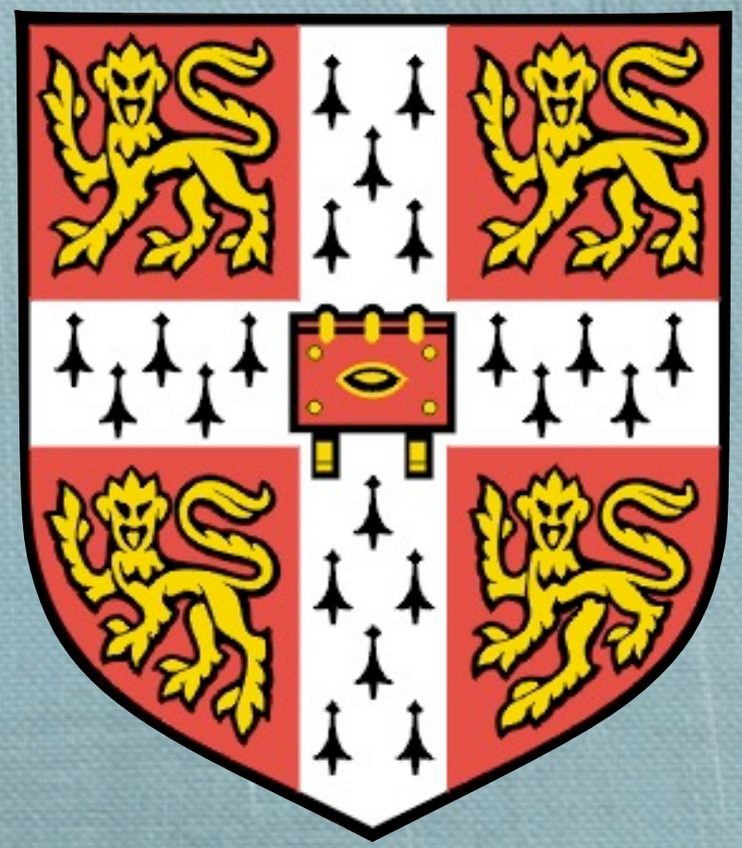
CCPN



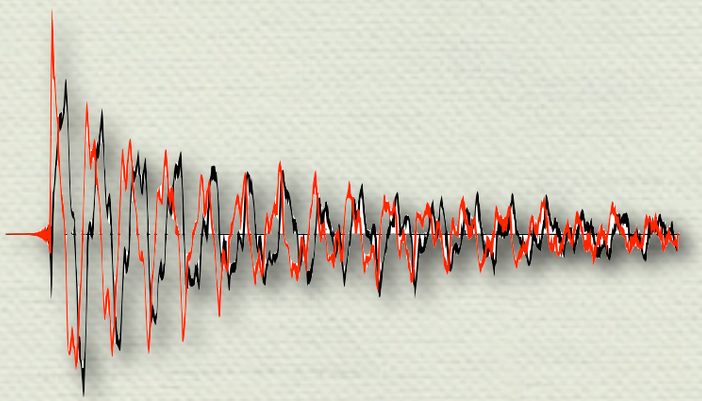
Non-Parametric Analysis of NMR data

Daniel O'Donovan

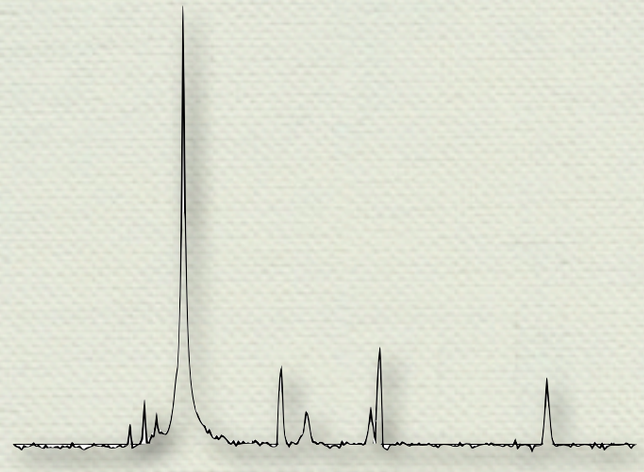
CCPN - Dept. of Biochemistry

Introduction

NMR Data



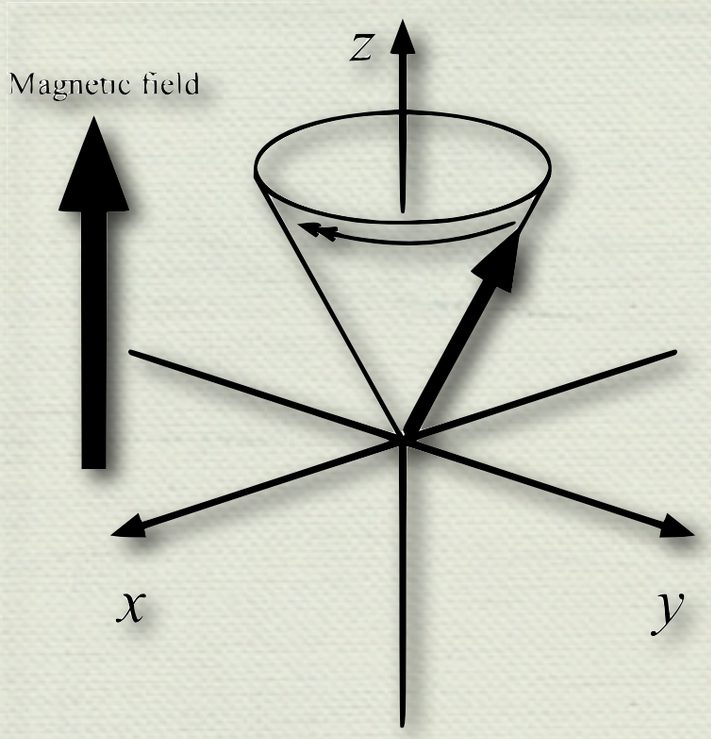
