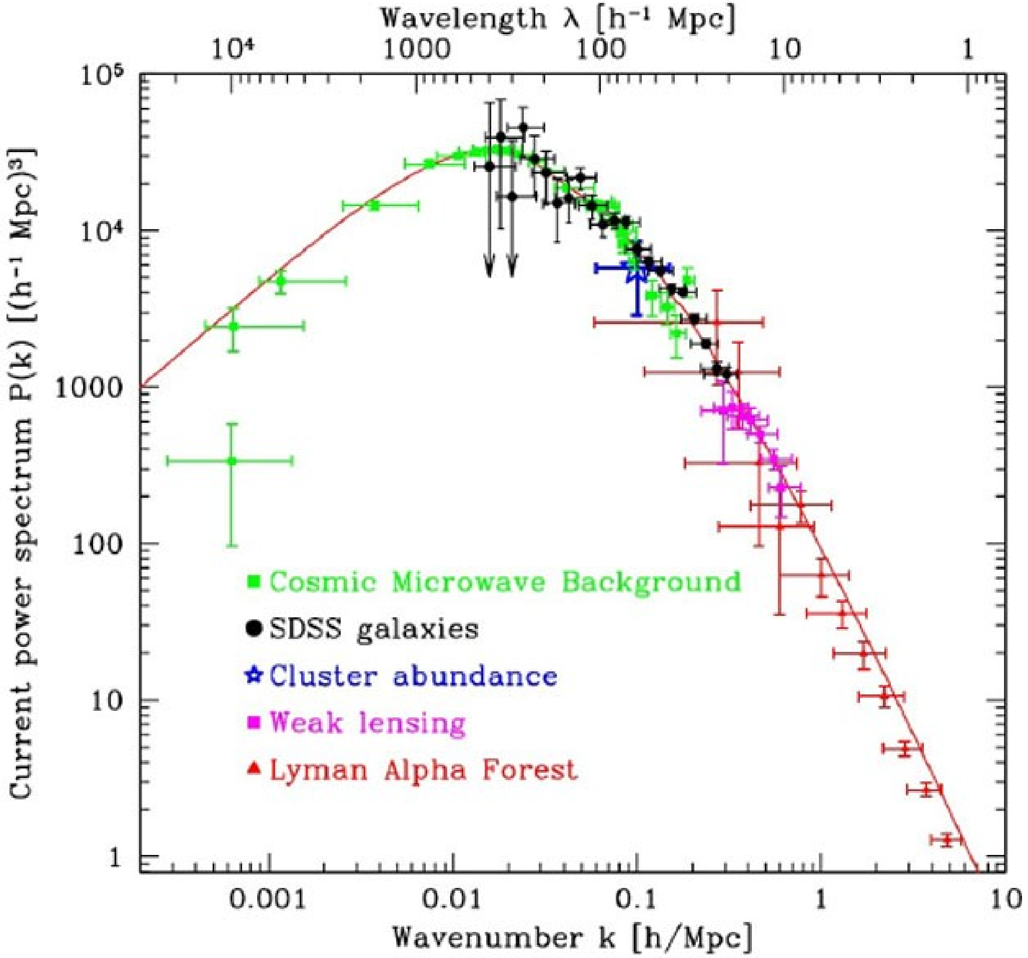
Data analysis in Cosmology



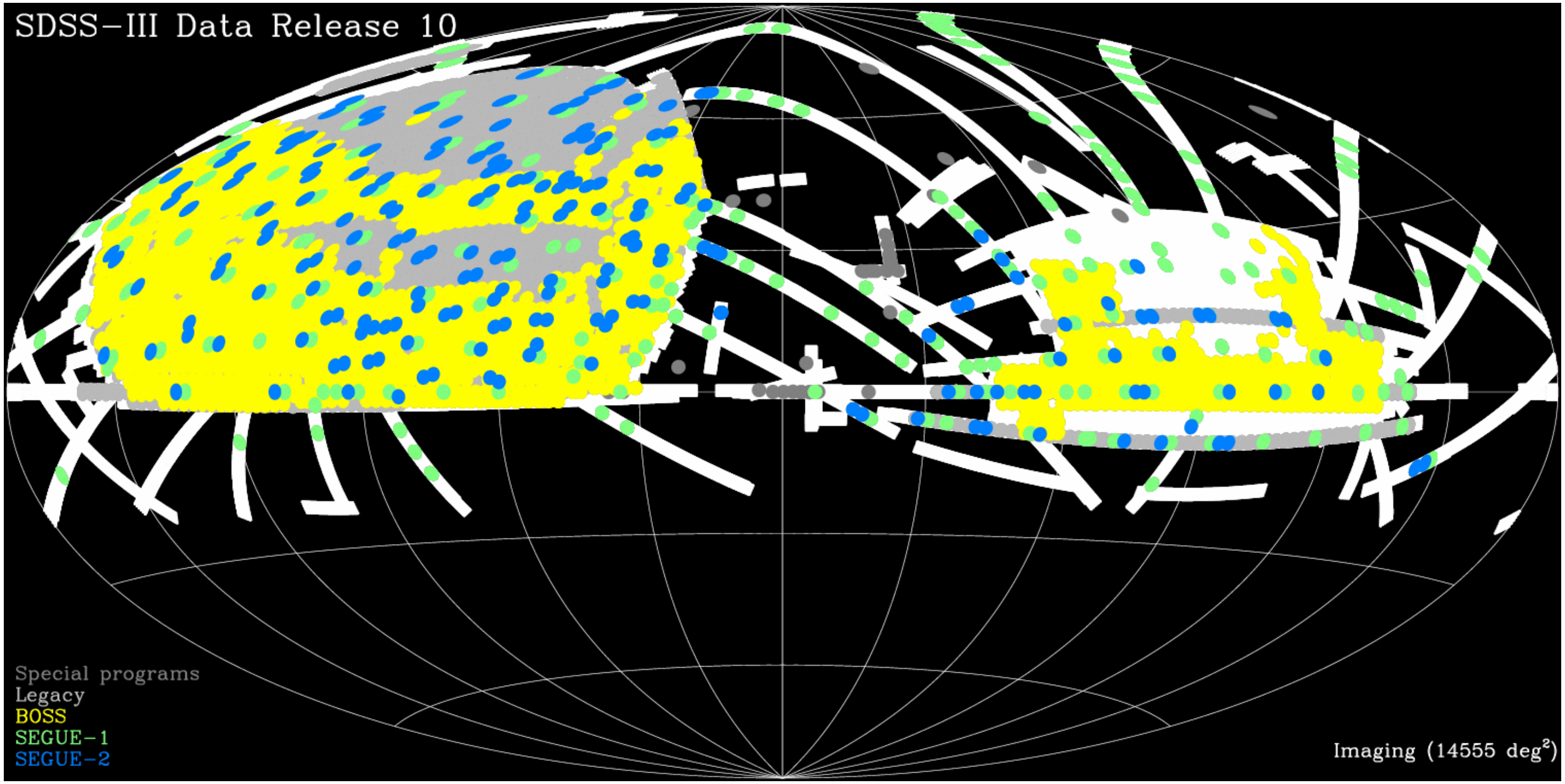
Gustavo Niz

U. de Guanajuato

2018

Survey volume

Observations --- Volume survey *Vs*



Volume V encloses Vs

Mask

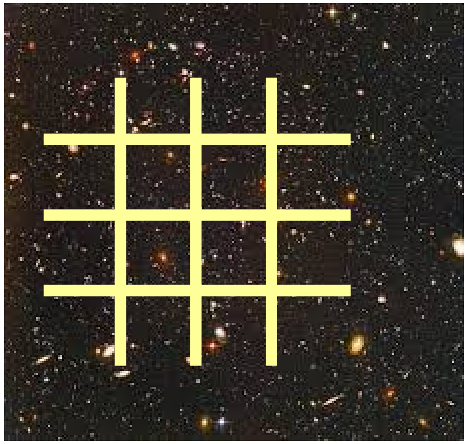
Larger volume enclosing survey area and use of a mask

$$W(\mathbf{x}) = 1 \quad \text{if } \mathbf{x} \in V_S,\ 0 \text{ otherwise.} \quad \Rightarrow \quad \int_V W(\mathbf{x}) d^d x = V_S\,.$$

Can also use the mask to assign weights (later) $0 \leq W(x) \leq 1$

Breaks statistical homogeneity $\rho^w(x) = W(x)\rho(x)$

Weighted density in microcell is $$\rho^w(\mathbf{x}) \equiv \frac{n_j^w}{\delta V} \equiv \frac{W(\mathbf{x}_j) n_j}{\delta V} = \frac{W_j n_j}{\delta V}$$



$$n_j^{w\,2} \neq n_j^w \qquad\qquad n_j^2 = n_j$$

Fourier $$\rho^w_{\mathbf{k}} = \frac{1}{V} \sum_j n_j W_j e^{-i\mathbf{k}\cdot\mathbf{x}_j}$$

