National Physical Laboratory

Sensor network topics

Peter Harris
Data Science Group, NPL

Bath SAMBA ITT, 29th January – 2nd February

Outline

- Background

- Three topics

  Prediction under uncertainty

  Source identification and tracking

  Network design

- Available data

  Real acoustic sensor network data

  Software to simulate sensor network data

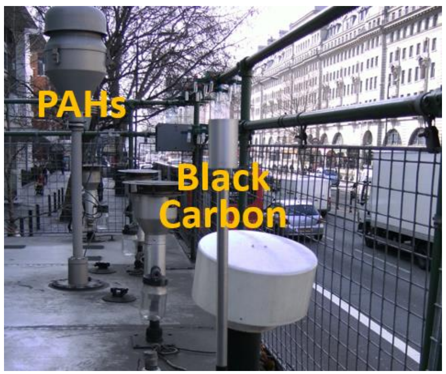
Sensor networks

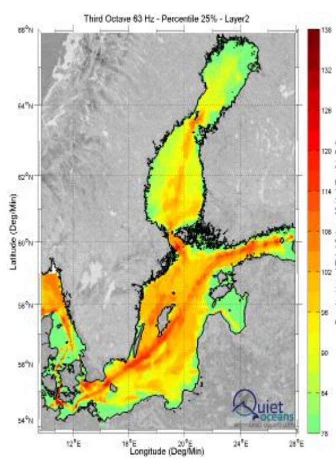
National Physical Laboratory

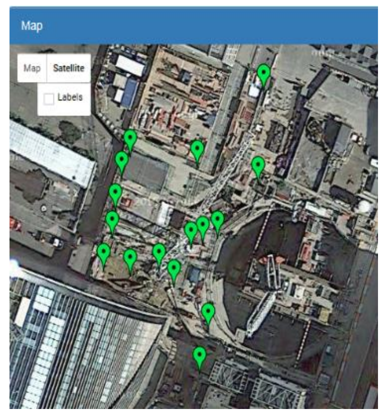
- Collection of (often) many low-cost sensors making discrete measurements (in space, time and frequency) of quantities that vary continuously (in space, time and frequency)
- Can operate at different scales (factory, neighbourhood, city, region, global)
- Source of large volumes of data
- Want to use data to make predictions and decisions, for source identification and attribution, all supported by statements of uncertainty

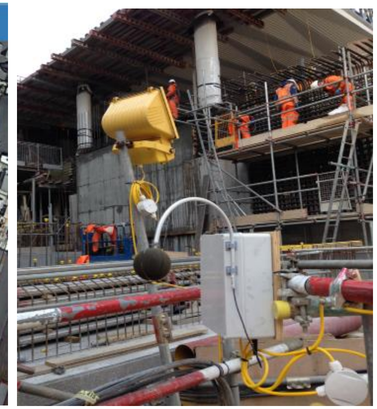
Examples

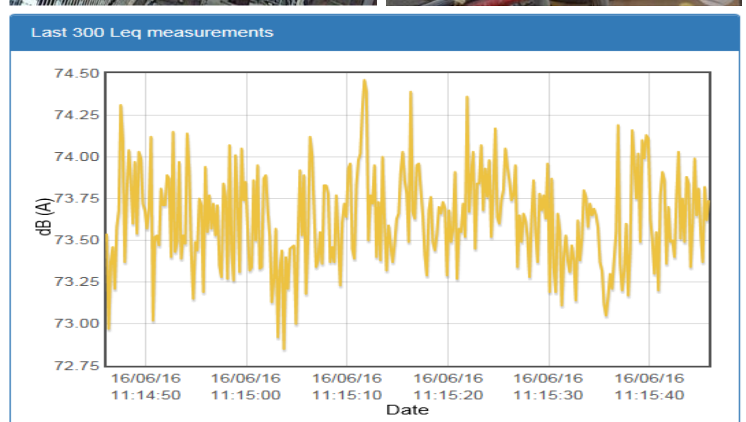
- Underwater acoustic noise in oceans
- Air quality and airborne noise in urban environments
- Radiometric measurements from satellites for EO





