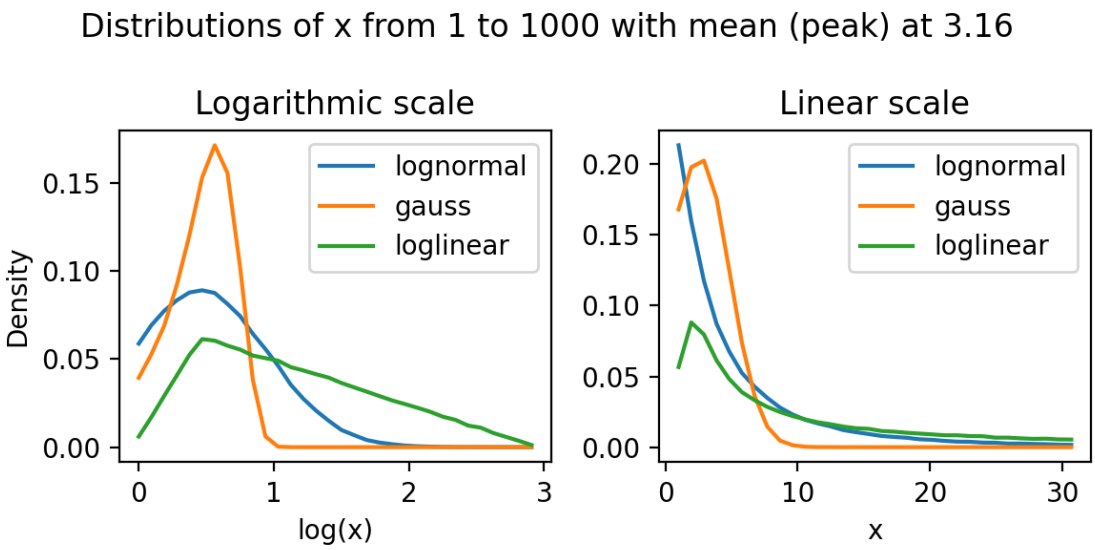
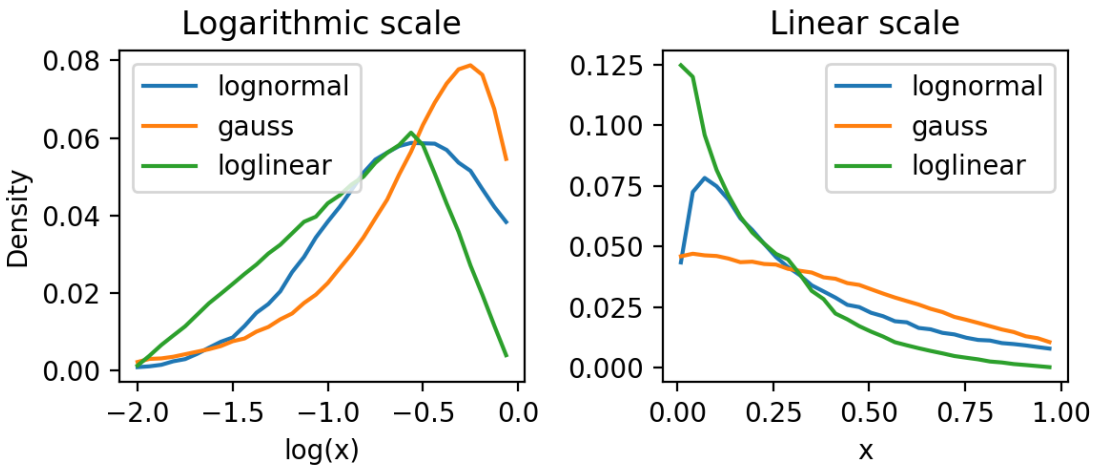
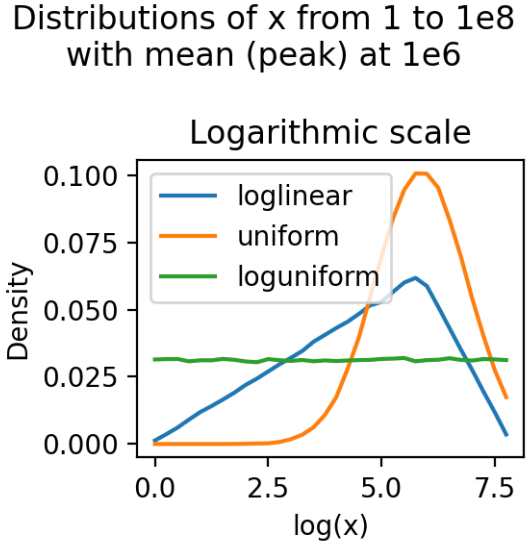
supermodel

January 13, 2022

[1]:
```
from pca import cluster, PCA
import numpy as np
import matplotlib.pyplot as plt
import warnings
import pandas as pd
from tabela_napake_p import modeli_napake
from mpl_toolkits.mplot3d import Axes3D
warnings.filterwarnings("ignore")

# primeri vseh treh porazdelitev v primeru povprečja večjega in manjšega od 1,
# ob tem so najprej predstavljene porazdelitve za N v primeru večje možnosti
# za širitev na druge planete
```



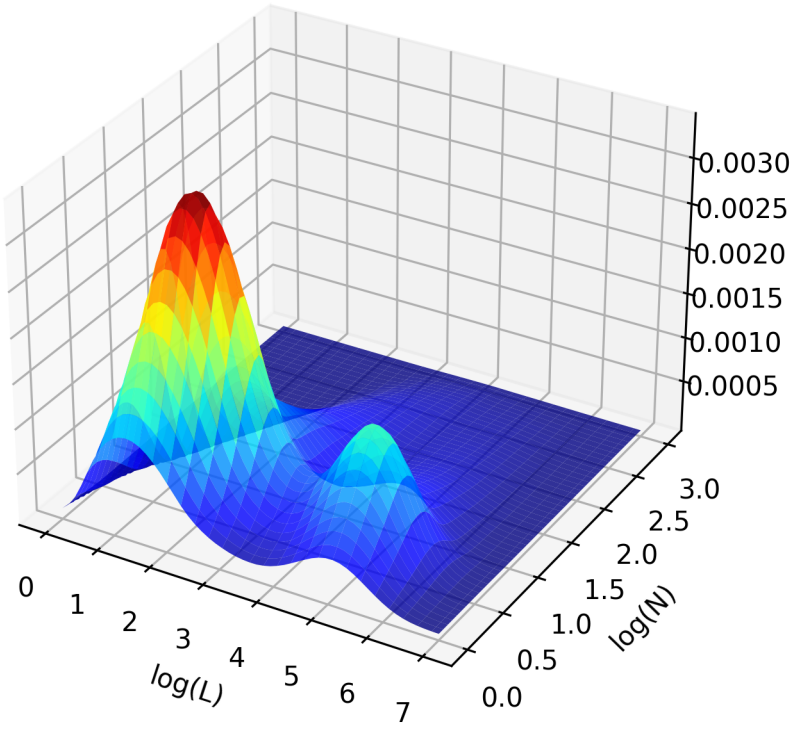




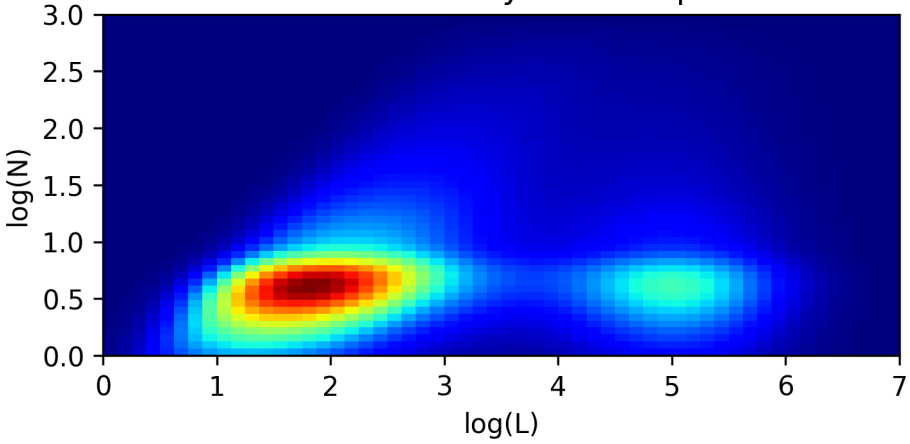
[2]:
```
cluster(model=0, ks=[3, 5], supermodel=1)  # supermodel
cluster(model=0, ks=[5], supermodel=2)  # supermodel 2
```