Fourier Transform

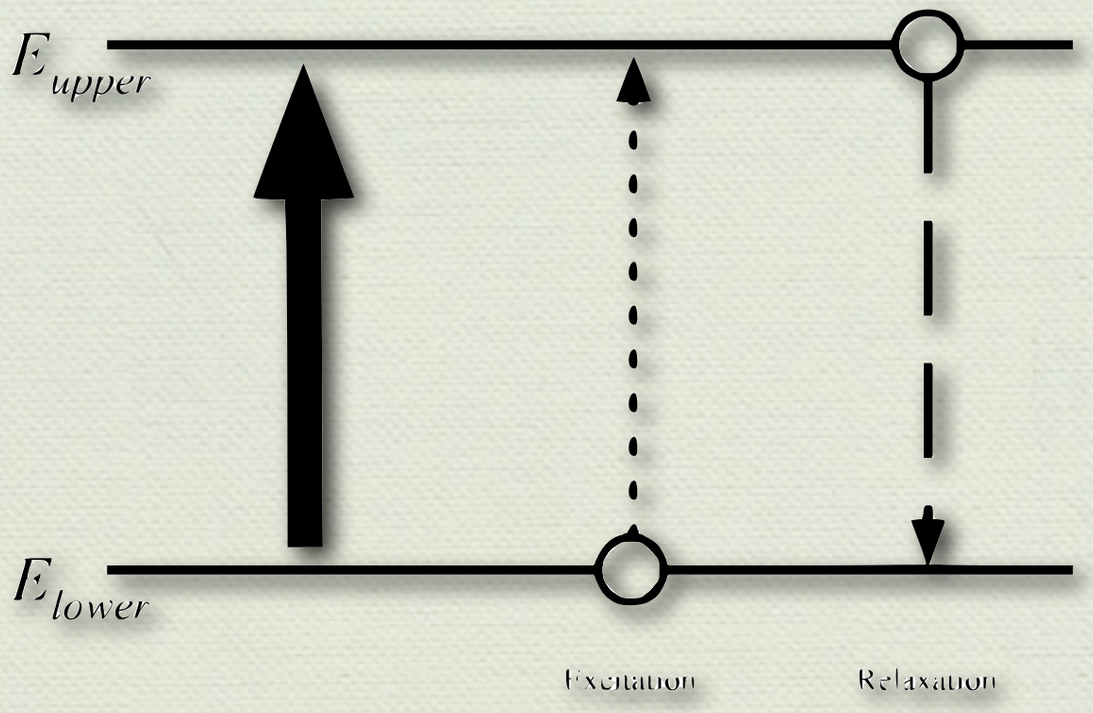


Non-Parametric Analysis

$$\text{maximise } S = -\sum p_i \log p_i$$

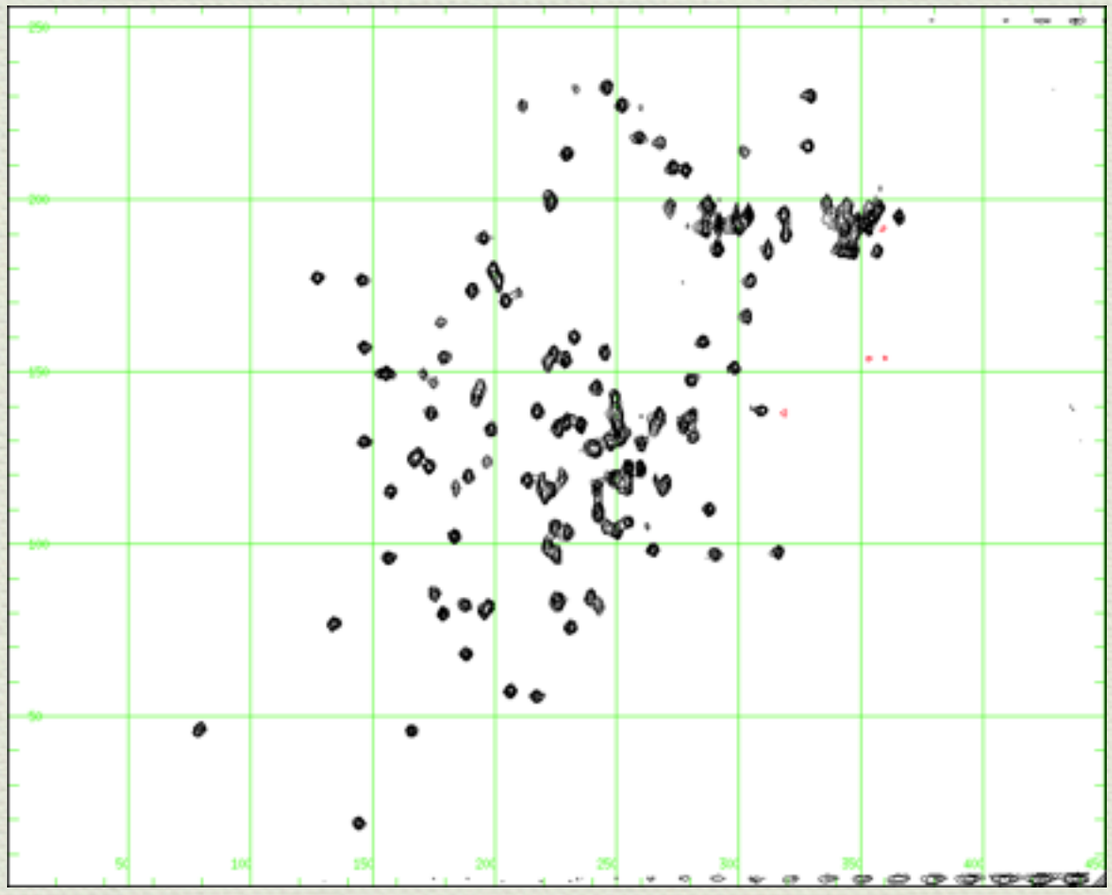
Intro: NMR





- Method for obtaining physical, chemical, electronic and structural information about a molecule.