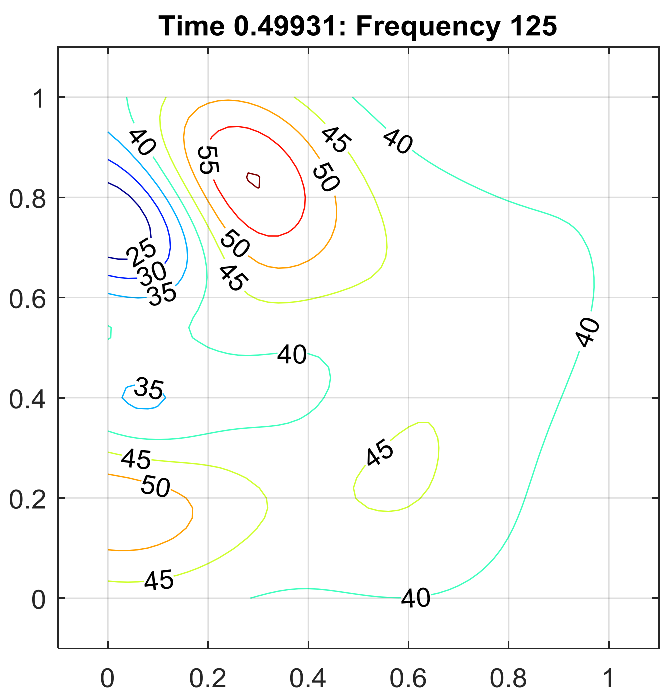


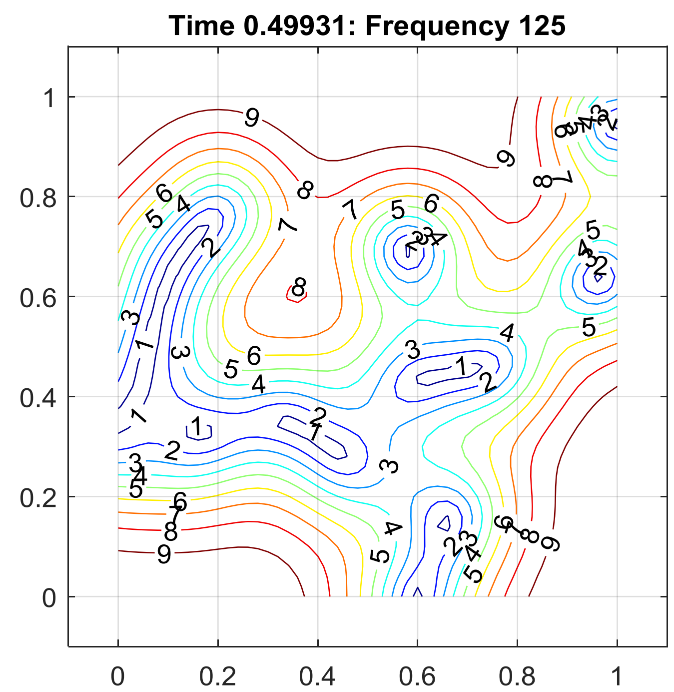




Prediction problem

- Create spatial-temporal-frequency maps of both estimates and uncertainties for measured quantities
- For single sensor network or sensor networks co-located and operating with different modalities (measuring different quantities that are dependent)





Prediction problem

- GPs are a useful tool but challenging to apply when the size of the training set exceeds a few thousand
  - Example of airborne acoustic network provided 500,000 measured values per day

- Given training set (X, y), a GP with covariance function $k(x_1, x_2)$, and noise variance σ_n^2

$$y_* = K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} y,$$

$$\text{cov}(y_*) = K(X_*, X_*) - K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} K(X, X_*)$$

Requires solution of large symmetric systems of equations

Prediction problem

National Physical Laboratory

- Learning about parameters of covariance function and σ_n^2 by maximising log-likelihood

$$\log p(y|K,\sigma_n^2) \propto -\frac{1}{2} y^{\mathrm{T}} [K(X,X) + \sigma_n^2 I]^{-1} y - \frac{1}{2} \log|K(X,X) + \sigma_n^2 I|$$

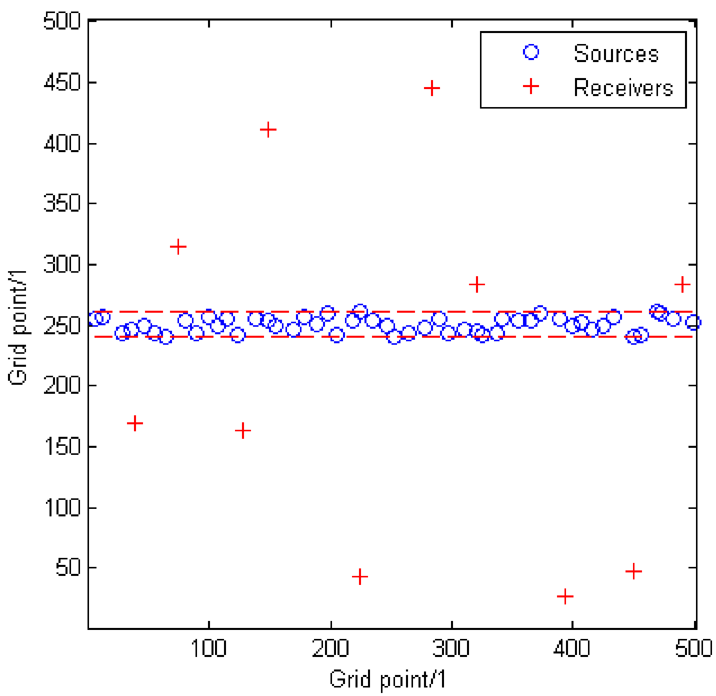
  Evaluation also requires eigenvalues of large symmetric matrix