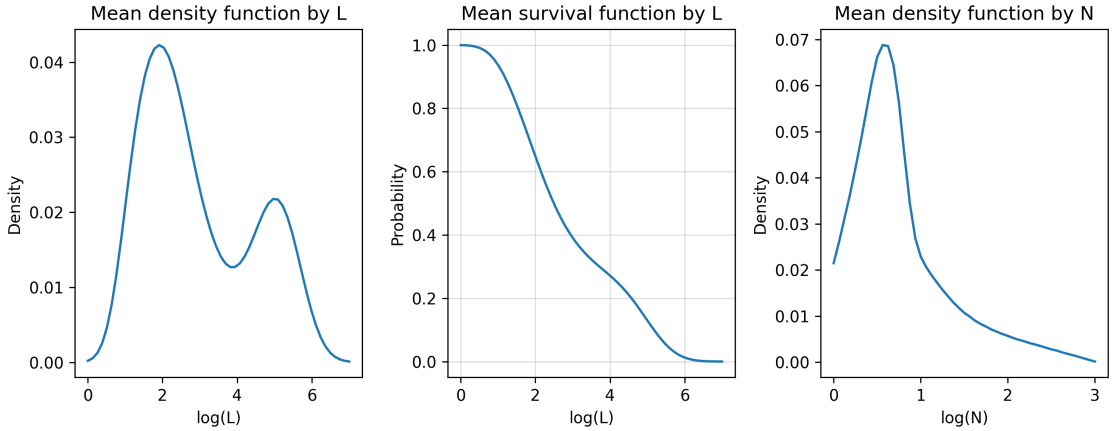
Supermodel

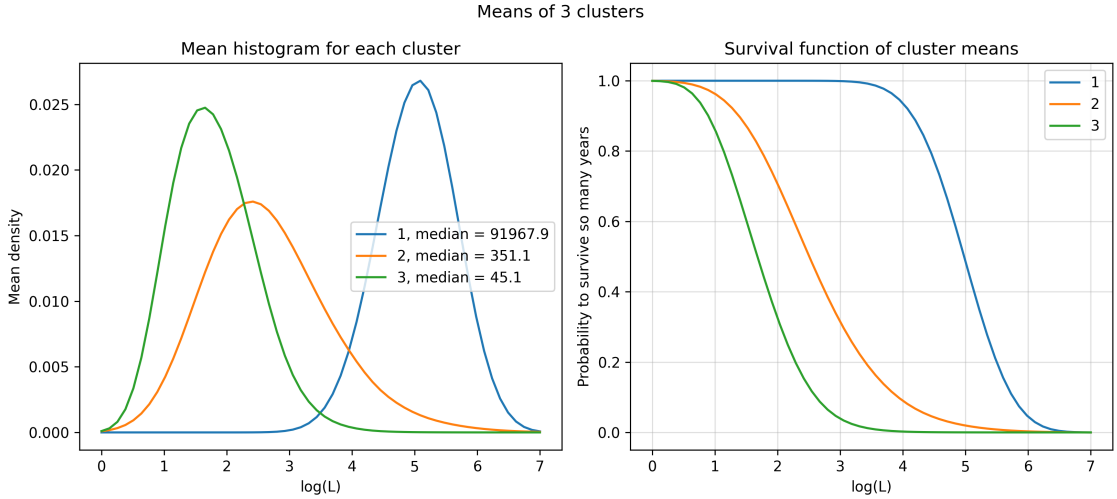
Mean density



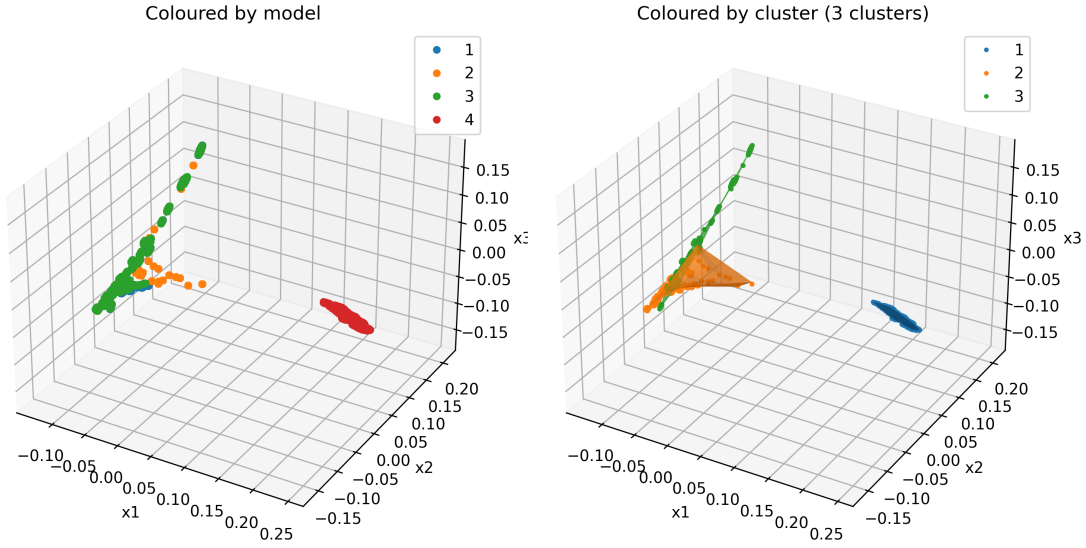
Mean density - heatmap



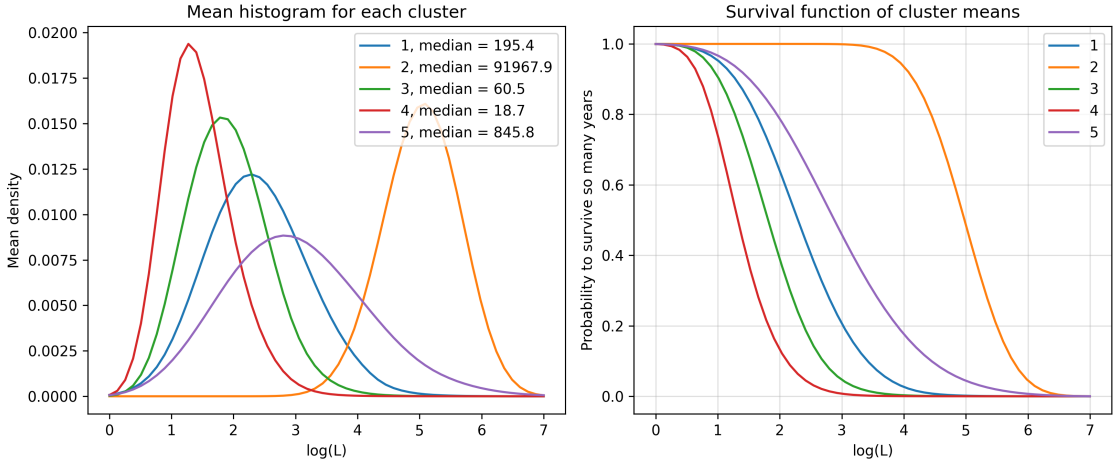

Mean density function by L
Density
log(L)
Mean survival function by L
Probability
log(L)
Mean density function by N
Density
log(N)

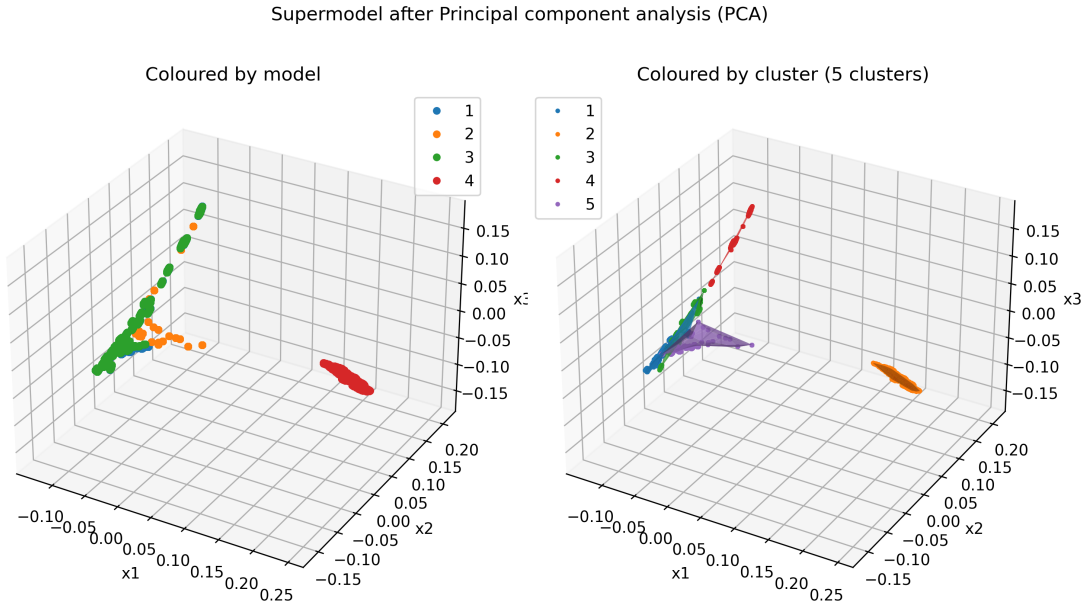

Means of 3 clusters
Mean histogram for each cluster
Mean density
log(L)
1, median = 91967.9
2, median = 351.1
3, median = 45.1
Survival function of cluster means
Probability to survive so many years
log(L)
1
2
3

Supermodel after Principal component analysis (PCA)



