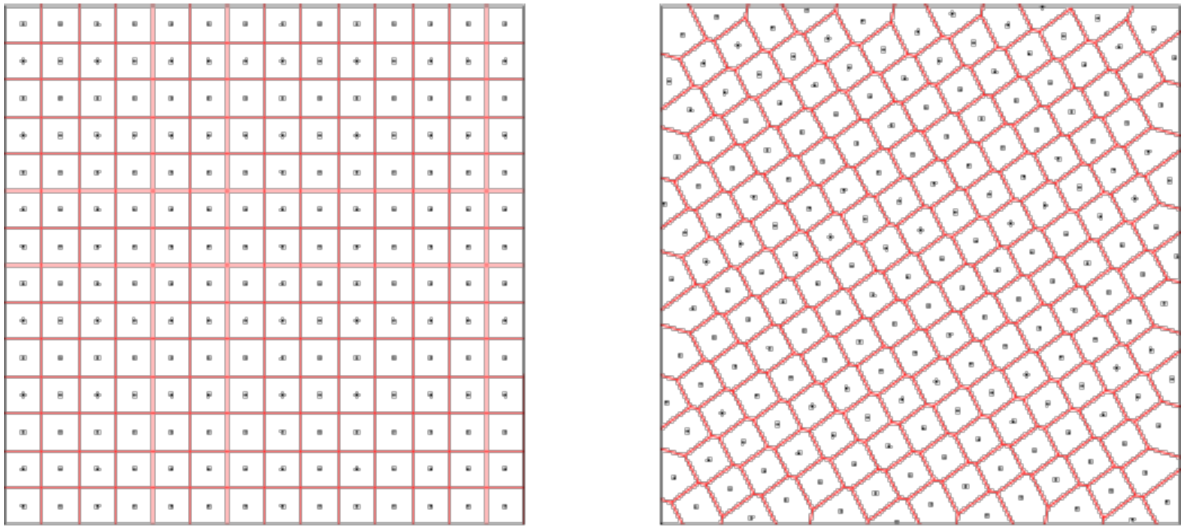
- Develop algorithms to extract non-measured contributing factors
- Estimation algorithms are applied to **every** single flight

Parameter Estimation Implementation during Ground Roll		
Parameter	Expected Value	Standard Deviation
$C_{D,G}$	0.1285	0.1517
$C_{D,GS}$	0.1373	0.0042
μ_{roll}	0.0197	0.0048
$\mu_{roll+brake}$	0.1123	0.0038

PARAMETER SAMPLING

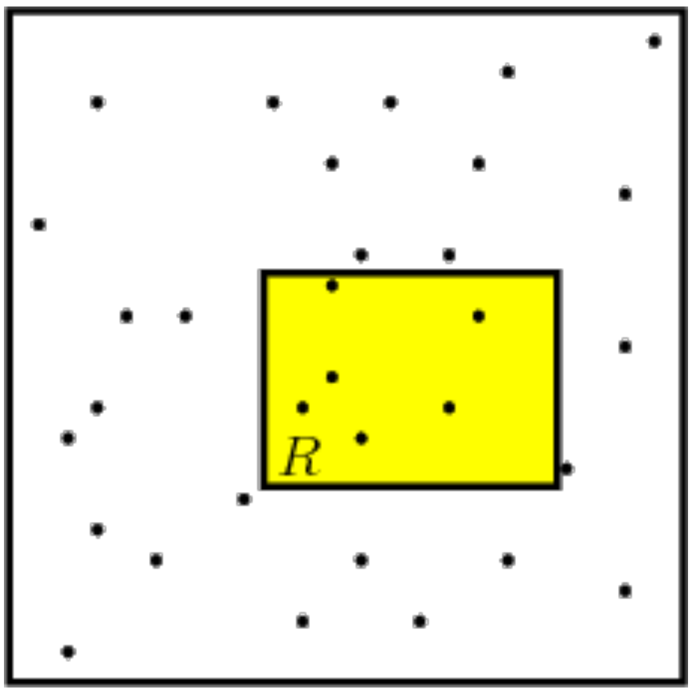
Uniform sampling?


