COMPASS
INSTITUTE OF PLASMA PHYSICS ASCR

# Statistically consistent Reynolds stress profiles

## Showcase of bootstrapping and kernel regression

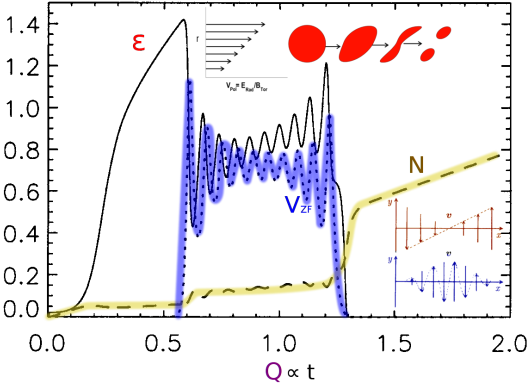
Ondřej Grover and the COMPASS team

$11^{th}$ FTTF Mariánská / 2019-01-10

# Outline

1. Motivation: H-mode, L-H transition and Zonal Flows
2. Experimental setup
3. Bootstrapped kernel regression of fluctuation statistics

**high confinement mode (H-mode):**
reference ITER scenario, $\sim 2 \times \tau_E, \beta_p$




E. Kim and P. H. Diamond
*Phys. Rev. Lett.* **90** 185006 (2003)

## quantities in the model

- $\epsilon$ turbulence intensity
- $V_{ZF} = \partial_r \tilde{v}_E$ Zonal Flow
- $N = \partial_r p$ pressure grad.
- mean flow $\partial_r \langle v_E \rangle \propto N$
- $Q$ external heating $\propto t$

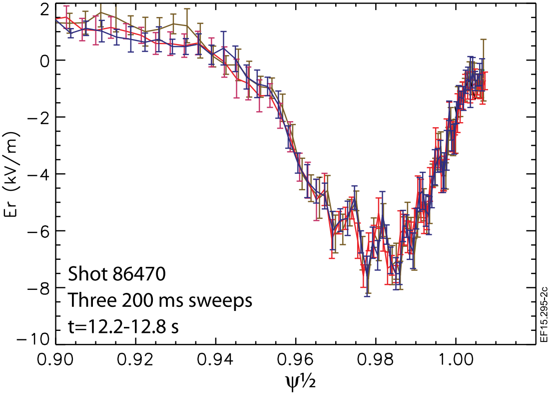
## Reynolds stress (RS)

$$\vec{v} = \langle \vec{v} \rangle + \tilde{\vec{v}}$$

$$\frac{\mathrm{d}\langle v_p \rangle}{\mathrm{d}t} = -\nabla_r \langle \tilde{v}_r \tilde{v}_p \rangle$$

$R_{rp} = \langle \tilde{v}_r \tilde{v}_p \rangle$ . . . RS

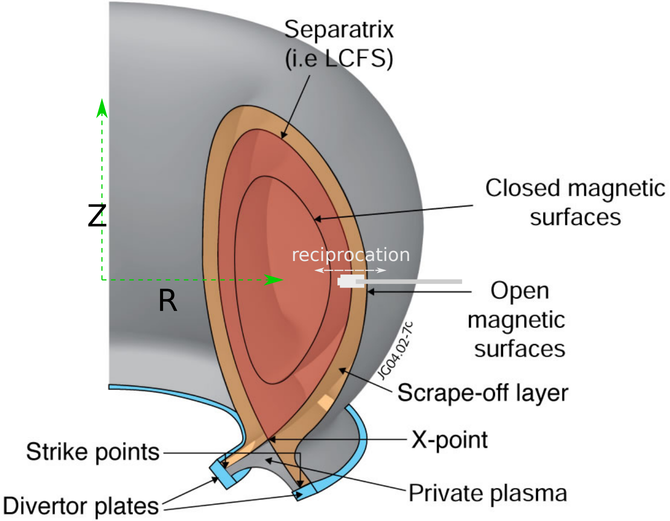
# Stationary zonal flows (ZF) in JET



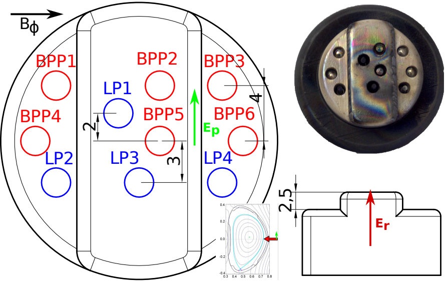

J.C. Hillesheim et al., *Phys. Rev. Lett.* **116**, 065002 (2016)