Power spectrum

Modes are not independent anymore (because statistical homogeneity is broken)

Before we had

$$
\begin{aligned}
\langle \rho^*_{\mathbf{k}} \rho_{\mathbf{k}'} \rangle &= \frac{1}{V^2} \sum_j \langle n_j \rangle e^{i(\mathbf{k}-\mathbf{k}')\cdot \mathbf{x}_j} + \frac{1}{V^2} \sum_{i \neq j} \langle n_i n_j \rangle e^{i\mathbf{k}\cdot\mathbf{x}_i} e^{-i\mathbf{k}'\cdot\mathbf{x}_j} \\
&= \frac{1}{V^2} \sum_j \langle \rho \rangle \delta V \, e^{i(\mathbf{k}-\mathbf{k}')\cdot \mathbf{x}_j} + \frac{1}{V^2} \sum_{i \neq j} \langle \rho \rangle^2 \delta V^2 \left[1 + \xi(\mathbf{x}_j - \mathbf{x}_i)\right] e^{i\mathbf{k}\cdot\mathbf{x}_i} e^{-i\mathbf{k}'\cdot\mathbf{x}_j} \\
&= \frac{\langle \rho \rangle}{V^2} \int_V e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}} d^d x + \frac{\langle \rho \rangle^2}{V^2} \int_V d^d x e^{i\mathbf{k}\cdot\mathbf{x}} \int_V d^d x' e^{-i\mathbf{k}'\cdot\mathbf{x}'} \\
&\quad + \frac{\langle \rho \rangle^2}{V^2} \int_V d^d x d^d x' \xi(\mathbf{x}' - \mathbf{x}) e^{i\mathbf{k}\cdot\mathbf{x}} e^{-i\mathbf{k}'\cdot\mathbf{x}'} \\
&= \frac{\langle \rho \rangle}{V} \delta_{\mathbf{k}\mathbf{k}'} + 0 + \langle \rho \rangle^2 \frac{1}{V} \delta_{\mathbf{k}\mathbf{k}'} P(\mathbf{k}) \, ,
\end{aligned}
$$

Power spectrum

$$
\begin{aligned}
\langle \rho^{w*}_{\mathbf{k}} \rho^{w}_{\mathbf{k}} \rangle &= \frac{1}{V^2} \sum_j \langle n_j \rangle W_j^2 + \frac{1}{V^2} \sum_{i \neq j} \langle n_i n_j \rangle W_i W_j e^{i\mathbf{k}\cdot(\mathbf{x}_i - \mathbf{x}_j)} \\
&= \frac{1}{V^2} \sum_j \langle \rho \rangle W_j^2 \delta V + \frac{1}{V^2} \sum_{i \neq j} \langle \rho \rangle^2 \delta V^2 \left[1 + \xi(\mathbf{x}_j - \mathbf{x}_i)\right] W_i W_j e^{i\mathbf{k}\cdot(\mathbf{x}_i - \mathbf{x}_j)} \\
&= \frac{\langle \rho \rangle}{V} \frac{1}{V} \int_V W(\mathbf{x})^2 d^d x + \langle \rho \rangle^2 \left| \frac{1}{V} \int_V d^d x W(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} \right|^2 \\
&\quad + \frac{\langle \rho \rangle^2}{V^2} \int_V d^d x d^d x' W(\mathbf{x}) W(\mathbf{x}') \xi(\mathbf{x}' - \mathbf{x}) e^{i\mathbf{k}\cdot(\mathbf{x} - \mathbf{x}')} \\
&= \frac{\langle \rho \rangle}{V} \sum_{\mathbf{q}} |W_{\mathbf{q}}|^2 + \langle \rho \rangle^2 |W_{\mathbf{k}}|^2 \\
&\quad + \langle \rho \rangle^2 \frac{1}{V} \sum_{\mathbf{k}'} P(\mathbf{k}') \frac{1}{V} \int_V d^d x W(\mathbf{x}) e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}} \frac{1}{V} \int_V d^d x' W(\mathbf{x}') e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}'}
\end{aligned}
$$

