Australian Government
Department of Defence
Science and Technology

25 years of particles and other random points

Neil Gordon, David Salmond and Adrian Smith

PERSONAL REFLECTIONS

Adrian Smith

University of London

Today: Commonplace approach

Complex applications

Graphical/hierarchical models

Bayesian computation via simulation

In particular, Particle Filtering

How did we get here?

- Mid-1960s Statistical Context?

CAMBRIDGE MATHEMATICS

- Mid-1960s
- Probability
- No statistics

POST-WAR DEBATES

- Fisher
- Neyman-Pearson

FISHER

- Maximum likelihood
- Significance tests
- Fiducial inference

NEYMAN-PEARSON

- Hypothesis tests
- MV unbiased estimation
- Properties of procedures

WALD (1950)

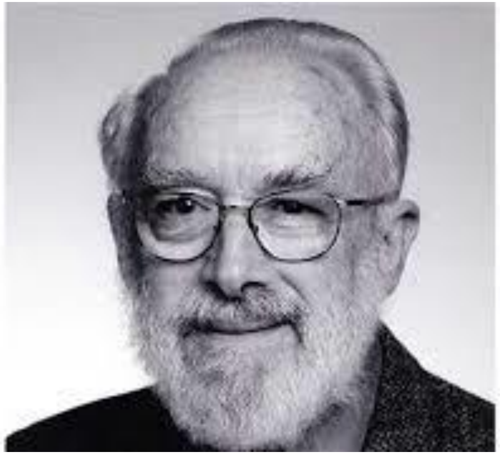
Decision theory

Complete class theorems

(Bayes as “mathematics”)

DENNIS LINDLEY

1923-2013



CLASH!

- Ideas
- Personalities

LINDLEY’S DISSATISFACTION

- Ad hoc approaches
- Cult of personality

MATHEMATICS IN GENERAL

- Desiderata
- Axioms
- Derived theory

(e.g. Kolmogorov/probability)

LINDLEY KNOWLEDGE BASE

- Bayes
- Laplace
- Gauss

- Jeffreys

(lectures)

AND …

- Wald!
- Decision theory
- Complete class theorems

EARLY 1950S

- Lindley’s intellectual goal
- Give statistics a firm axiomatic foundation

KEY 1953 PAPER
‘Statistical Inference’

Emphasises Bayes rule aspects of Wald

(Still a frequentist!)

1954

- Went to USA
- Work with L. J. Savage

ASTONISHING OUTCOME

Axiomatic attempts to underpin classical statistics inevitably led to Bayes!

LINDLEY/SAVAGE

Acknowledged debts to

- Ramsey
- De Finetti

1967
Head of Statistics Department
University College London

Opportunity to build Bayesian School

LINDLEY/SMITH (1972)

- Hierarchical models
- Structured priors
- Minimal use of vague priors
  (latter in low dimensions)

BAYES DECEPTIVELY SIMPLE

Posterior = Constant x Likelihood x Prior

Sequential learning: Posterior (t) = Prior (t+1)

BUT …

Big problem of computation!

NOTTINGHAM GROUP (1980s)
(efficient computation up to 8 dimensions)

Gauss-Hermite Quadrature

Reparameterization

Quasi Monte Carlo

BUT 8-D STILL WOEFULLY INADEQUATE!

- Eg Pharmacokinetics/dynamics
- 100’s of individual 5-D non-linear regressions
- Population distribution of 5-vectors

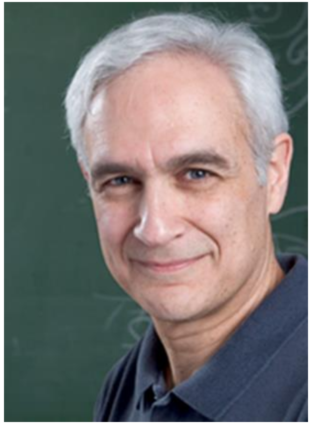
FOCUSSED ATTACK ON COMPUTATION (1988/89)

Review of related ideas

- EM algorithm (Dempster, Laird, Rubin, 1977)
- Image Analysis (Geman and Geman, 1984)
- Importance sampling (Rubin, 1987/1988)
- Laplace approximation (Tierney and Kadane, 1986)
- Data augmentation (Tanner and Wong, 1987)

Gibbs Sampler

Donald Geman



Stuart Geman

(1984)


RECENT ADVANCES
BIOSTATISTICS

Breakthrough (with Alan Gelfand)

Sampling-Based Approach to Calculating Marginal Densities

(Technical Report 1988: JASA 1990)

Illustration paper with Amy Racine-Poon, Sue Hills

Explosion of …

- MCMC methodological developments
- Application to complex problems

Don Rubin

Importance sampling
(Rubin, 1987/1988)

Sampling-Importance-Resampling

Replace Prior/Posterior Functions

By Sampled Point Clouds

(Non-Linear Signal Processing)

Goodbye‘tweaks’ to Kalman Filters!

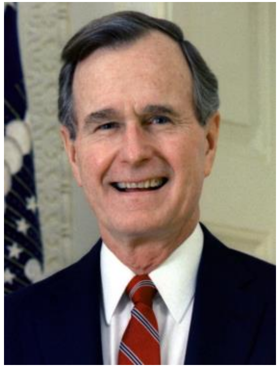
THE REST IS HISTORY!

Australian Government
Department of Defence
Science and Technology

David Salmond

Back in 1991, 1992 …

John Major was the British Prime Minister



George Bush Senior was in White House (soon to be replaced by Bill Clinton)

Boris Yeltsin was in the Kremlin

Recursive Bayesian Estimation:

Available Information (discrete case)

- System model (dynamics)

$$\boldsymbol{x}_k = \boldsymbol{f}_{k-1}(\boldsymbol{x}_{k-1}, \boldsymbol{w}_{k-1}) \quad \leftrightarrow \quad p(\boldsymbol{x}_k \mid \boldsymbol{x}_{k-1})$$

Transition Density

- Measurement model

$$\boldsymbol{z}_k = \boldsymbol{h}_k(\boldsymbol{x}_k, \boldsymbol{v}_k) \quad \leftrightarrow \quad p(\boldsymbol{z}_k \mid \boldsymbol{x}_k)$$

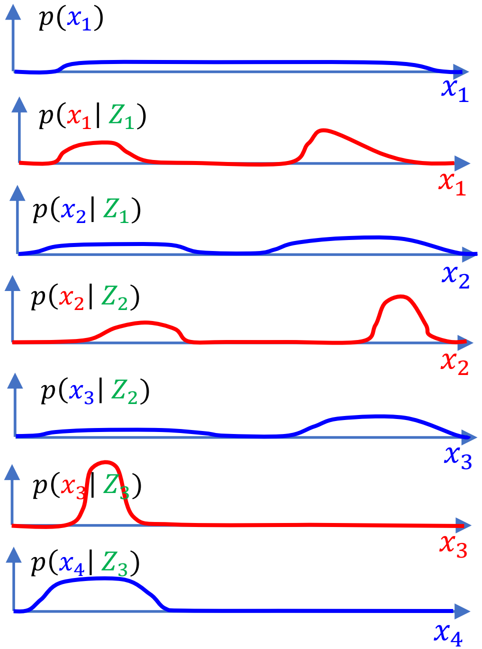
Likelihood

Prior at k=0

- Initial information (prior): $p(\boldsymbol{x}_0)$
- Measurements: $Z_k = \{\boldsymbol{z}_1, \dots, \boldsymbol{z}_k\}$
- Require: $p(\boldsymbol{x}_k \mid Z_k)$ **Posterior pdf**

(at each time step k)

The recursive Bayesian estimator can be viewed as a sequence of static problems (updates) interspersed by dynamic transitions (predictions)



$p(\boldsymbol{x}_0) \rightarrow p(\boldsymbol{x}_1)$ Prediction from k=0 to k=1

$p(\boldsymbol{x}_1) \rightarrow p(\boldsymbol{x}_1 \mid Z_1)$ Update at k=1 using measurement $\boldsymbol{z}_1$ ($Z_1 = \{\boldsymbol{z}_1\}$)

$p(\boldsymbol{x}_1 \mid Z_1) \rightarrow p(\boldsymbol{x}_2 \mid Z_1)$ Prediction from k=1 to k=2

$p(\boldsymbol{x}_2 \mid Z_1) \rightarrow p(\boldsymbol{x}_2 \mid Z_2)$ Update at k=2 using measurement $\boldsymbol{z}_2$ ($Z_2 = \{\boldsymbol{z}_1, \boldsymbol{z}_2\}$)

$p(\boldsymbol{x}_2 \mid Z_2) \rightarrow p(\boldsymbol{x}_3 \mid Z_2)$ Prediction from k=2 to k=3

$p(\boldsymbol{x}_3 \mid Z_2) \rightarrow p(\boldsymbol{x}_3 \mid Z_3)$ Update at k=3 using measurement $\boldsymbol{z}_3$ ($Z_3 = \{\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3\}$)