Intro: NMR

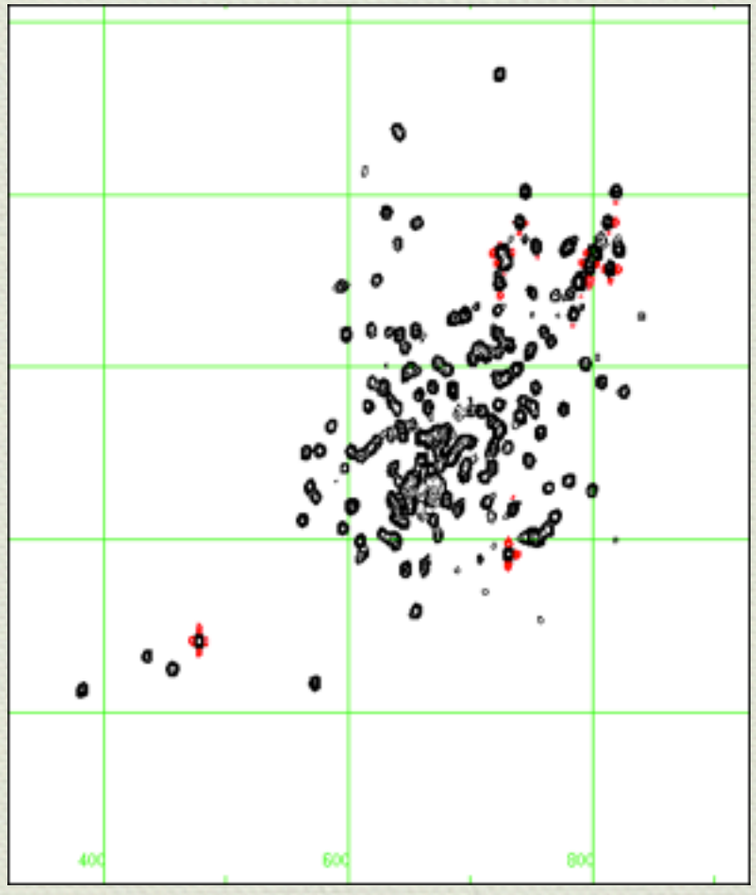


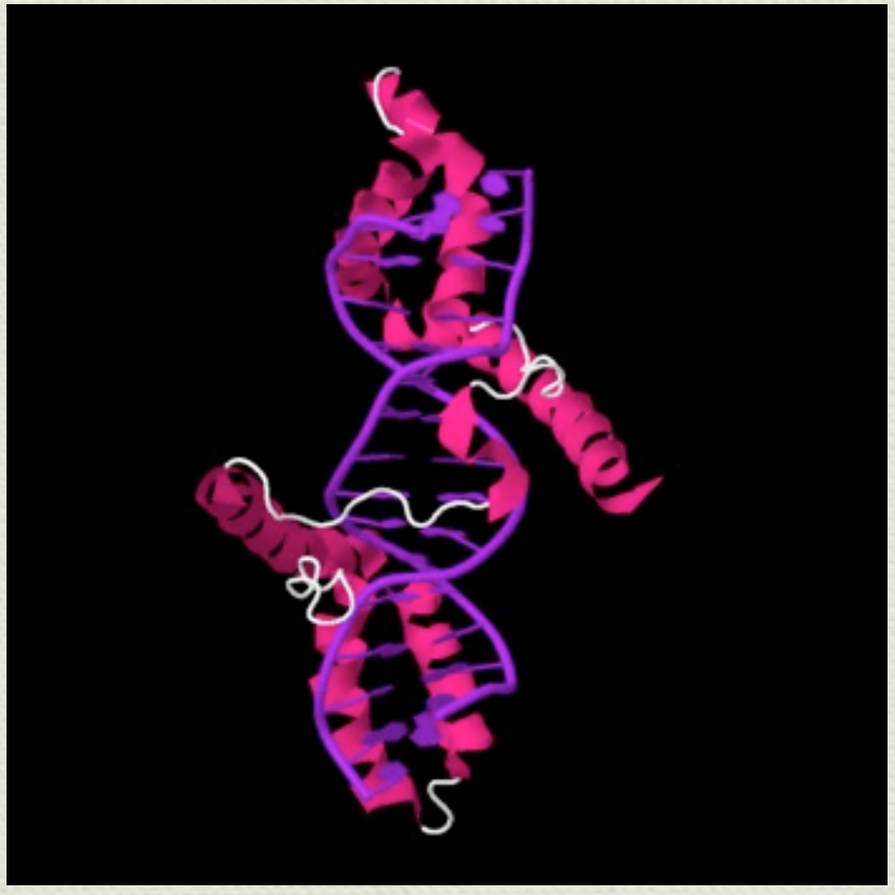
(1UST - Dr. Tim Stevens et al.)

- Small molecules - often easily determined

Intro: NMR

- Larger, more complex molecules require increased resolution and dimensionality.



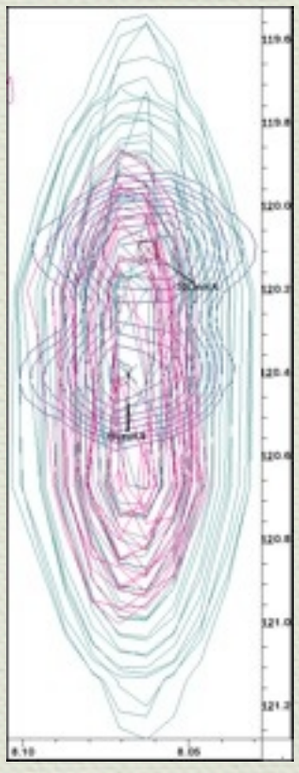


(2GZK - Dr. Katherine Stott et al.)