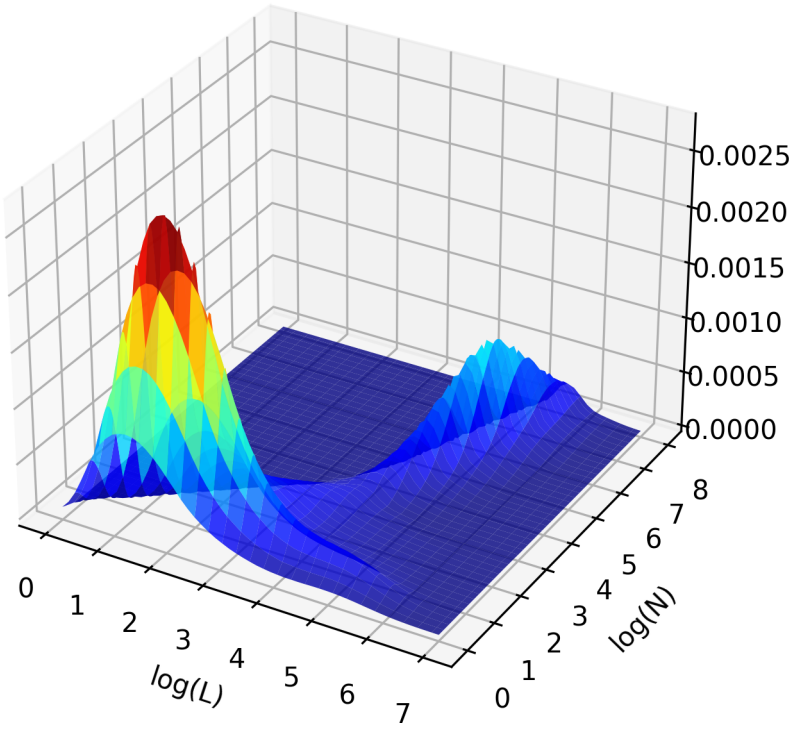
Means of 5 clusters

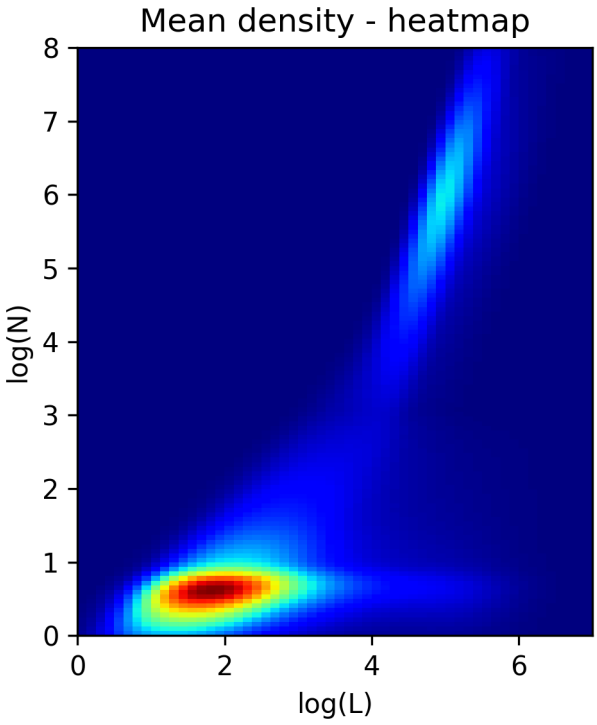


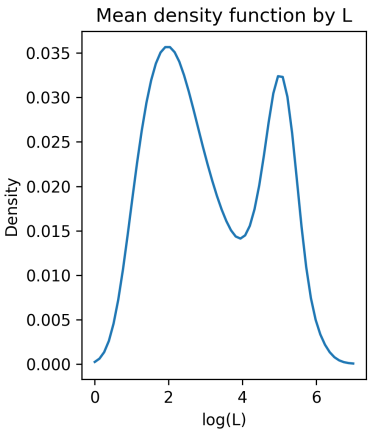


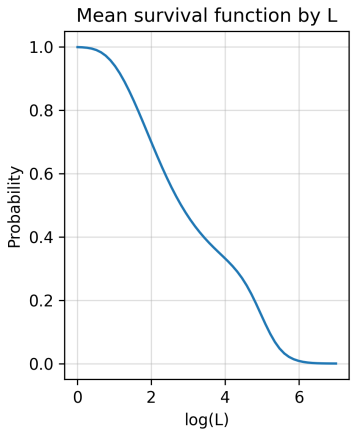
Supermodel 2

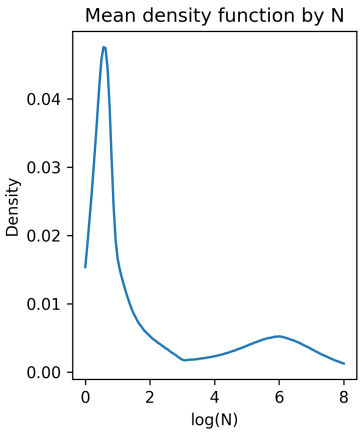
Mean density

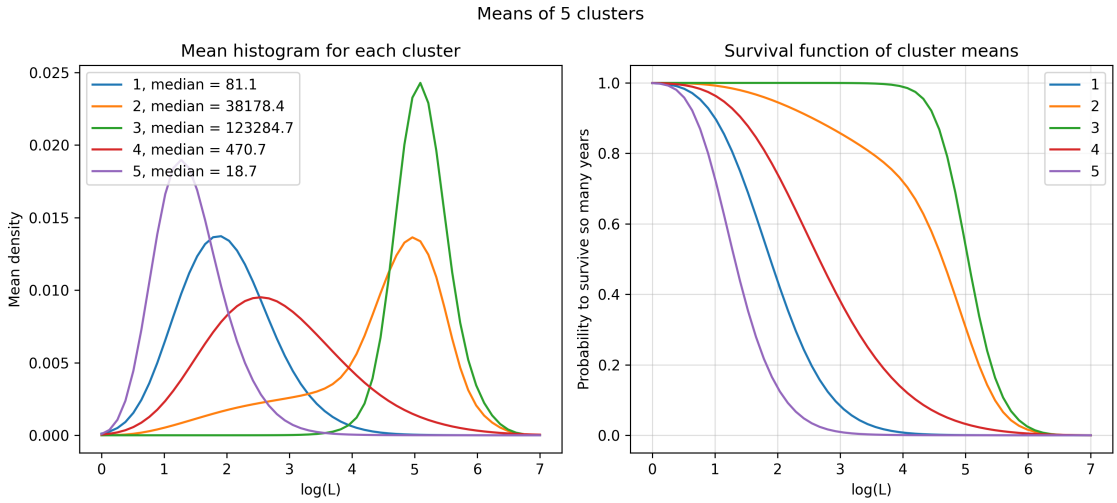


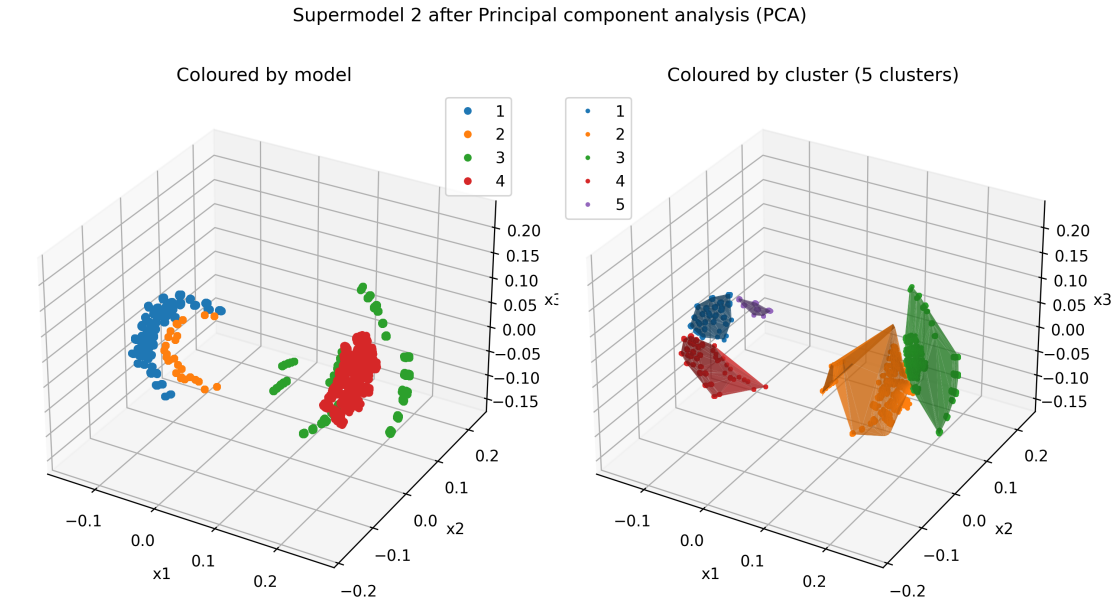

Mean density - heatmap
log(N)
log(L)
0
1
2
3
4
5
6
7
8
0
2
4
6


Mean density function by L
Density
log(L)
0.000
0.005
0.010
0.015
0.020
0.025
0.030
0.035
0
2
4
6


Mean survival function by L
Probability
log(L)
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6