$p(\boldsymbol{x}_3 \mid Z_3) \rightarrow p(\boldsymbol{x}_4 \mid Z_3)$ Prediction from k=3 to k=4

General (Formal) Bayesian Recursive Estimator

TRANSITION DENSITY
(SYSTEM DYNAMICS)

Prediction:

$$\boxed{p(\mathbf{x}_{k+1} \mid Z_k)} = \int \overbrace{p(\mathbf{x}_{k+1} \mid \mathbf{x}_k)}\, \overbrace{p(\mathbf{x}_k \mid Z_k)}\, d\mathbf{x}_k$$

LIKELIHOOD
(MEASUREMENT MODEL)

$k := k + 1$

Update:
(Bayes rule)

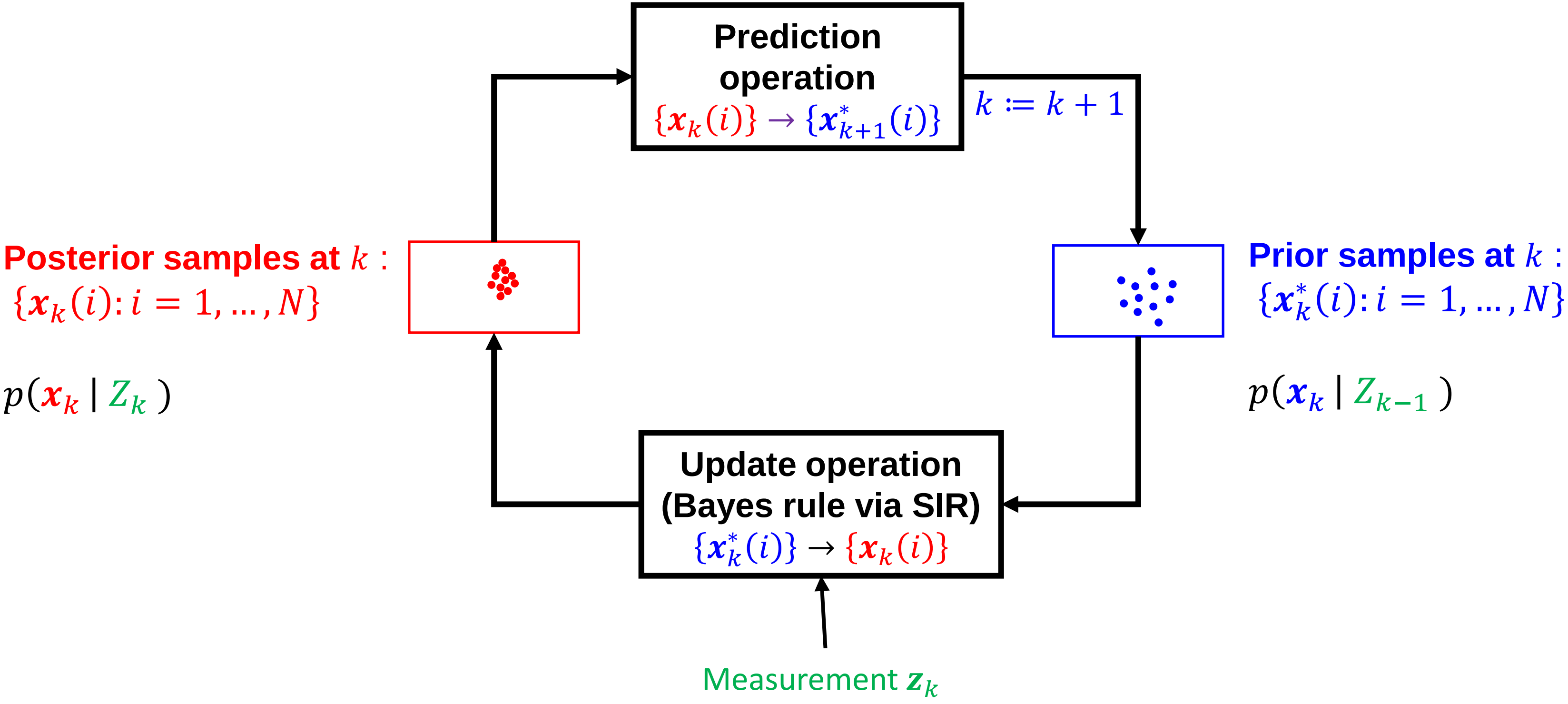
$$\boxed{p(\mathbf{x}_k \mid Z_k)} \propto \overbrace{p(\mathbf{z}_k \mid \mathbf{x}_k)}\, \overbrace{p(\mathbf{x}_k \mid Z_{k-1})}$$

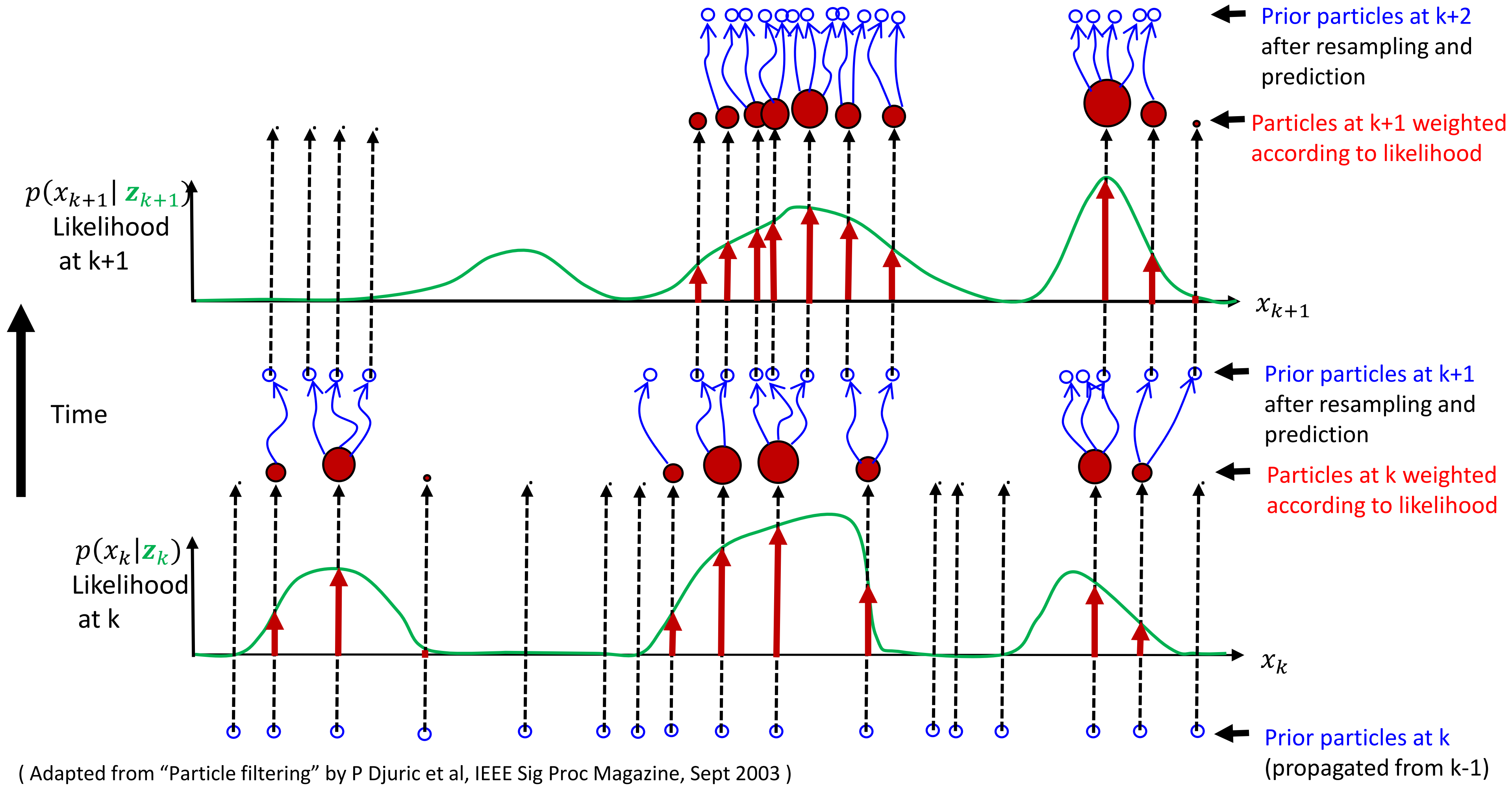
POSTERIOR

PRIOR

Ho and Lee, "A Bayesian approach to problems in stochastic estimation and control", IEEE Trans AC-9, Oct 1964, pp 333-339.