- Doppler backscattering reflectometry
- $\lambda_{ZF} \sim 10\rho_i$
- observed before H-mode

# Experiment geometry



# “Reynolds stress” probe head




design details[1,2]

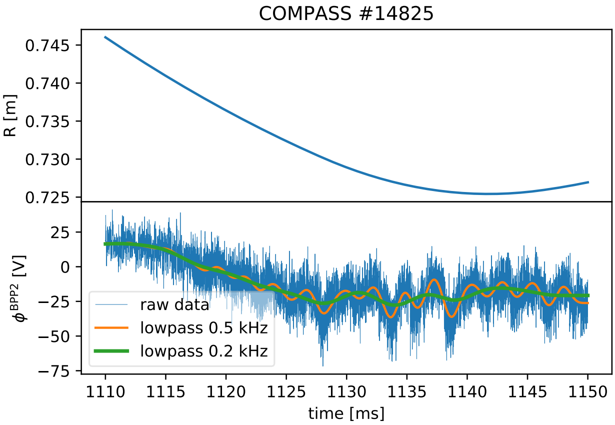
## Electrostatic turbulence

- 2D in drift plane $\perp B_\phi$
- $\vec{v} \approx \vec{v}_E = \frac{\vec{E}\times\vec{b}_\phi}{B_\phi}$
- $v_p = \frac{E_r}{B_\phi},\ v_r = \frac{E_p}{B_\phi}$
- $\tilde{B}_\phi = 0 \Rightarrow \tilde{\vec{v}} \propto \tilde{\vec{E}} = -\nabla\tilde{\phi}$

$$V_{fl}^{\mathrm{LP}} \approx \phi - 2.8\, T_e[\mathrm{eV}]$$
$$\phi^{\mathrm{BPP}} \approx \phi - 0.6\, T_e[\mathrm{eV}]$$ [a]

[a] J Adamek et al 2016 RSI DOI:10.1063/1.494

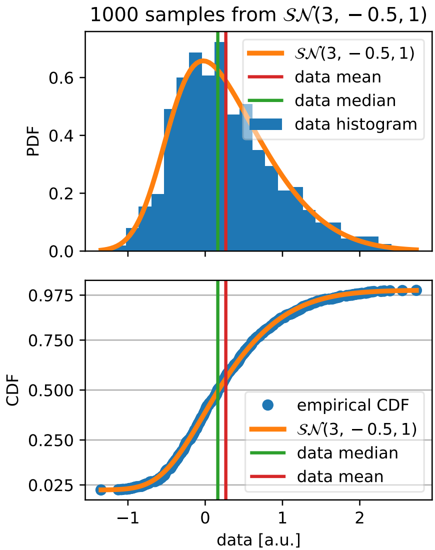
[1] O Grover et al 2018 Nucl. Fusion DOI:10.1088/1741-4326/aabb19
[2] O Grover et al 2017 RSI DOI:10.1063/1.4984240



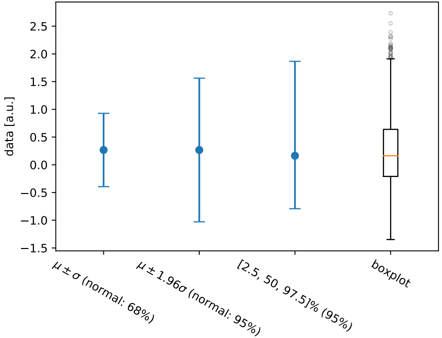

- only fluctuation **statistics** sought
- $\langle\cdot\rangle$ by lowpass filter: $f_{cutoff} = ?$ ergodicity?
- errorbars for lowpass?

smoothing in time (with varying reciprocation speed) **X** smoothing in space

# Errorbars for non-gaussian statistics



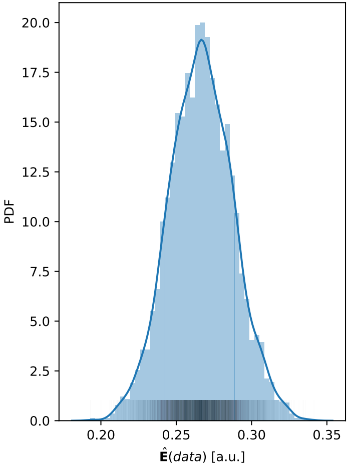

- real data never gaussian
- always check what errorbars mean (and their origin)
- empirical CDF (ECDF) robustly constructed **X** histogram (bins?)