(a) 196 Halton points (b) 196 Hammersley points


(a) 196-point Sukharev grid (b) 196 lattice points

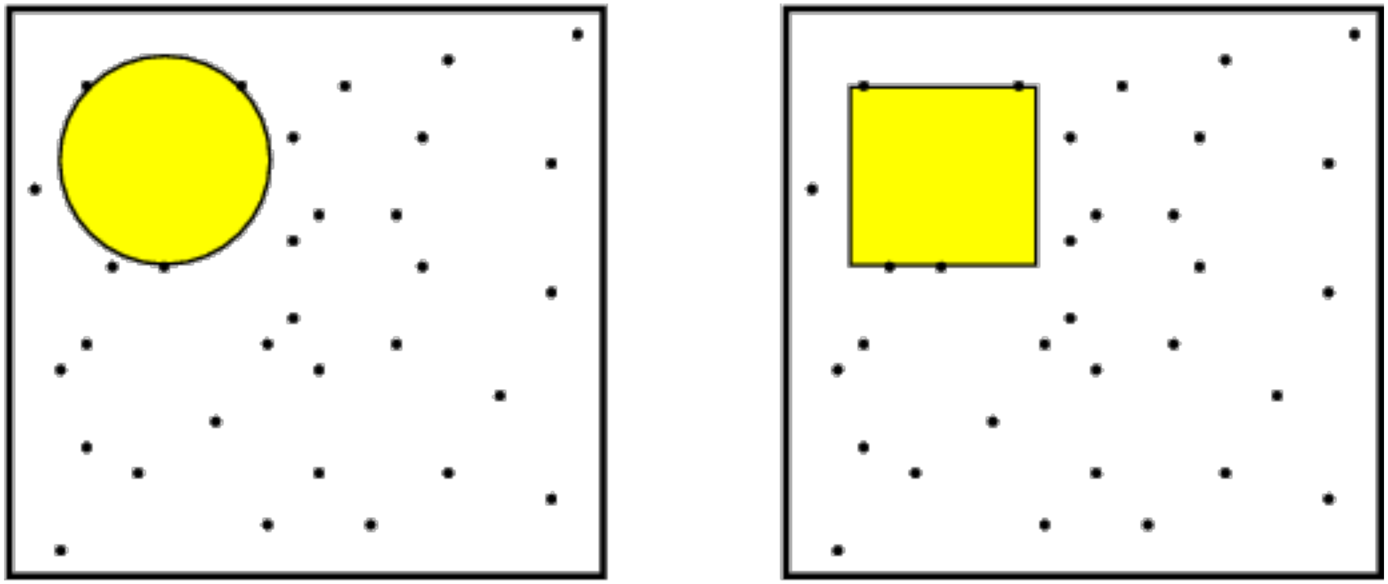
PARAMETER SAMPLING

Discrepancy measures whether the right number of points fall into boxes



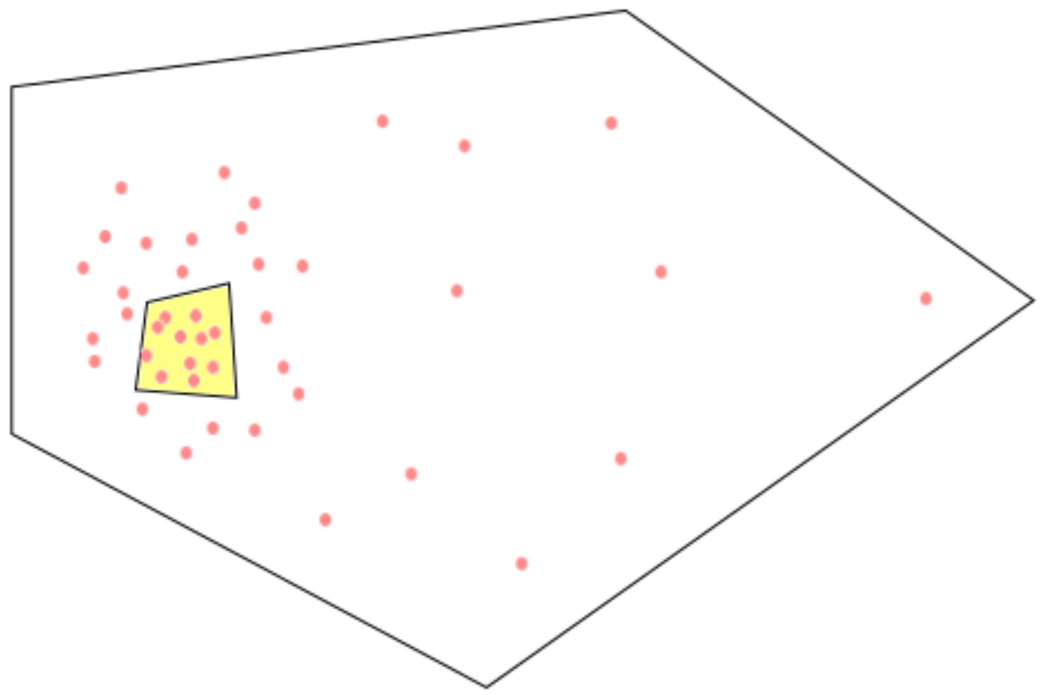
$$\frac{1}{N}\sum_{i=1}^{N} I_{\{S(\boldsymbol{x}_i)\geq\gamma\}}$$

