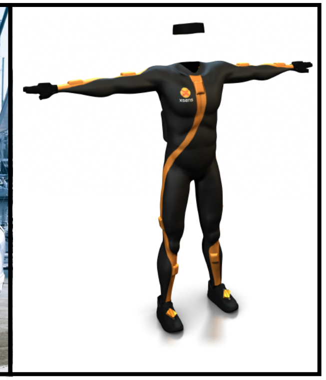
# Representing and working with uncertainty in dynamical systems

Thomas Schön
Division of Automatic Control
Linköping University
Sweden

www.control.isy.liu.se/~schon

**Special thanks to:** Jeroen Hol (*Xsens Technologies*), Lennart Ljung (*LiU*), Johan Kihlberg (*Xdin*), Fredrik Lindsten (*LiU*), Henk Luinge (*Xsens Technologies*), Michael Jordan (*Berkeley*), Brett Ninness (*University of Newcastle*), Per-Johan Nordlund (*Saab*), Simon Tegelid (*Xdin*) and Adrian Wills (*University of Newcastle*).

# Dynamical systems

What is a dynamical system?

State → $x_{t+1}$; Known input → $u_t$

$$x_{t+1} \mid x_t \sim f_\theta(x_{t+1} \mid x_t, u_t), \qquad \text{Dynamics}$$

Measurements → $y_t$

$$y_t \mid x_t \sim h_\theta(y_t \mid x_t, u_t), \qquad \text{Measurements}$$

$$x_1 \sim \mu_\theta(x_1). \qquad \text{Initial state}$$

Static parameters → $\theta$

*"The present state of a dynamical system depends on its history."*

# Outline

**1. Representation - probabilistic state space models (SSM's)**

**2. State inference**

a) General solution
b) LGSS models and the Kalman filter
c) Sensor fusion example
d) Particle filter for general SSM's via positioning examples

**3. Parameter inference**

a) Problem formulation
b) Bayesian solution - particle MCMC
c) System identification example - semiparametric Wiener model

$$
\begin{aligned}
x_{t+1} \mid x_t &\sim f_\theta(x_{t+1} \mid x_t, u_t), \\
y_t \mid x_t &\sim h_\theta(y_t \mid x_t, u_t), \\
x_1 &\sim \mu_\theta(x_1).
\end{aligned}
$$

# Probabilistic models of dynamical systems

We often model a dynamical system using **probability density functions (pdf's)**

$$
\begin{aligned}
x_{t+1} \mid x_t &\sim f_\theta(x_{t+1} \mid x_t, u_t), && \text{Dynamics} \\
y_t \mid x_t &\sim h_\theta(y_t \mid x_t, u_t), && \text{Measurements} \\
x_1 &\sim \mu_\theta(x_1). && \text{Initial state}
\end{aligned}
$$

State: $x_t$; Known input: $u_t$; Measurements: $y_t$; Static parameters: $\theta$.

## Model = pdf

The state process is hidden (latent) and it is observed indirectly via the measurement process.

This type of model is referred to as a **state space model (SSM)** or a **hidden Markov model (HMM)**.

# Probabilistic models of dynamical systems

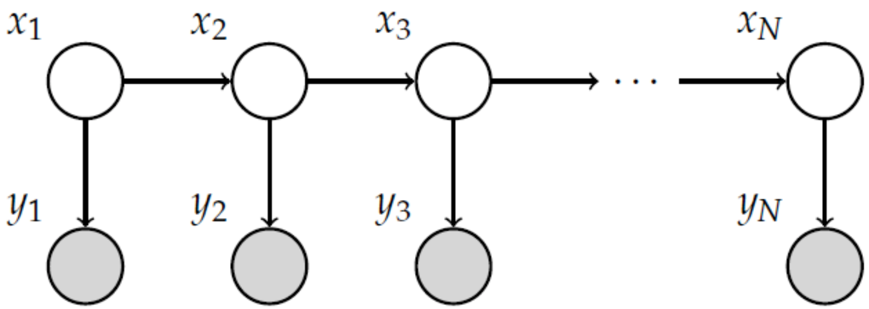
Alternative model formulation 1 (common in engineering):

Uncertainty in the model

$$\begin{aligned} x_{t+1} &= \widetilde{f}_\theta(x_t, u_t) + v_{\theta,t}, \\ y_t &= \widetilde{h}_\theta(x_t, u_t) + e_{\theta,t}, \\ x_1 &\sim \mu_\theta(x_1). \end{aligned}$$

Uncertainty in the measurements

Alternative model formulation 2 (graphical model):




The state is a variable that contains all information about the past and the present of a system, which is needed in order to predict the future.

It is the Markov property

$$p(x_{t+1} \mid x_1, \ldots, x_t) = p(x_{t+1} \mid x_t)$$

that allows for this.