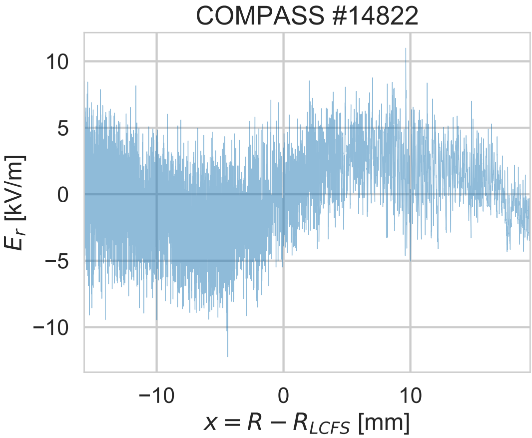
Uncertainty in **E**stimator (mean, average, fit, ...) results:
population $\mathbf{E}(\cdot)$ **vs.** sample $\hat{\mathbf{E}}(\cdot) \Rightarrow$ errorbars for $\hat{\mathbf{E}}(\cdot)$ ?

## Bootstrapping procedure

1. assume ECDF $\approx$ population CDF
2. sample new datasets from ECDF
   - same length as original dataset
   - "virtual new experiments"
   - in practice: random resampling of data with replacement
   - use $\sim$1000...10000 resamples
   - tests robustness to adding/omitting points
3. for each re-sample calculate $\hat{\mathbf{E}}(\cdot)$
4. statistics on samples of $\hat{\mathbf{E}}(\cdot)$



# Raw profile data



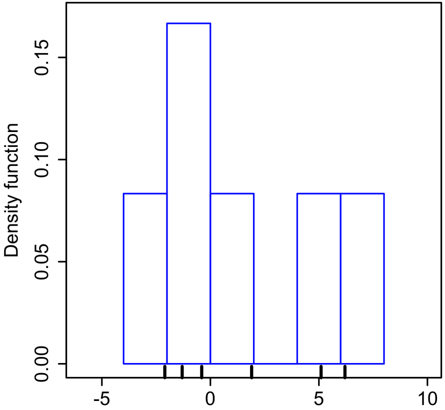

- points $[x(t), E_r(t)]$
- assume $p(x, E_r)$

serial auto-correlation in time series ⇒ actual bootstrapping procedure: blocks (not points) with autocorrelation time width

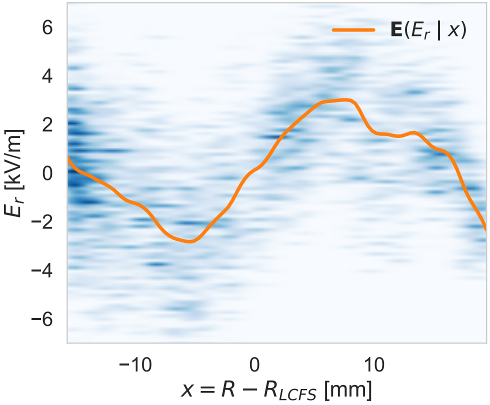
# Kernel density estimate

$$\text{PDF} \approx \hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$ [Wikipedia: Kernel density estimation]

kernel. . . e.g. $K(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})$, bandwidth $h = ?$



# Kernel regression (Nadaraya-Watson 1964)



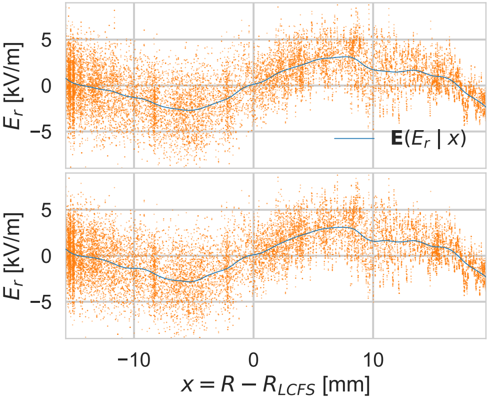

- assume smooth $p(x, E_r)$ $\Rightarrow$ kernel density estimate[a]
- kernel $K_i(x) \propto \exp\left(-\left(\frac{x-x_i}{h}\right)^2\right)$
- conditional expectation on dataset $\mathcal{D} = \{(x_i, E_{r,i})\}_{i=0}^{N}$

$$\mathbf{E}(E_r \mid x) = \frac{\sum K_i(x) \cdot E_{r,i}}{\sum K_i(x)}$$

- bandwidth $h \leftarrow$ cross-valid.