Reducing the dispersion means reducing the radius of the largest empty ball

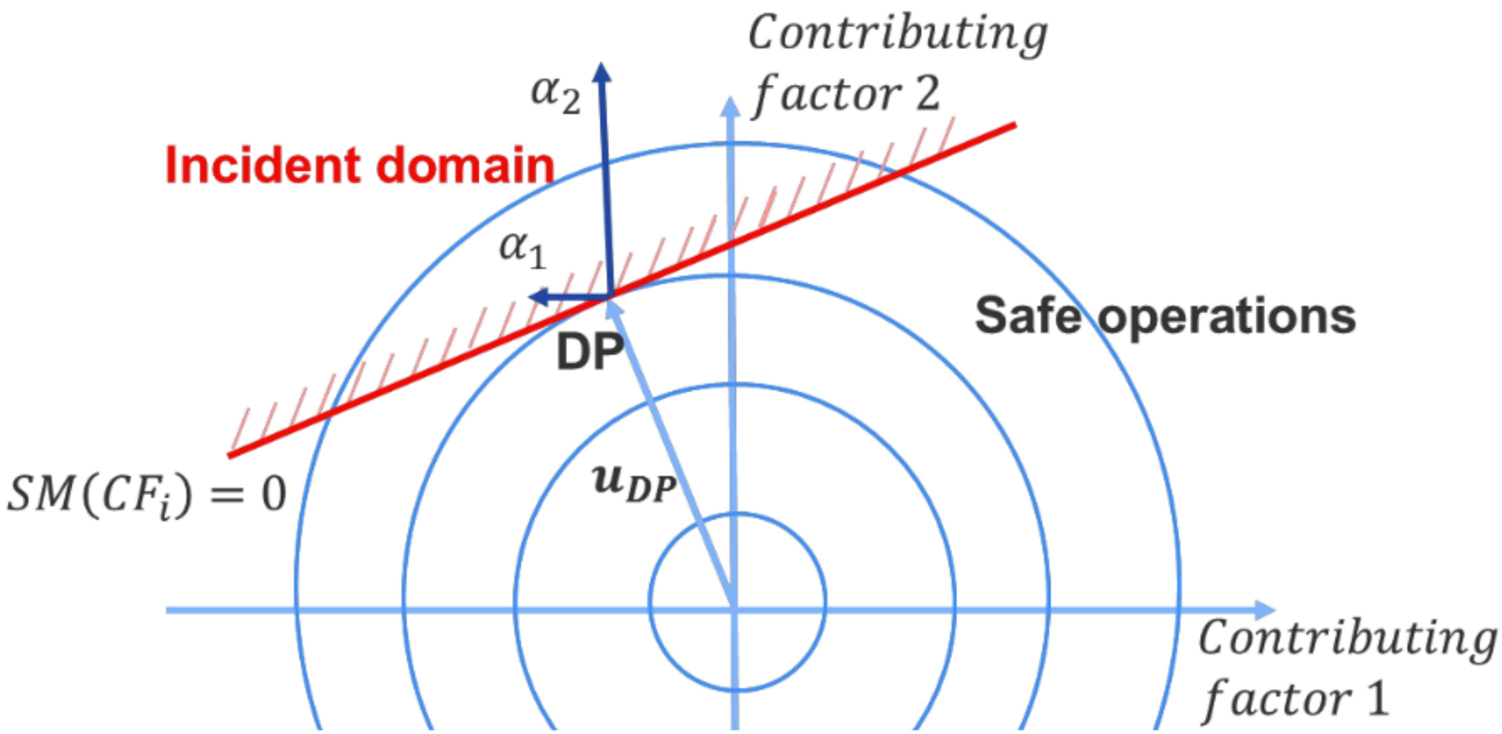


(a) L_2 dispersion   (b) L_∞ dispersion

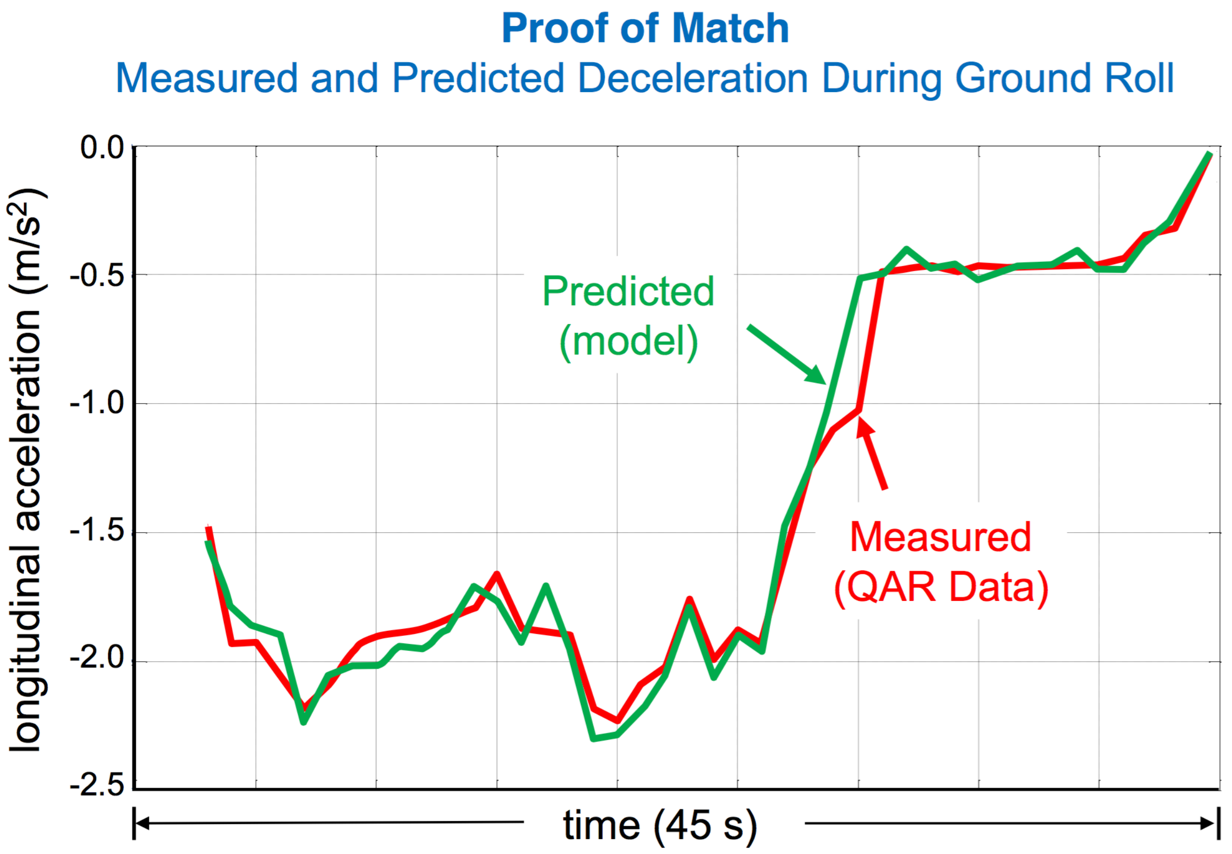
Importance Sampling



