to do

- illustrator how acentric Wilson follows from bivariate normal
- illustrate how conditional histogram arises from 2d histogram as a slice
- illustrate folded normal and bivariate Rice

Informative priors in scaling & merging

Doeke Hekstra

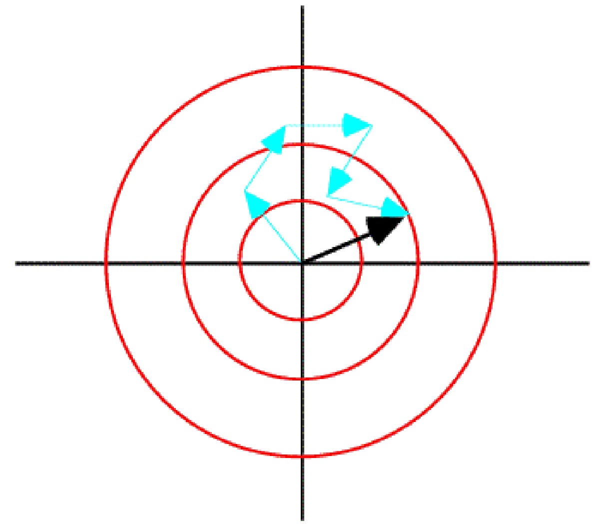
February 22, 2021

The CLT & the Wilson distribution

$$F_{hkl} = \sum_{i=1}^{n} f^{i} \cdot e^{2\pi i(hx_i + ky_i + lz_i)}$$

Electronic property of atom

Structural property (position)



Building up a structure in real space quickly looks like a random walk in the complex plane!

As a result, **the distribution of acentric structure factors looks like a bivariate normal distribution!**

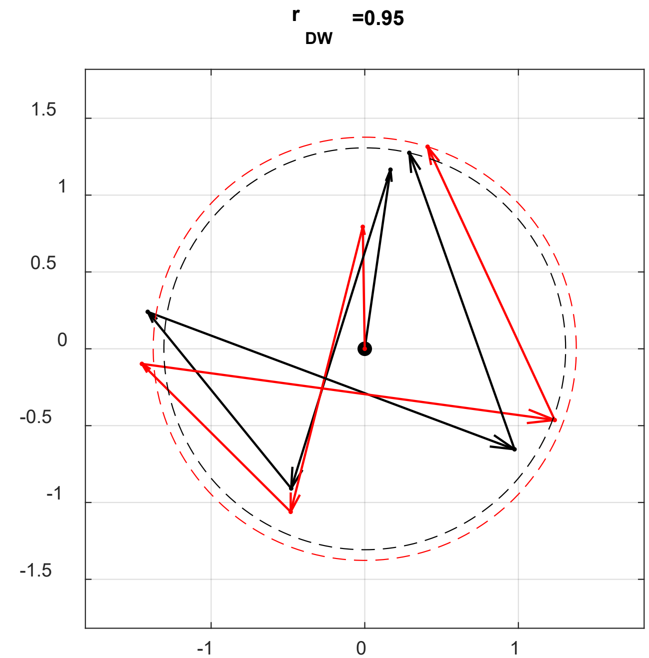
(For centrics, it looks like a univariate normal distribution.)

Related structures

$$F_{hkl} = \sum_{i=1}^{n} f^{i} \cdot e^{2\pi i(hx_i + ky_i + lz_i)}$$

Electronic property of atom

Structural property (position)



Acta Cryst. (1990). A46, 900–912

Structure-Factor Probabilities for Related Structures

By Randy J. Read

Department of Medical Microbiology and Infectious Diseases, University of Alberta, Edmonton, Alberta, Canada T6G 2H7

(*Received* 17 *January* 1990; *accepted* 9 *May* 1990)

Aim: Using informative priors in scaling & merging

For time-resolved crystallography data, we have two kinds of useful information
- High-quality synchrotron reference data, E_{ref}
- Knowledge that E_{on} and E_{off} tend to be highly correlated

For complex structure factors, these correlations are easily expressed as extensions of the Wilson model. For example,

$$P(E_1, E_2, E_3) = P\left(E_{1x}, E_{2x}, E_{3x}, E_{1y}, E_{2y}, E_{3y}\right) = N(0, C)$$

$$C = \frac{1}{2}\begin{bmatrix} 1 & r_x & r & 0 & 0 & 0 \\ r_x & 1 & r & 0 & 0 & 0 \\ r & r & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & r_x & r \\ 0 & 0 & 0 & r_x & 1 & r \\ 0 & 0 & 0 & r & r & 1 \end{bmatrix}$$

Aim: Using informative priors in scaling & merging

For time-resolved crystallography data, we have two kinds of useful information

- High-quality synchrotron reference data, E_{ref}
- Knowledge that E_{on} and E_{off} tend to be highly correlated

For complex structure factors, all of these correlations are easily expressed as extensions of the Wilson model. For example,

$$P(E_1, E_2 | E_3) = P\left(E_{1x}, E_{2x}, E_{1y}, E_{2y} | E_{3x}, E_{3y}\right) = N\left(rE_3, C_{1,2|3}\right)$$

$$C_{1,2|3} = \frac{1}{2}\begin{bmatrix} 1-r^2 & r_x - r^2 & 0 & 0 \\ r_x - r^2 & 1-r^2 & 0 & 0 \\ 0 & 0 & & \\ 0 & 0 & & \end{bmatrix}$$

$$r = \frac{Cov(E_{1x}, E_{3x})}{\sqrt{Var(E_{1x})Var(E_{3x})}}, \textit{etc.}$$

https://en.wikipedia.org/wiki/M

If there is a sufficient number of independent differences between $\mathfrak{F}$ and $\mathfrak{G}$, the distribution of **F** will be a Gaussian with variance $\varepsilon\sigma_\Delta^2$ about **DG**. In the non-centric case, the variance is distributed in the complex plane, giving rise to the following conditional probability distribution:

$$p_N[\mathbf{F}|\mathbf{G}] = \frac{1}{\pi\varepsilon\sigma_\Delta^2}\exp\left(-\frac{|\mathbf{F}-D\mathbf{G}|^2}{\varepsilon\sigma_\Delta^2}\right). \qquad (7)$$

Aim: Using informative priors in scaling & merging

For time-resolved crystallography data, we have two kinds of useful information

- High-quality synchrotron reference data, E_{ref}
- Knowledge that E_{on} and E_{off} tend to be highly correlated

$$
\begin{tikzcd}
P(E_1)P(E_2) \arrow[r, "\text{Correlation w\textbackslash reference}"] \arrow[d, "\text{Correlation among on, off}"'] & P\left(E_1|E_{ref}\right)P(E_2|E_{ref}) \arrow[d] \\
P(E_1, E_2) \arrow[r] & P(E_1, E_2|E_{ref})
\end{tikzcd}
$$

Aim: Using informative priors in scaling & merging

For now:

1. Normalization for better E_{ref}
2. Suitability of the Rice and Folded-Normal distributions

Pending

1. Use of the $\text{Bivariate Non}-\text{central } \chi^2$ distribution (implemented*, but don't know how to pick parameters).

Acentric

- $P(E_1) \sim Wilson$
- $P(E_1|E_2) \sim Rice$
- $P\left(E_1 \middle| E_{ref}\right) \sim Rice$
- $P(|E_1|, |E_2|)$ tbd
- $P(|E_1|, |E_2| \mid E_{ref}) \sim Bivariate\ Rice$ (done!)

  $P(|E_1|^2, |E_2|^2 \mid E_{ref}) \sim \text{Bivariate Non}-\text{central } \chi^2$ (unfinished)

* https://www.researchgate.net/publication/220557408

Aim: Using informative priors in scaling & merging

Sampling for the bivariate Rice dist: draw from

$$P(E_1, E_2 | E_3) = P\left(E_{1x}, E_{2x}, E_{1y}, E_{2x} | E_{3x}, E_{3y}\right) = N\left(rE_3, C_{1,2|3}\right)$$

ıditional covariance matrix

$$C_{1,2|3} = \frac{1}{2}\begin{bmatrix} 1-r^2 & r_x - r^2 & 0 & 0 \\ r_x - r^2 & 1-r^2 & 0 & 0 \\ 0 & 0 & 1-r^2 & r_x - r^2 \\ 0 & 0 & r_x - r^2 & 1-r^2 \end{bmatrix}$$

sample as $\sqrt{E_{1x}^2} + E_{1y}^2$ etc.

https://en.wikipedia.org/wiki/Gaussian_integral#n-dimensional_and_functional_generalization

Aim: Using informative priors in scaling & merging

Sampling for the bivariate Von Mises dist

$$P(E_1, E_2 | E_3) = P\left(E_{1x}, E_{2x}, E_{1y}, E_{2x} | E_{3x}, E_{3y}\right) = N\left(rE_3, C_{1,2|3}\right)$$

ıditional covariance matrix

$$C_{1,2|3} = \frac{1}{2}\begin{bmatrix} 1-r^2 & r_x - r^2 & 0 & 0 \\ r_x - r^2 & 1-r^2 & 0 & 0 \\ 0 & 0 & 1-r^2 & r_x - r^2 \\ 0 & 0 & r_x - r^2 & 1-r^2 \end{bmatrix}$$

https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1541-0420.2006.00682.x

Normalizing structure factors

Three steps

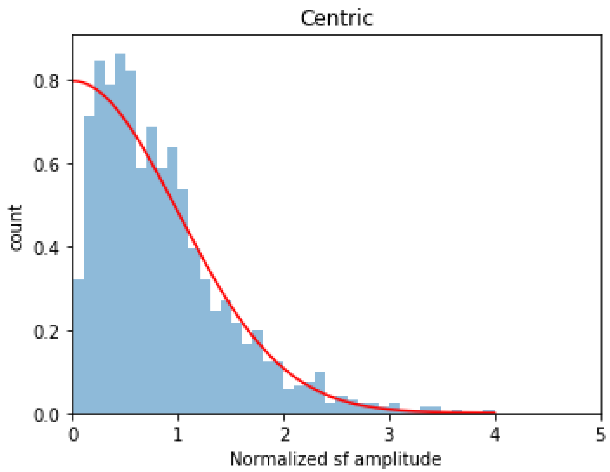
1. Fit $\boxed{E_h = f e^{-s^T \boldsymbol{B} s} \frac{F_h}{\sqrt{\varepsilon}},}$ using the Wilson distributions as the loss function, with $s = 1/d_h$ and h short for (h, k, l).

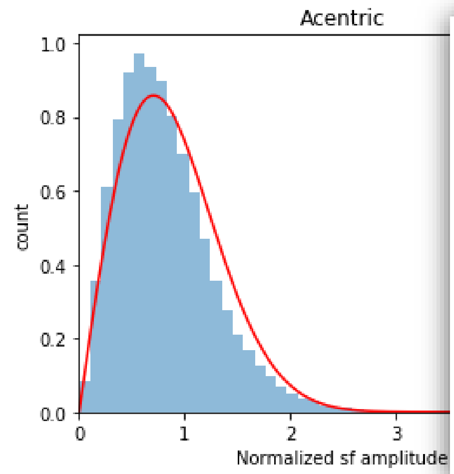
2. Fit $\boxed{E'_h = E_h\left(\sum_m A_m \cos\left(2\pi\tilde{h} \cdot m\right) + B_m \sin\left(2\pi\tilde{h} \cdot m\right)\right),}$ with $\tilde{h} = \left(\frac{h}{N_h}, \frac{k}{N_k}, \frac{l}{N_l}\right)$ and m short for (m, n, p), and the same loss function for increasing $m_{max} = n_{max} = p_{max}$. Pick best Fourier order by cross-validation.

3. Perform k-nearest neighbor regression on the $|E'|^2$. Obtains Σ as the local estimate of $\langle E'^2 \rangle$. Then $E_{knn} = E'/\sqrt{\Sigma}$.