[a] J. S. Racine, Nonparametric Econometrics: A Primer (2008)

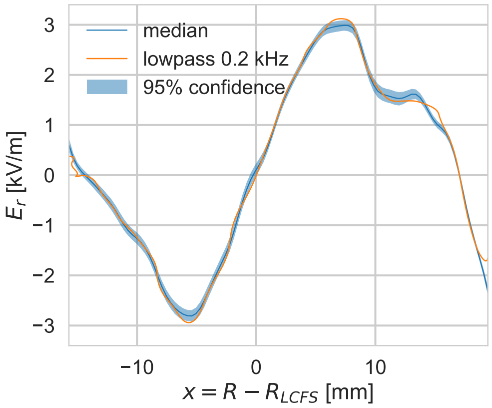
# Bootstrapping (Efron 1979): virtual repetitions of the experiment




- assume $p(x, E_r)$ the same if experiment repeated
- bootstrap (re-)sample: $\mathcal{D}^* = \{(x_{I_i}, E_{r,I_i})\}_{i=0}^{N}$ where $I_i$ random $\in \{0 \ldots, N\}$
- regression on $\sim 1000$ "virtual experiments"

$\Rightarrow$ test of robustness to omission/duplication of values

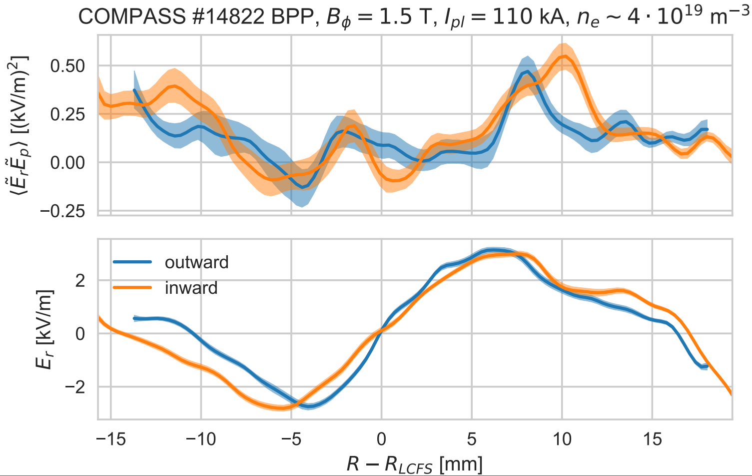
# Confidence bounds compared with lowpass



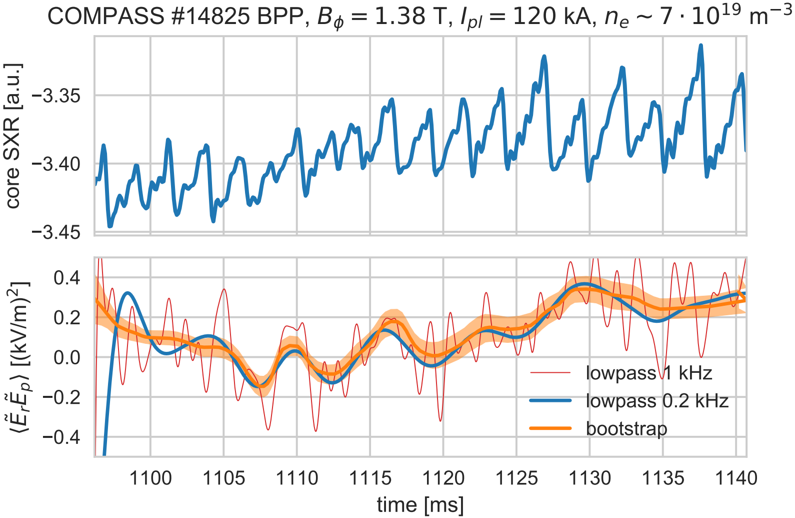

- point-wise quantiles of $\mathbf{E}(E_r \mid x)$ on $\left\{\mathcal{D}_|{}^*\right\}_{j=1}^{1000}$
- statistical significance of features

Assume $p(x, E_r, E_p)$ and calculate

$$\langle \tilde{E}_r \tilde{E}_p \rangle = \mathrm{Cov}(E_r, E_p \mid x) = \mathbf{E}(E_r \cdot E_p \mid x) - \mathbf{E}(E_r \mid x) \cdot \mathbf{E}(E_p \mid x)$$



# Robustness to saw-teeth crashes



## Statistics take-away

- always ask and/or make clear what errorbars mean
- bootstrapping for quantifying estimator uncertainty
- kernel regression useful for smoothing with good statistical sense

## Physics take-away

- Reynolds stress gradient can drive zonal flows whcih can decorrelate turbulence ($\rightarrow$ L-H transition)
- zonal-like structures observed, but did not produce (observable) zonal flows $\Rightarrow$ possibly collisional damping too high