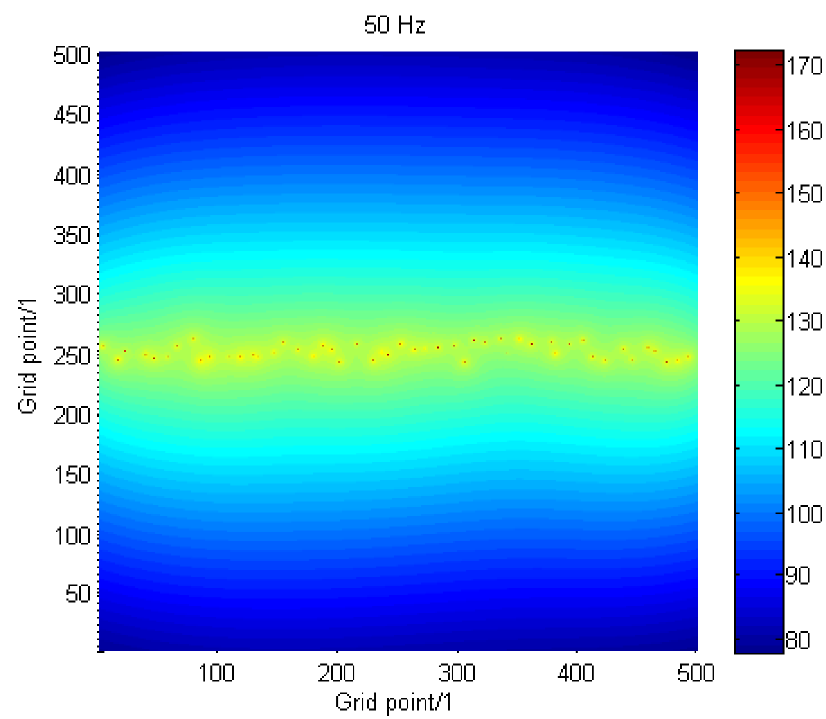
- Exploit structure of data (gridded) and structure of covariance function (separable) to decompose matrices as tensor matrix products … then, e.g., combine with iterative methods to solve equations

  Copes with large problems but computations are slow

Source identification and tracking problem

- Add knowledge that the measured quantity depends on a small number of sources and a model for how the quantity varies with distance from a source
  - For an acoustic problem, this could be a model of propagation loss





Source identification and tracking problem

- For a single time instant, vector s represents source factors at possible source locations

- Find solution $s \geq 0$ to minimize

$$\|As - y\|_2^2 + \tau\|s\|_1$$

First term matches model to sensor observations y: highly under-determined system: maybe ~100 observations but ~10,000 unknowns

Second term regularises solution by forcing sparsity

Parameter τ used to control balance between "fitting" data and level of sparsity

Columns of design matrix A are highly coherent, which limits the "spatial resolution" to which the source locations can be estimated

Source identification and tracking problem

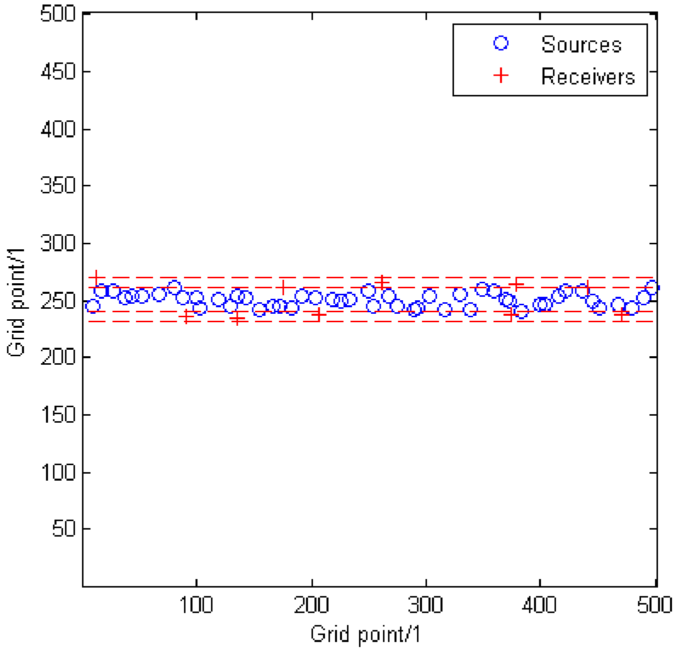
- Account for additional knowledge that locations of some of the sources are known (from GPS transponder data), but not their source levels