# The use of probabilistic models

The SSM can be used to answer many questions, where two of the most common are:

1. **State inference:** Infer the states from the available measurements.
2. **Parameter inference:** Infer the static model parameters from the available measurements.

Answering the second question typically involves solving various state inference problems.

**System identification** deals with the problem of finding a dynamical model based on measurements of the input signal and the output signal,

$$u_{1:N} = \{u_1, \dots, u_N\}, \qquad y_{1:N} = \{y_1, \dots, y_N\}.$$

**Sensor fusion** is the process of using information from **several different** sensors to **learn (estimate)** what is happening (this typically includes states of various dynamical systems and various static parameters).

# State inference

The **aim** is to compute a probabilistic representation of our knowledge of the state, based on information that is present in the measurements.

The **filtering probability density function (pdf)** provides a good representation of the uncertainty about the state at time t, given the measurements up to time t,

$$p(x_t \mid y_{1:t})$$

The obvious question is now, how do we compute this object?

Bayes' theorem

$$p(x_t \mid y_{1:t}) = p(x_t \mid y_t, y_{1:t-1}) = \frac{p(y_t \mid x_t, y_{1:t-1})p(x_t \mid y_{1:t-1})}{p(y_t \mid y_{1:t-1})}$$

$$= \frac{h(y_t \mid x_t)p(x_t \mid y_{1:t-1})}{p(y_t \mid y_{1:t-1})}$$

Markov property

# State inference

Need an expression also for the prediction pdf

$$p(x_t \mid y_{1:t-1})$$

Let us start by noting that by marginalization we have

$$p(x_t \mid y_{1:t-1}) = \int p(x_t, x_{t-1} \mid y_{1:t-1}) \mathrm{d}x_{t-1}$$

$$p(x_t, x_{t-1} \mid y_{1:t-1}) = p(x_t \mid x_{t-1}, y_{1:t-1}) p(x_{t-1} \mid y_{1:t-1})$$
$$= f(x_t \mid x_{t-1}) p(x_{t-1} \mid y_{1:t-1})$$

Markov property

Hence, the prediction pdf is given by

$$p(x_t \mid y_{1:t-1}) = \int f(x_t \mid x_{t-1}) p(x_{t-1} \mid y_{1:t-1}) \mathrm{d}x_{t-1}$$

# State inference - summarizing the development

We have now showed that for the nonlinear SSM

$$
\begin{aligned}
x_{t+1} \mid x_t &\sim f(x_t \mid x_{t-1}), \\
y_t \mid x_t &\sim h(y_t \mid x_t),
\end{aligned}
$$

the uncertain information that we have about the state is captured by the filtering pdf, which we compute sequentially using a **measurement update**

$$
p(x_t \mid y_{1:t}) = \frac{\overbrace{h(y_t \mid x_t)}^{\text{measurement model}} \overbrace{p(x_t \mid y_{1:t-1})}^{\text{prediction pdf}}}{p(y_t \mid y_{1:t-1})},
$$

and a **time update**

$$
p(x_t \mid y_{1:t-1}) = \int \underbrace{f(x_t \mid x_{t-1})}_{\text{dynamic model}} \underbrace{p(x_{t-1} \mid y_{1:t-1})}_{\text{filtering pdf}} \mathrm{d}x_{t-1},
$$

# State inference - simple special case (LGSS)

Consider the following special case (Linear Gaussian State Space (LGSS) model)

$$x_{t+1} = Ax_t + Bu_t + v_t, \qquad v_t \sim \mathcal{N}(0, Q),$$
$$y_t = Cx_t + Du_t + e_t, \qquad e_t \sim \mathcal{N}(0, R).$$

or, equivalently,

$$x_{t+1} \mid x_t \sim f(x_{t+1} \mid x_t) = \mathcal{N}(x_{t+1} \mid Ax_t + Bu_t, Q),$$
$$y_t \mid x_t \sim h(y_t \mid x_t) = \mathcal{N}(y_t \mid Cx_t + Du_t, R).$$

Gaussian variables and linear transformation implies that the mean and the covariance captures everything there is to know.

# State inference - simple special case (LGSS)

## Measurement update

$$p(x_t \mid y_{1:t}) = \frac{h(y_t \mid x_t)p(x_t \mid y_{1:t-1})}{p(y_t \mid y_{1:t-1})}$$

$$p(x_t \mid y_{1:t}) = \mathcal{N}\left(x_t \mid \widehat{x}_{t|t}, P_{t|t}\right)$$

innovation

$$\widehat{x}_{t|t} = \widehat{x}_{t|t-1} + K_t(y_t - C\widehat{x}_{t|t-1} - Du_t),$$
$$K_t = P_{t|t-1}C^T(CP_{t|t-1}C^T + R)^{-1},$$
$$P_{t|t} = P_{t|t-1} - K_t C P_{t|t-1}$$

decrease uncertainty

## Time update

$$p(x_{t+1} \mid y_{1:t}) = \int f(x_{t+1} \mid x_t)p(x_t \mid y_{1:t})\mathrm{d}x_t$$