Mean density function by N
Density
log(N)
0.00
0.01
0.02
0.03
0.04
0
2
4
6
8

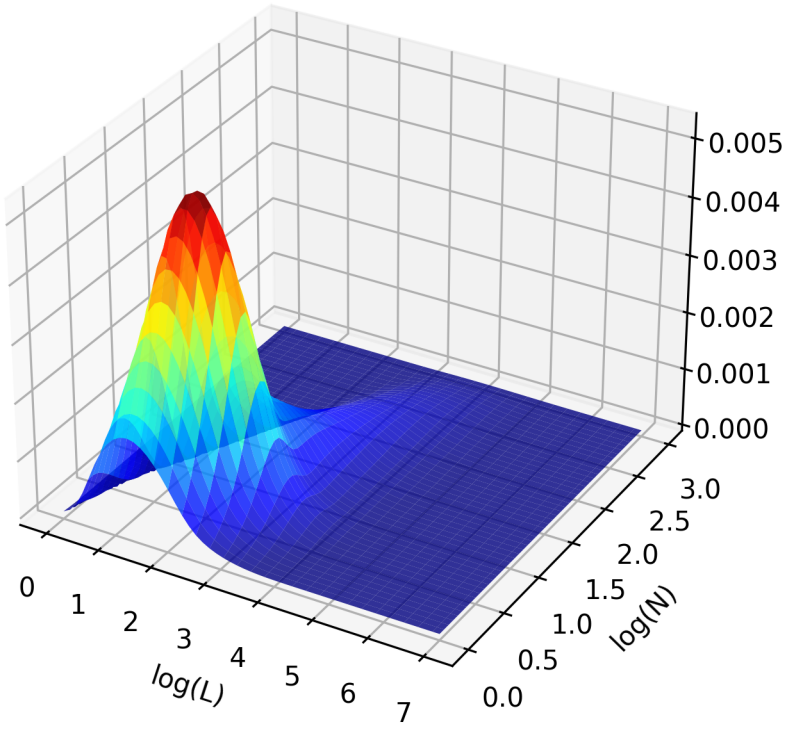




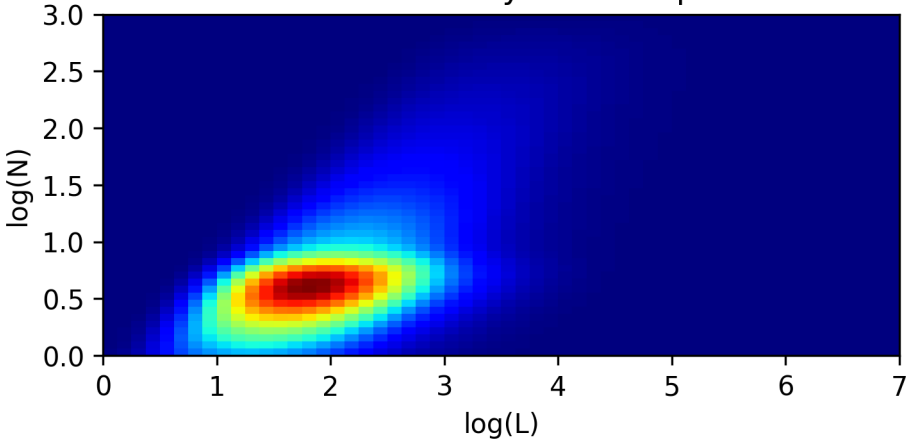
[3]:
```
cluster(model=1, ks=[4])  # model 1
```