Basic particle filter: manipulate sets of random samples
(a mechanisation of the formal Bayesian recursive filter)





(Adapted from "Particle filtering" by P Djuric et al, IEEE Sig Proc Magazine, Sept 2003)

1-D example, “nonstationary growth model”

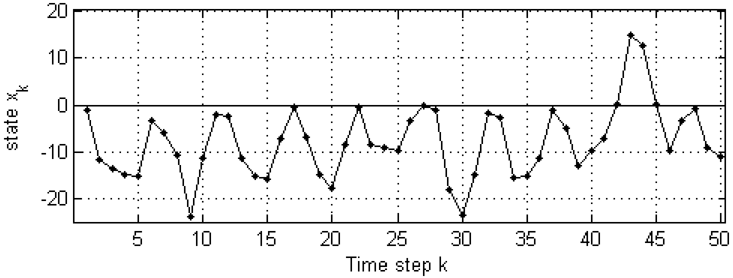
[From KITTAGAWA, G.: ’Rejoinder to Non-Gaussian state-space modelling of nonstationary time series’, *J . Amer. Statistical Assoc.,* 1987, Vol 82, pp 1060-1063]

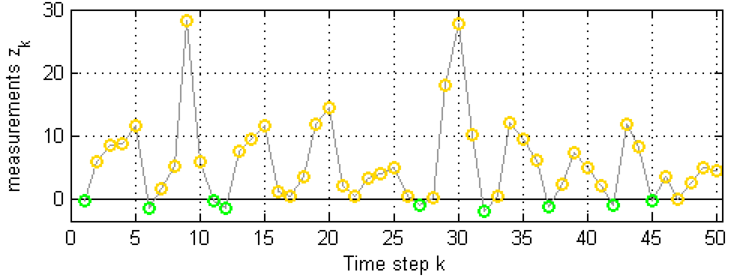
Dynamics model: $$x_k = \frac{x_{k-1}}{2} + \frac{25\, x_{k-1}}{1+ x_{k-1}^2} + 8\cos(1.2(k-1)) + w_k$$

Measurement model: $$z_k = \frac{x_k^2}{20} + v_k$$

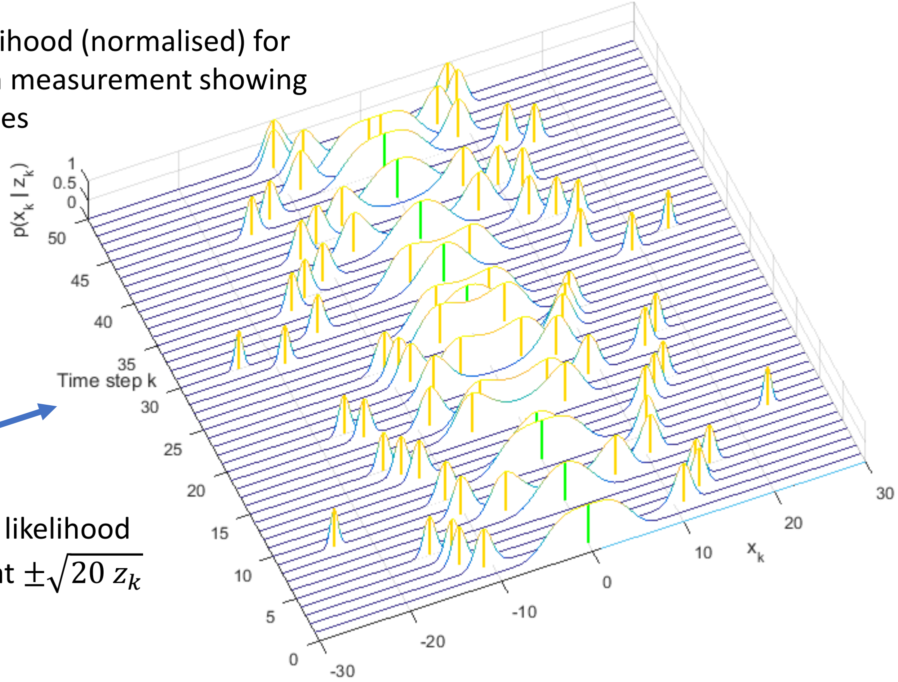
where $w_k \sim N(0,10)$ and $v_k \sim N(0,1)$.

A realisation of the time series and associated measurements :