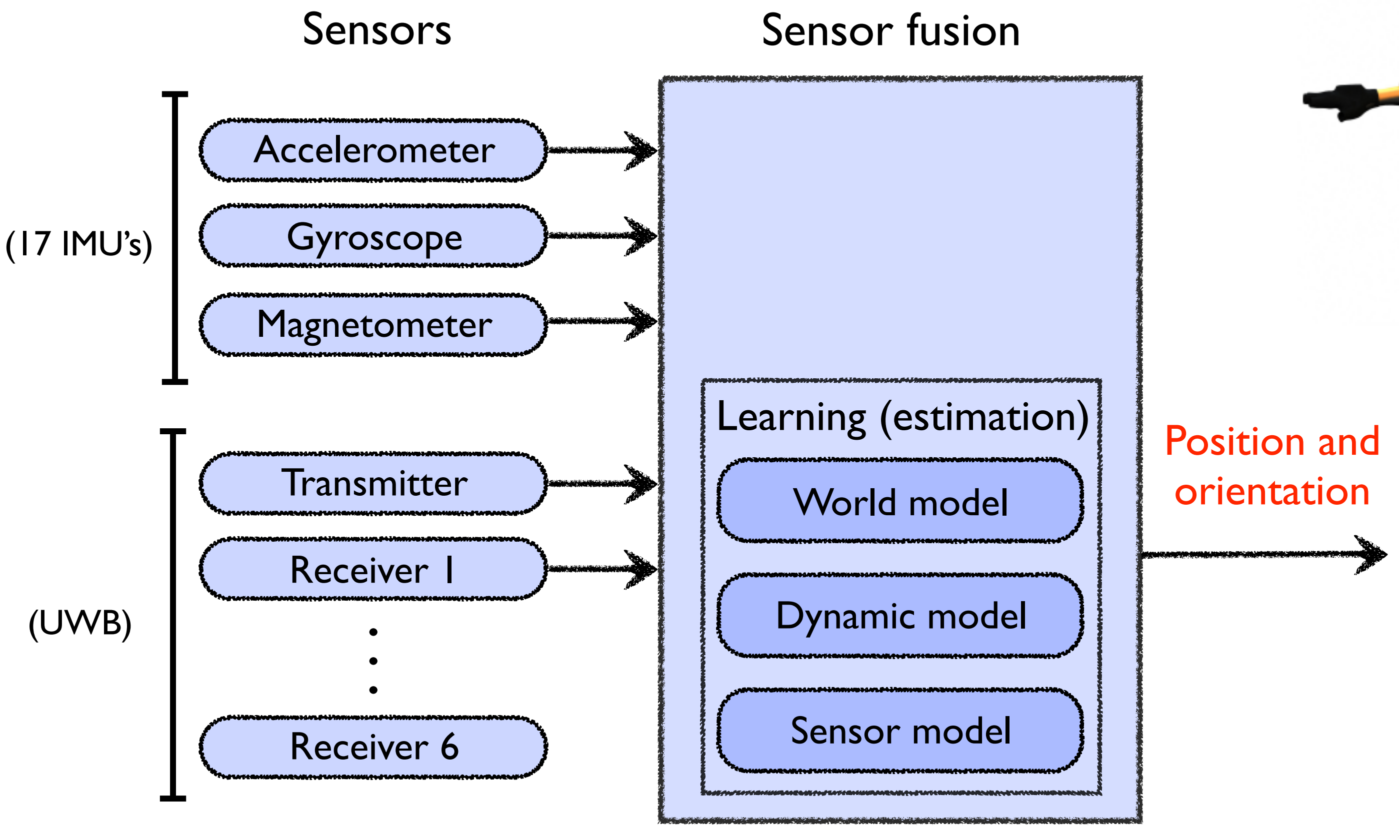
$$p(x_{t+1} \mid y_{1:t}) = \mathcal{N}\left(x_{t+1} \mid \widehat{x}_{t+1|t}, P_{t+1|t}\right)$$