Intro: NMR

- Limit where time restrains the amount of data that it is reasonable to collect.

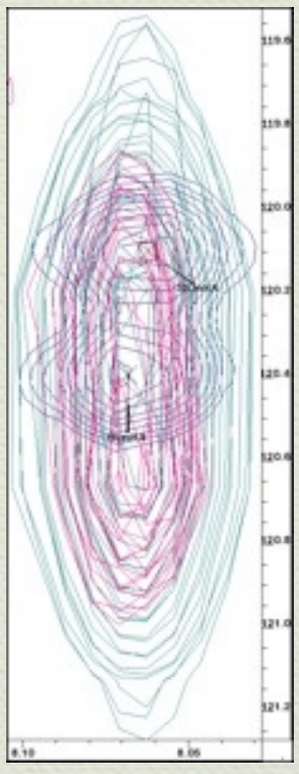
Before



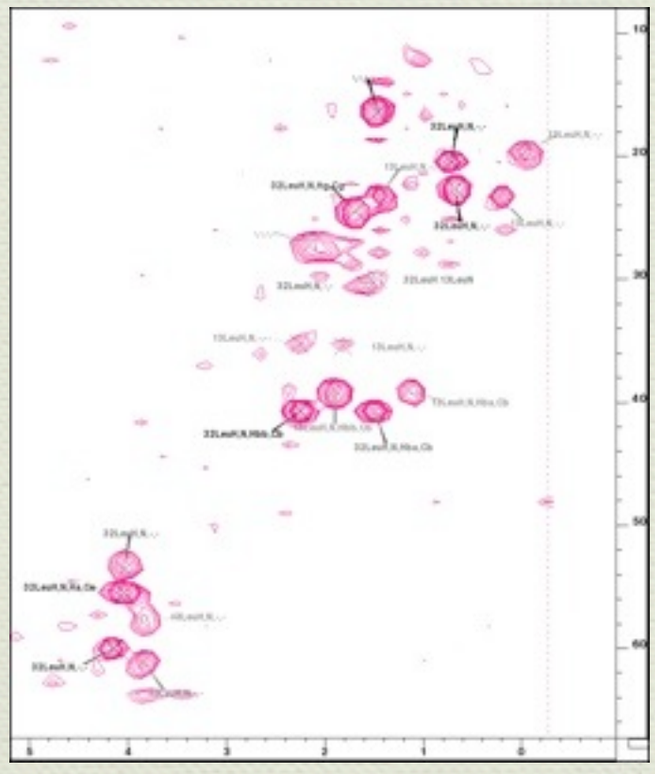
Intro: NMR

- Limit where time restrains the amount of data that it is reasonable to collect.
- New methods allow for increased dimensionality and resolution.