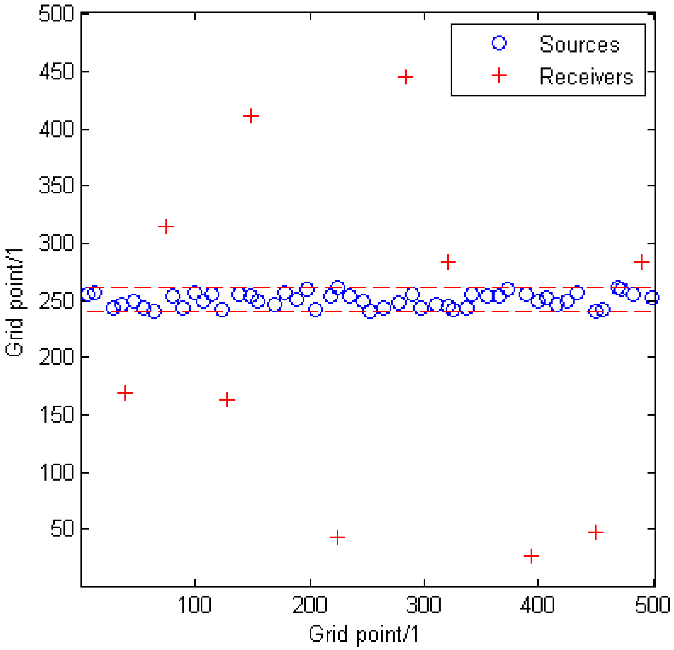
- Use knowledge of the precision of the sensors to provide uncertainty evaluation

- Use data recorded at consecutive times to track sources
  - The sources can be expected to move along continuous paths … initially straight-lines … and to have a constant source level
  - Real-time solution in the form of a filter?

Design problem

- To deploy a sensor network, need to decide
  - Number of sensors
  - Locations of sensors
  - "Quality of sensors": short-term (noise characteristics) and long-term (drift characteristics)





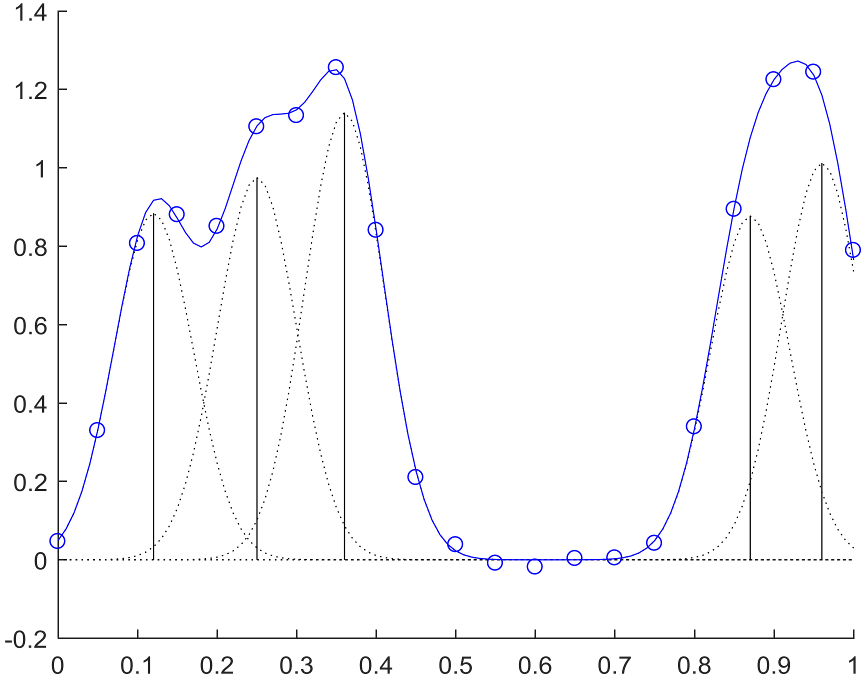
Design problem

- Optimise a measure of information (gain)
  - E.g., trace of covariance matrix for predictions made at a number of spatial locations
- But balanced against constraints
  - On cost (dependent on number and quality of sensors)
  - On where the sensors may be located (e.g., shipping lane is inaccessible)
- May also be a design specification
  - Requirement to make a prediction that meets a target uncertainty