Example: normalizing 1OTB

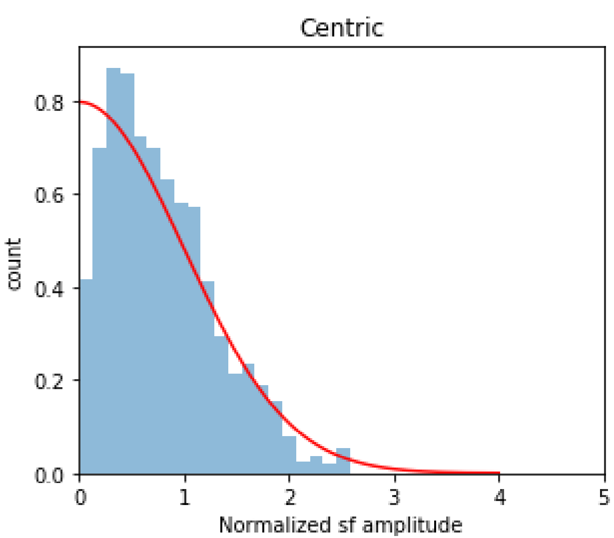
After simple anisotropic scaling of 1OTB:

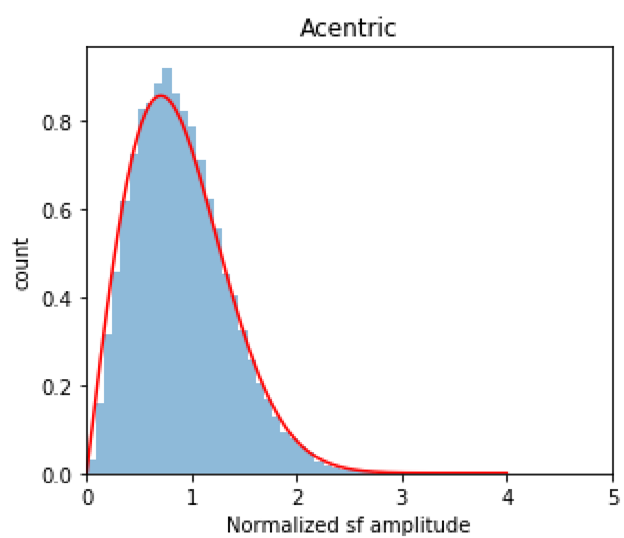


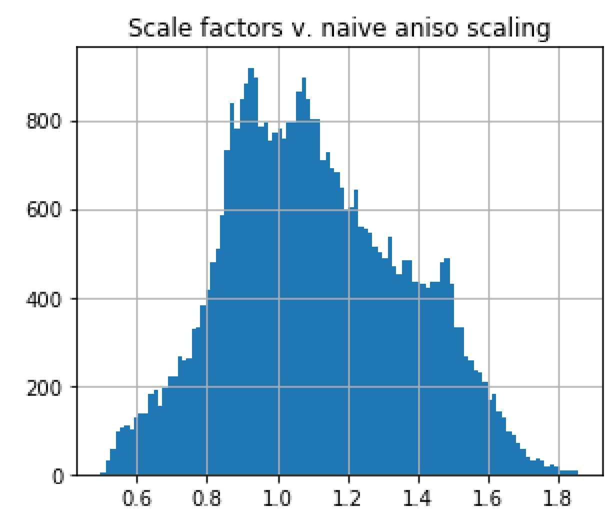



```
For n = 1 the test loss = 6180.78
Elapsed time: 3.284 s
For n = 2 the test loss = 5248.84
Elapsed time: 5.117 s
For n = 3 the test loss = 5125.22
Elapsed time: 37.18 s
For n = 4 the test loss = 5075.84
Elapsed time: 183.1 s
For n = 5 the test loss = 5083.3
Elapsed time: 896.2 s
```

After anisotropic scaling with Fourier corrections of 1OTB:

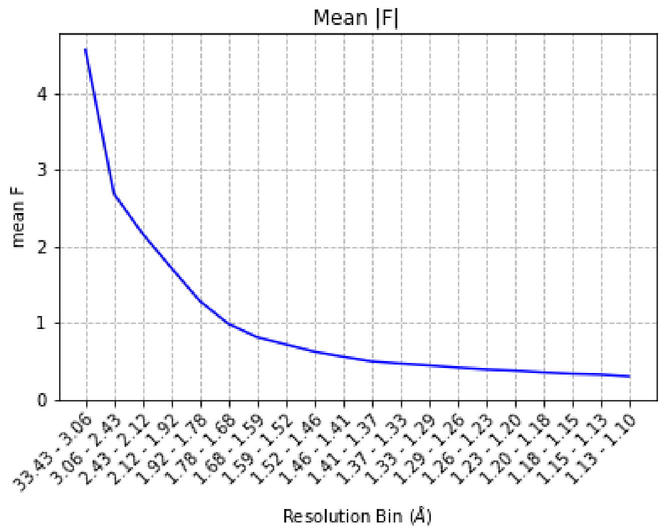


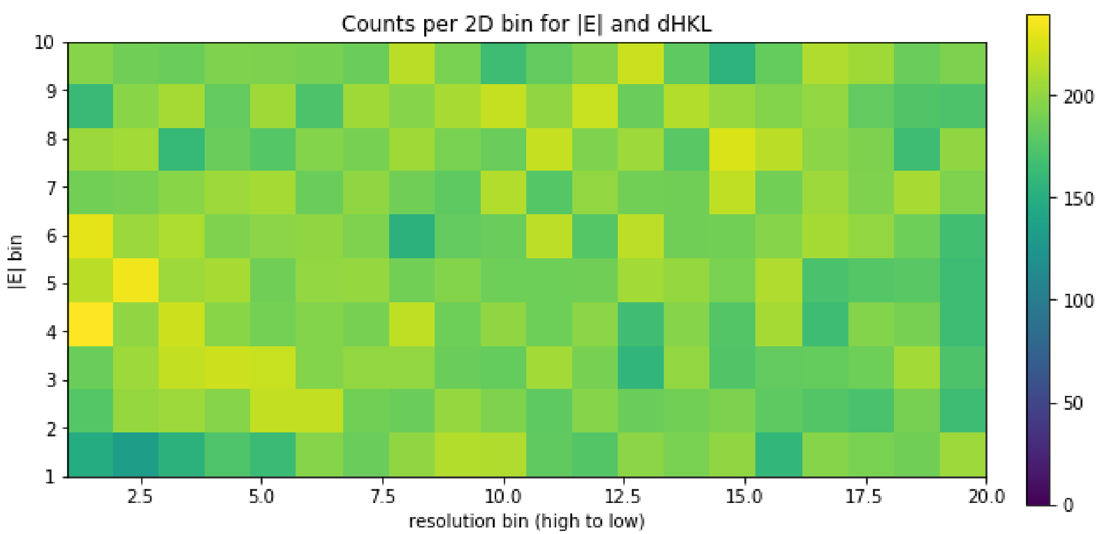




Example: normalizing 1OTB

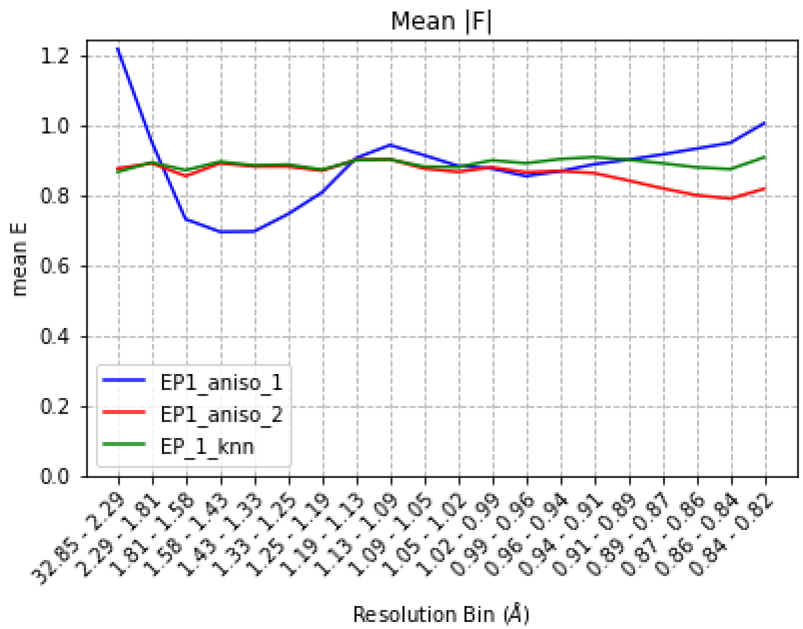
Naïve structure
factor amplitudes:

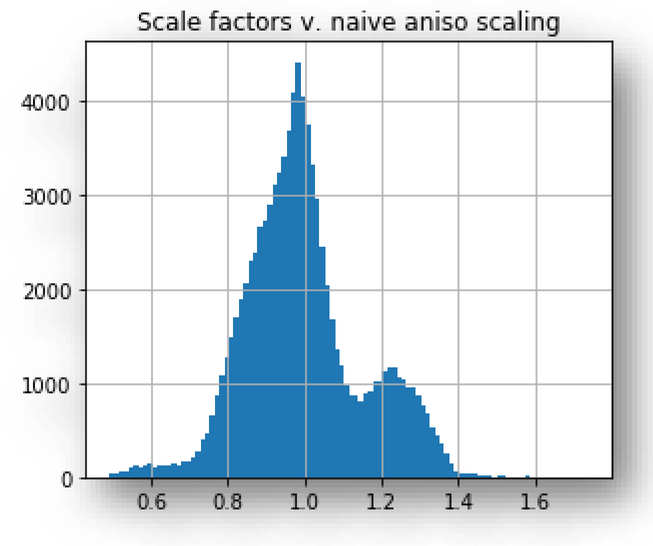


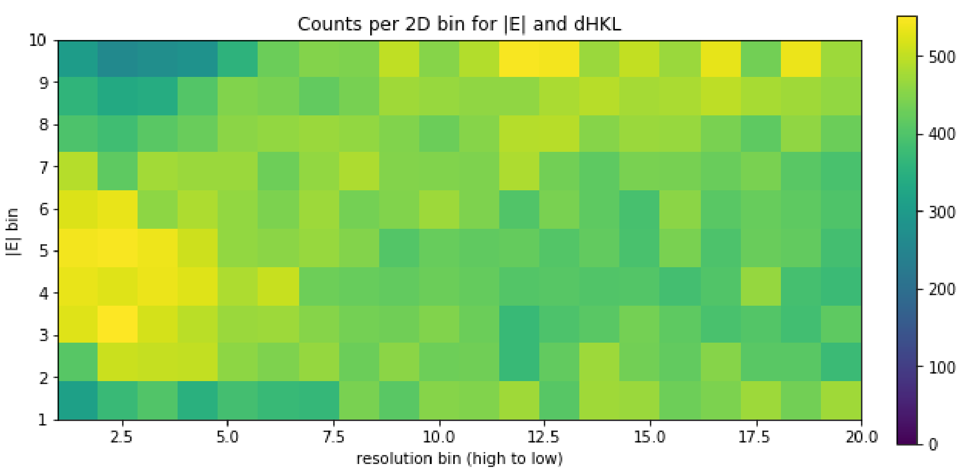


Example: normalizing 1NWZ

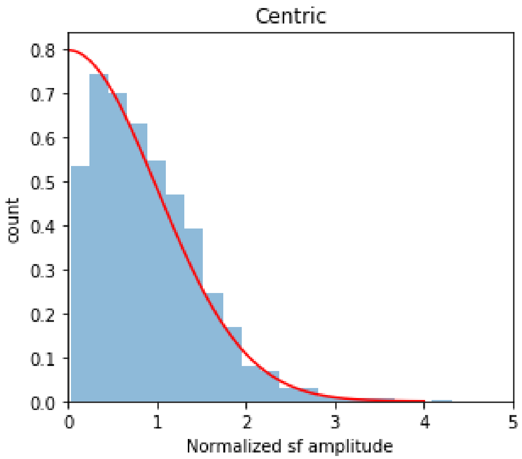
After anisotropic scaling with Fourier corrections of 1NWZ:

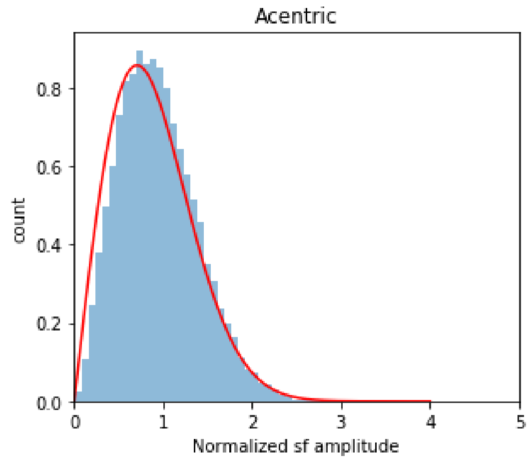


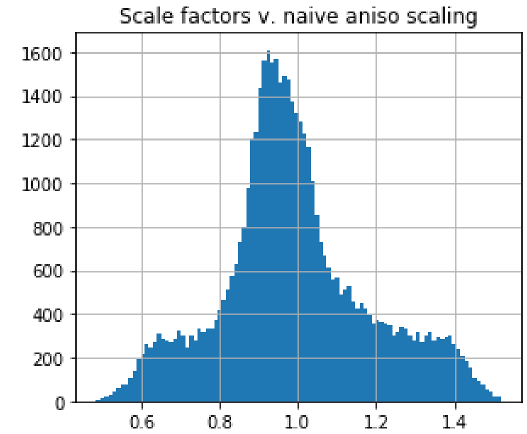


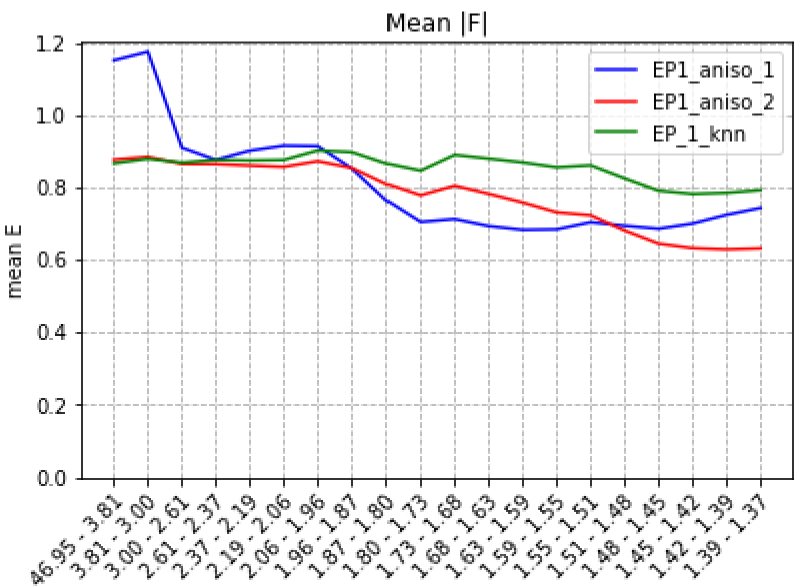


A troublemaker: **GFP@RT**

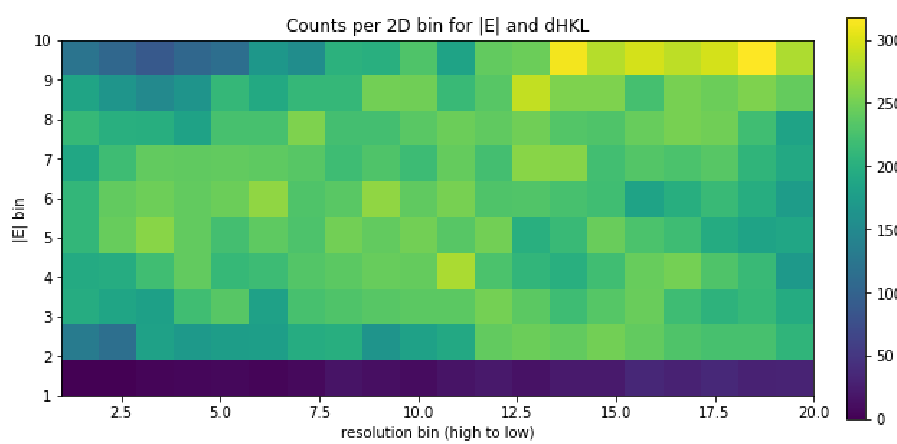








This dataset (here for *n* = 4) did not scale very well. *k*-NN may be more appropriate. Perhaps reflects truncation approach…



Normalizing HEWL anomalous

```
# For simplicity...
ds1 = ds1[(ds1["I(+)"]>=0) & (ds1["I(-)"]>=0)]
```

1A_Anom_dataset_prep_and_scaling

Normalizing HEWL anomalous

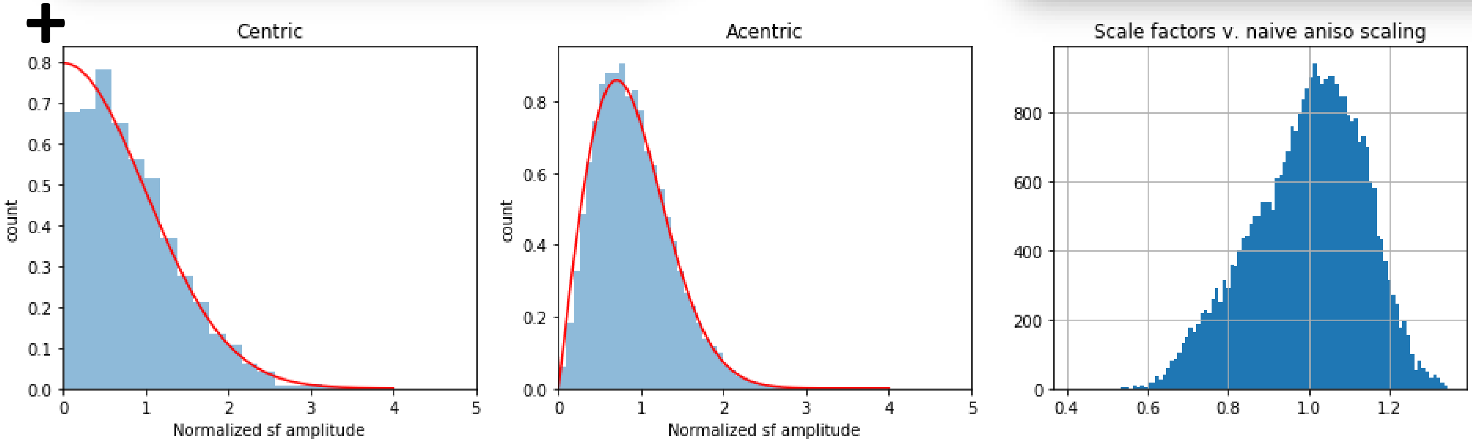
+

```
For n = 1 the test loss = 5849.07
Elapsed time: 0.4382 s
For n = 2 the test loss = 5667.69
Elapsed time: 4.897 s
For n = 3 the test loss = 5628.02
Elapsed time: 37.26 s
For n = 4 the test loss = 5580.48
Elapsed time: 176.7 s
For n = 5 the test loss = 5568.9
Elapsed time: 898.9 s
```

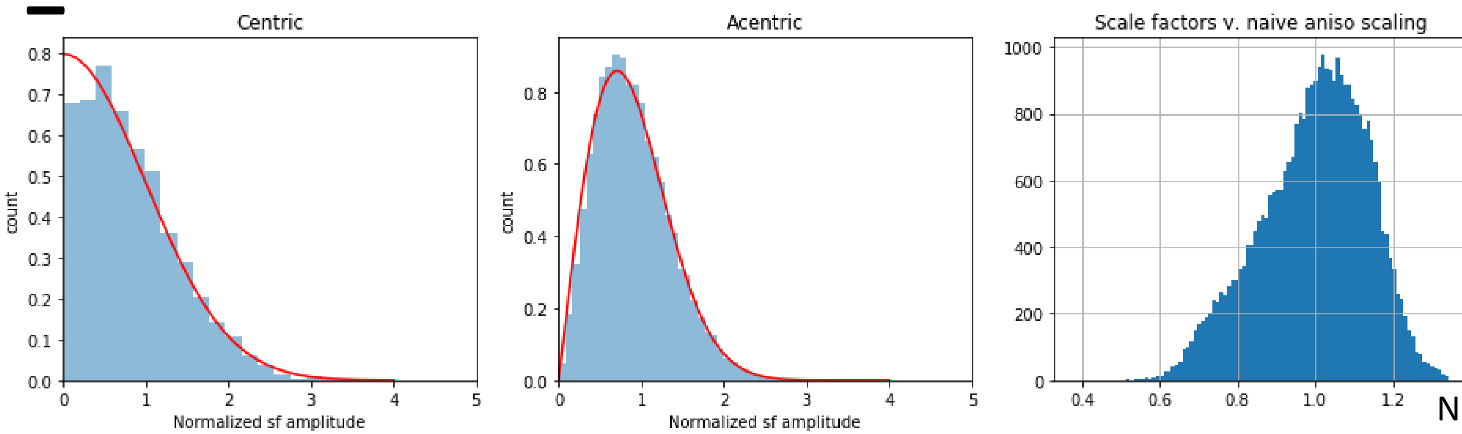
−

```
For n = 1 the test loss = 5840.85
Elapsed time: 0.2482 s
For n = 2 the test loss = 5657.45
Elapsed time: 4.063 s
For n = 3 the test loss = 5619.43
Elapsed time: 28.2 s
For n = 4 the test loss = 5574.26
Elapsed time: 155.6 s
For n = 5 the test loss = 5565.04
Elapsed time: 830.5 s
```