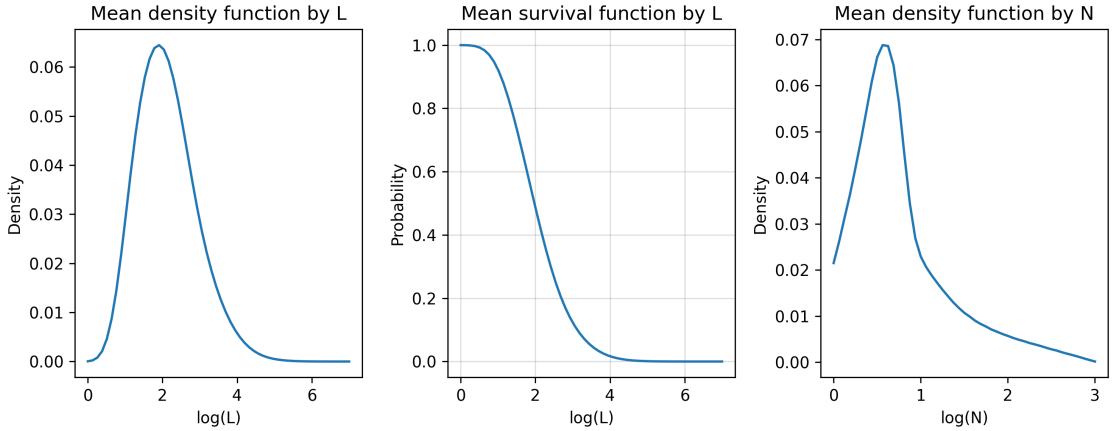
Model I

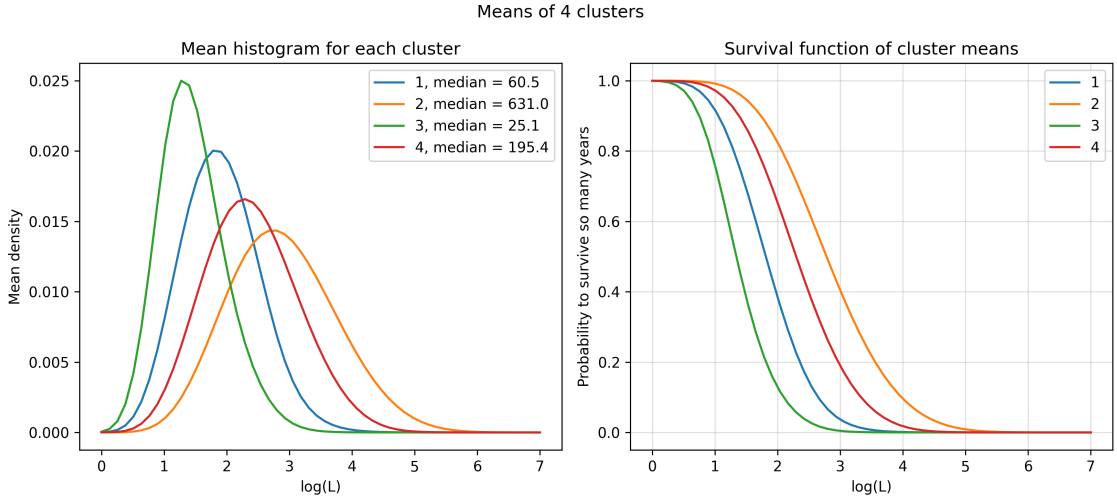
Mean density

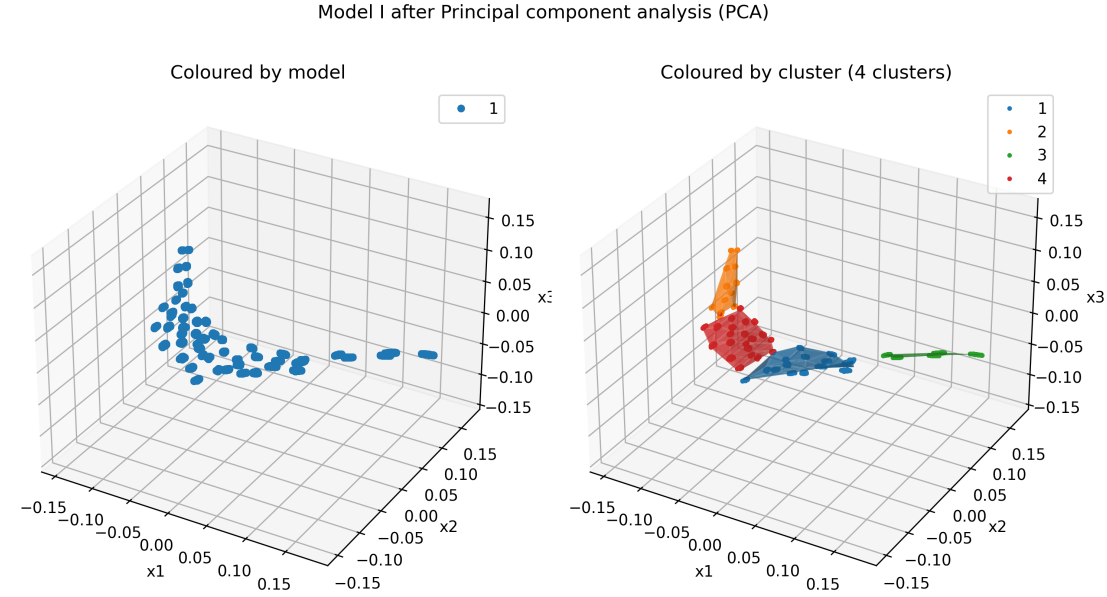


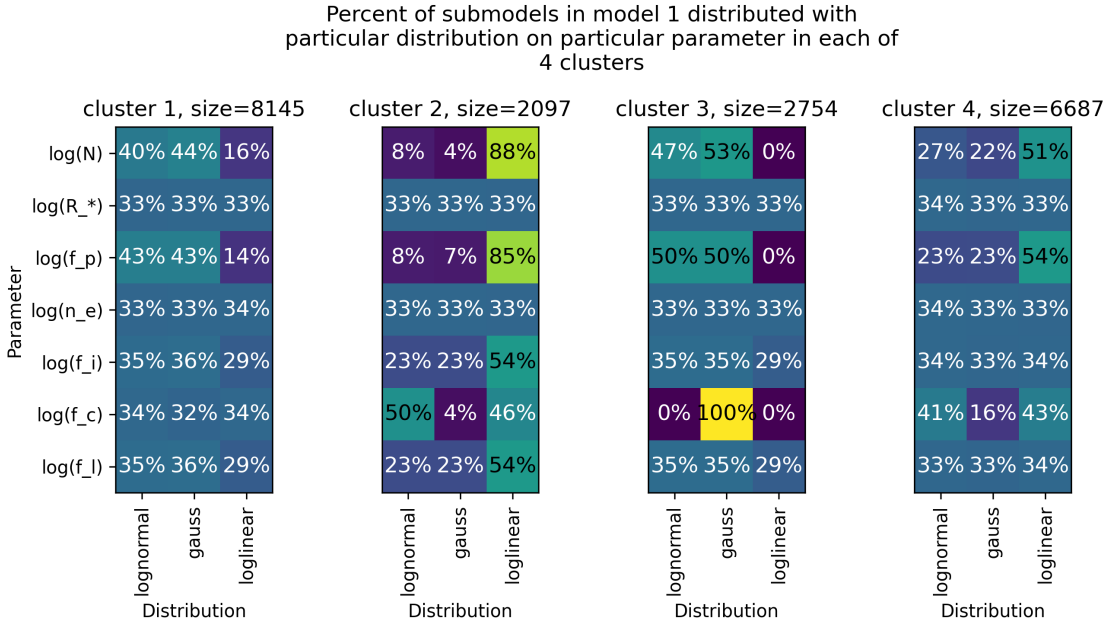
Mean density - heatmap

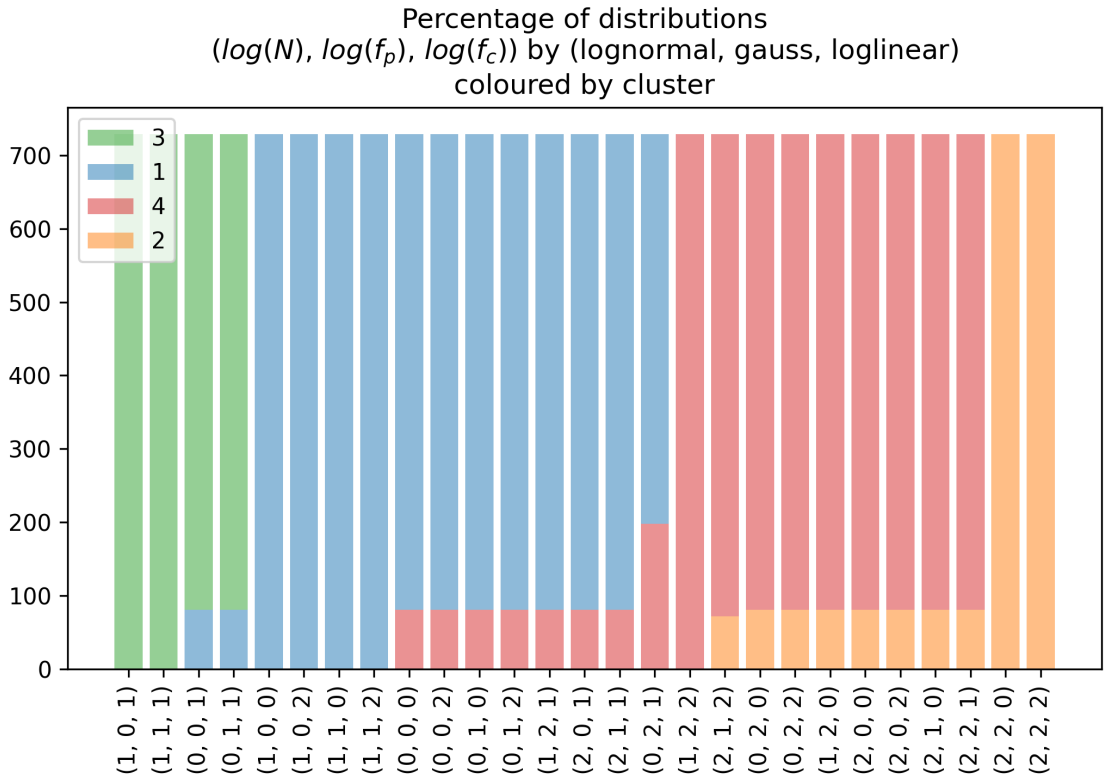



Mean density function by L
Density
log(L)
Mean survival function by L
Probability
log(L)
Mean density function by N
Density
log(N)


Means of 4 clusters
Mean histogram for each cluster
Mean density
log(L)
1, median = 60.5
2, median = 631.0
3, median = 25.1
4, median = 195.4
Survival function of cluster means
Probability to survive so many years
log(L)
1
2
3
4

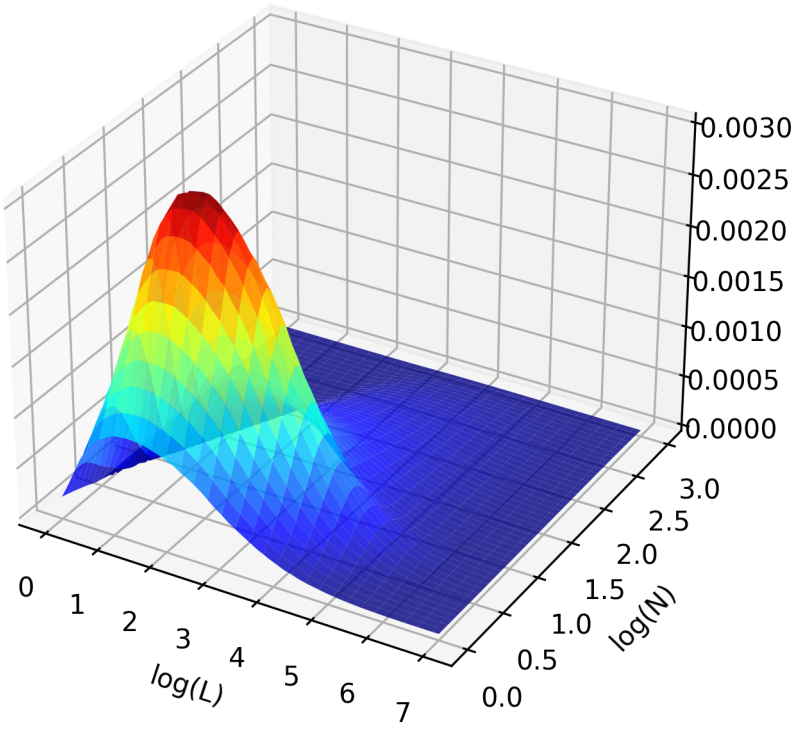

Model I after Principal component analysis (PCA)
Coloured by model
1
Coloured by cluster (4 clusters)
1
2
3
4
x1
x2
x3

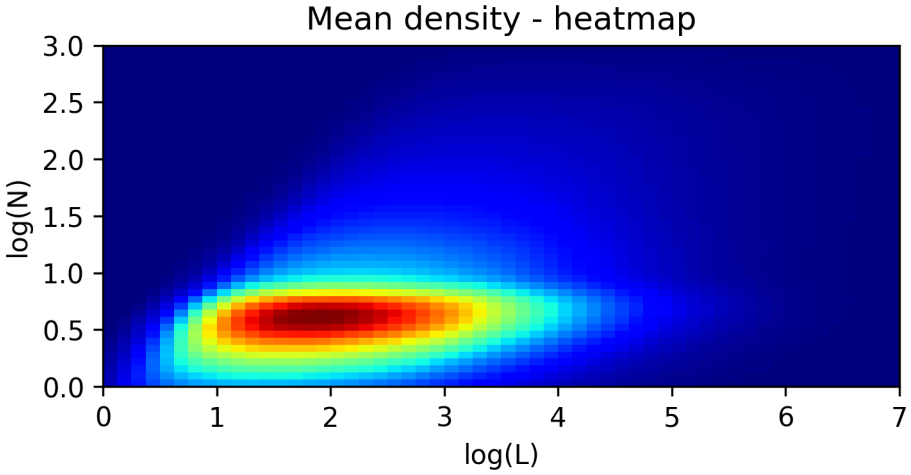

Percent of submodels in model 1 distributed with particular distribution on particular parameter in each of 4 clusters
cluster 1, size=8145
cluster 2, size=2097
cluster 3, size=2754
cluster 4, size=6687
Parameter
log(N)
log(R_*)
log(f_p)
log(n_e)
log(f_i)
log(f_c)
log(f_l)
lognormal
gauss
loglinear
Distribution