Before

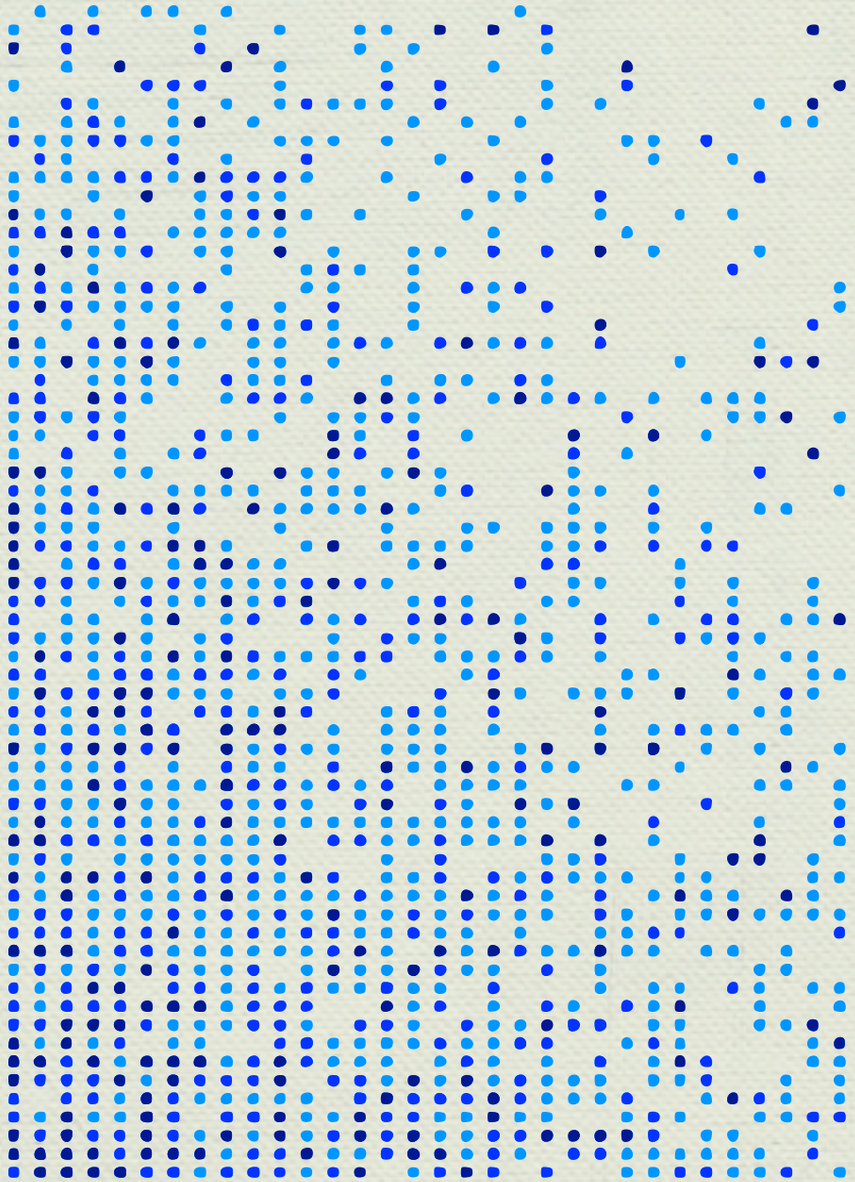


After



Intro: NMR

- Non-Uniform Sampling (NUS) method for increasing resolution with out increasing time involves recording non-continuously.
- NUS data cannot be processed using regular Fourier Transforms.
- Require new methods to process this data.



NUS Sample Grid

Intro: Non-Parametric Methods

- Non-parametric; make no (few) assumptions about model.
- Favourite non-parametric method:

Intro: Non-Parametric Methods

- Non-parametric; make no (few) assumptions about model.
- Favourite non-parametric method:

Maximum Entropy

Intro: Non-Parametric Methods

- Non-parametric; make no (few) assumptions about model.
- Favourite non-parametric method:

Maximum Entropy

- *Entropy* equivalent to *(negative) information*

Intro: Non-Parametric Methods

- Non-parametric; make no (few) assumptions about model.
- Favourite non-parametric method:

Maximum Entropy

- *Entropy* equivalent to *(negative) information*
- Maximising Entropy *minimises* information

Intro: Non-Parametric Methods

- Non-parametric; make no (few) assumptions about model.
- Favourite non-parametric method:

Maximum Entropy

- *Entropy* equivalent to *(negative) information*
- Maximising Entropy *minimises* information
- Analogy... *(courtesy of Ray Freeman)*

Non-Parametric Methods

- Terrible crime committed in Cambridge at night

King's College Cambridge at night
Photo by "jgraham"

Non-Parametric Methods

Non-Parametric Methods

Description from witness 1:

Hat

Non-Parametric Methods

Description from witness 1:

Hat

Description from witness 2:

Glasses

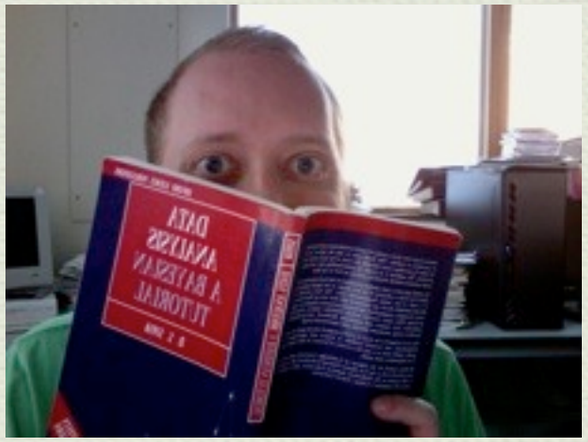
Non-Parametric Methods

Description from witness 1:

Hat

Description from witness 2:

Glasses



Description from witness 3:

Book

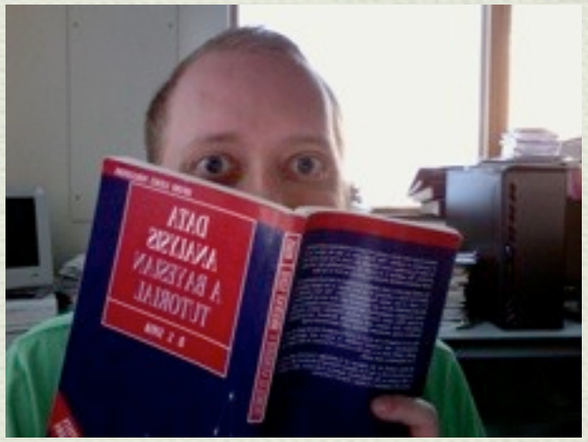
Non-Parametric Methods

Description from witness 1:

Hat

Description from witness 2:

Glasses



Description from witness 3:

Book

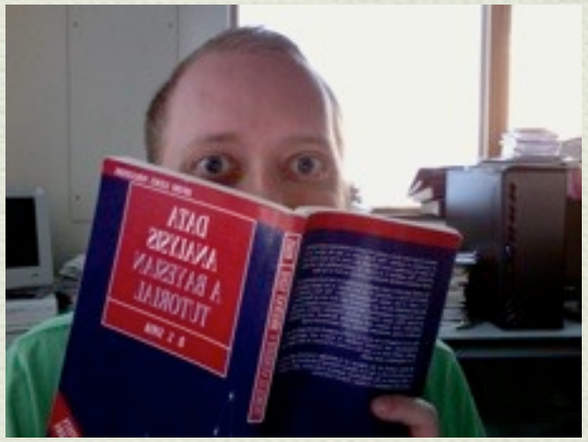
Description from witness 4:

Beard

Non-Parametric Methods

Description from witness 1:
Hat

Description from witness 2:
Glasses



Description from witness 3:
Book

Description from witness 4:
Beard

Least Committal Solution:

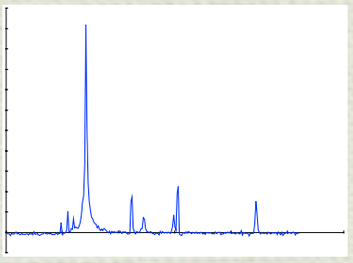
Maximum Entropy:
Simplest

Non-Parametric Methods

- In NMR, equivalent to having data set with parts missing.
- Task for Maximum Entropy is to reconstruct these missing parts.

Non-Parametric Methods

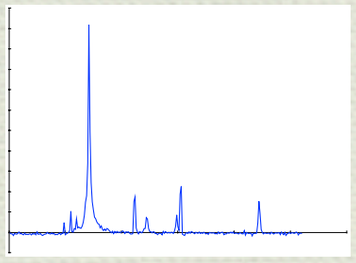
- In NMR, equivalent to having data set with parts missing.
- Task for Maximum Entropy is to reconstruct these missing parts.



Maximal Entropy spectrum

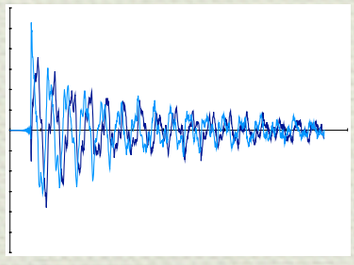
Non-Parametric Methods

- In NMR, equivalent to having data set with parts missing.
- Task for Maximum Entropy is to reconstruct these missing parts.



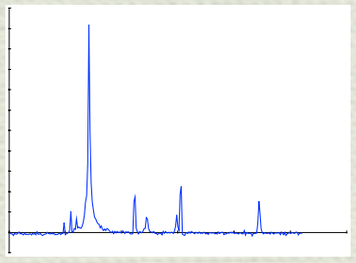
Maximal Entropy spectrum

Inverse Fourier Transform



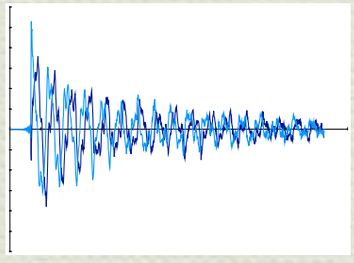
Non-Parametric Methods

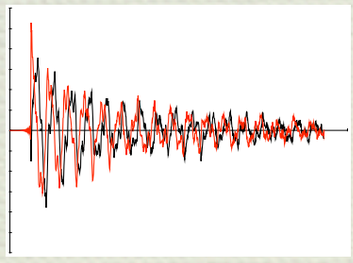
- In NMR, equivalent to having data set with parts missing.
- Task for Maximum Entropy is to reconstruct these missing parts.