+

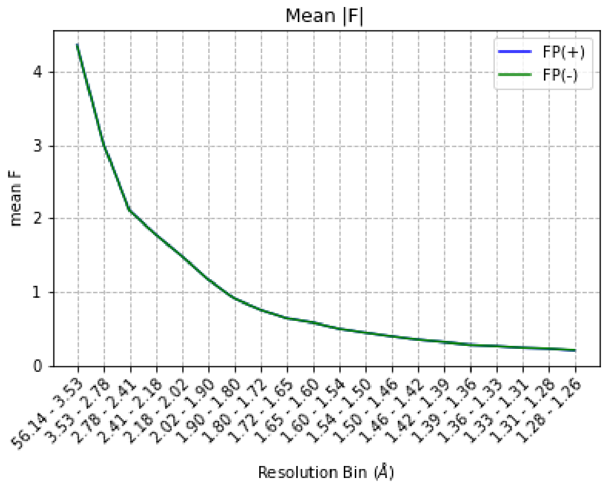


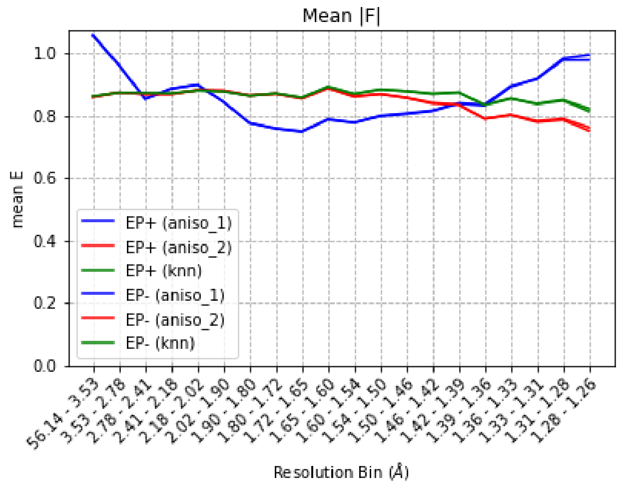
−

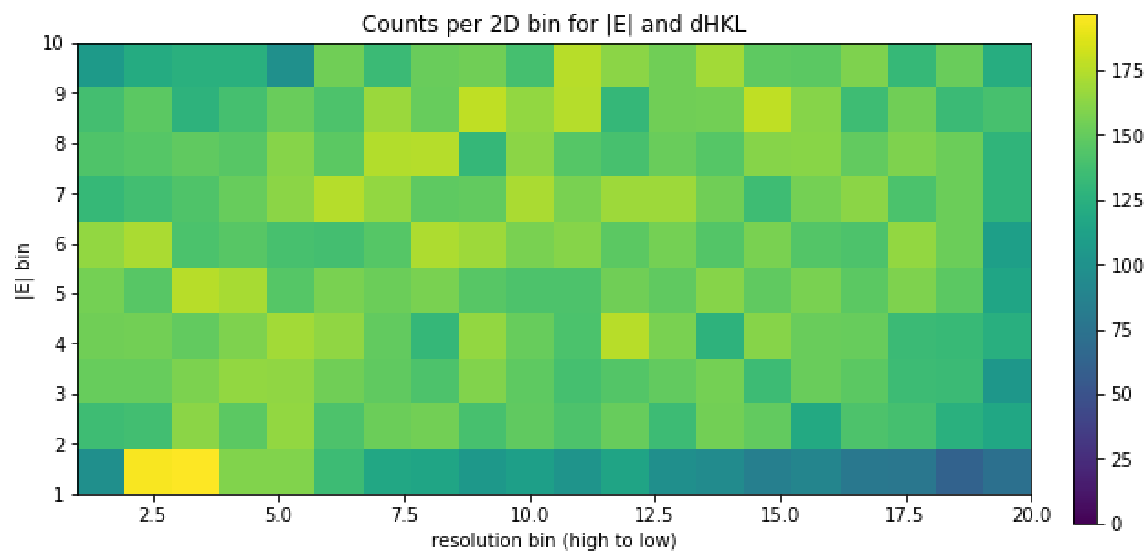


NECAT_HEWL_RT_NaI_82_XDS

Normalizing HEWL anomalous







1A_Anom_dataset_prep_and_scaling

Double-Wilson model

In the DW model, the real and imaginary components of two data sets are both modeled as correlated random walks:

$$\begin{bmatrix} Re(F^A) \\ Im(F^A) \\ Re(F^B) \\ Im(F^B) \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \frac{1}{2}\Sigma \begin{bmatrix} 1 & 0 & r & 0 \\ 0 & 1 & 0 & r \\ r & 0 & 1 & 0 \\ 0 & r & 0 & 1 \end{bmatrix}\right)$$

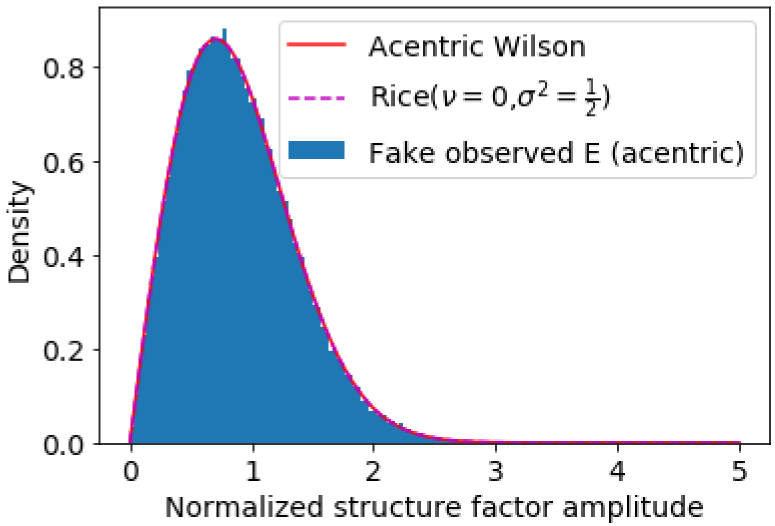
where $\Sigma = \langle |F|^2 \rangle = \mathrm{N}\sigma^2$ for a 2D random walk with N steps each with variance $\frac{1}{2}\sigma^2$ along each dimension. $r = r_{DW}$ governs the correlation between datasets.

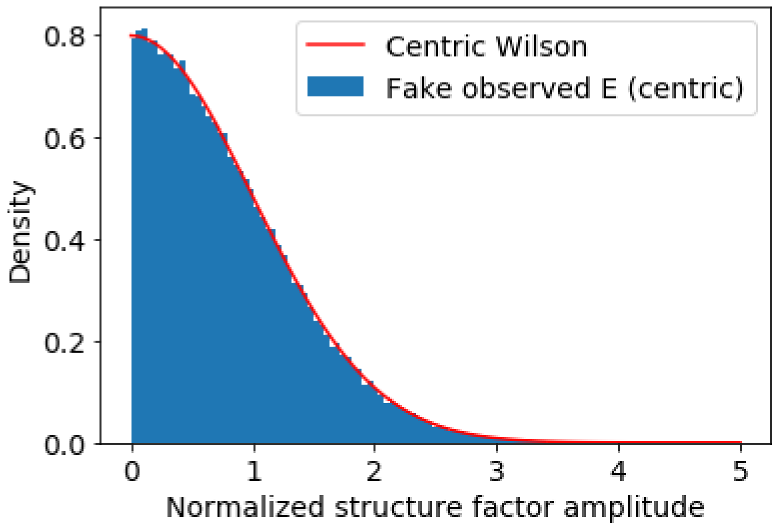
$$\begin{bmatrix} F^A \\ F^B \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}\right)$$

Note that the ½ disappears, because F^A can be thought of as the sum of a random walk in the complex plane added to its own complex conjugate.

Double-Wilson model

The amplitudes of the centric and acentric reflections follow the Wilson distribution:

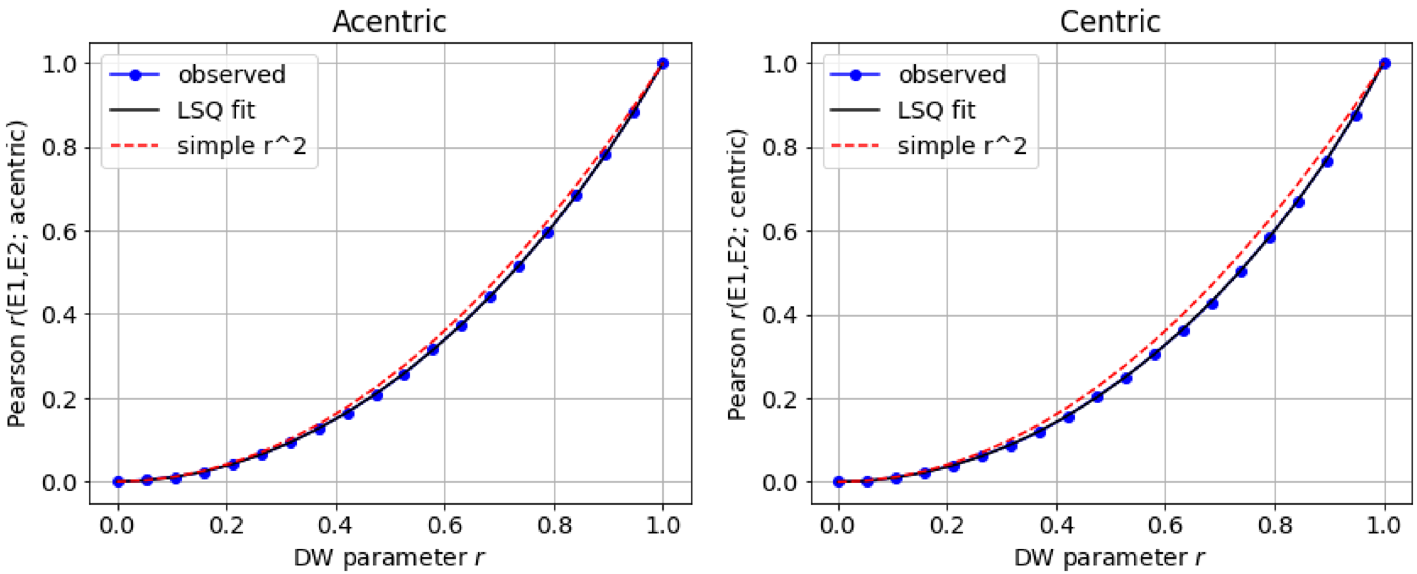




2_Surrogate_data_example

Double-Wilson model

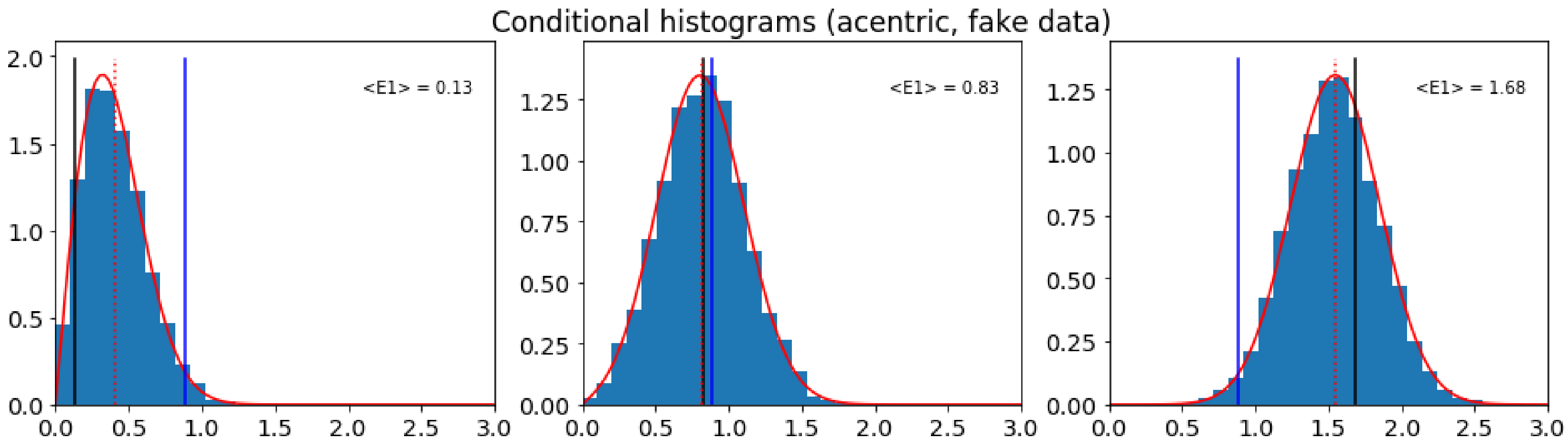
The Pearson correlations between structure factor amplitudes from two correlated data sets almost equal r_{DW}^2.



2_Surrogate_data_example

Double-Wilson model

The conditional distributions of structure factor amplitudes of one data set given the other are described by the Rice distribution (acentric) and Folded Normal (centric).



Conditional mean, $\mathbf{E}(|E_2| \mid |E_1|) = r_{DW}|E_1|$ (centric & acentric)

Conditional variance, $Var(|E_2|||E_1|) = \begin{cases} \frac{1}{2}(1 - r_{DW}^2) & \text{(acentric)} \\ (1 - r_{DW}^2) & \text{(centric)} \end{cases}$

2_Surrogate_data_example

Double-Wilson model

The conditional distributions of structure factor amplitudes of one data set given the other are described by the Rice distribution (acentric) and Folded Normal (centric).