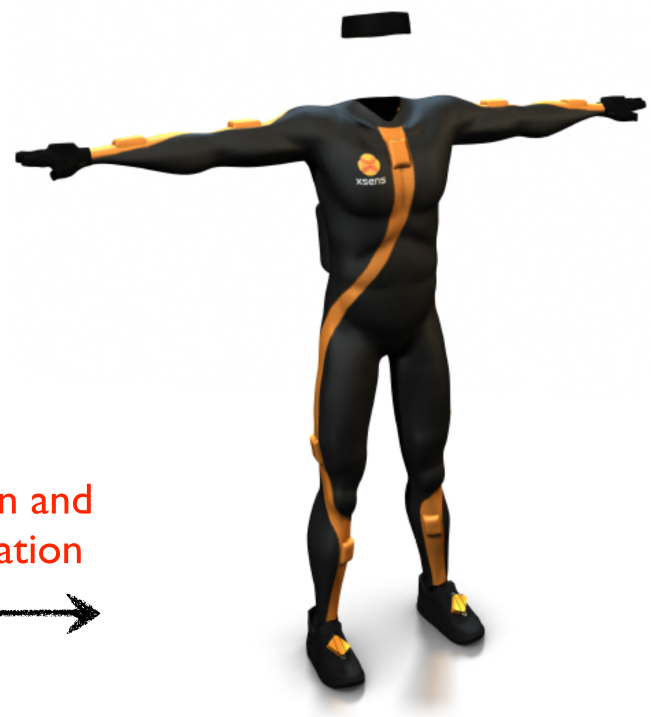
$$\widehat{x}_{t+1|t} = A\widehat{x}_{t|t} + Bu_t,$$
$$P_{t+1|t} = AP_{t|t}A^T + Q$$

increase uncertainty

**Aim:** Estimate the position and orientation of a human (i.e. human motion) using measurements from inertial sensors and ultra-wideband (UWB).

**Industrial partner:** Xsens Technologies



# State inference - a sensor fusion example (II/III)

## The sensors

Sensor unit integrating an IMU and a UWB transmitter into a single housing.


- Inertial measurements @ 200 Hz
- UWB measurements @ 50 Hz
- Mobile transmitter and 6 stationary, synchronized receivers at known positions.
- Time-of-arrival (TOA) measurements

# Outline

**1. Representation - probabilistic state space models (SSM's)**

**2. State inference**

a) General solution
b) LGSS models and the Kalman filter
c) Sensor fusion example
d) Particle filter for general SSM's via positioning examples

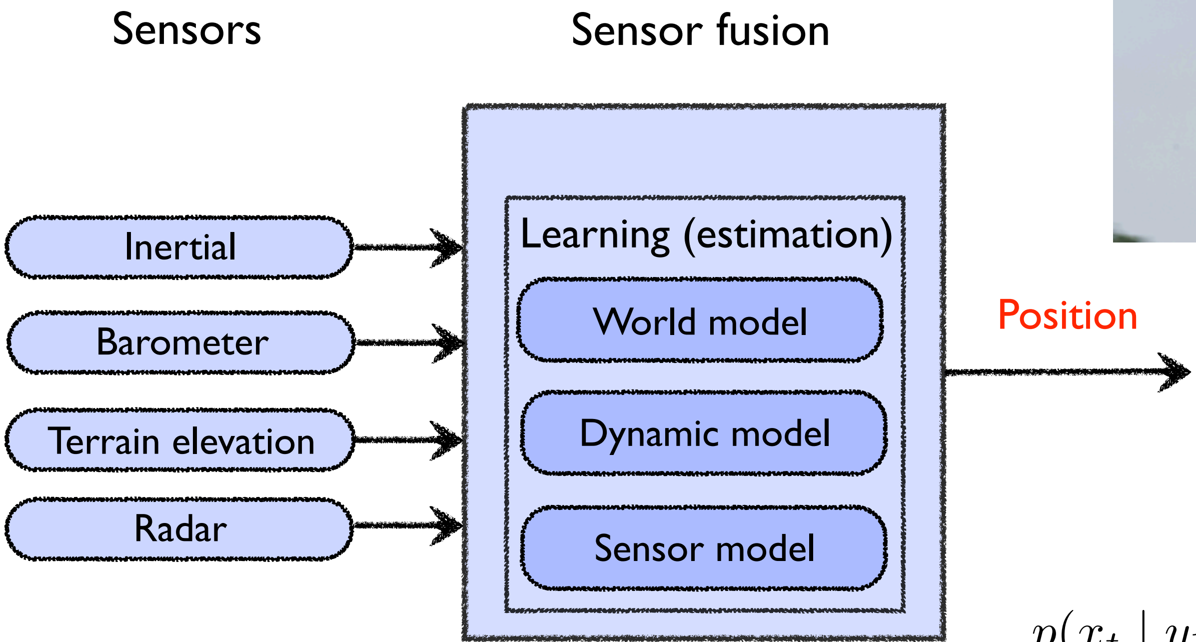
**3. Parameter inference**

a) Problem formulation
b) Bayesian solution - particle MCMC
c) System identification example - semiparametric Wiener model