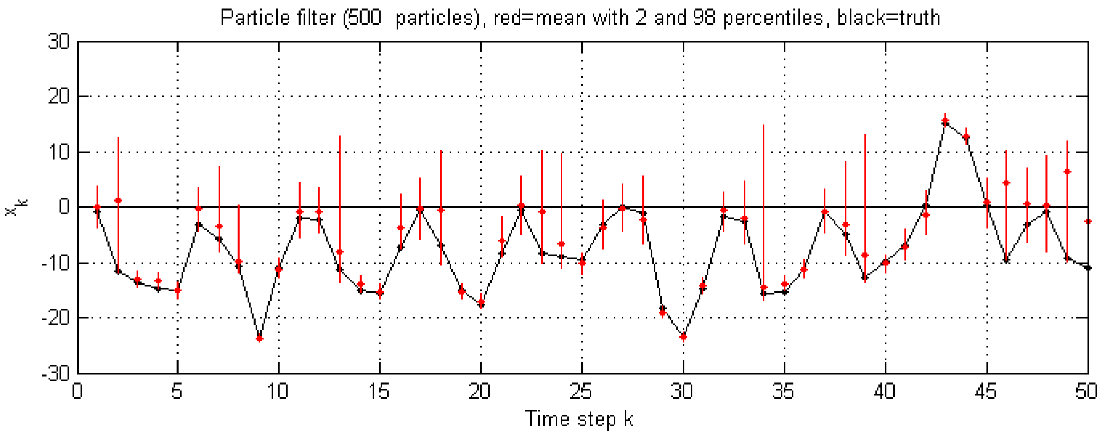


Likelihood (normalised) for each measurement showing modes

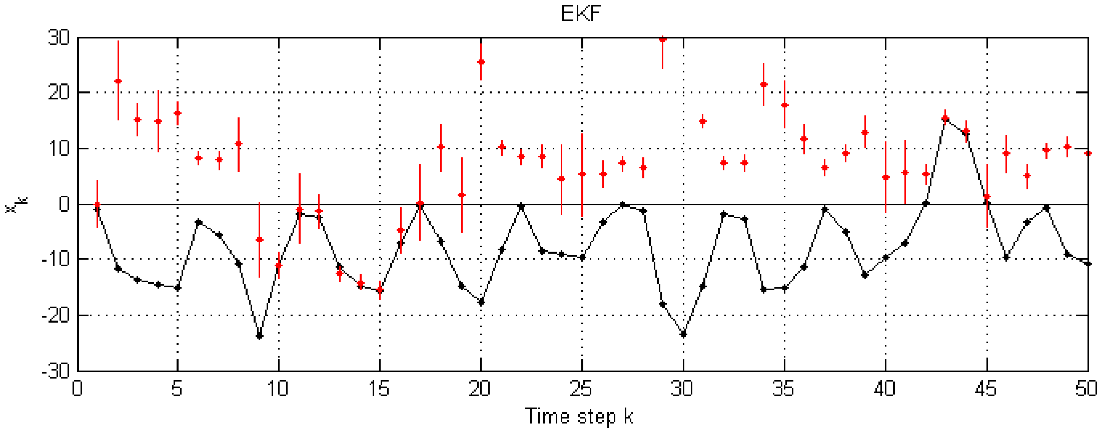


For $z_k > 0$, likelihood has modes at $\pm\sqrt{20\, z_k}$

Point estimates (posterior means) from particle filter and EKF



Particle filter: truth maintained between 2 and 98 percentile interval

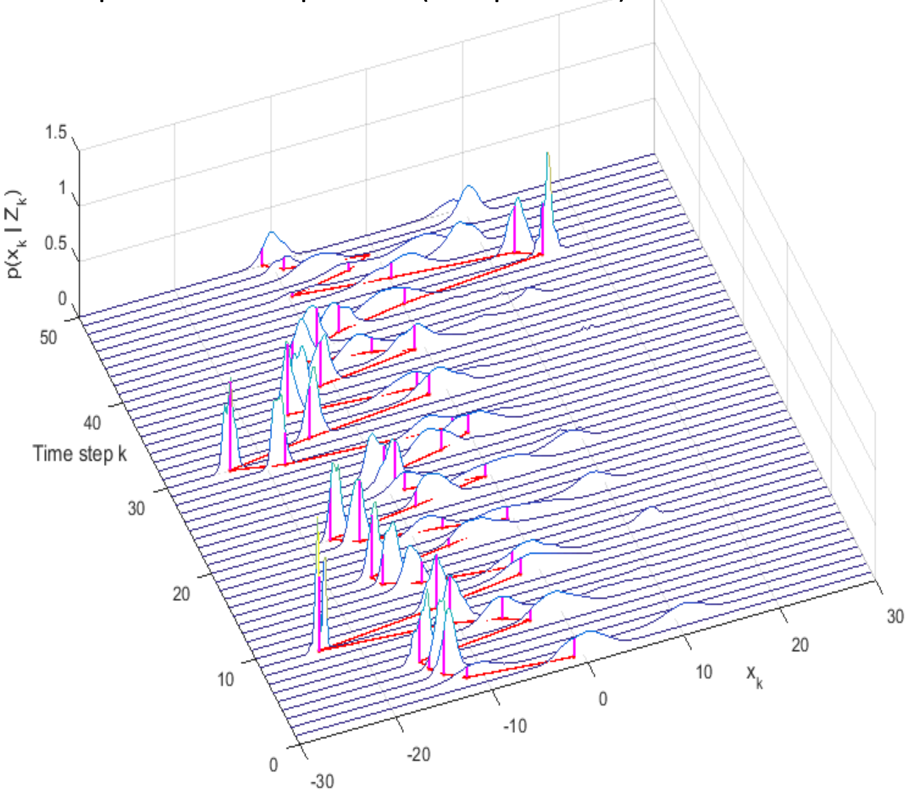


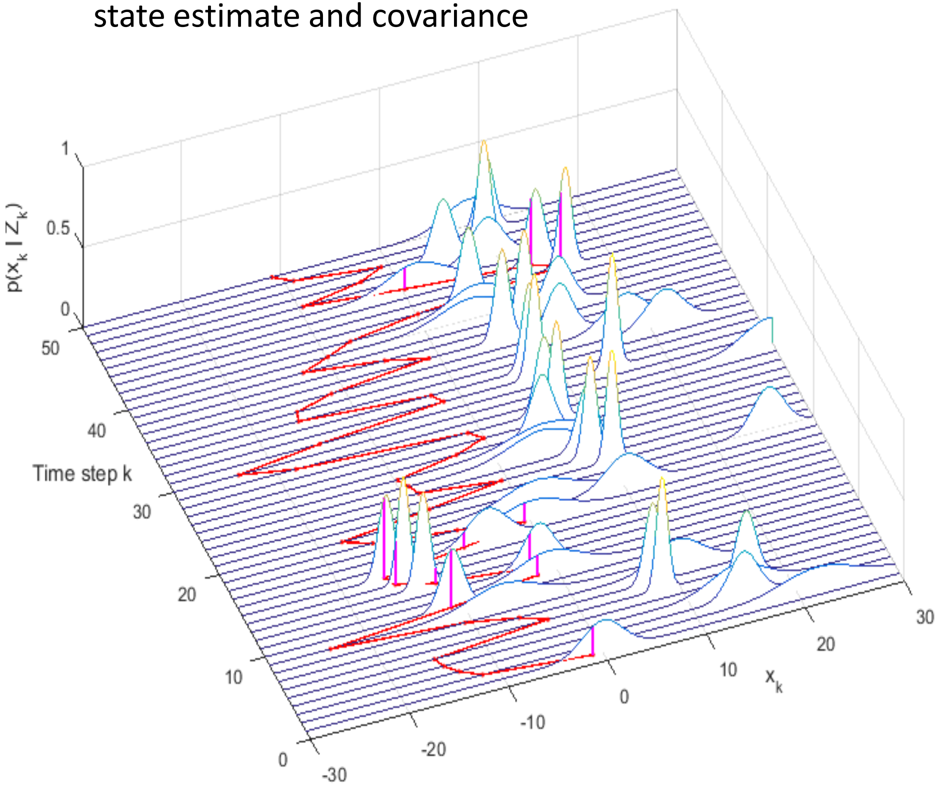
EKF shows serious divergence for most of the time

Evolution of posterior pdf from the filters

Particle filter – kernel density pdf constructed from posterior sample sets (500 particles)

EKF – posterior Gaussian pdf from filter state estimate and covariance



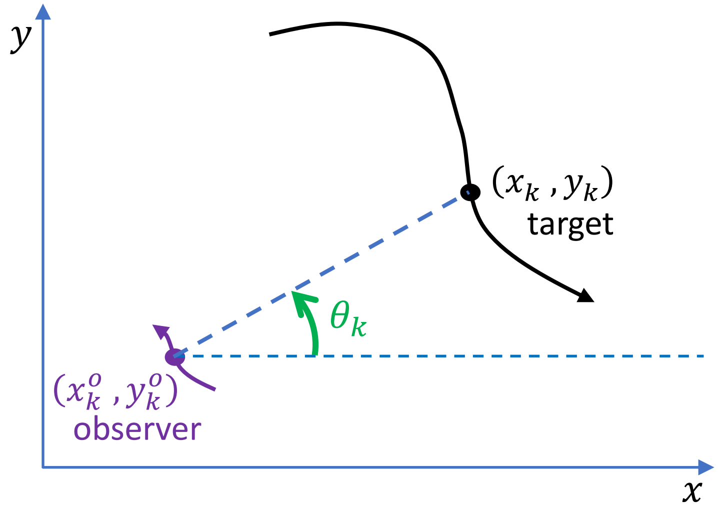


Red = actual evolution of state

Bearings-only tracking

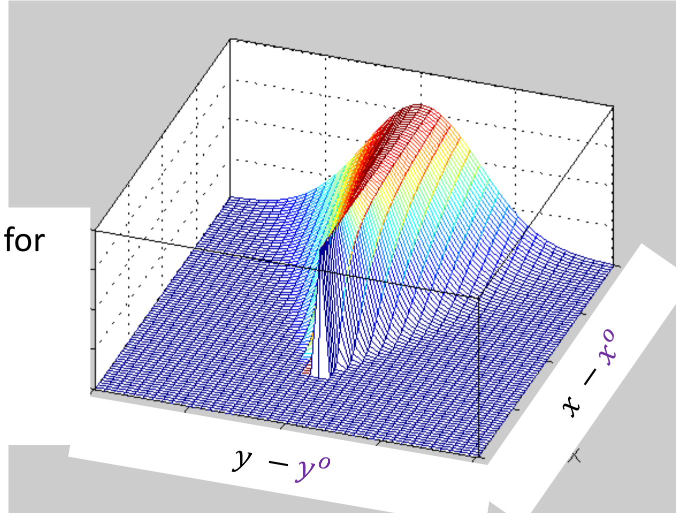
Noisy measurements of target bearing relative to observer:

$$z_k = \theta_k + v_k = \underbrace{\tan^{-1}\left(\frac{y_k - y_k^o}{x_k - x_k^o}\right)}_{\text{Nonlinear measurement function}} + v_k$$



Likelihood of (x, y) for the measurement $z = 0$

$p(z = 0\,|x, y)$



Estimate target trajectory in position and velocity $(x\,, \dot{x}\,, y\,, \dot{y}\,)_k$ given knowledge of observer's positions $(x^o, y^o)_k$ (assumed perfect).

Bearings-only tracking example:

Tracking a gently manoeuvring target (near constant velocity model) which passes close to a stationary observer

Gaussian prior on initial target position (SD≅ 0.3) and velocity (SD≅ 0.005)