Parseval formula

$$\int d^d x W_i e^{-i\mathbf{k}\cdot\mathbf{x}} = V W_{\mathbf{k}}$$

$$V\xi(\mathbf{r}) = \sum_{\mathbf{k}'} e^{i\mathbf{k}'\cdot\mathbf{r}} P(\mathbf{k}')$$

Power spectrum

Before we obtained $$\tilde{P}(\mathbf{k}) \equiv \frac{V}{\langle\rho\rangle^2}\langle|\rho_{\mathbf{k}}|^2\rangle = \frac{1}{\langle\rho\rangle} + P(\mathbf{k}),$$

and now $$\langle\rho_{\mathbf{k}}^{w*}\rho_{\mathbf{k}}^{w}\rangle - \langle\rho\rangle^2|W_{\mathbf{k}}|^2 = \frac{\langle\rho\rangle}{V}\sum_{\mathbf{q}}|W_{\mathbf{q}}|^2 + \frac{\langle\rho\rangle^2}{V}\sum_{\mathbf{k}'}|W_{\mathbf{k}-\mathbf{k}'}|^2 P(\mathbf{k}').$$

Was zero before

$$\langle|\rho_{\mathbf{k}}^{w} - \langle\rho_{\mathbf{k}}^{w}\rangle|^2\rangle$$

Convolution

$$\sum_{\mathbf{k}'}|W_{\mathbf{k}-\mathbf{k}'}|^2 P(\mathbf{k}') = \sum_{\mathbf{q}}|W_{\mathbf{q}}|^2 P(\mathbf{k}+\mathbf{q})$$

Assuming P does not change much over range of q

$$= P(\mathbf{k})\sum_{\mathbf{q}}|W_{\mathbf{q}}|^2$$

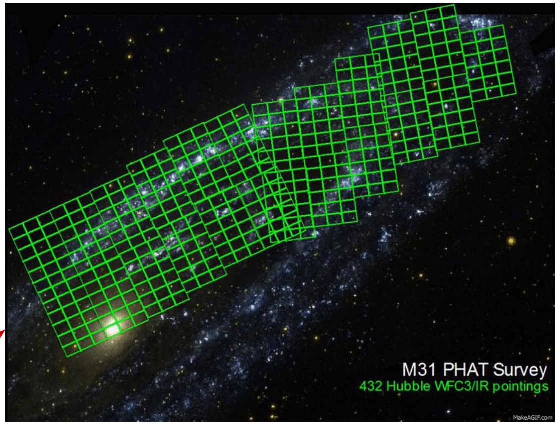
$$P(\mathbf{k}) \approx \frac{V\langle|\rho_{\mathbf{k}}^{w} - \langle\rho_{\mathbf{k}}^{w}\rangle|^2\rangle - \langle\rho\rangle\sum_{\mathbf{q}}|W_{\mathbf{q}}|^2}{\langle\rho\rangle^2\sum_{\mathbf{q}}|W_{\mathbf{q}}|^2} \equiv \frac{V^2\langle|\rho_{\mathbf{k}}^{w} - \langle\rho_{\mathbf{k}}^{w}\rangle|^2\rangle - N_{\text{eff}}}{\langle\rho\rangle^2 V\sum_{\mathbf{q}}|W_{\mathbf{q}}|^2}$$

Where $$N_{\text{eff}} = \langle\rho\rangle V\sum_{\mathbf{q}}|W_{\mathbf{q}}|^2.$$

Power spectrum

Turn into estimator using *ni*

$$\sum \langle n_j \rangle W_j = \langle \rho \rangle \sum W_j \delta V \quad \Rightarrow \quad \langle \rho \rangle = \frac{\sum \langle n_j \rangle W_j}{\sum W_j \delta V}$$



Replace with definition

$$\bar{\rho} = \frac{\sum_{j=cells} n_j W_j}{\sum_{j=cells} W_j \delta V} = \frac{\sum_{i=gal} W_j}{V_S} \equiv \frac{N_S}{V_S}$$

Hence $\rho^w_{\mathbf{k}} - \langle \rho^w_{\mathbf{k}} \rangle \;\rightarrow\; \frac{1}{V} \sum_j n_j W_j e^{-i\mathbf{k}\cdot\mathbf{x}_j} - \frac{N_S}{V_S} W_{\mathbf{k}} \;=\; \frac{1}{V} \sum_i W_i e^{-i\mathbf{k}\cdot\mathbf{x}_i} - \frac{N_S}{V_S} W_{\mathbf{k}}\,.$

$$\boxed{\hat{P}(\mathbf{k}) = \frac{V}{\sum_{\mathbf{q}} |W_{\mathbf{q}}|^2} \left| \frac{V_S}{N_S V} \sum_{i=gal} W_i e^{-i\mathbf{k}\cdot\mathbf{x_i}} - W_{\mathbf{k}} \right|^2 - \frac{N_S}{V_S}}$$