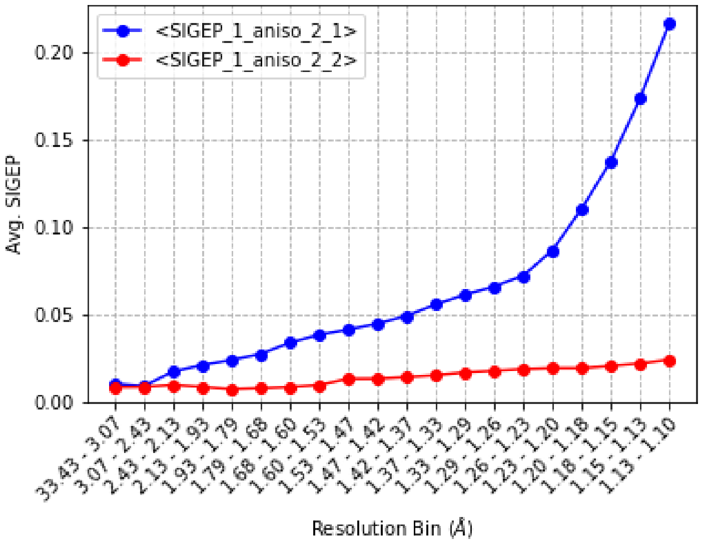
Conditional mean, $\mathbf{E}(|E_2| \mid |E_1|) = r_{DW}|E_1|$ (centric & acentric)

Conditional variance, $Var(|E_2|||E_1|) = \begin{cases} \frac{1}{2}(1 - r_{DW}^2) & \text{(acentric)} \\ (1 - r_{DW}^2) & \text{(centric)} \end{cases}$

```
rice.pdf(    x, cond_mean/np.sqrt(cond_var), scale=np.sqrt(cond_var))
foldnorm.pdf(x, cond_mean/np.sqrt(cond_var), scale=np.sqrt(cond_var))
```

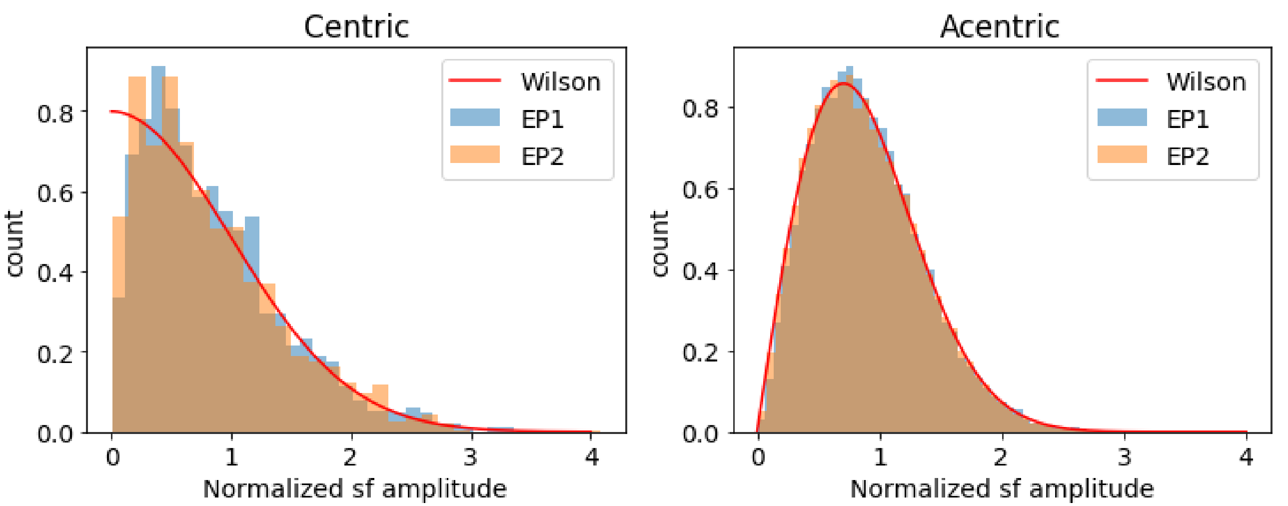
2_Surrogate_data_example

Comparison: 1OTB v 1NWZ



Data quality
Based on Anisotropic + Fourier
For the rest of the analysis,
we'll cut the datasets to 1.2 Å.

dataset 1: 1OTB
dataset 2: 1NWZ



3_Fitting_DW_to_paired_data

Comparison: 1OTB v 1NWZ

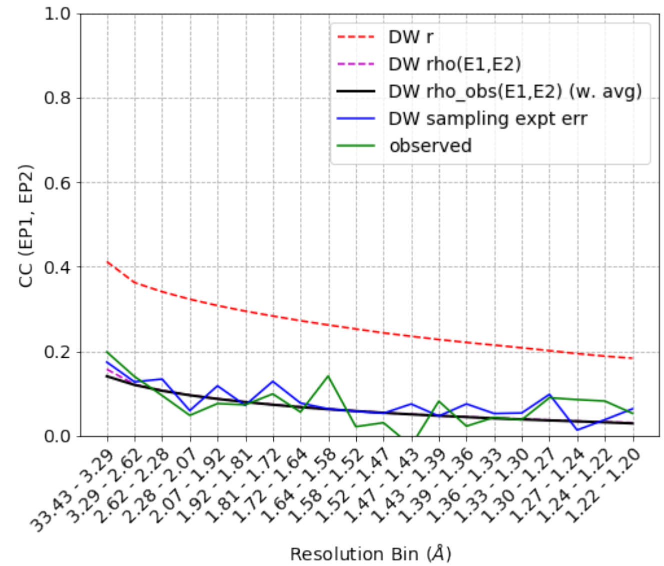
```
(a,b) = fitting_dw.fit_ab(ds1_2,labels=[EP1_label,EP2_label],\
                          dHKL_label=dHKL_label, dHKL_bin_label=dHI
print(f"a: {a:.3}")
print(f"b: {b:.3}")
```

```
`ftol` termination condition is satisfied.
Function evaluations 9, initial cost 2.6541e-02, final cost 9.907
8e-03, first-order optimality 2.06e-08.
a: 0.388
b: 0.887
```

$$r_{DW} = a \cdot e^{-bs^2}$$

This is the inferred true r_{DW} after correcting for measurement error.

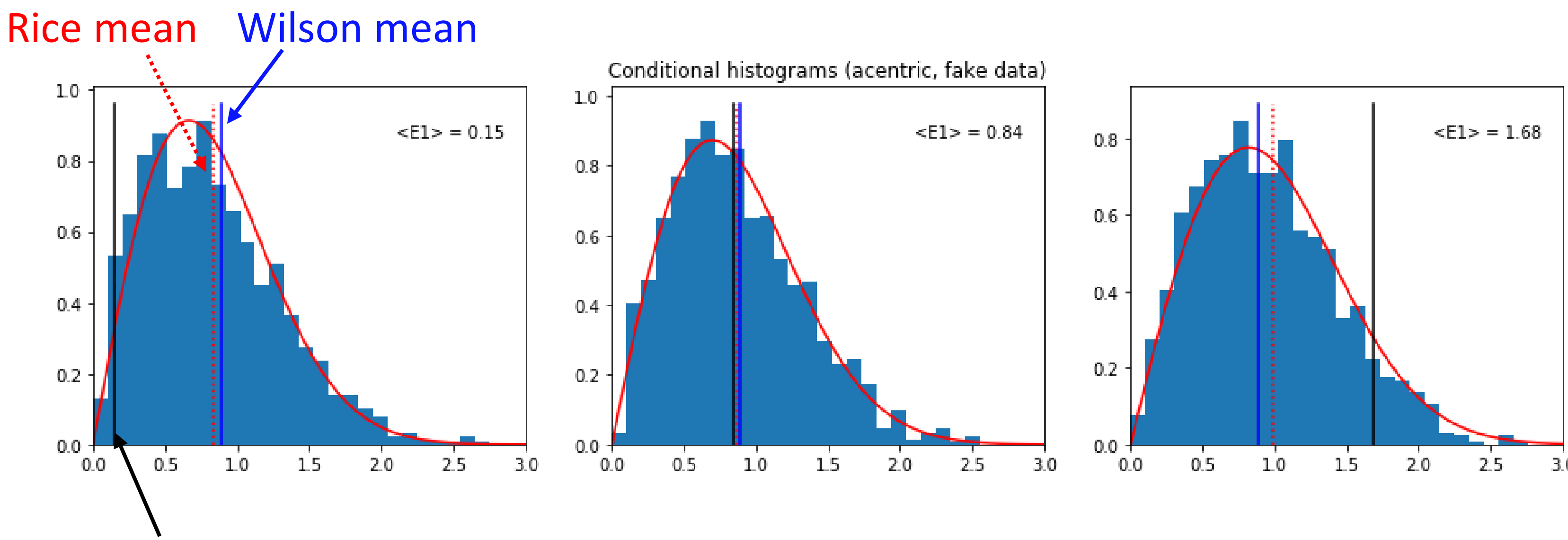
Note that the effective r_{DW} produced by **`fitting_dw.eff_r_dw_per_hkl`** is lower as it includes measurement error



2_Surrogate_data_example
3_Fitting_DW_to_paired_data

Comparison: 1OTB v 1NWZ

At this low r_{DW}, the changes in prior are already quite small!