# State inference - particle filters

**Aim:** Find the position, velocity and orientation of a fighter aircraft.

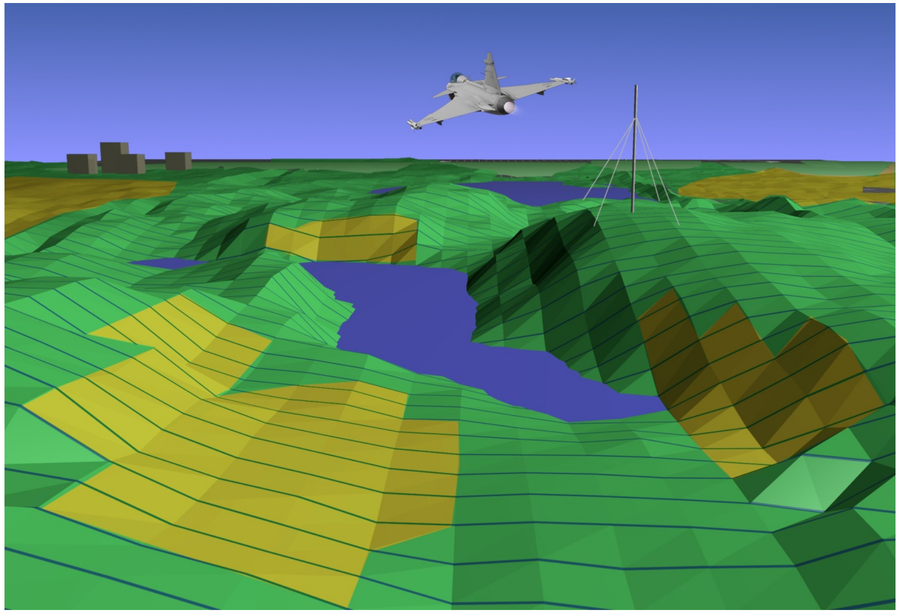
**Industrial partner:** Saab

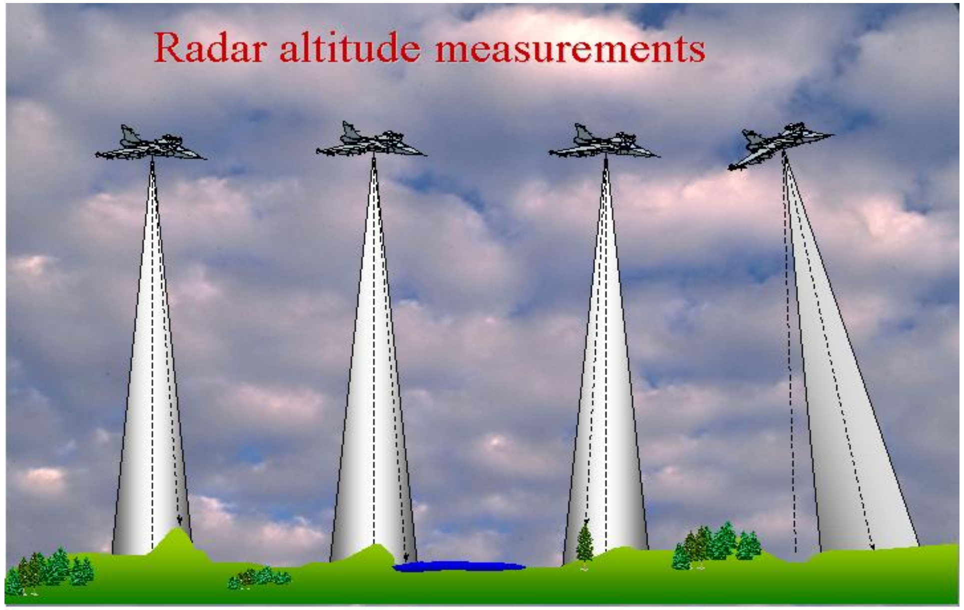



$$p(x_t \mid y_{1:t}) = \frac{h(y_t \mid x_t) p(x_t \mid y_{1:t-1})}{p(y_t \mid y_{1:t-1})}$$

$$p(x_{t+1} \mid y_{1:t}) = \int f(x_{t+1} \mid x_t) p(x_t \mid y_{1:t}) \mathrm{d}x_t$$

# State inference - particle filters






*"Think of each particle as one simulation of the system state (in the movie we are visualizing the horizontal position). Only keep the good ones."*

Show movie

# State inference - particle filters

The **idea** in the particle filter (member of the larger family of Sequential Monte Carlo (SMC) methods) is to use the following nonparametric representation of the filtering pdf

$$p(x_t \mid y_{1:t}) \approx \sum_{i=1}^{N} w_t^i \delta_{x_t^i}(x_t), \qquad \sum_{i=1}^{N} w_t^i = 1, \quad w_t^i \geq 0, \forall i$$

The weights and the particles are then updated as new measurements becomes available.