Data

- Real acoustic sensor network data
  - 17 sensors at known locations (longitude and latitude), but sensor 9 recorded no data
  - Each sensor makes a measurement every minute
  - Each sensor records sound level at 26 frequencies (also aggregated into single broadband level)
  - Data for 8 consecutive days (in 8 separate files)
- Two spatial variables, temporal variable and can choose to ignore log-frequency variable (by using broadband level)
- Drop-outs in data (treated as missing values)

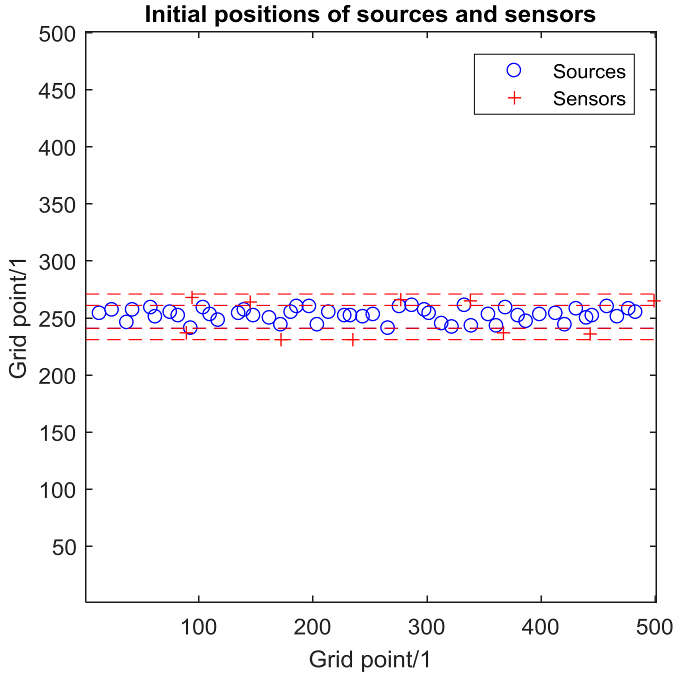
Simulation with single spatial variable

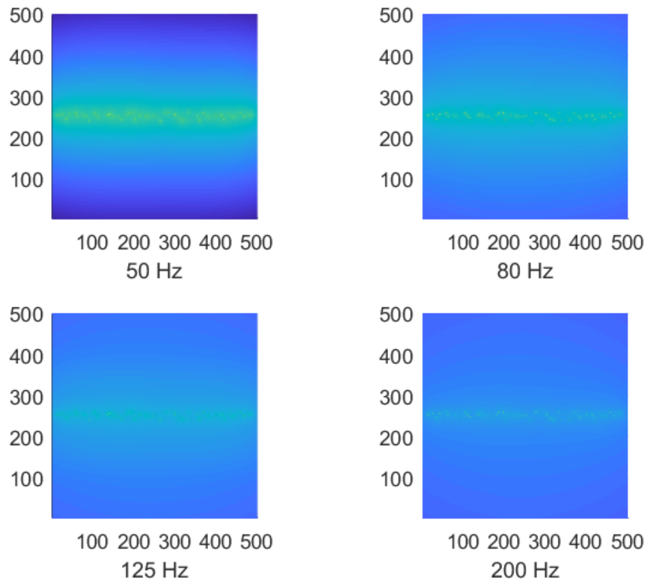


- Control number of sources, locations and source levels
- Control number of sensors, locations and noise variance
- Control Gaussian propagation loss function

- Matlab script to simulate matrix A and measurement vector y, and runs LASSO and Bayesian LASSO

Simulation with two spatial, temporal and frequency variables





- Extended to two spatial variables, temporal and frequency variable, with (more) realistic propagation loss function
- Matlab script to simulate design matrix A and measurement vectors y (as sources move)

National Physical Laboratory

Department for
Business, Energy
& Industrial Strategy

FUNDED BY BEIS

The National Physical Laboratory is operated by NPL Management Ltd, a wholly-owned company of the Department for Business, Energy and Industrial Strategy BEIS).