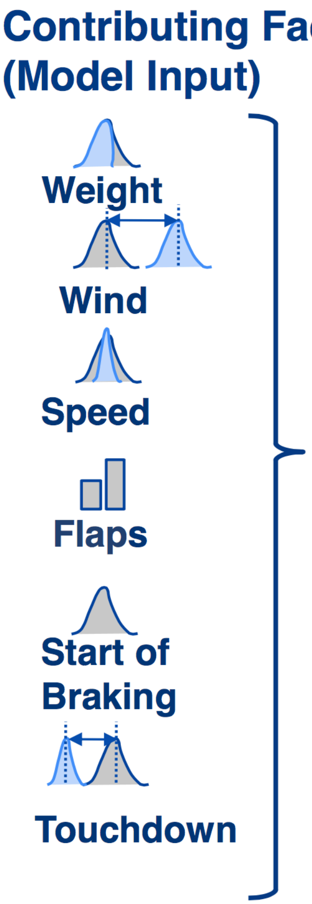
INCIDENT DOMAIN

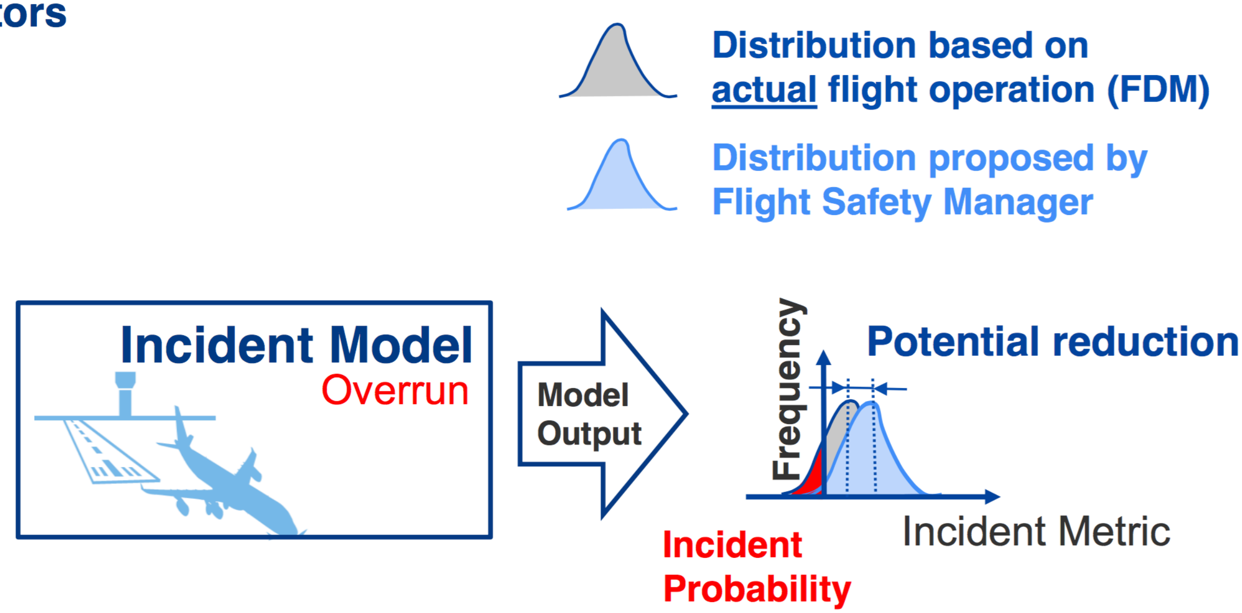


Proof of Model



CHANGE MANAGEMENT

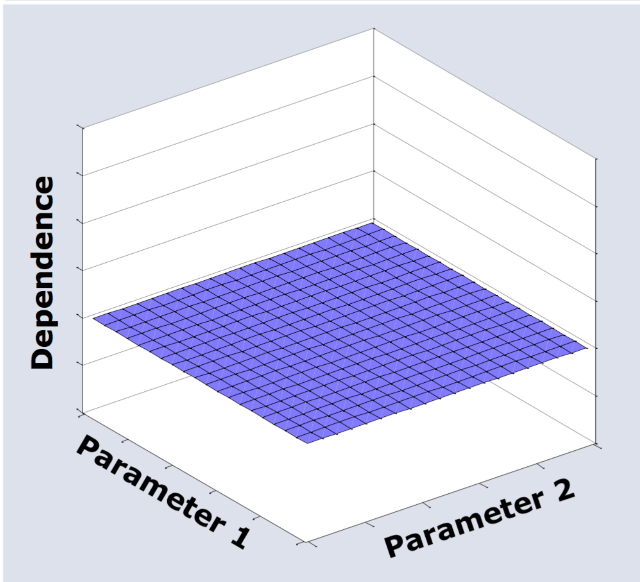




- Predictive analysis allows the assessment of the impact of mitigation actions **BEFORE** implementing them
- Impact of mitigation actions to **OTHER** incidents automatically considered (e.g. runway overrun vs. hard landing vs. tail strike)

