Initial Results from an International MHD Data Mining Collaboration

Datamining – extract new information from databases – (old and new)

First steps:
1) design database: one entry per mode, per timestep
2) preprocess raw data: each shot (~100MB) condensed by >100x
Preprocessing Stage 1: SVD

Divide each time signal into 1ms pieces, then within these,

**Singular value decomposition** “separates variables” time and space” for each mode

27/18 probe signals → one time function (chronos) \( C(t) \)

one spatial function (topos) \( \mathcal{I}(x) \)  \( F_{\text{mode}} = C(t) \cdot \mathcal{I}(x) \)

**per mode** *(actually 2 or 3 in practice, sin-like and cos-like, → travelling wave)*
Preprocessing 2: SVDs grouped into “flucstrucs”

**Singular value decomposition** “separates variables” time and space” for each mode

27/18 probe signals $\rightarrow$ one time function (chronos) $C(t)$

one spatial function (topos) $\mathcal{F}(x)$

$F_{\text{mode}} = C(t) \cdot \mathcal{F}(x)$

**per mode** (actually 2 or 3 in practice, sin-like and cos-like, $\rightarrow$ travelling wave)

Group Singular Vectors with matching spectra

$$\gamma_{c_1, c_2} = \frac{G(c_1, c_2)^2}{G(c_1, c_1)G(c_2, c_2)}; \quad > 0.7$$

$$G(a, b) = \langle |\mathcal{F}(a)|\mathcal{F}^*(b)\rangle$$
Data Mining Classification by Clustering

Cluster in delta-phase

Full dataset

freq. [kHz]

$\kappa_h$
Real Time Mode identification

- Identify by cluster probability density functions
- Multivariate nature produces huge range in values

**Solution:** modes are represented as **multivariate von Mises** distributions

- trivially compute the likelihood of any new data being of a certain type of documented mode.

**H-1 application:**

Mode is clearly the dashed green mode.
<table>
<thead>
<tr>
<th></th>
<th>H-1 Heliac</th>
<th>TJ-II</th>
<th>Heliotron-J</th>
<th>LHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>R (m)</td>
<td>1.0</td>
<td>1.5</td>
<td>1.2</td>
<td>~3.6</td>
</tr>
<tr>
<td>&lt;a&gt; (m)</td>
<td>0.2</td>
<td>~0.22</td>
<td>~0.25</td>
<td>0.5-0.65</td>
</tr>
<tr>
<td>B₀ (T)</td>
<td>0.5T (&lt;1.0T)</td>
<td>1.2T</td>
<td>1.2T (~&lt;1.5T)</td>
<td>0.4-3T</td>
</tr>
<tr>
<td>ℓ</td>
<td>1.1-1.5 (0.9-2)</td>
<td>1-2.5 (1.44)</td>
<td>0.4-0.6 (0.3-0.8)</td>
<td>0.3-1</td>
</tr>
<tr>
<td>Heating</td>
<td>ECH, ICRF[ω_ci], Helicon</td>
<td>ECH, 2xNBI</td>
<td>ECH, ICRF[ω_ci], 2xNBI</td>
<td>ECH, ICRF[ω_ci], NBI 2-20MW</td>
</tr>
<tr>
<td>nₑ</td>
<td>0.3-2×10¹⁸</td>
<td>2-16×10¹⁸</td>
<td>5-25×10¹⁸</td>
<td>20-200×10¹⁸</td>
</tr>
<tr>
<td>Temperature</td>
<td>50-100eV</td>
<td>Te~&gt;1keV, Ti~1-300eV</td>
<td>Te<del>2-400eV, Ti</del>1-300eV</td>
<td>Te<del>0.3-3.5keV, Ti</del>0.3-8.6keV</td>
</tr>
<tr>
<td>Vₐlfvén</td>
<td>~5-10×10⁶</td>
<td>~5-10×10⁶</td>
<td>~5-10×10⁶</td>
<td>~1.5-15×10⁶</td>
</tr>
<tr>
<td>Periods</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>m=10/l=2</td>
</tr>
<tr>
<td>Bₘ,ₙ(r-a)</td>
<td>Tor. 0.2, Helical 0.2</td>
<td>Tor 0.12, Helical 0.16</td>
<td>Tor 0.15, Helical 0.2</td>
<td>Tor 0.11, Helical 0.1</td>
</tr>
<tr>
<td>Config.</td>
<td>Well, Low shear</td>
<td>Well, Low shear</td>
<td>V. Low shear</td>
<td>Med-High shear</td>
</tr>
<tr>
<td>Features</td>
<td>Hi-res ne profile, Fine configuration scan</td>
<td>Flexible configuration, NBI and ECH</td>
<td>Variable bumpiness, RF, NBI fast ions</td>
<td>Highest power, β long pulse</td>
</tr>
</tbody>
</table>

![Plasma](image1.png)
H-1: Identification with Alfvén Eigenmodes: \( n_e \)

- Coherent mode near iota = 1.4, 26-60kHz, Alfvénic scaling with \( n_e \)
- \( m \) number resolved by bean array of Mirnov coils to be 2 or 3.

\[
V_{\text{Alfvén}} = \frac{B}{\sqrt{\mu_0 \rho}} \propto \frac{B}{\sqrt{n_e}}
\]

- Scaling in \( \sqrt{n_e} \) in time (right) and over various discharges (below)

Critical issue in fusion reactors:

\( V_{\text{Alfvén}} \sim \) fusion alpha velocity

\( \Rightarrow \) fusion driven instability!
H-1: Alfvén eigenmode “Zoo”
Identification with Alfvén eigenmodes: $k_{||}, \iota$

$$\omega_{\text{res}} = k_{||} V_A = \frac{(m/R_0)(\iota - n/m)}{B/\sqrt{\mu_0 \rho}}$$

- $k_{||}$ varies as the angle between magnetic field lines and the wave vector
  $$k_{||} \propto \iota - n/m$$
- $\iota$ resonant means $k_{||}, \omega \to 0$

*Expect $F_{\text{res}}$ to scale with $\delta \iota$*
Overall fit assuming radial location

Better fit of frequency to iota, $n_e$ obtained if the location of resonance is assumed be either at **the zero shear radius**, or at an outer radius if the associated resonance is not present.

Assumed mode location

$I \sim 5/4$
Heliotron J: Poloidal Modes from m = -4 to 4

Data set of > 2,000 shots, including both directions of $B_0$

Clusters – Freq. vs time

Corresponding phase variation
TJII: Alfvénic/Non Alfvénic Scalings distinguished by Kullback-Leibler divergence
LHD: spectra complex, huge data volume

Toroidal angle

Frequency (kHz)

Time (ms)

N=2
Conclusions/Future Work

Discovery of new information
• promising, but needs either very high quality of data, or human intervention (ideally both!)

Real time identification
• Works well using Von Mises distribution to reduce problems in probability density function

Incorporation into IEA Stellarator CWGM MHD database
• Needs further reduction to be most useful – several methods
  – Store cluster statistics for a concise overview
  – Store more complete data for some “canonical” shots
  – Develop importance criteria – relationship to transitions, confinement loss

➢ Time dependence important! – (Detering, Blackwell, Hegland, Pretty)
➢ Currently adding W7-AS data – tokamak data?

[ Open source python code “PYFUSION” http://pyfusion.googlecode.com ]
Supported an Australian Research Council grant and Kyoto University Visiting Professorship