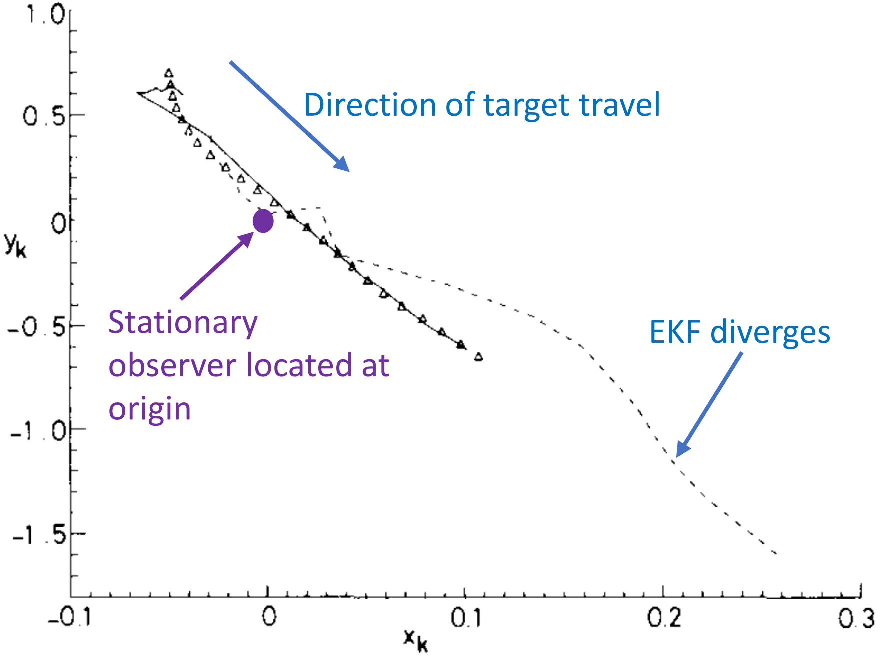


Fig. 6 *Estimated posterior mean of the target track in the x-y plane*

△ Target position
——— Bootstrap estimate
- - - - EKF estimate

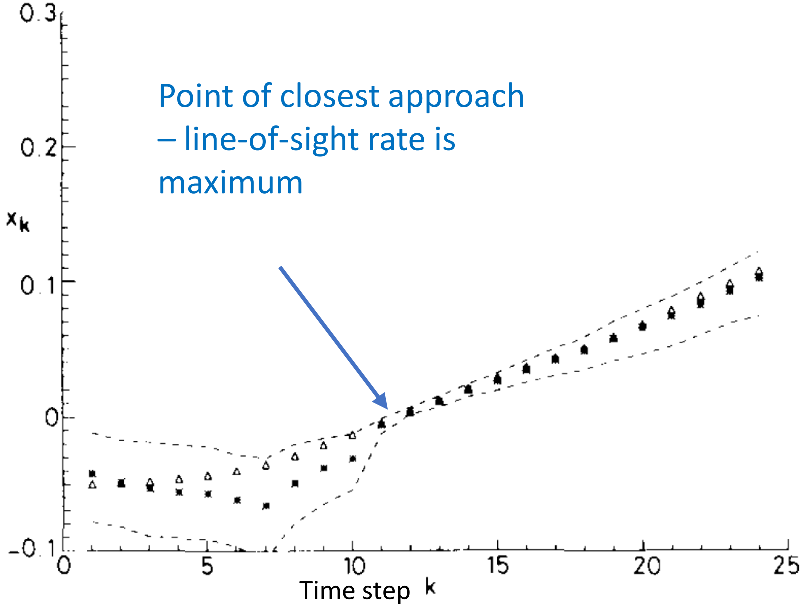


Fig. 8 *Bootstrap estimate of the posterior mean and 95% probability region: x co-ordinate*

△ True target state
✱ Bootstrap estimated target state
- - - - Estimated limits of 95% probability region

Why we thought the scheme might be significant:

It provides the posterior distribution of the state (in sample representation) rather than just the mean and covariance

It works for essentially any distribution / likelihood – multimodal, fragmented over the state space, with hard edges, with restricted domains etc

Basic filter is very simple

Optimal performance can be approached just by increasing the number of particles (admittedly only practical for low dimensional problems with reasonable system noise)

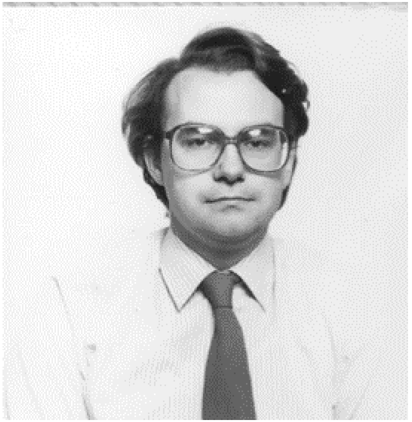
Algorithmically it does not resemble the Kalman recipe

Only need to evaluate the likelihood for the update (at possibly many points) – no need to derive Jacobians

How things looked in 1992

Adrian Smith

Neil Gordon, working towards a PhD, supervised by Adrian



Salmond

Professor of Statistics at Imperial College

At the Defence Research Agency, Farnborough (previously the Royal Aircraft Establishment)

Theoretical developments	Foundations: P Del Moral (1994), T Lyons (1997), J Liu, R Chen (1998) Convergence: D Crisan, A Doucet (2000) Smoothing: S Godsill, T Clapp (1999), G Kitagawa (1996) Rao-Blackwellisation: A Doucet (1998), J de Freitas (2000) Regularisation: C Musso, N Oudjane (1998) Implementation: P Clifford, P Fearnhead (1999)
Target tracking	Bearings-only / Range-only: B Ristic, S Arulampalam (2000) Manoeuvring (multiple model): S McGinnity, G Irwin (1998) Multiple targets: D Avitzour (1995), C Hue, J-P LeCadre, P Perez (2002), S Maskell Group tracking: M Moreland (2002) Track-before-detect: Y Boers (2001) Radar applications: H Driessen, M Rutten
Robotics / navigation	Road networks / terrain aided: F Gustafsson, N Bergman (1999) Robot localisation: D Fox, F Dellaert, S Thrun (1998)
Signal/ image processing	Contour tracking in images (“condensation”): A Blake, M Isard (1996) Model selection, communications applications: P Djuric (1999) Signal reconstruction: S Godsill (1997) Nonlinear time series: G Kitagawa (1996) Econometrics: M Pitt, N Shepard (1997) Semiconductor composition estimation: A Marrs (2001)

Australian Government

Department of Defence
Science and Technology

Neil Gordon

Bayesian Stats in the 80s

- Late 70s-early 80s focus was efficient numerical integration tools
- Gelfand & Smith (90) showed how sample based methods could revolutionise Bayesian statistics
- MCMC methods
  - Gibbs
  - Metropolis-Hastings
- Great for off-line batch analysis
- We wanted on-line recursive …

Important seminar

- Adrian's seminar "Bayesian statistics without tears"
- Had idea … kept quiet!
- Discussed in "office" (to avoid phone calls)
- Carefully wrote down on train going home

- Important :
  - pay attention in seminars …
  - no sleeping …
  - you never know!

Source : tripadvisor.com.au

What computing was available in 1989?

OK … not quite that bad … But still …

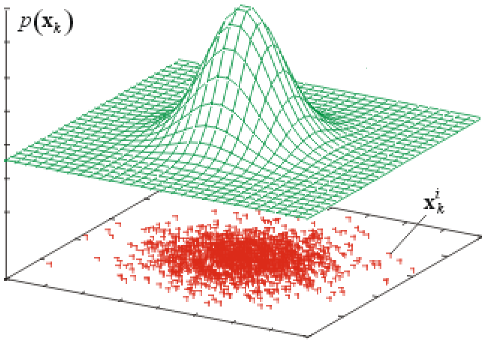
Source : en.wikipedia.org

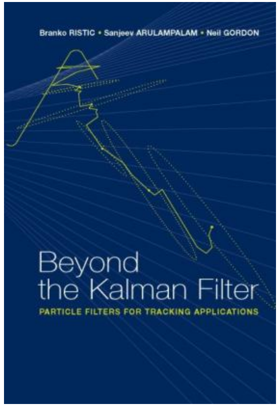
Source : www.publicdomainimages.net

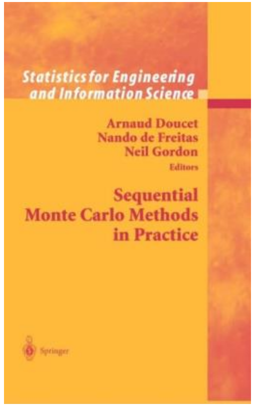
1990s Google image search “particle filter”

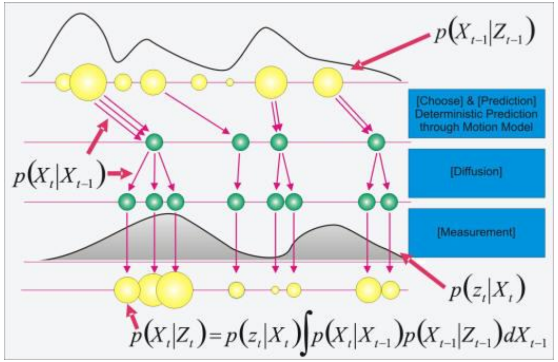
Source : en.wikipedia.org

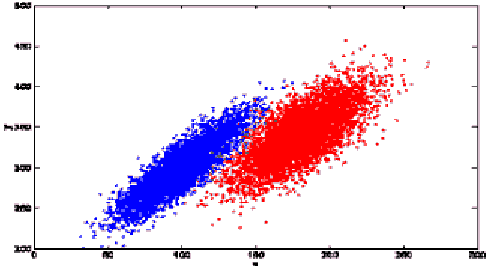
And now ….


$p(\mathbf{x}_k)$
$\mathbf{x}_k^i$


Branko RISTIC • Sanjeev ARULAMPALAM • Neil GORDON
Beyond the Kalman Filter
PARTICLE FILTERS FOR TRACKING APPLICATIONS


Statistics for Engineering and Information Science
Arnaud Doucet
Nando de Freitas
Neil Gordon
Editors
Sequential Monte Carlo Methods in Practice
Springer


$p(X_{t-1}|Z_{t-1})$
[Choose] & [Prediction] Deterministic Prediction through Motion Model
$p(X_t|X_{t-1})$
[Diffusion]
[Measurement]
$p(z_t|X_t)$
$p(X_t|Z_t) = p(z_t|X_t)\int p(X_t|X_{t-1})p(X_{t-1}|Z_{t-1})dX_{t-1}$



My favourite developments and applications

- ASIR
- Rao-Blackwellisation
- VRPF

- GPS free navigation on HP ipaq
- Track before detect
- Migrating birds
- Finding AE1

Future trends?

