Large Scale Data Analysis Case Study: NSTX-U Langmuir Probes

MA Jaworski

Princeton Plasma Physics Laboratory, Princeton NJ

Princeton Plasma Physics Laboratory
Graduate Summer School
Guest Lecture – August 17th, 2018

*Work supported by DOE contract DE-AC02-09CH11466
Outline of Material

- Will you have too much data and why that means you have to think
- Strategies for attacking your data
- Langmuir Probes as a case study of where things fall apart
- Extracting Electron Distribution Functions: practical aspects and recommended practices
Diverted tokamaks are the most developed concept for MFE

- Magnetic fields confine plasma in toroidal geometry
- Divertors developed to separate eroded material from core plasma

A lot of processes happen in dense region at target plate – we need to diagnose it!
Array of plasma diagnostics utilized during experiments

- Local plasma conditions and PFC currents measured with Langmuir probe system
- 2D fast-cameras provide nearly full toroidal coverage of divertor
How much data is this?

• Each I-V characteristic can be a single “data set”
  – 250 kHz sampling frequency on digitizer
  – 500 Hz sweep rate, 2 V-sweeps per cycle
  – 1000 sweeps per 1s discharge per probe
  – 2000 shots per year

• $2000 \times 1000 \times 4 = 8$ Million data sets in 1 year

• What are the bigger picture questions and strategies needed to pare this down?
What if I just analyze all the data by hand?

• Just not practical, but even if it were: would you want to?

• Scientific reporting is to provide sufficient information to replicate the work
  – Hand analysis means *you* are making decisions during analysis
  – If *you* are integral to analysis, no one else can repeat it!

• Systematic analysis allows study of the method itself

• Avoids “cherry-picking” if its systematic
Outline of Material

• Will you have too much data and why that means you have to think

• Strategies for attacking your data

• Langmuir Probes as a case study of where things fall apart

• Extracting Electron Distribution Functions: practical aspects and recommended practices
1. Know what you want to get from the data!
   – E.g. I want the most accurate Ne and Te in divertor
   – E.g. I want the post-ELM evolution in the divertor

2. Know what you don’t want in your data sets!
   – E.g. I don’t want to analyze turbulent plasmas
   – E.g. I *do* want to analyze turbulent plasmas

3. Be Selective using 1 and 2!
   – Large data sets means you can “afford” to throw some things away
   – Don’t slow yourself down analyzing every single case

4. Build in consistency checks (recommended practice)
   – Ensure your systematic analysis is being correctly implemented
Outline of Material

• Will you have too much data and why that means you have to think

• Strategies for attacking your data

• Langmuir Probes as a case study of where things fall apart

• Extracting Electron Distribution Functions: practical aspects and recommended practices
Langmuir probes as case study

• The (almost) original plasma diagnostic (I. Langmuir 1881-1957)

• “Simple to implement; hard to interpret”

• Integrated measure of charged particles in plasma

• LP theory is deep and complex, lots of regimes of operation (see Demidov 2002)

• Perturbative diagnostic (at least locally)

Irving Langmuir (source: wikipedia)
Many assumptions in “text-book” theory fall down

• Typical method:
  – Subtract ion saturation; fit exponential curve to electron current
  – Derive Ne from Te + Isat + plasma species

• Most derivations do not include magnetization
• Collisionality can be accounted for (see Laframboise method)
• Assumptions of Maxwellian distribution not always good to make
Classical interpretation often yields higher temperatures relative to other diagnostics

- Classical interpretation makes use of data up to floating potential
  - Assumes single Maxwellian distribution
  - Only uses ~5% of distribution

- Independent measurements often indicate lower temperatures
  - Thomson scattering on ASDEX had some indications of non-Max. pops.
  - Thomson scattering on DIII-D consistently lower $T_e$
  - Anomalously low sheath heat transmission coeff. on numerous machines
Why expect a Maxwellian distribution?