Maximal Entropy spectrum

Inverse Fourier Transform

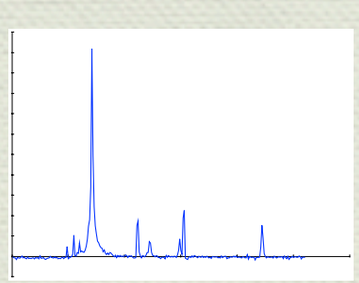




Calculate fit (Chi^2)

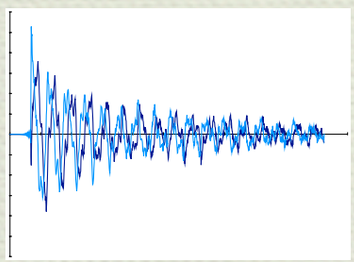
Non-Parametric Methods

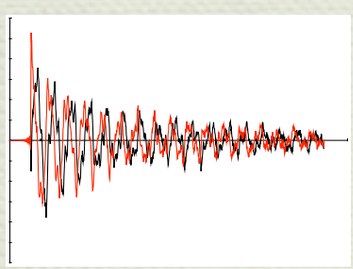
- In NMR, equivalent to having data set with parts missing.
- Task for Maximum Entropy is to reconstruct these missing parts.



Maximal Entropy spectrum

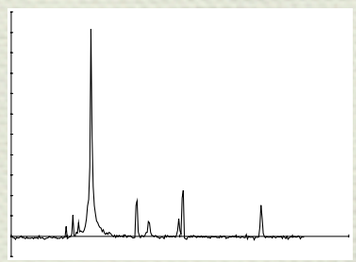
Inverse Fourier Transform





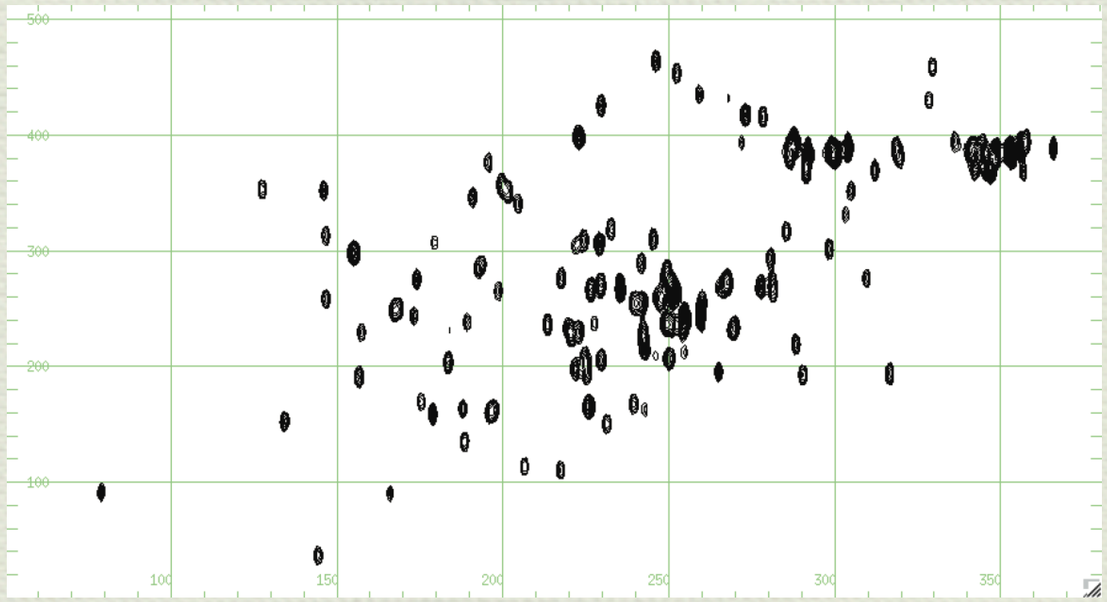
Calculate fit (Chi^2)