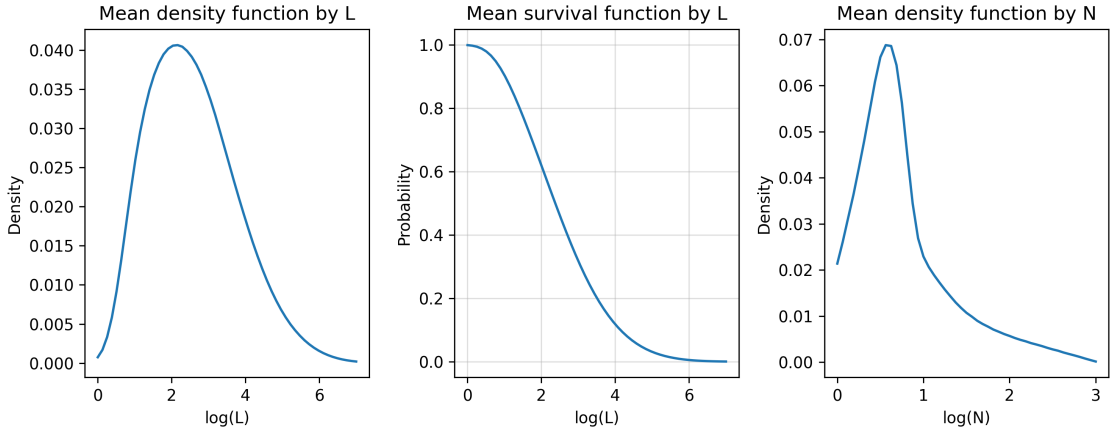


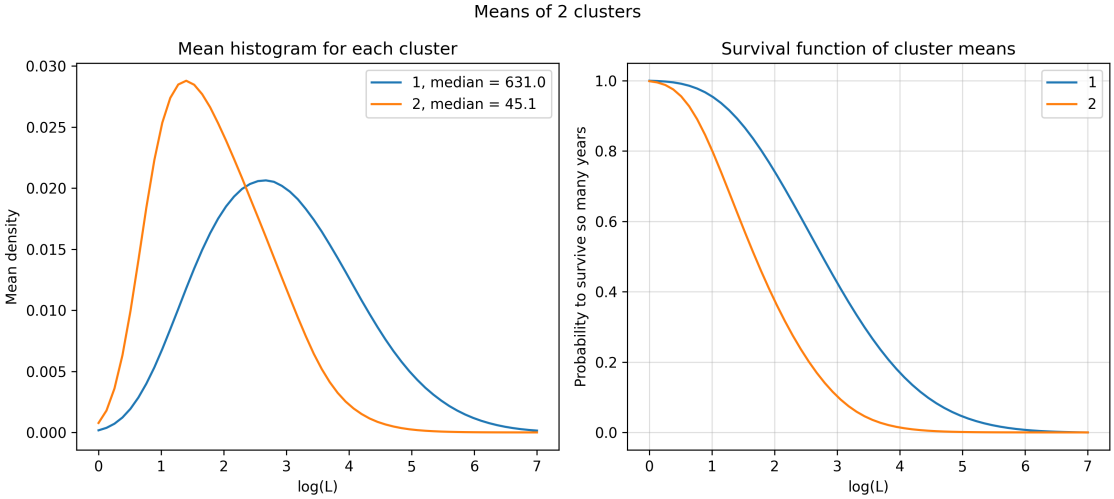
[4]:
```
cluster(model=2, ks=[2])  # model 2
```

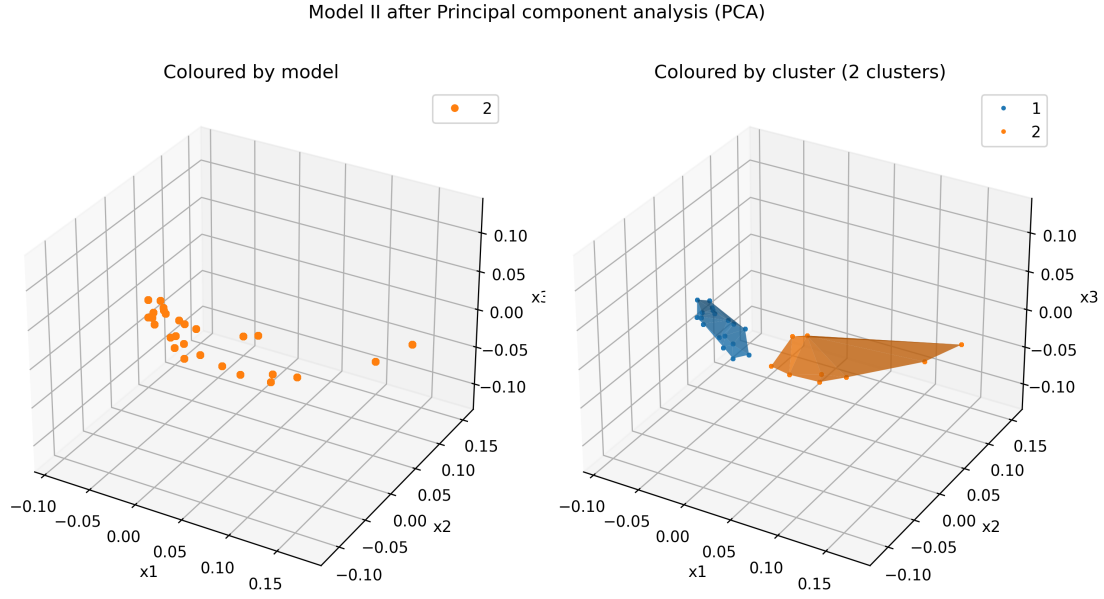
Model II

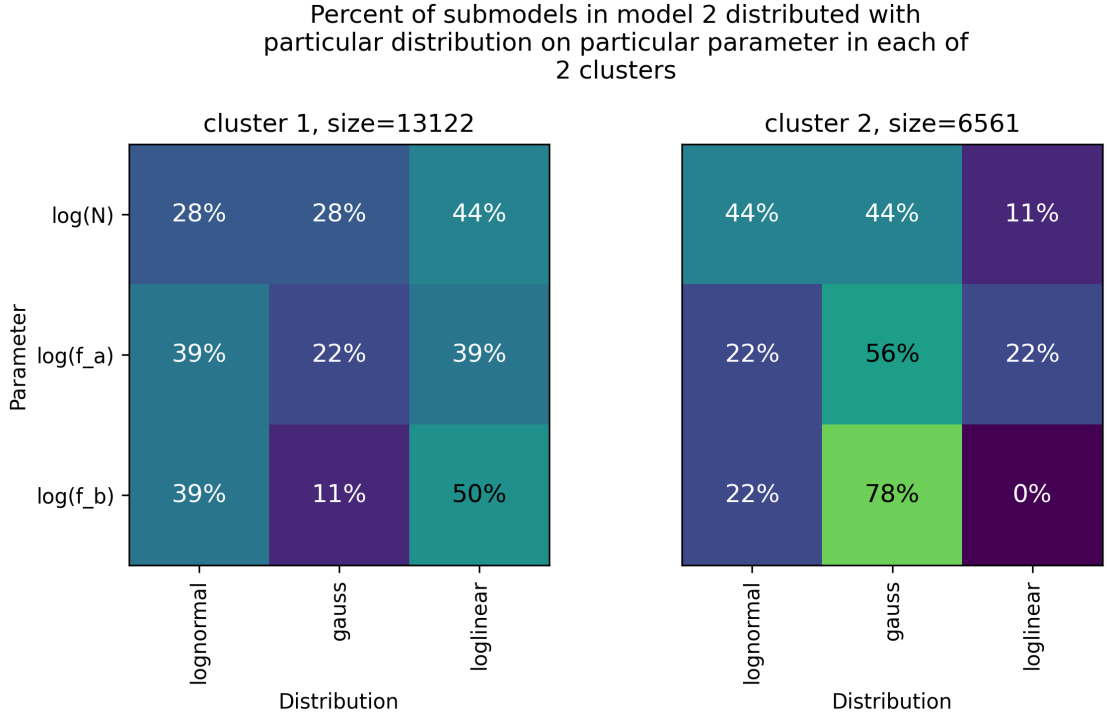

Mean density
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
0
1
2
3
4
5
6
7
log(L)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
log(N)

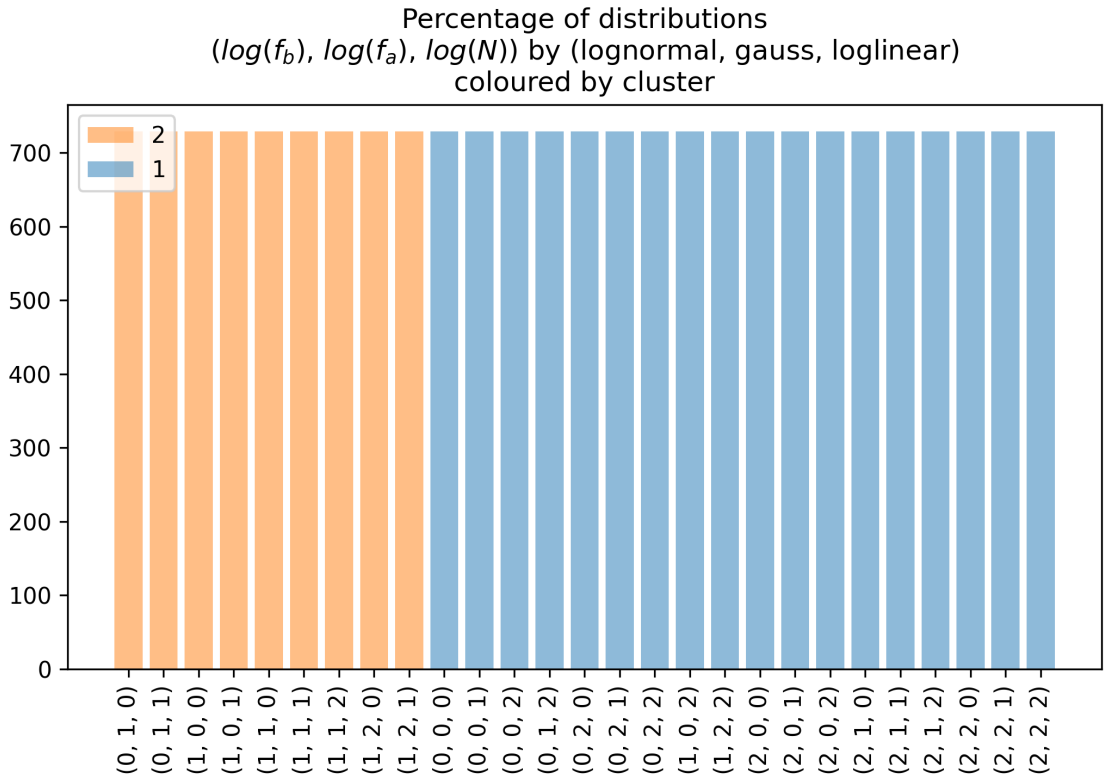

Mean density - heatmap
3.0
2.5
2.0
1.5
1.0
0.5
0.0
log(N)
0
1
2
3
4
5
6
7
log(L)