Maxwellian plasmas assumed due to plasma collisionality
- Collisionality often calculated based on system length

Numerous modeling studies indicate non-Maxwellian distributions
- Target plasmas result in low $T_e$ and high $N_e$ – yield large collisionalities in the divertor
- Non-Maxwellian distributions still obtained

Temperature scale length requires consideration as well

$\lambda_{ee} \approx 1 \times 10^{16} \frac{T_e^2}{N_e}$

$\nu^* = \frac{L}{\lambda_{ee}}$

$L \parallel$ vs. $\lambda_{Te}$

Detached plasma
High $N_e$, low $T_e$

Other examples:
- Fokker-Planck: Chodura, CPP, 1992
- PIC modeling: Tskhakaya, JNM, 2011

Collisionality Must Be Calculated with the Correct Scale Length

- Chodura, in 1992, pointed out the importance of local temperature scale length

- Application of Chodura criterion suggests simple limits for thermal conduction
  - Based on moments of electron distribution function
  - Most heat-carrying electrons have 3-5x thermal velocity

- NSTX divertor plasmas indicate $T_e > 15 \text{eV}$ for fluid conduction to hold

Temperature gradient scale length

$$L_{T_e} = \frac{T_e}{T'_e} \approx \frac{\kappa_{0e} T_e^{7/2}}{q_e}$$

Chodura criteria due to energetic electrons carrying heat

$$\frac{\lambda_{ee}}{L_{T_e}} < \frac{1}{100}$$

Scaling of minimum temperature to satisfy Chodura collisionality req.

$$T_e^{3/2} \geq \frac{10^{18} q_e}{\kappa_{0e} N_e}$$

Jaworski JNM 2013
Kinetic probe interpretation provides more complete analysis of IV characteristics

- When electron energy scale length much longer than probe perturbation scale, velocity “diffusion” term negligible
  - $f(r,v) \rightarrow f(x,W)$
  - $W$ is total energy
  - $f_0$ is distribution far from probe

- Solution for probe characteristic determined by geometry and diffusivity
  - In magnetized plasma, cross-field diffusivity scales with Larmor radius
  - Diffusivity parameter takes form $\psi(W) = \psi_0 W^{-1/2}$ in this case

- When $\psi_0 >> 1$, first derivative becomes proportional to distribution function

- Demonstrated on CASTOR tokamak

\[
x = r, \quad W = \frac{1}{2}mv^2 + e\phi(x)
\]
\[
\lambda_e \gg r_s
\]
\[
\nabla_x D_x(W) \nabla_x f_0 = 0
\]
\[
j_e(V) = \frac{8\pi e}{3m^2\gamma} \int_{eV}^{\infty} \frac{(W - eV)f_0(W)dW}{1 + \frac{W-eV}{W}\psi(W)}
\]
\[
\psi(W) = \frac{1}{\gamma\lambda(W)} \int_\alpha^{\infty} \frac{D(W)dr}{(\frac{r}{a})D(W - e\phi)}
\]
\[
\frac{dj_e(V)}{dV} \propto \left(\frac{eV}{\psi_0}\right)^{3/2} f_0(eV)
\]

Arslanbekov, PSST, 1994
Demidov, PoP, 1999
Popov, PPCF, 2009
Non-Maxwellian distributions likely: motivates kinetic Langmuir probe interpretation

- Fluid-based reconstruction (OEDGE) indicates conduction-limited regime violates Chodura conditions

- Kinetic Langmuir probe interpretation theory developed over 30 years ago
  - Golubovsky 1981 – first application to high-pressure discharges
  - Arslanbekov 1994 – application to high-pressure and magnetized discharges
  - Popov 2009 – application to tokamak edge region at midplane

- In the right conditions, first derivative becomes proportional to distribution function

\[ \nabla_x D_x(W) \nabla_x f_0 = 0 \]

\[ \frac{dj_e(V)}{dV} \propto \frac{(eV)^{3/2}}{\psi_0} f_0(eV) \]

Popov, PPCF, 2009; Jaworski FED 2012, Jaworski JNM 2013
Bi-modal distributions observed in NSTX divertor

- Typical distributions shown
  - Scrape-off layer plasma where classical $T_e \sim 15\text{eV}$
  - Private plasma example demonstrating $T_e \sim 1\text{eV}$
- Ion current effects due to sheath growth estimated to avoid including in fits$^{1,2}$
- Bi-modal distribution often “best” model
- Total density calculated from $I_{\text{sat}}$
  - Sound speed calculated using mixture of both plasma populations$^3$