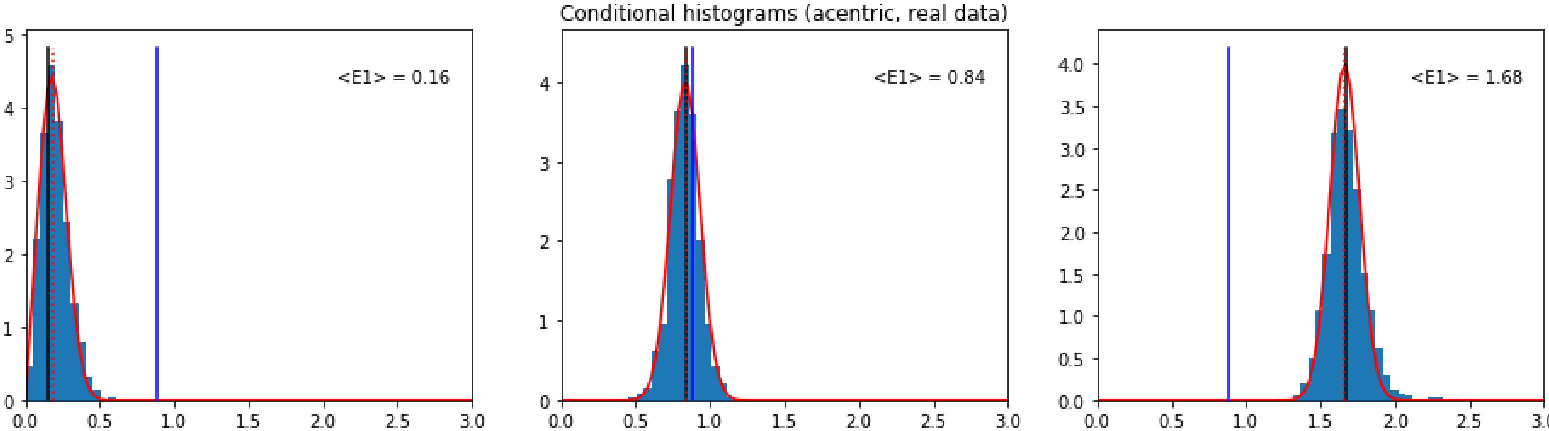


Mean $|E_1|$ for a bin with ~1,200 smallest $|E_1|$

middle bin of $|E_1|$

one-but-largest bin of $|E_1|$

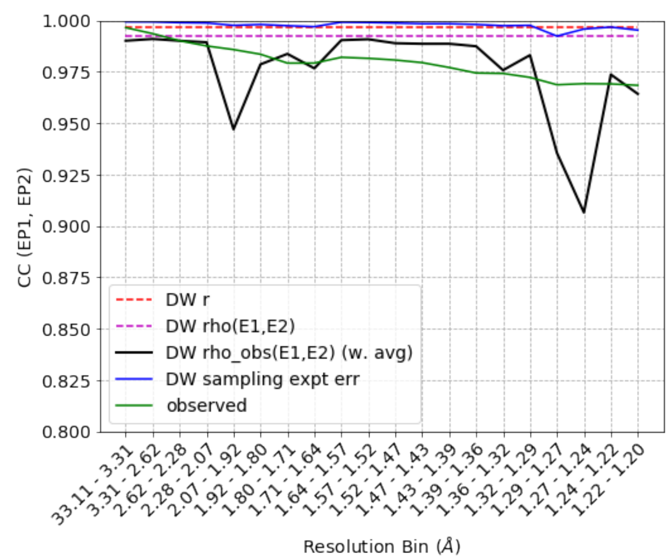
Comparison: **3PYP** v 1NWZ



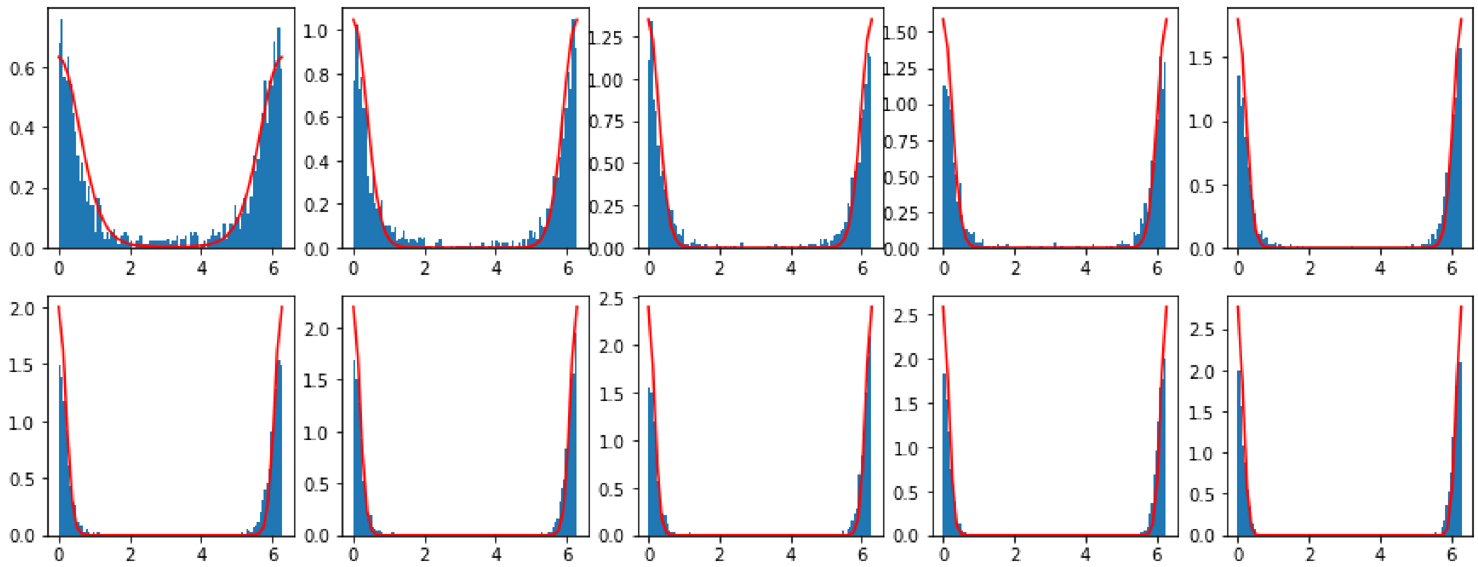
In this case, the priors are highly informative!
(shown for $r_{DW} = 0.99$)

In this case, the experimental errors dominate the correlation coefficient.

Variability in the black line suggests that expt error estimates are conservative, but not perfect.



Comparison: **3PYP** v 1NWZ



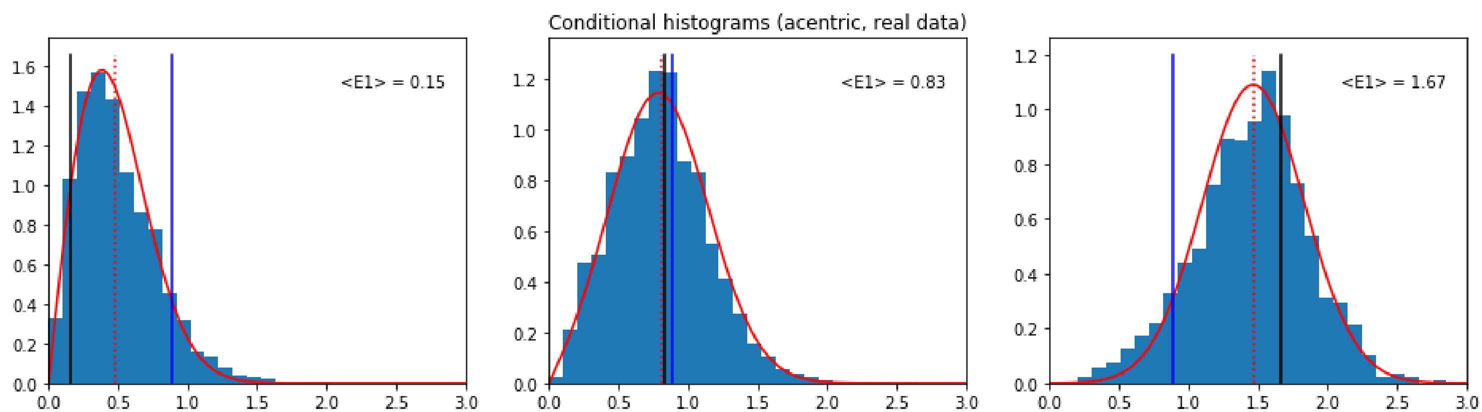
Phase differences are well-described by the corresponding Von Mises distribution.
(shown here for the 10 smallest bins of $|E_1| \cdot |E_2|$)
Red fit for $r_{DW} = 0.99$.

```
vonmises.pdf(x, E1*E2/cond_var)
```

3_Fitting_DW_to_paired_data

Comparison: DHFR (RT, cryo)

	4KJK	4KJJ	4PST	4PSS	
4KJK	1,0	0.90, 0.43	0.95, 0.01	0.88, 0.51	2013 RT
4KJJ		1,0	0.88, 0.26	0.93, 0.16	2013 cryo
4PST			1,0	0.89, 0.22	2005 RT
4PSS				1,0	2015 cryo



fit using $r_{DW} = 0.85$

Keedy, *Structure*, 2013
https://www.sciencedirect.com/science/article/pii/S0969212614001403

4KJK, 4KJJ cut to 1.5Å
all else cut to 1.2Å

Summary of (*a*, *b*) estimates

n	Dataset pair	*a*	*b*	Res. range	Details
0	(5KVX, 5KW3)	0.94	0.79	Cut to 1.7Å	Thaumatin (100K, 278K)
1	(2VWR, 5E1Y)	0.93	0.15	To 1.35Å	LNX2/PDZ2 (100K, 277K)
2	(3PYP, 1NWZ)	1.00	0.00	Cut to 1.1Å	PYP (cryotrapped lit, dark, both 100K)
3	(1NWZ, 1OTB)	0.39	0.89	Cut to 1.2Å	PYP (100K, 295K)*
4	(4EUL, GFP_1.37A) (4EUL, GFP_{PHENIX})	0.66 0.67	0.00 0.4	Cut to 1.8Å Cut to 1.6Å	GFP (100K, 277K**)
5	DHFR	~0.9	0-0.5	Cut to 1.2Å	(see previous slide)
-	HEWL/NaI anom.	1.00	0.00	To 1.26Å	NECAT_HEWL_RT_NaI_82_XDS

*31.9 v 35.3% solvent (1NWZ/1OTB)
**RT data set looks rather crappy; second row using "Filtered" FPs from PHENIX refinement MTZ

could look at correlation of these *a* and *b* with uc's in Mpro dataset.

Specifying priors

- Ultimately, we do not know *a priori* the correlation between a reference data set and a target data set which is to be scaled and merged.
- To parametrize priors, we need to know:
  - The normalized structure factor amplitudes of the reference
  - Initial (a, b) or r_{DW} calculated per s.f. using **`eff_r_dw_per_hkl`** **(**in **`fitting_dw.py`)**

4_Making_Priors
4A_Making_Priors_Anom

Specifying priors

- **5_Parsing_DW_parameters** summarizes how to formulate priors based on the provided r_{DW} and $|E|$.

Using informative priors in scaling & merging

For time-resolved crystallography data, we have two kinds of useful information

- High-quality synchrotron reference data, E_{ref}
- Knowledge that E_{on} and E_{off} tend to be highly correlated

$$\begin{tikzcd}
P(E_1)P(E_2) \arrow[r, "\text{Correlation w\textbackslash reference}"] \arrow[d, "\text{Correlation among on, off}"'] & P\left(E_1|E_{ref}\right)P(E_2|E_{ref}) \arrow[d] \\
P(E_1, E_2) \arrow[r] & P(E_1, E_2|E_{ref})
\end{tikzcd}$$

Using informative priors in scaling & merging

For time-resolved crystallography data, we have two kinds of useful information

- High-quality synchrotron reference data, E_{ref}
- Knowledge that E_{on} and E_{off} tend to be highly correlated

$$P(E_1, E_2, E_3) = P\left(E_{1x}, E_{2x}, E_{3x}, E_{1y}, E_{2y}, E_{3y}\right) = N(0, C)$$

$$C = \frac{1}{2}\begin{bmatrix} 1 & r_x & r & 0 & 0 & 0 \\ r_x & 1 & r & 0 & 0 & 0 \\ r & r & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & r_x & r \\ 0 & 0 & 0 & r_x & 1 & r \\ 0 & 0 & 0 & r & r & 1 \end{bmatrix}$$

$$P(R_1, R_2) = \frac{(1+K)^2 R_1 R_2}{2\pi\beta^2(1-\upsilon^2)} \exp\left(\frac{-2K}{1+\upsilon} - \frac{(1+K)(R_1^2+R_2^2)}{2(1-\upsilon^2)\beta}\right) \times \int_0^{2\pi} \exp\left(\frac{\upsilon(1+K)R_1 R_2 \cos\theta}{(1-\upsilon^2)\beta}\right) \times I_0\left(\sqrt{\frac{2K(1+K)(R_1^2+R_2^2+2R_1R_2\cos\theta}{\beta(1+\upsilon)^2}}\right) d\theta$$

with $R_1 = |E_1|$ (etc.), $K = R_3^2/\Sigma(1-r^2)$, and ${}^{(1+K)}/_{\beta} = 2/\Sigma(1-r^2)$.
This works as long as $|E_1|$ and $|E_2|$ have the same variance.
A version exists for unequal variance but is giving me issues.

$$\upsilon = \frac{r_x - r^2}{1 - r^2}$$

Using informative priors in scaling & merging

For time-resolved crystallography data, we have two kinds of useful information

- High-quality synchrotron reference data, E_{ref}
- Knowledge that E_{on} and E_{off} tend to be highly correlated

$$\begin{tikzcd}
P(E_1)P(E_2) \arrow[r, "\text{Correlation w\textbackslash reference}"] \arrow[d, "\text{Correlation among on, off}"'] & P\left(E_1|E_{ref}\right)P\left(E_2|E_{ref}\right) \arrow[d] \\
P(E_1, E_2) \arrow[r] & P\left(E_1, E_2|E_{ref}\right)
\end{tikzcd}$$

Using informative priors in scaling & merging

For time-resolved crystallography data, we have two kinds of useful information

- High-quality synchrotron reference data, E_{ref}
- Knowledge that E_{on} and E_{off} tend to be highly correlated

$$
\begin{tikzcd}
P(E_1)P(E_2) \arrow[r, "\text{Correlation w\textbackslash reference}"] \arrow[d, "\text{Correlation among on, off}"'] & P\big(E_1|E_{ref}\big)P(E_2|E_{ref}) \arrow[d] \\
P(E_1, E_2) \arrow[r] & P(\boldsymbol{E}_1, \boldsymbol{E}_2|\boldsymbol{E}_{ref})
\end{tikzcd}
$$

Aim: Using informative priors in scaling & merging

One more super-interesting prior: the multivariate normal on the full structure factors $(E_{1x}, E_{1y}, E_{2x}, E_{2y})$.