Mean density function by L
Density
log(L)
Mean survival function by L
Probability
log(L)
Mean density function by N
Density
log(N)


Means of 2 clusters
Mean histogram for each cluster
1, median = 631.0
2, median = 45.1
Mean density
log(L)
Survival function of cluster means
1
2
Probability to survive so many years
log(L)

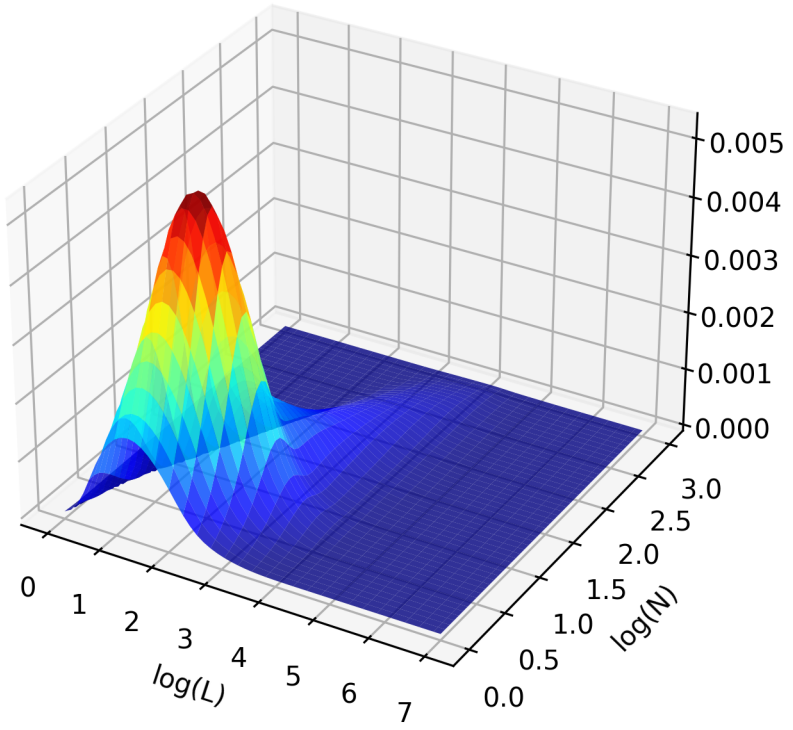

Model II after Principal component analysis (PCA)
Coloured by model
2
Coloured by cluster (2 clusters)
1
2
x1
x2
x3

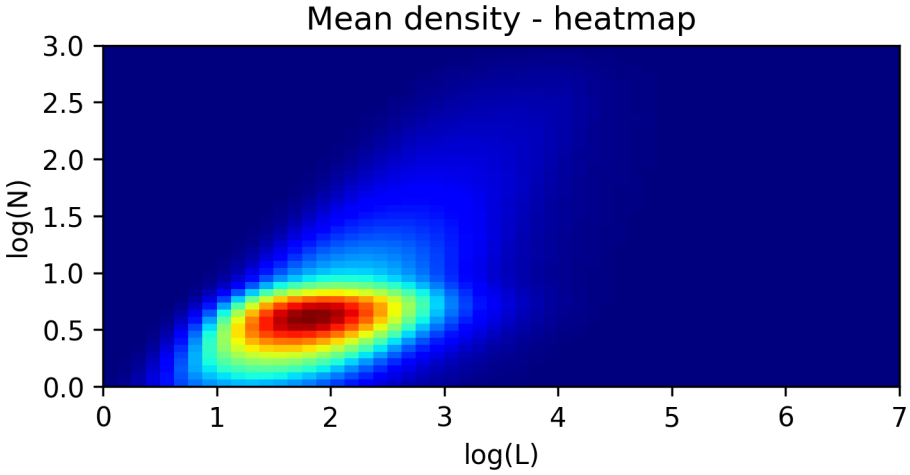

Percent of submodels in model 2 distributed with particular distribution on particular parameter in each of 2 clusters
cluster 1, size=13122
cluster 2, size=6561
Parameter
log(N)
log(f_a)
log(f_b)
28%
28%
44%
39%
22%
39%
39%
11%
50%
44%
44%
11%
22%
56%
22%
22%
78%
0%
lognormal
gauss
loglinear
Distribution

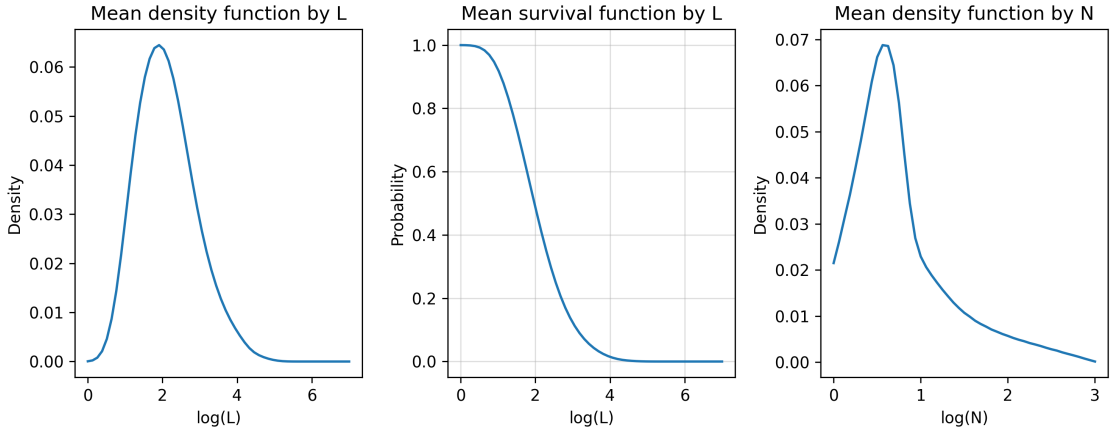


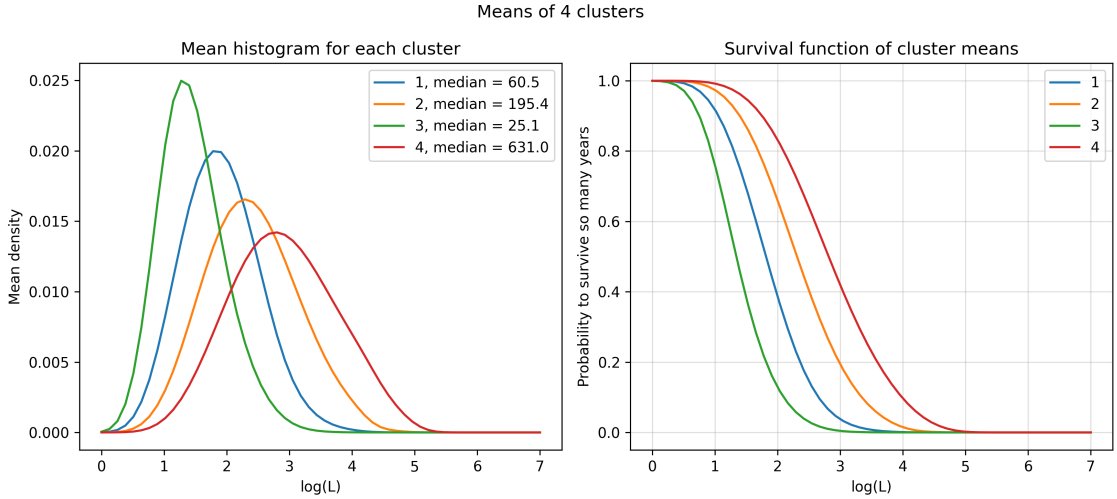
[5]:
```
for s in [1, 2]:
    cluster(model=3, ks=[4], supermodel=s)  # model 3
```