Repeat until convergence

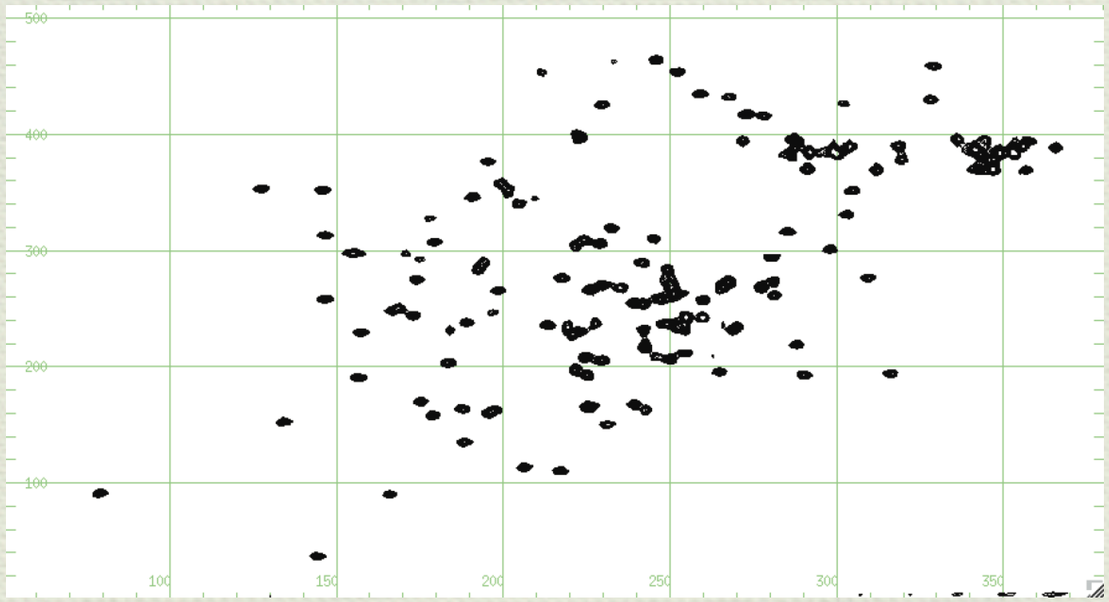


MaxEnt

- Typically have sharper peaks and less noise than regular FFT spectra.



Fourier Transform



Maximum Entropy

MaxEnt

- Maximum Entropy Advantages
  - Current state-of-the-art Maximum Entropy uses Bayesian methods to explore spectra.

MaxEnt

- Maximum Entropy Advantages
  - Current state-of-the-art Maximum Entropy uses Bayesian methods to explore spectra.
- Maximum Entropy Disadvantages
  - Fundamentally flawed: quality of result depends on resolution - should be resolution independent

MaxEnt

- Maximum Entropy Advantages
  - Current state-of-the-art Maximum Entropy uses Bayesian methods to explore spectra.
- Maximum Entropy Disadvantages
  - Fundamentally flawed: quality of result depends on resolution - should be resolution independent
- Alternatives
  - Considering *Independent Component Analysis* (ICA)
  - Sophisticated *Markov Chain Monte Carlo* methods (MCMC)

Code

- MemSys C Library (*MaxEnt Data Consultants, Drs. S. Gull and John Skilling (Dept. Astronomy)*)
  - Powerful exploration algorithms
    - Largely linear algebra operations
- Vector processing optimised using OpenMP, (also MPI and CUDA)
- Optimised Fourier Transform Library FFTW3
- Python (ctypes) wrapping code

Code: OpenMP

- Available in Fortran and C, through GCC 4.2+ and ICC
- In C, implemented with `#pragma` statements
- Trivial example:

Code: OpenMP

- Available in Fortran and C, through GCC 4.2+ and ICC
- In C, implemented with `#pragma` statements
- Trivial example:

```
for ( i = 0; i < HUGE; i++ )
{
    x[i] = a * y[i];
}
```

Code: OpenMP

- Available in Fortran and C, through GCC 4.2+ and ICC
- In C, implemented with `#pragma` statements
- Trivial example:

```
for ( i = 0; i < HUGE; i++ )
{
    x[i] = a * y[i];
}
```

```
#pragma omp parallel for \
        shared( HUGE, x, a, y) \
        private( i )
for ( i = 0; i < HUGE; i++ )
{
    x[i] = a * y[i];
}
```

Code: OpenMP

- Available in Fortran and C, through GCC 4.2+ and ICC
- In C, implemented with `#pragma` statements
- Trivial example:

```
for ( i = 0; i < HUGE; i++ )
{
    x[i] = a * y[i];
}
```

```
#pragma omp parallel for \
        shared( HUGE, x, a, y) \
        private( i )
for ( i = 0; i < HUGE; i++ )
{
    x[i] = a * y[i];
}
```

- Invoked (GCC) :
  - `$ gcc -fopenmp file.c -o file.o -lgomp`

Code: FFTW3

- FFTW3 the *Fastest Fourier Transform in the West*
  - Highly optimised C library for calculating discrete Fourier Transforms
  - Very easy to use and very fast
  - Available from http://www.fftw.org/ and package managers

Code: Python

- Python: general purpose and high level
  - Great for prototyping VERY quickly
    - Has plotting and math libraries
  - Great for calling optimised code libraries
    - Several ways to interface with foreign libraries
  - CCPN API and software written in Python
    - Available from http://www.ccpn.ac.uk/

Code: Python

- Example of Python with ctypes

```
# Load Python module ctypes
>>> import ctypes

# Load regular C shared library 'MaxEnt.so'
>>> cLib = ctypes.cdll.LoadLibrary( 'MaxEnt.so' )

# Run C function 'runUnitTests' with no arguments
>>> cLib.runUnitTests( None )
Tests Completed Successfully!
>>>
```

- Also possible to pass and return ints, floats, arrays, pointers and structures

With Thanks

- CCPN - Collaborative Computing Project for NMR
  - Prof. Ernest Laue, Tim Stevens, Wayne Boucher, Rasmus Fogh, John Ionides and Alan da Silva

Thursday, 24 June 2010