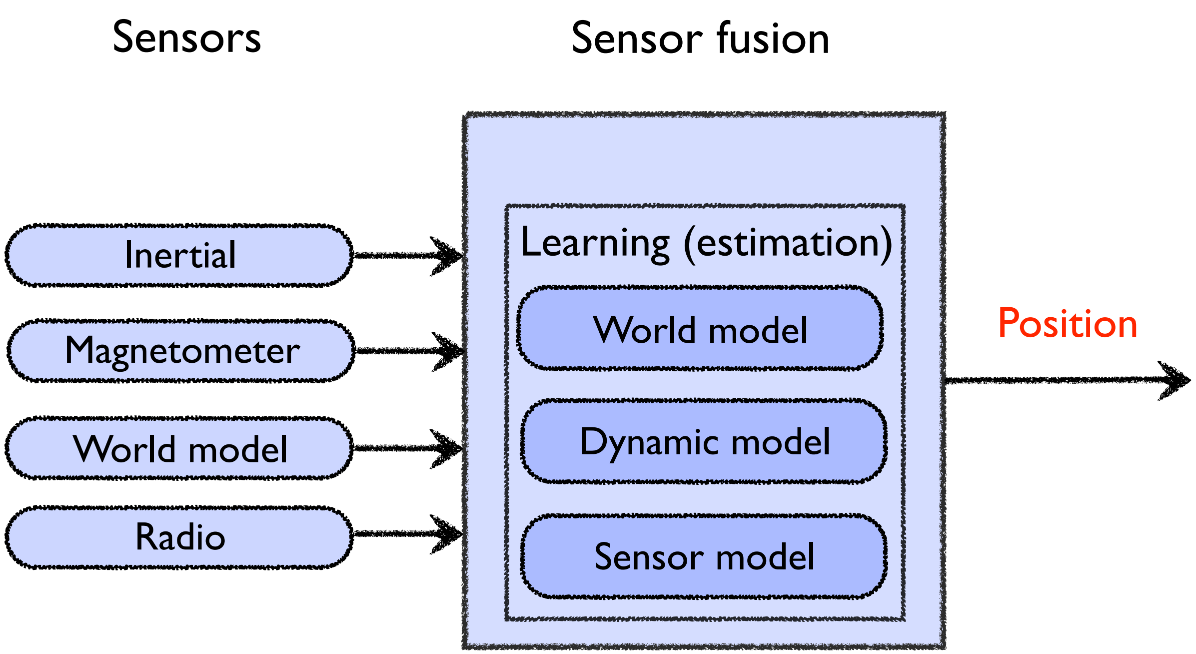
This implies that the multidimensional integrals are replaced by finite sums, which is manageable,

$$\text{"}\delta + \int \to \sum\text{"}$$

# State inference - particle filters

**Aim:** Compute the position of a person moving around indoors using sensors located in an ID badge.

**Industrial partner:** Xdin




Show movie

# Parameter inference - system identification

**System identification** deals with the problem of estimating a dynamical model based on measurements of the input signal and the output signal,

$$u_{1:N} = \{u_1, \dots, u_N\}, \qquad y_{1:N} = \{y_1, \dots, y_N\}.$$

This involves **parameter inference** (among other things). Two approaches:

1. **Maximum Likelihood (ML):** Computes the point estimate of the parameters that makes the observed measurements as likely as possible,

$$\widehat{\theta}^{\text{ML}} = \arg\max_{\theta} p_\theta(y_{1:N})$$

2. **Bayesian:** All variables are now assumed to be stochastic, hence the parameters are no longer deterministic variables. Compute

$$p(\theta \mid y_{1:N})$$

# Monte Carlo and Markov chain Monte Carlo (MCMC)

Monte Carlo methods provides computational solutions, where the obtained accuracy is only limited by our computational resources.

*An MCMC method simulates a Markov chain where the stationary distribution is given by the target distribution of interest.*

*These samples can then be used to compute various estimates.*

There are **constructive strategies** for doing this and some of the most popular are the Gibbs sampler and the Metropolis Hastings sampler.