- Higher power computing - better approximation to Bayesian solution
- Streaming data at scale
- Model based vs data based
- Context enhanced processing
- Likelihood-free and digital twin

Australian Government
Department of Defence
Science and Technology

The search for MH370

Source : ATSB

MH370 Flight Path Reconstruction Team

Australian Government
Australian Transport Safety Bureau

Australian Government
Department of Defence
Science and Technology

Air Accidents Investigation Branch

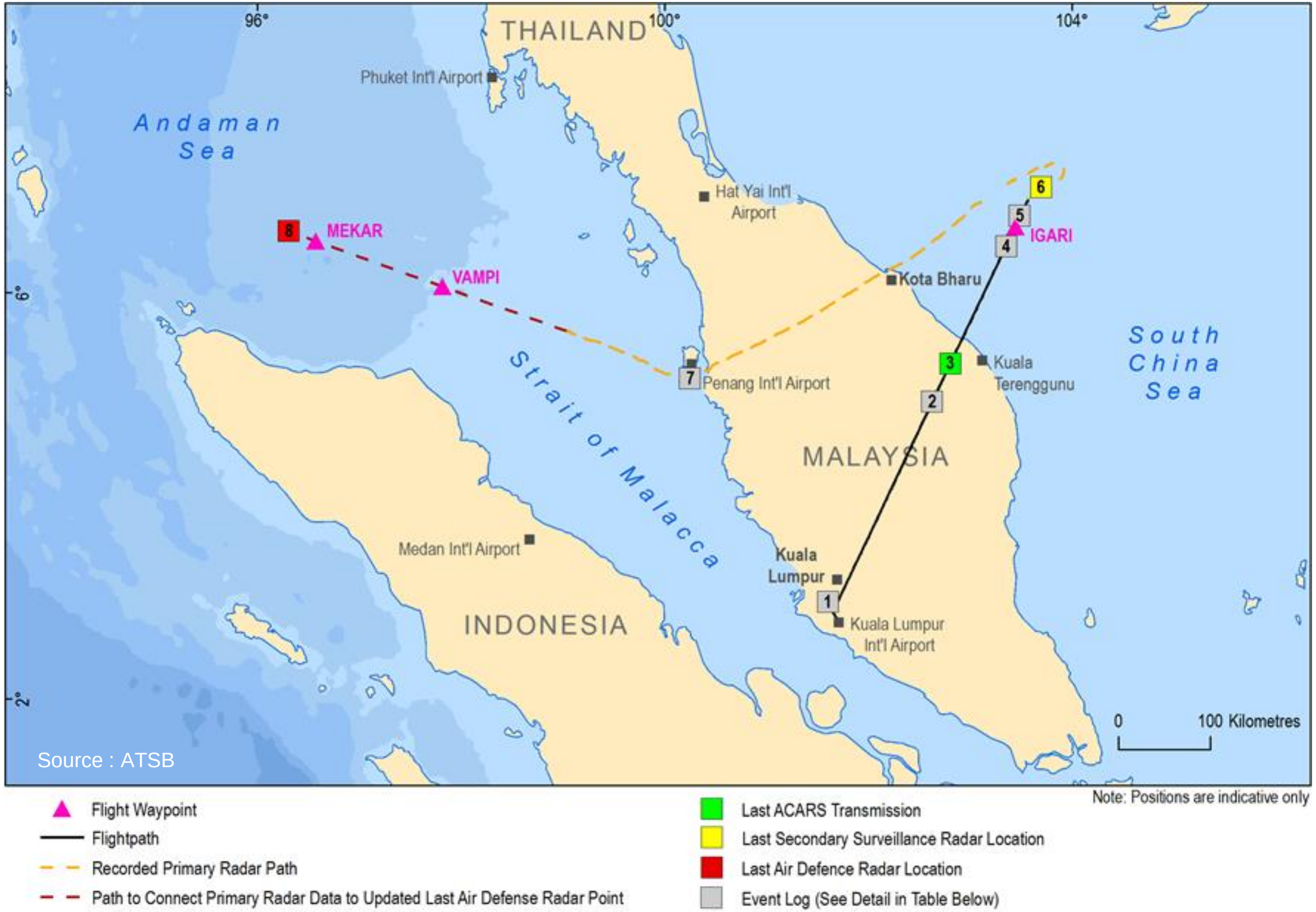
National Transportation Safety Board

inmarsat

THALES

BOEING

CSIRO



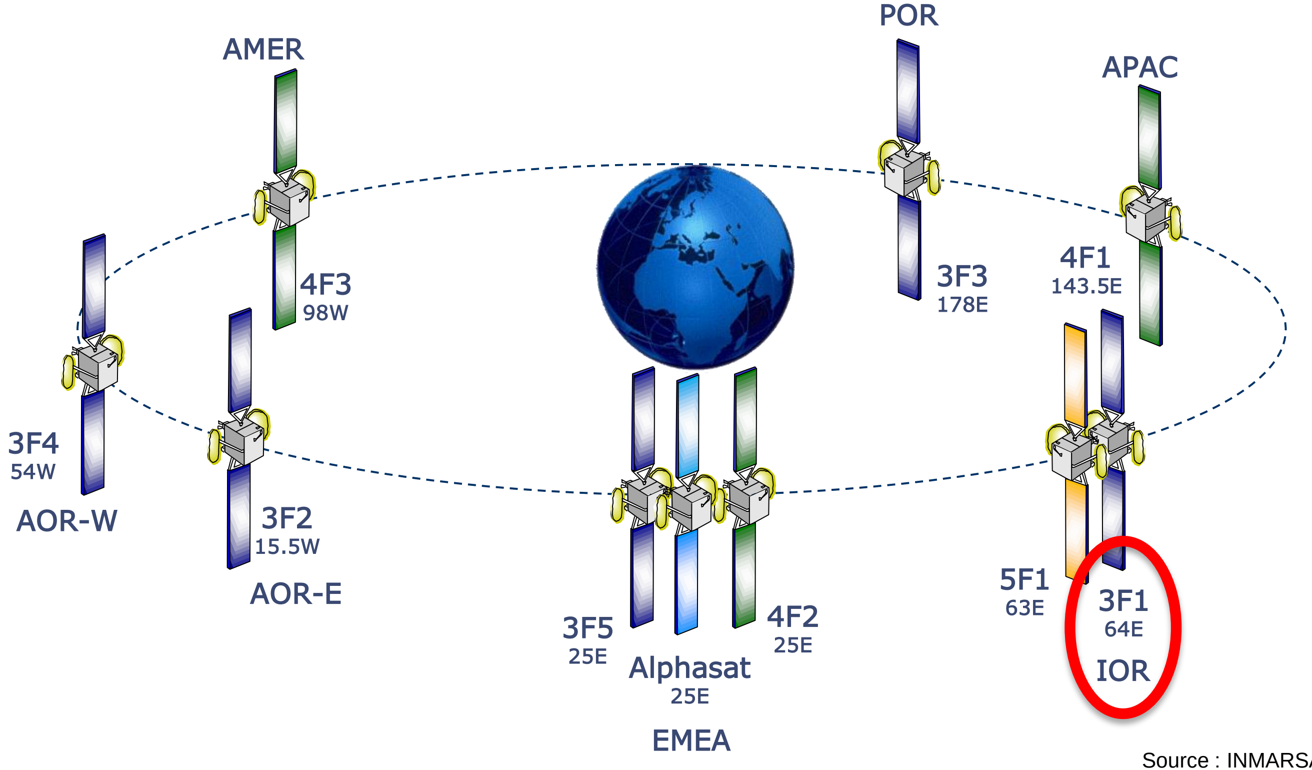
THAILAND
96°
100°
104°
6°
2°
Phuket Int'l Airport
Andaman Sea
Hat Yai Int'l Airport
MEKAR
VAMPI
IGARI
Kota Bharu
Penang Int'l Airport
Kuala Terengganu
South China Sea
Strait of Malacca
MALAYSIA
Medan Int'l Airport
Kuala Lumpur
Kuala Lumpur Int'l Airport
INDONESIA
0
100 Kilometres
1
2
3
4
5
6
7
8
Source : ATSB
Note: Positions are indicative only
Flight Waypoint
Flightpath
Recorded Primary Radar Path
Path to Connect Primary Radar Data to Updated Last Air Defense Radar Point
Last ACARS Transmission
Last Secondary Surveillance Radar Location
Last Air Defence Radar Location
Event Log (See Detail in Table Below)

Bayesian Approach

- Prior
  - Radar data
- Likelihood
  - Inmarsat metadata (BTO, BFO)
- Dynamics
  - Cruise and manoeuvre
  - Environmental data

- GOAL : PDF at time of final electronic communication
  - Descent scenarios defined by ATSB