Selection function

We assumed $\langle n_i \rangle = \langle \rho \rangle \delta V$

Because selection effects it is not true for observed galaxies

Selection function *S(x)* $\langle n_j \rangle \;=\; S(\mathbf{x}_j)\langle \rho \rangle \delta V \quad \Rightarrow \quad \left\langle \frac{n_j}{S_j} \right\rangle \;=\; \langle \rho \rangle \delta V \,,$

Corrected number in microcell $n_j^c \;\equiv\; A\frac{W(\mathbf{x}_j)}{S(\mathbf{x}_j)} n_j \;=\; A\frac{W_j}{S_j} n_j \,,$

$$W_j \longrightarrow \frac{W_j}{S_j}$$

Normalization
such that $\sum n_j^c = N$

Hence

$$\rho^c_{\mathbf{k}} \;=\; \frac{A}{V} \sum_j n_j \frac{W_j}{S_j} e^{-i\mathbf{k}\cdot\mathbf{x}_j}$$

$$\langle \rho^c_{\mathbf{k}} \rangle \;=\; \frac{1}{V} \sum_j \left\langle \frac{n_j}{S_j} \right\rangle W_j e^{-i\mathbf{k}\cdot\mathbf{x}_j} \;=\; \frac{A\langle \rho \rangle}{V} \int_V W(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} d^d x \;=\; A\langle \rho \rangle W_{\mathbf{k}} \,,$$

Selection function

We get

$$P(\mathbf{k}) \approx \frac{V^2\langle|\rho^c_{\mathbf{k}} - \langle\rho^c_{\mathbf{k}}\rangle|^2\rangle - A^2\langle\rho\rangle \int_V \frac{W(\mathbf{x})^2}{S(\mathbf{x})} d^d x}{\langle\rho\rangle^2 A^2 V \sum_{\mathbf{q}} |W_{\mathbf{q}}|^2} .$$

As opposed to (before)

$$P(\mathbf{k}) \approx \frac{V\langle|\rho^w_{\mathbf{k}} - \langle\rho^w_{\mathbf{k}}\rangle|^2\rangle - \langle\rho\rangle \sum_{\mathbf{q}} |W_{\mathbf{q}}|^2}{\langle\rho\rangle^2 \sum_{\mathbf{q}} |W_{\mathbf{q}}|^2}$$

Optimal Weights - observe fewer galaxies per unit volume at high redshifts

If W=S, then each galaxy is observed equally. Feldman showed an optimal weight given by

$$W(\mathbf{x}) = \frac{S(\mathbf{x})}{1+\bar{n}(\mathbf{x})P(k)} = \frac{S(\mathbf{x})}{1+S(\mathbf{x})\langle\rho\rangle P(k)} ,$$

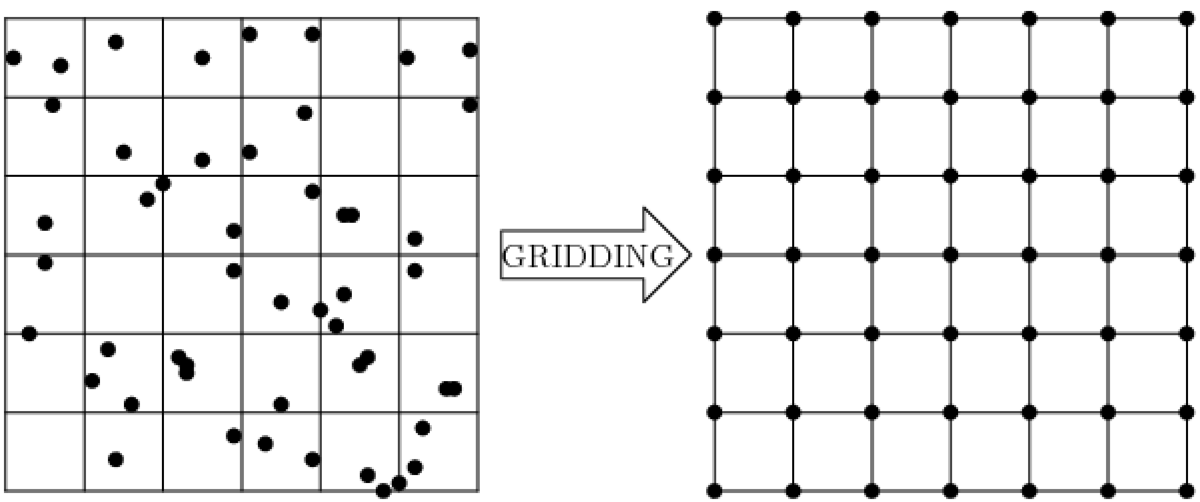
Cyclic argument. Weights with P to calculate P. May iterate or use synthetic catalogues