# Particle MCMC (PMCMC)

The **aim** in particle Markov chain Monte Carlo (PMCMC) is to compute

$$p(\theta, x_{1:T} \mid y_{1:T})$$

or some of its marginals distributions, e.g.,

$$p(\theta \mid y_{1:T})$$

$$p(x_{1:T} \mid y_{1:T})$$

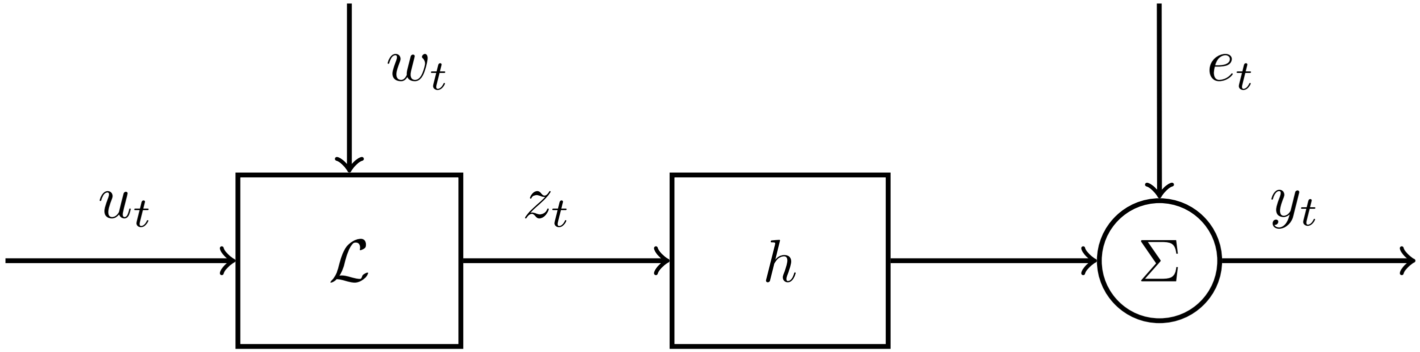
when the model is given by

$$\begin{aligned} x_{t+1} \mid x_t &\sim f_\theta(x_{t+1} \mid x_t, u_t), \\ y_t \mid x_t &\sim h_\theta(y_t \mid x_t, u_t). \end{aligned}$$

The **fundamental idea** is to make use of a sequential Monte Carlo (SMC) sampler to construct a proposal for an MCMC sampler.

# Semi-parametric Wiener model

Rather than describing a general solution, let us be very specific and consider an example,




This is a Wiener model (= a dynamic LGSS model followed by a static nonlinearity),

$$
\begin{aligned}
x_{t+1} &= Ax_t + Bu_t + w_t, & w_t &\sim \mathcal{N}(0, Q), \\
z_t &= Cx_t, & & \\
y_t &= h(z_t) + e_t, & e_t &\sim \mathcal{N}(0, r).
\end{aligned}
$$

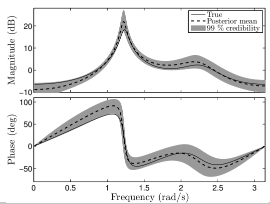
# Semi-parametric Wiener model

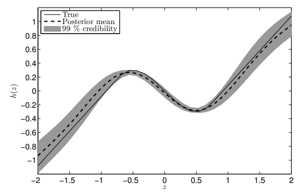
Recall that the task is to find the dynamical model based on measurements of the input signal and the output signal,

$$u_{1:N} = \{u_1, \dots, u_N\}, \qquad y_{1:N} = \{y_1, \dots, y_N\}.$$

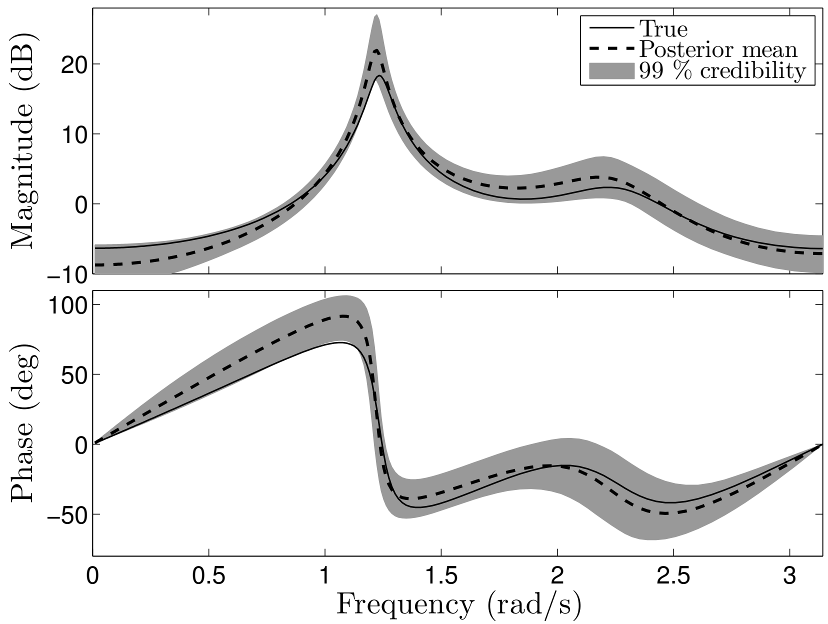
The red parts of the model below are inferred from data.

$$\begin{aligned} x_t &\in \mathbb{R}^{n}, & & \\ x_{t+1} &= A x_t + B u_t + w_t, & w_t &\sim \mathcal{N}(0, Q), \\ z_t &= \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} x_t, & & \\ y_t &= h(z_t) + e_t, & e_t &\sim \mathcal{N}(0, r) \end{aligned}$$