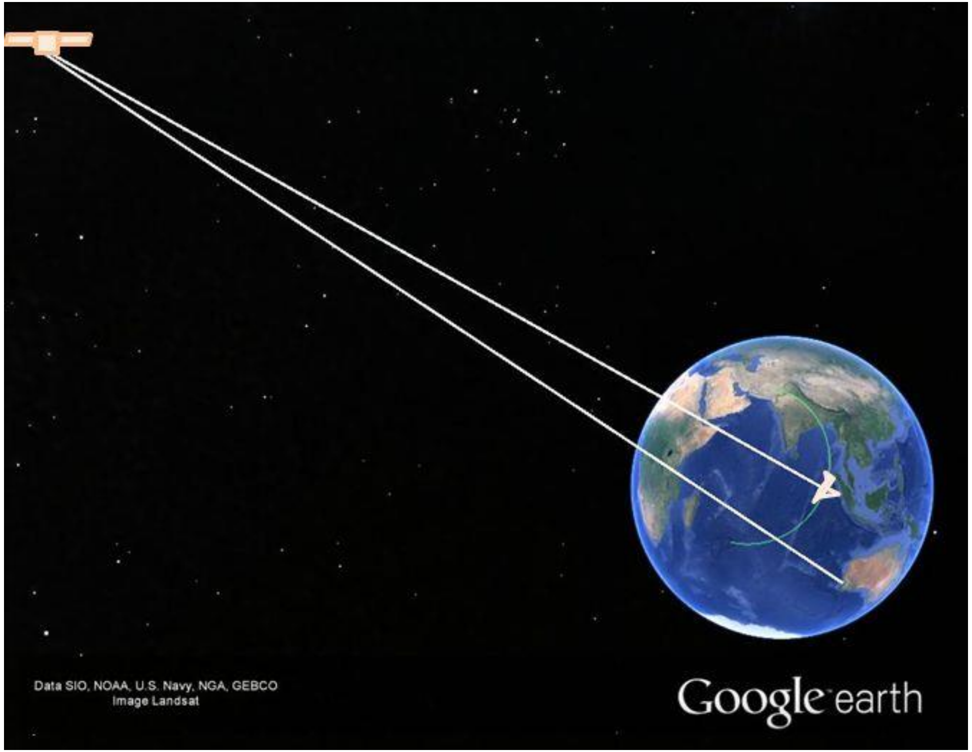
Inmarsat Satellite Fleet



Source : INMARSAT

Burst Timing Offset

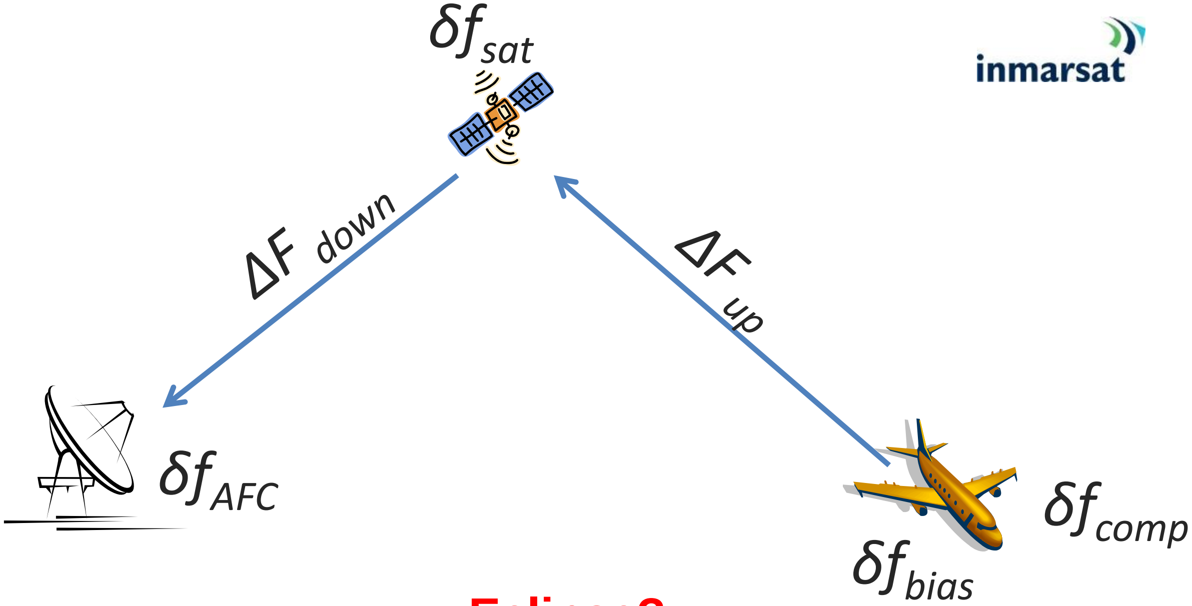
Inmarsat 3
IOR Satellite



Inmarsat Ground Earth Station
Perth (Australia)

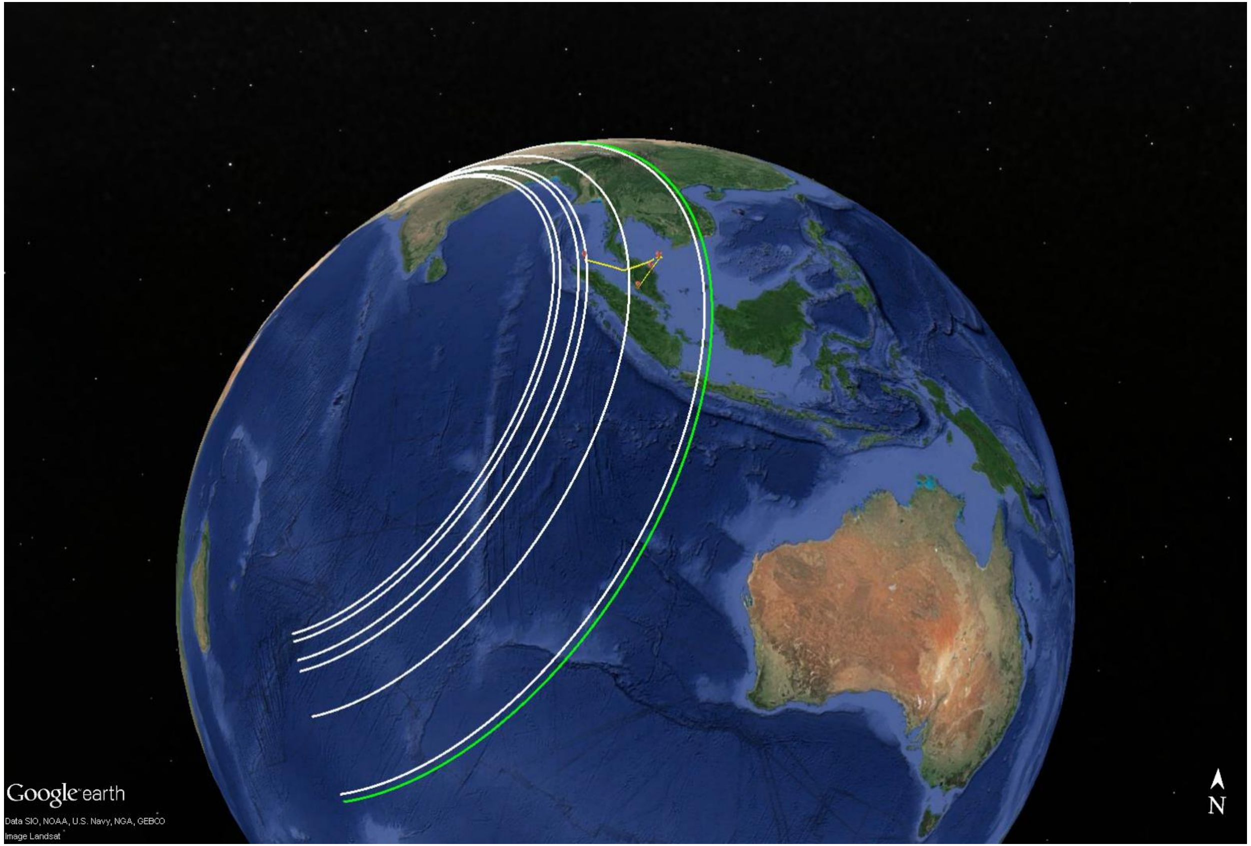
Burst Frequency Offset (residual Doppler)

(Measured – Expected) frequency at GES



Eclipse?

Source : INMARSAT



18:28
18:41
19:41
20:41
21:41
22:41
23:13
00:11
00:19

BTO and BFO : Simple summary

- BTO
  - Constrains allowable locations at time of transmission
  - Uncertainty calibrated from flight data
- BFO
  - Constrains allowable (speed,heading) at time of transmission
  - Highly sensitive to vertical speed
  - Uncertainty calibrated from flight data

- Sparse set of data
- What is possible between transmission times?