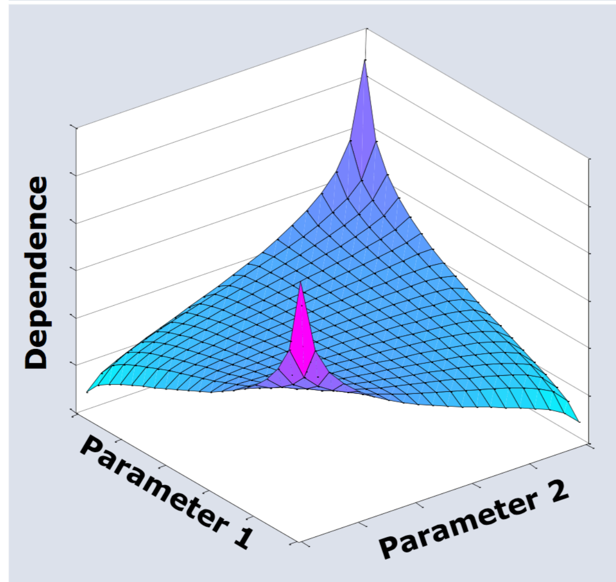
IDENTIFYING UNKNOWNS

Correlation Coefficient



Only captures CONSTANT dependency between two parameters

Copula



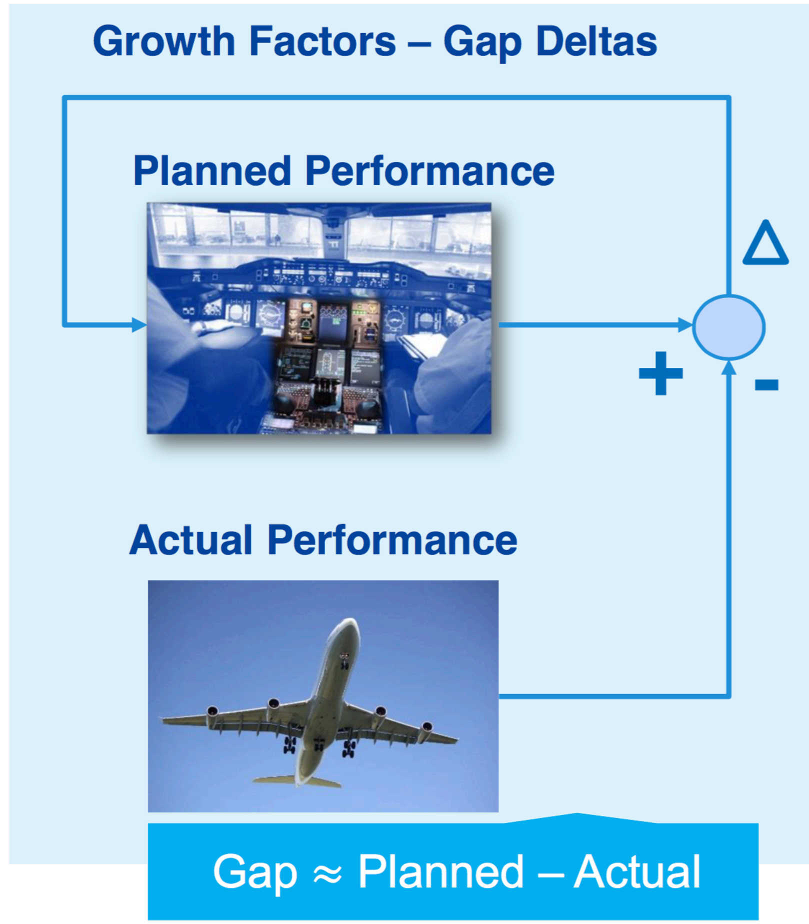
Capable of capturing VARIABLE (nonlinear) dependencies between more than two parameters

GAP ANALYSIS

1. Comparison between planned and actual performance
   - Takeoff planning
   - Landing distances
   - Fuel consumption
   - …

Mismatch can be expressed to quantify growth factors

2. Exploitation and correlation of further data sources:
   - ATM data
   - Weather data
   - Training data
   - Maintenance records
   - …



Predictive Analysis enables airlines:

PREDICTIVE ANALYSIS

- To QUANTIFY airline-specific incident and accident probabilities BEFORE things go wrong.
- To IDENTIFY and QUANTIFY HIDDEN and UNKNOWN contributing factors.
- To QUANTIFY the main drivers behind incidents.
- To QUANTIFY the effectiveness potential mitigation actions BEFORE implementing them.