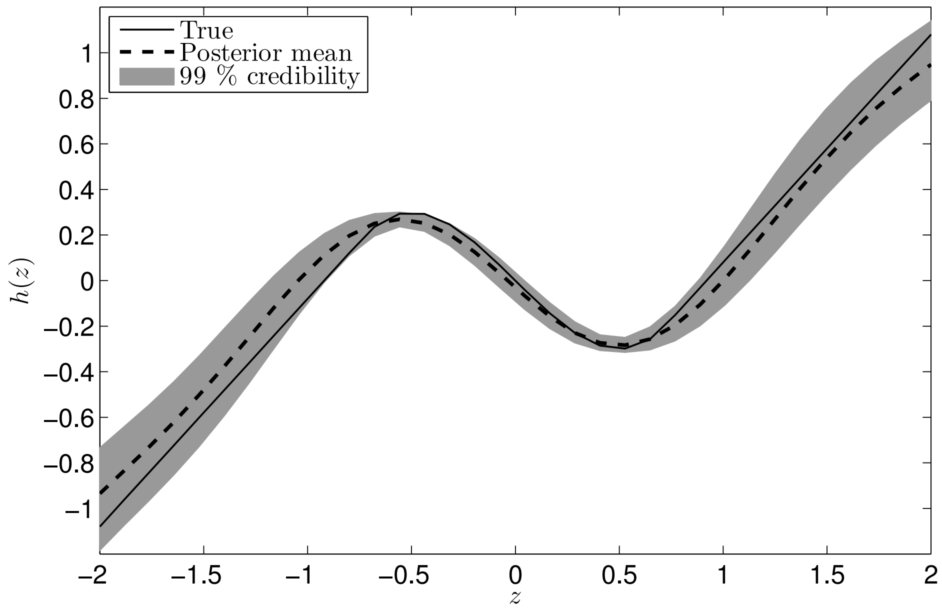


# Semi-parametric Wiener model - representing uncertainty




Linear system




Static nonlinearity

We used a nonparametric model for the static nonlinearity, more specifically a Gaussian process.

Show convergence using a movie

How do we represent uncertainty for a nonlinear model?!

# Conclusions

- **Take home message:** Given the computational tools that we have today it can be rewarding to resist the linear Gaussian convenience!

- There are by now a lot of tools that allows us to do this (e.g., SMC, PMCMC).

- There is a lot of **interesting research** that remains to be done!
- The **industrial utility** of the sensor fusion technology is **growing** as we speak!

---

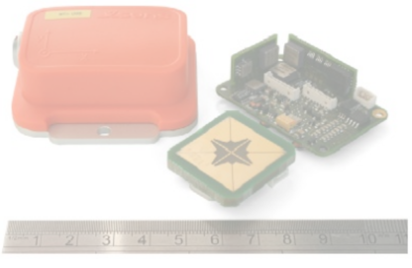
Throughout the talk I have touched upon a lot of methods that clearly deserves much more time than I gave them in this tutorial presentation.

Tomorrow and on Thursday I am giving an intensive course on this in Brussels, for details, see

http://www.rt.isy.liu.se/~schon/CourseBrussels2012/index.html

# Thank you for your attention!!

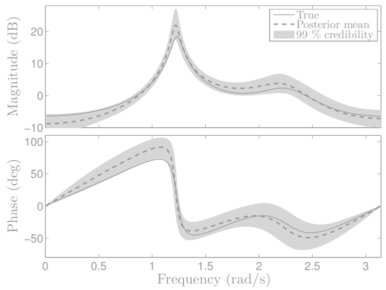
$$p(\theta, x_{1:T} \mid y_{1:T})$$

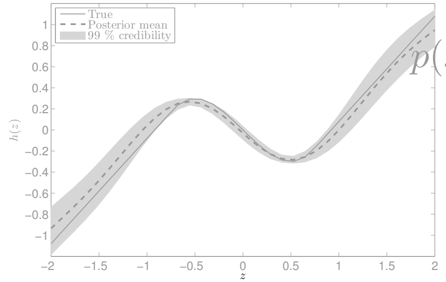


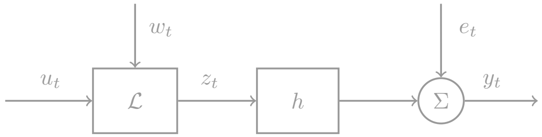
$$\begin{aligned} x_{t+1} \mid x_t &\sim f_\theta(x_{t+1} \mid x_t, u_t), \\ y_t \mid x_t &\sim h_\theta(y_t \mid x_t, u_t), \\ x_1 &\sim \mu_\theta(x_1). \end{aligned}$$

$$p(x_t \mid y_{1:t}) = \frac{h(y_t \mid x_t)p(x_t \mid y_{1:t-1})}{p(y_t \mid y_{1:t-1})},$$

$$p(x_{t+1} \mid y_{1:t}) = \int f(x_{t+1} \mid x_t)p(x_t \mid y_{1:t})\mathrm{d}x_t$$




**Special thanks to:** Jeroen Hol (*Xsens Technologies*), Lennart Ljung (*LiU*), Johan Kihlberg (*Xdin*), Fredrik Lindsten (*LiU*), Henk Luinge (*Xsens Technologies*), Michael Jordan (*Berkeley*), Brett Ninness (*University of Newcastle*), Per-Johan Nordlund (*Saab*), Simon Tegelid (*Xdin*) and Adrian Wills (*University of Newcastle*).