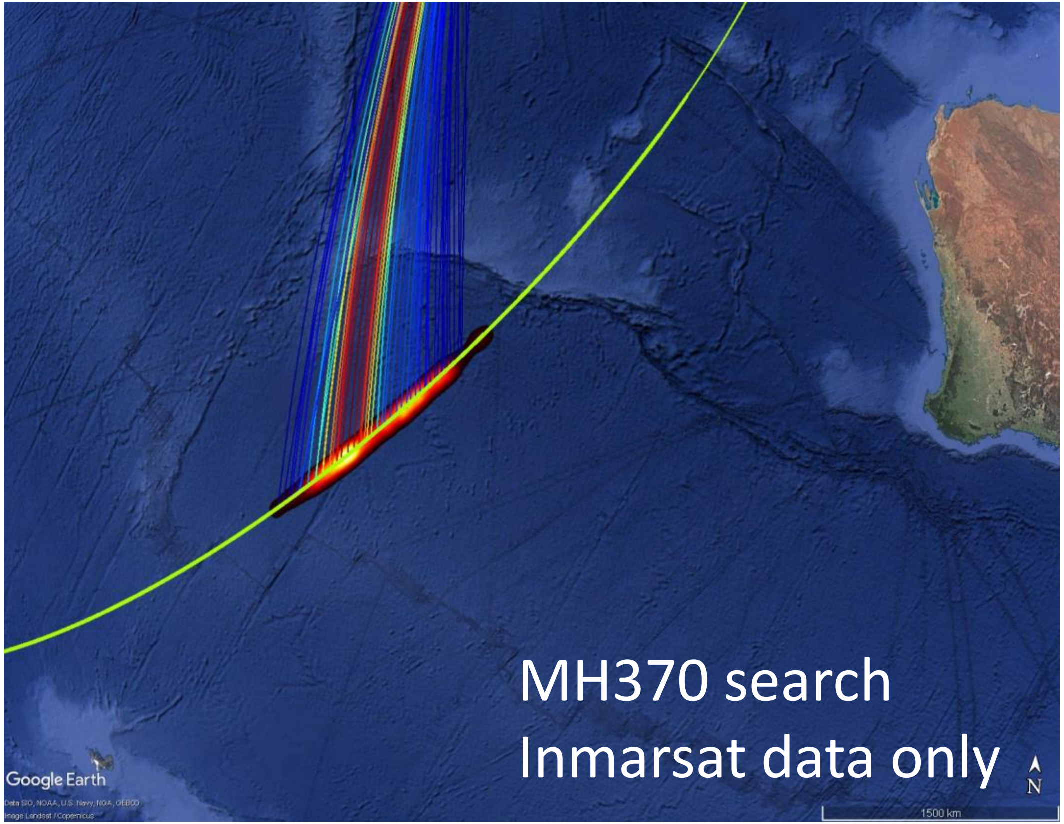
How are commercial aircraft flown … Autopilot

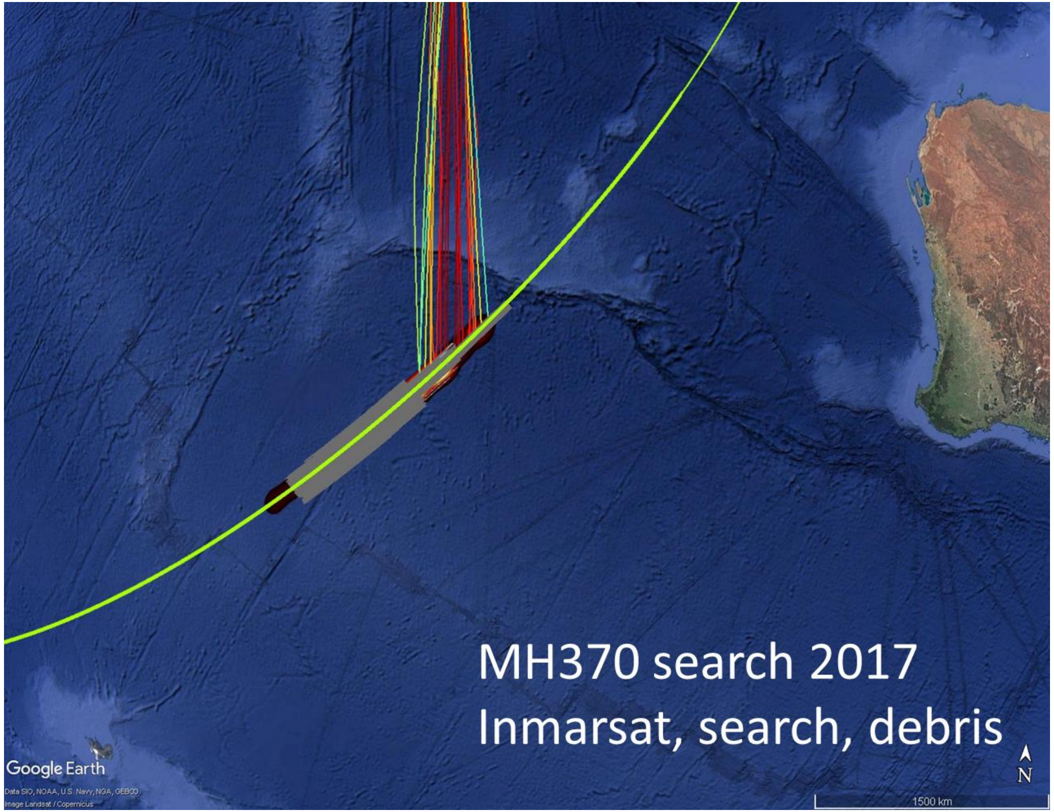
- Lateral Navigation
- Constant - Magnetic/True - Heading/Track

Aircraft dynamics : Manoeuvre/Cruise

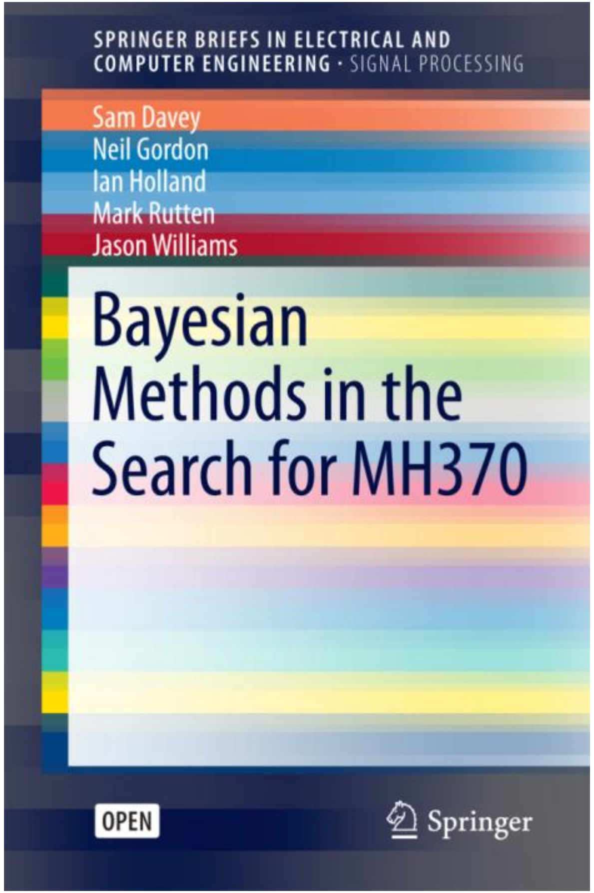
Sequence of deliberate manoeuvres interspersed with periods of cruise

- Deliberate manoeuvres commanded via autopilot
- Cruise : OU model
- Manoeuvre : Speed, direction, altitude
- Unknown autopilot mode
  - CMH, CTH, CMT, CTT, LNAV, CI
- Validate parameters of model with known flight data
  - But retain flexibility on rate of manoeuvre occurrence
- Wind and air temperature is important
  - Calibrate BOM error model with known flight data


MH370 search
Inmarsat data only
Google Earth
N
1500 km


MH370 search 2017
Inmarsat, search, debris
Google Earth
N
1500 km

- Problem involves many sources of uncertainty
  - Measurement error in BTO, BFO
  - Uncertainty in BFO offset, aircraft dynamics, possible manoeuvres, wind speed/direction, eclipse calibration
- Calculate PDF at time of final transmission
- Procedure validated with data from previous flights reflecting all above uncertainties
- Descent PDF defined by ATSB informed by
  - Descent rate bounds
  - Boeing flight simulator
  - Condition of debris



Bayesian Methods in the Search for MH370

Sam Davey
Neil Gordon
Ian Holland
Mark Rutten
Jason Williams

ISBN 978-981-10-0378-3
Springer Briefs in Signal Processing
May 2016

Open Access

Challenge : Produce your solution to pdf

MH370@dsto.defence.gov.au