The 2PCF equivalent is $W(r;z) = \dfrac{1}{1+4\pi\bar{n}(z)J_3(r)} ,$

From particles to the grid

Can calculate P(k) using FFT. For that we need to assign a density field into a regular grid.

Nearest grid point (others are: cloud-in-cell, triangular-shape-cloud)



Python: scipy.interpolate.griddata

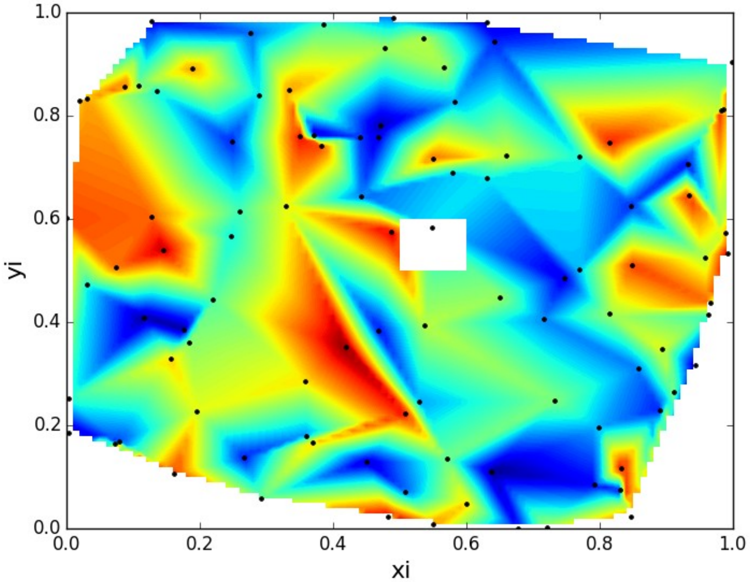
use “nearest” method

From particles to the grid

```
import matplotlib.pyplot as plt
import numpy as np
from scipy.interpolate import griddata

# data coordinates and values
x = np.random.random(100)
y = np.random.random(100)
z = np.random.random(100)

# target grid to interpolate to
xi = yi = np.arange(0,1.01,0.01)
xi,yi = np.meshgrid(xi,yi)

# set mask
mask = (xi > 0.5) & (xi < 0.6) & (yi > 0.5) & (yi < 0.6)

# interpolate
zi = griddata((x,y),z (xi,yi),method='linear')

# mask out the field
zi[mask] = np.nan
```

nearest

```
# plot
fig = plt.figure()
ax = fig.add_subplot(111)
plt.contourf(xi,yi,zi,np.arange(0,1.01,0.01
))
plt.plot(x,y,'k.')
plt.xlabel('xi',fontsize=16)
plt.ylabel('yi',fontsize=16)
plt.savefig('interpolated.png',dpi=100)
plt.close(fig)
```



Power spectrum using FFT

Construct Fourier of Grid density using Discrete Fourier Transform (DFT) → FFTW

Power spectrum is

$$P(k_F n_1) = \frac{V}{N^6} \left\langle \left| \delta^{FFTW}(\boldsymbol{n}_1) \right|^2 \right\rangle = \frac{V}{N^6} \left(\frac{1}{N_k} \sum_{|n_k - n_1| \leq \frac{1}{2}} |\delta^{FFTW}(\boldsymbol{n}_k)|^2 \right),$$

Where *V*=volume, *N*=no. of 1d grid cells, $H^3 = \frac{V}{N^3}$ and $k_F^3 = \frac{(2\pi)^3}{V}$

We sum over all Fourier modes within $k_1 - k_F/2 < |\boldsymbol{k}| < k_1 + k_F/2$
To estimate the power spectrum at $k = k_1 = k_F n_1$.