---

Bi-modal distributions observed in NSTX divertor

- Typical distributions shown
  - Scrape-off layer plasma

- Ion current growth effects estimated to avoid including in fits\(^1,2\)
- Bi-modal distribution often "best" model
- Total density calculated from \(I_{\text{sat}}\)
  - Sound speed calculated using mixture of both plasma populations\(^3\)

\(^1\) Gunn, RSI, 1997; \(^2\) Godyak, Demidov, J.Appl.Phys.D, 2011; \(^3\) PC Stangeby, PPCF, 1995

DON'T TRUST THE PROBES!
Empirical plasma reconstruction provides framework for checking consistency between diagnostics

- Utilizes measured data points as starting point in constraining plasma models
- Solution improves as more and more data constrains background
- OEDGE code suite used here: Onion-Skin Method (OSM2) + EIRENE + DIVIMP
  - OSM2 fluid solver
  - EIRENE neutral hydrogen
  - DIVIMP Monte Carlo impurities
- Utilized here to provide fluid background and identify candidate diagnostics for comparison
Spectroscopy provides independent checks on density and temperature

- Divertor spectrometer viewing strike-point region during discharge
- Deuterium Balmer lines shown in this spectra
- Pressure broadening analysis indicates density of $3.6 \times 10^{20} \text{ m}^{-3}$ (mean, 2.1-5.5x$10^{20}$m$^{-3}$ min/max)
  - Existence of high-n Balmer lines indicates low temperature

![Graph showing spectroscopy results with Balmer lines and pressure broadening analysis](image)
Broadening measurement and CR modeling of hydrogen spectrum consistent with kinetic interpretation

- Pressure broadening yields density
- OEDGE plasma+neutral solution provides local parameters
- CR model calculates excited state populations for given background (Maxwellian!)
- Brightness ratios normalized to B6 consistent with 3-5eV

Jaworski JNM 2013
Outline of Material

• Will you have too much data and why that means you have to think

• Strategies for attacking your data

• Langmuir Probes as a case study of where things fall apart

• Extracting Electron Distribution Functions: practical aspects and recommended practices
Not all models have the same knowns and unknowns!

- **Classical interpretation**
  - $T_e$, $N_e$, and $V_f$ come out
  - $V_f$ easily identified in I-V characteristic
  - $T_e$ extracted from exponential fit
  - $N_e$ derived from $I_{sat}$ and $T_e$

- **Kinetic interpretation for EEDF**
  - Entire distribution function is up for grabs!
  - Energy is referenced to plasma potential (not known ahead of time)
  - Psi transport function not really known ahead of time (but can be guessed)

- If you don’t know/trust the value, find consistency checks!
The EEDF algorithm for NSTX-U

1. Get data from central storage
2. Determine simple IV characteristics for starting point (classical analysis)
3. Perform data smoothing (first-derivative noise)
4. Distribution function loop
   1. Construct model curve based on best guess of f(E)
   2. Determine best Vp from model curve chi-square
   3. Calculate new f(E) using new Vp (also solve Psi0)
   4. Calculate new I-V based on f(E) and check chi-square
5. Calculate derived parameters (bi-modal Te, Ne, etc)
6. Write data files
The value of chi-square ("goodness-of-fit")

- Provides a quantitative value relating the model curve to the data set
- Not necessarily valuable as an absolute value
  - Should be ~1 if uncertainty correctly defined
  - Less free-parameters is better
  - Uncertainty not always well defined!
  - Should be minimized for best fits
- Be wary of "black box" fitting algorithms and traps
Consistency checks are critical!

- Make lots of plots during process development, verbose output is helpful

- Bug checking is critical – are you really applying the model you intended?

- Constantly ask yourself: is my model really better than another one?

- CHECK IT, BE QUANTITATIVE
Other recommendations

• COMMENT YOUR CODE
  – You won’t know when you’ll get back to it or put it down
  – Comment while it’s fresh in your mind

• Remember what research is: you get to learn when you are wrong!
Thank you!

- Will you have too much data and why that means you have to think
- Strategies for attacking your data
- Langmuir Probes as a case study of where things fall apart
- Extracting Electron Distribution Functions: practical aspects and recommended practices