- Conditioning on E_3 breaks the phase degeneracy, so all phases are relative to the phase of E_3. Without loss of generality, we can set that one to 0 for now.
- We can let $E_{on} = (1-p)E^{gs} + pE^{es}$ and $E_{off} = E^{gs}$.

Two approaches:

(1) Calculate the posterior analytically per h:

$$P\left(E_1, E_2 \,\middle|\{I_{on,hi}, I_{off,hi}\}, \Sigma, p\right)$$
$$= \prod_i P\left(I_{on,hi} | E_1, E_2, \Sigma_{hi}, p, \sigma_{I_{on,hi}}\right) \prod_i P\left(I_{off,hi} | E_1, E_2, \Sigma_{hi}, p, \sigma_{I_{on,hi}}\right)$$
$$\times N(\mu = r_{DW} E_3, C_{1,2|3}) / \iint dE_{1x} dE_{1y} dE_{2x} dE_{2y} (same)$$

if I'm not mistaken, the integral in the denominator is analytically tractable when $P\left(I_{on,hi} | \cdots\right)$ is a multivariate Student *t*.

https://en.wikipedia.org/wiki/Gaussian_integral#n-dimensional_and_functional_generalization

Aim: Using informative priors in scaling & merging

The multivariate normal on structure factors $(E_{1x}, E_{1y}, E_{2x}, E_{2y})$ as a prior:

Two approaches:

2) Variational inference
We do obtain some information on the phases of E^{gs} and E^{es} from measurement of I_{on} and I_{off}, but not nearly as much as about their magnitudes. It is not clear that the multivariate normal is a good variational distribution for $(E_{1x}, E_{1y}, E_{2x}, E_{2y})$.

Regardless of approach, if we can calculate a posterior distribution on the complex (E^{gs}, E^{es}), we have, in principle, a rigorous way to calculate difference maps and a starting point for more disciplined refinement of excited states.

Naïve approach:

$$P(E^{gs}, E^{es}) = P(|E^{gs}|, |E^{es}|) \times P(\varphi^{gs}, \varphi^{es})$$

Variational distributions for phases

For phasing of single structures, the Hendrickson-Lattman distribution is in common use.

The developments below modify and extend the ideas of Rossmann & Blow. By redefinition of the error in the isomorphous replacement method a simplified representation of the phase probability,

$$P(\alpha) = N \exp(A \cos \alpha + B \sin \alpha + C \cos 2\alpha + D \sin 2\alpha),$$

is found without approximation and its validity is established by computational tests. A, B, C and D are

This approach extends the Von Mises distribution:

The von Mises probability density function for the angle x is given by:[2]

$$f(x \mid \mu, \kappa) = \frac{e^{\kappa \cos(x-\mu)}}{2\pi I_0(\kappa)}$$

where $I_0(\kappa)$ is the modified Bessel function of order 0.

The parameters μ and $1/\kappa$ are analogous to μ and σ^2 (the mean and variance) in the normal distribution:

https://scripts.iucr.org/cgi-bin/paper?s0567740870002078

Variational distributions for phases

Journal of Applied Statistics
Vol. 37, No. 3, March 2010, 515–528

Bayesian analysis for bivariate von Mises distributions

Kanti V. Mardia*

Department of Statistics, University of Leeds, Leeds LS2 9JT, UK

(*Received 8 October 2009; final version received 11 December 2009*)

There's a distinct literature on bivariate extensions of the Von Mises distribution.

Table 1. Submodels
(unless stated); $(A)_i$

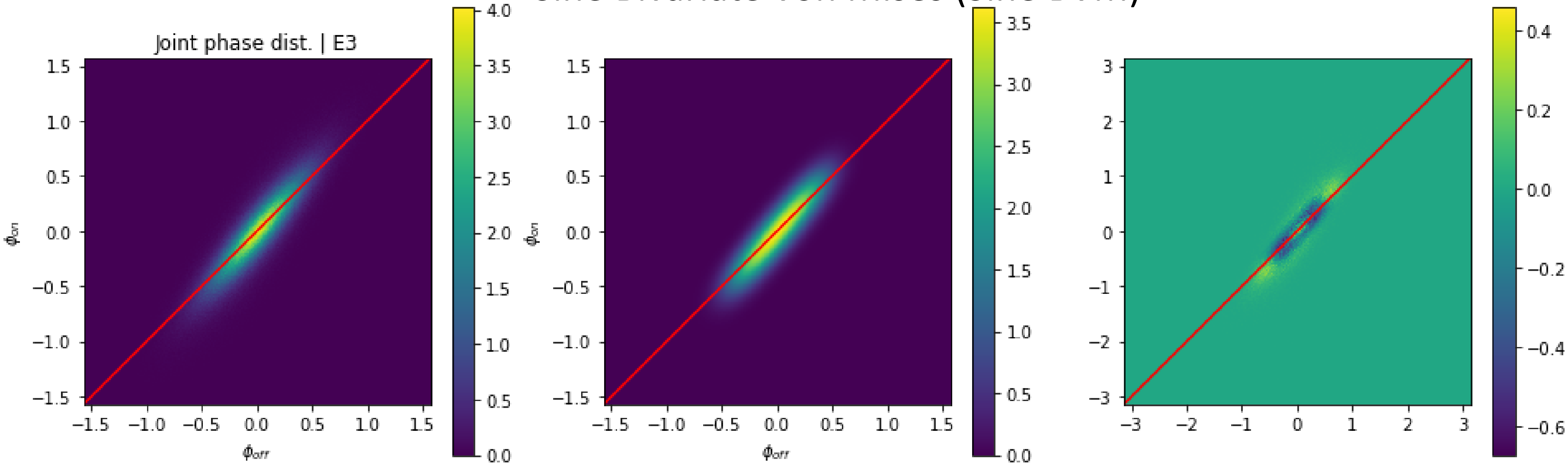
0.	$A = \kappa_3 R(\psi$
1.	$\alpha \neq 0, \beta \neq$
2.	$\alpha = \beta = -$
3.	$\alpha = -\beta =$
4.	$\alpha = 0, \beta =$
5.	$\alpha = \cosh \lambda$
6.	$\kappa_1 = \kappa_2 = ($

Table 2. Normalizing constants $\{c(\kappa_1, \kappa_2, A)\}^{-1}$ of submodels (see for notation Table 1).

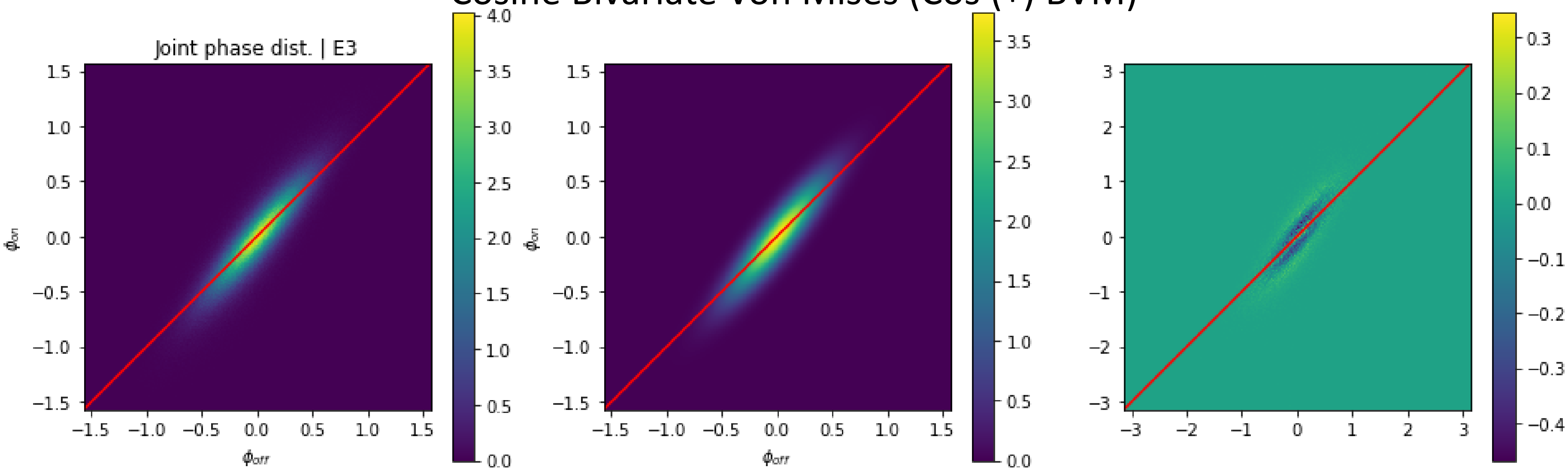
Model	Normalizing constant
General cosine	$I_0(\kappa_1)I_0(\kappa_2)I_0(\kappa_3) + 2\sum_{k=1}^{\infty}(\cos k\psi)I_k(\kappa_1)I_k(\kappa_2)I_k(\kappa_3)$
Rivest	$4\pi^2 \Sigma_k \Sigma_\ell I_k(\kappa_1)I_\ell(\kappa_2)I_{(k+\ell)/2}(u)I_{(k-\ell)/2}(v)$ and the sums are over $k+\ell$ even, $u = (\alpha+\beta)/2$, $v = (\alpha-\beta)/2$
Cosine +ve/cosine −ve	$(2\pi)^2\{I_0(\kappa_1)I_0(\kappa_2)I_0(\kappa_3) + 2\sum_{k=1}^{\infty} I_k(\kappa_1)I_k(\kappa_2)I_k(\kappa_3)\}$
Sine	$4\pi^2 \sum_{k=0}^{\infty}\binom{2k}{k}\left(\frac{\lambda^2}{4\kappa_1\kappa_2}\right)^k I_k(\kappa_1)I_k(\kappa_2)$
Diffused	$4\pi^2 I_0\left(\frac{1}{2}(\alpha+\beta)\right) I_0\left(\frac{1}{2}(\alpha-\beta)\right)$

Variational distributions for phases

Sine Bivariate Von Mises (Sine BVM)



Cosine Bivariate Von Mises (Cos (+) BVM)



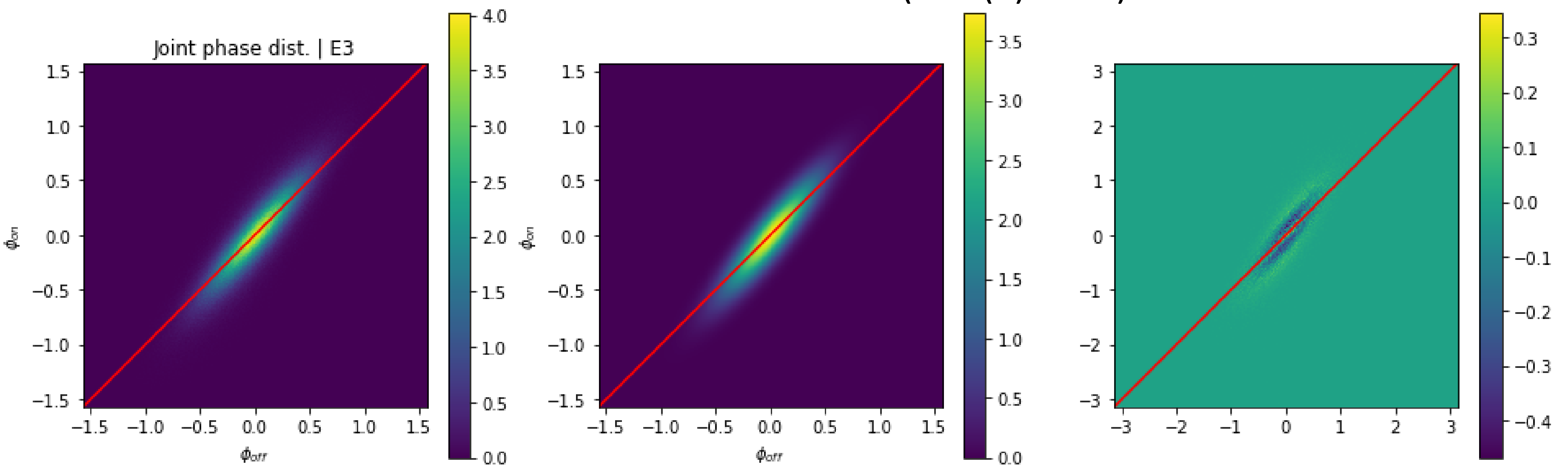
6 Bivariate_conditional_amplitude_phase_dists

Variational distributions for phases

Evaluations of both distributions are fast.

Somehow, I can't get the Cos(+) BVM to integrate to 1.

Cosine Bivariate Von Mises (Cos (+) BVM)

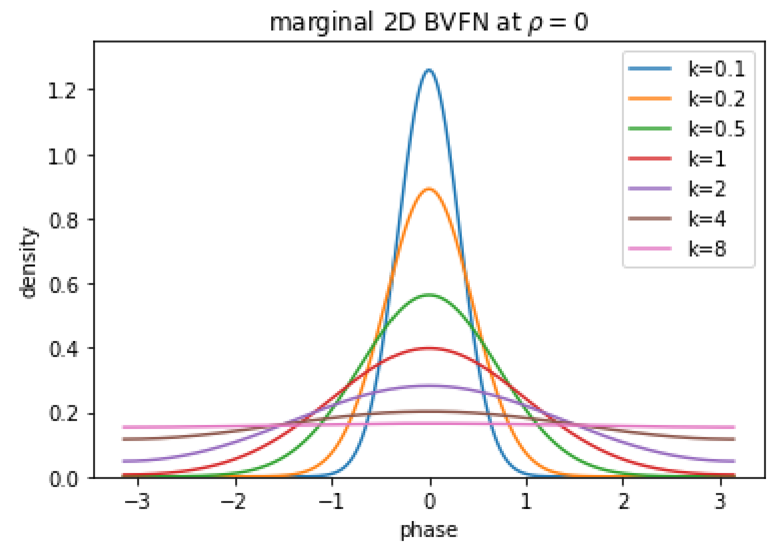


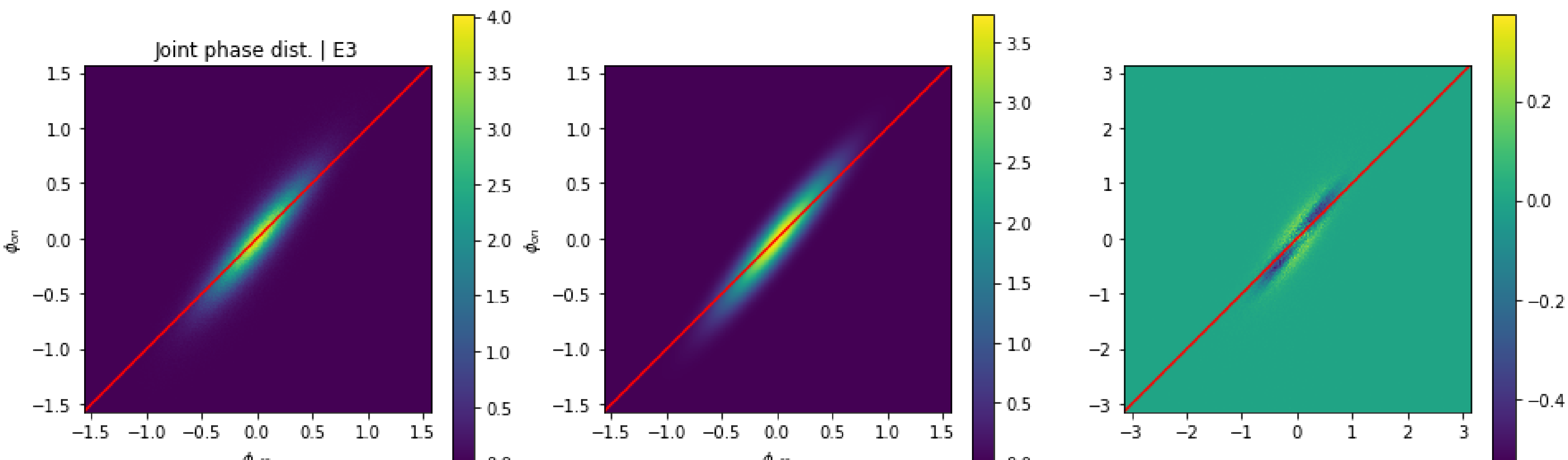
Variational distributions for phases

Evaluations of both distributions are fast.

Somehow, I can’t get the Cos(+) BVM to integrate to 1.

A bivariate normal folded onto the $[-2\pi, \pi)^2$ toroid seems to do just as well and is more intuitive



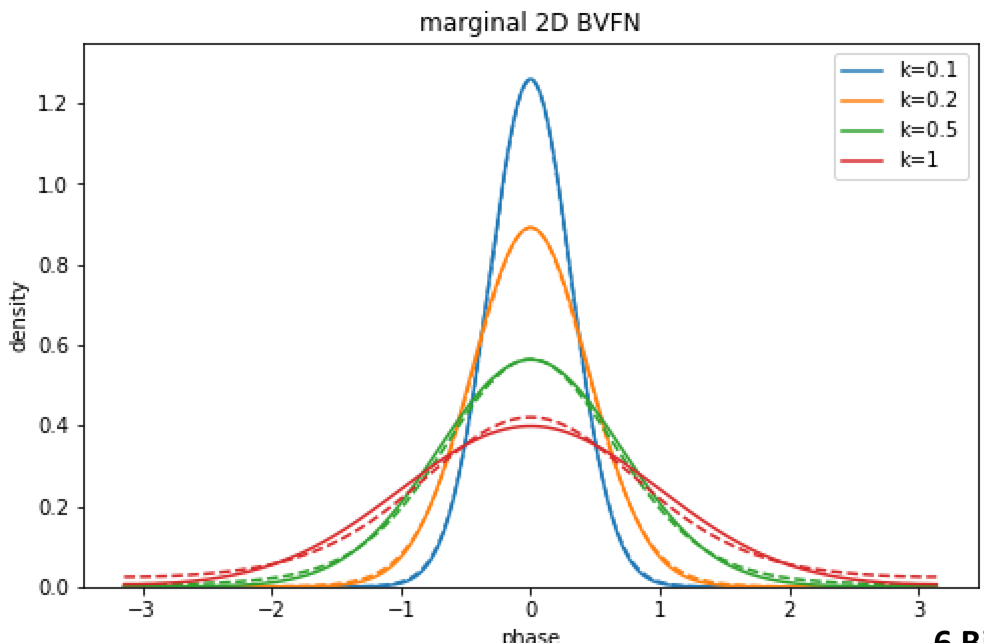


Variational distributions for phases

Evaluations of both distributions are fast.

Somehow, I can't get the Cos(+) BVM to integrate to 1.

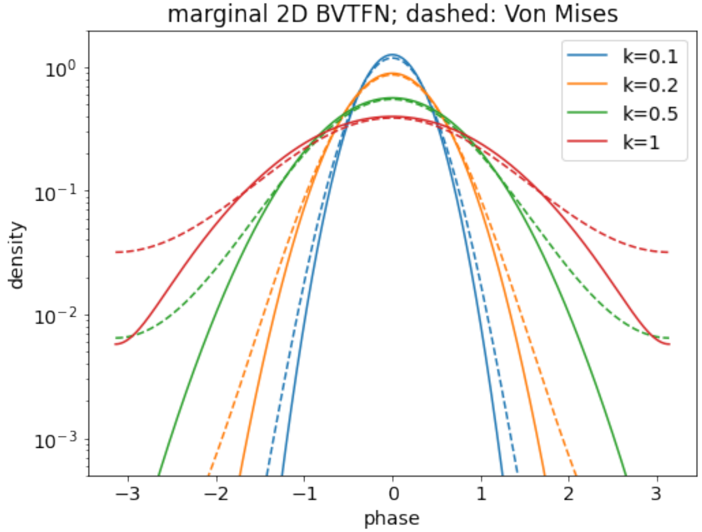
A bivariate normal folded onto the $[-2\pi, \pi)^2$ toroid seems to do just as well and is more intuitive



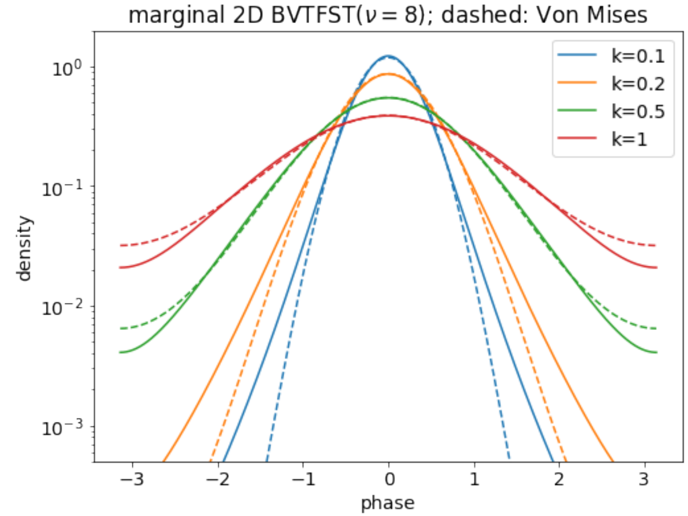
Comparison to Von Mises
As long as $\sigma(\phi) \leq 0.7\ rad$ (40 degrees), the correspondence is very good. The correspondence is again good at large $\sigma(\phi)$.

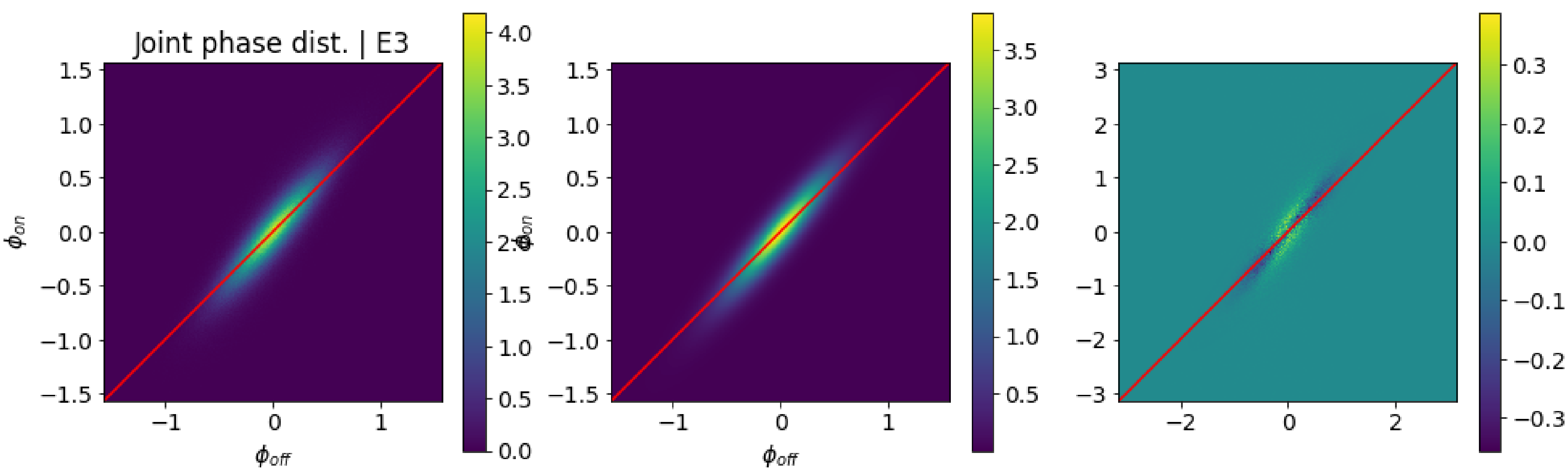
Variational distributions for phases

bivariate torus-folded normal



Torus-folded Student *t*





6 Bivariate_conditional_amplitude_phase_dists