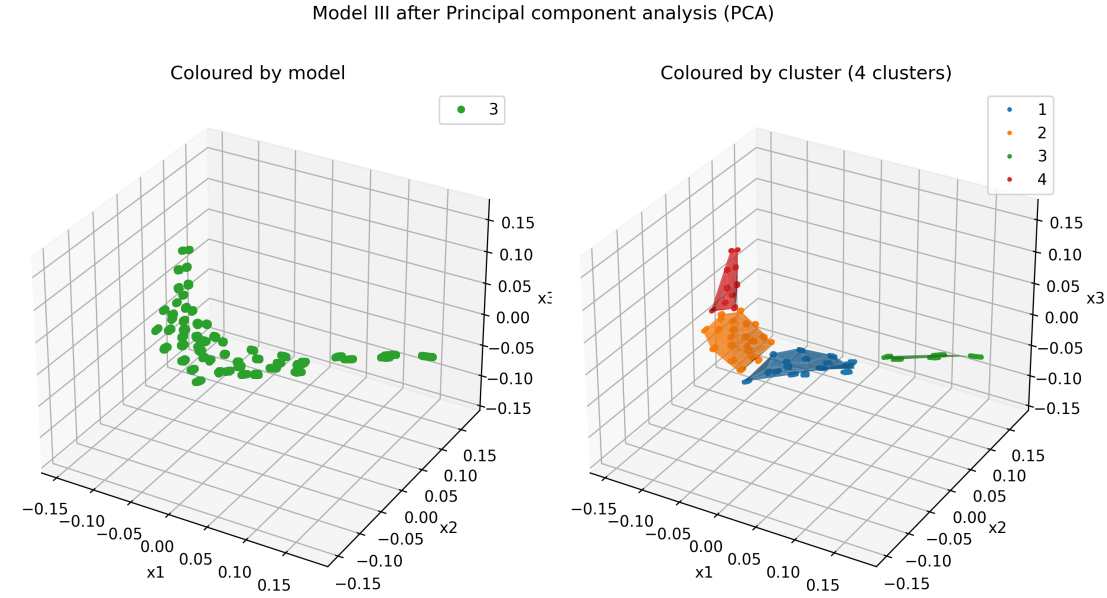
Model III

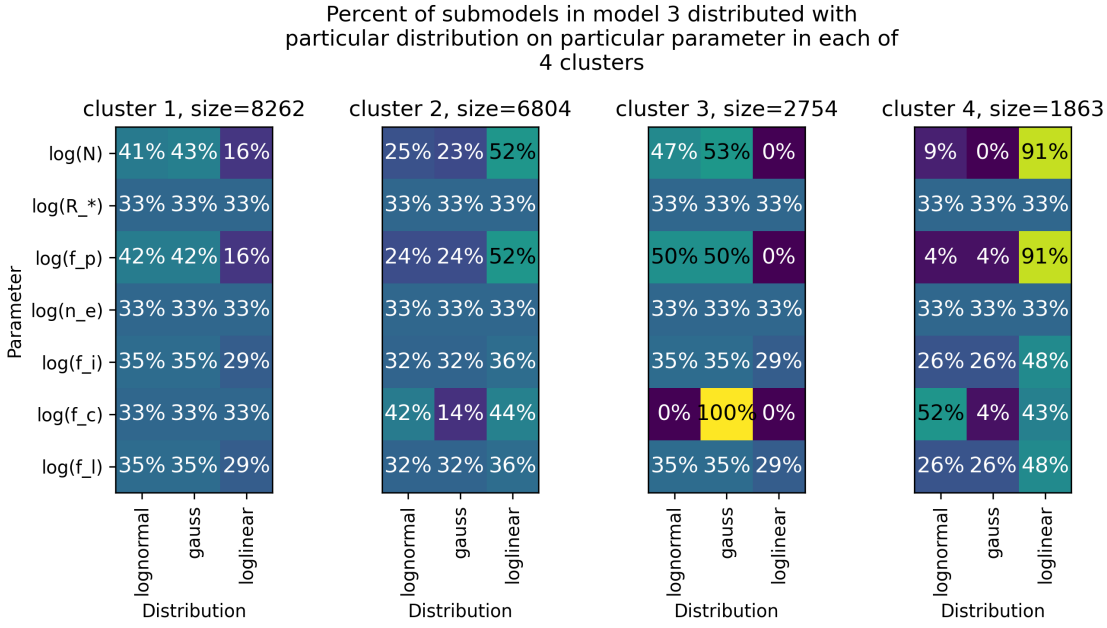

Mean density
0.005
0.004
0.003
0.002
0.001
0.000
0
1
2
3
4
5
6
7
log(L)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
log(N)

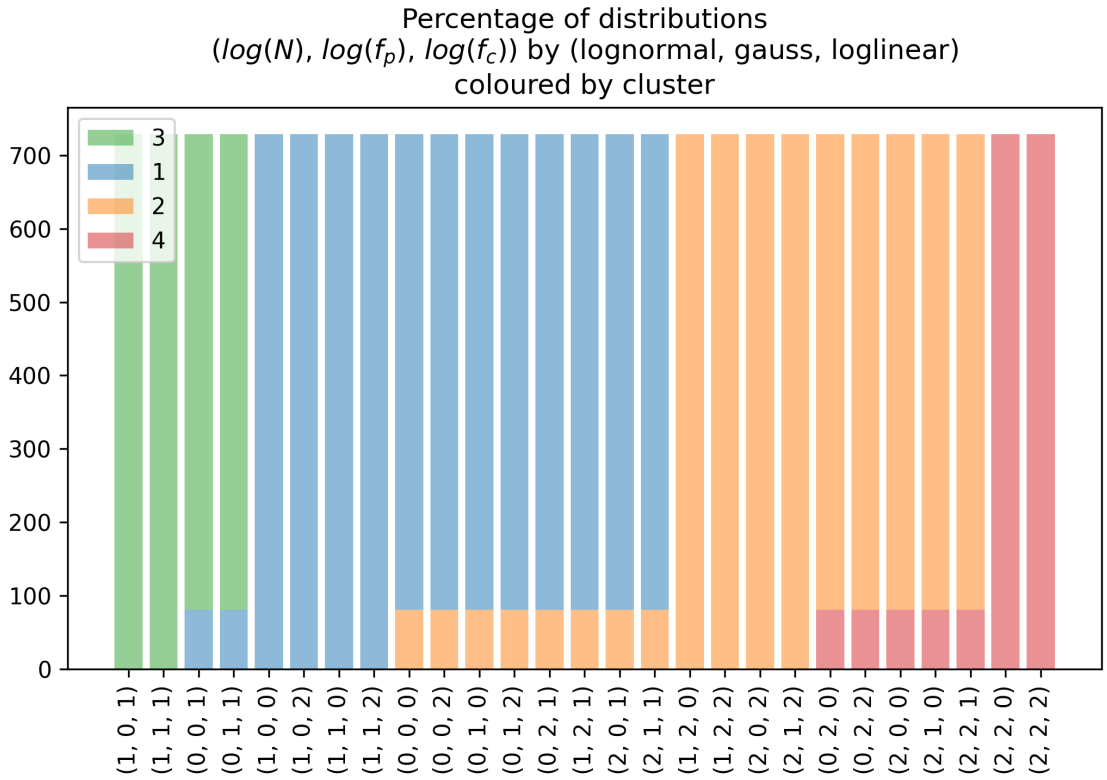

Mean density - heatmap
3.0
2.5
2.0
1.5
1.0
0.5
0.0
log(N)
0
1
2
3
4
5
6
7
log(L)


Mean density function by L
Density
log(L)
Mean survival function by L
Probability
log(L)
Mean density function by N
Density
log(N)


Means of 4 clusters
Mean histogram for each cluster
Mean density
log(L)
1, median = 60.5
2, median = 195.4
3, median = 25.1
4, median = 631.0
Survival function of cluster means
Probability to survive so many years
log(L)
1
2
3
4

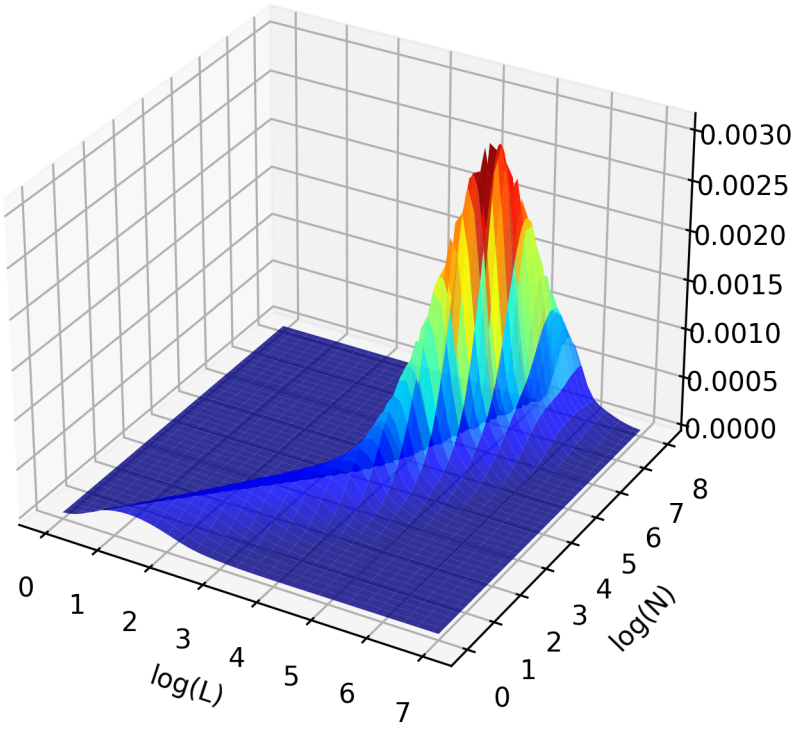

Model III after Principal component analysis (PCA)
Coloured by model
3
Coloured by cluster (4 clusters)
1
2
3
4
x1
x2
x3

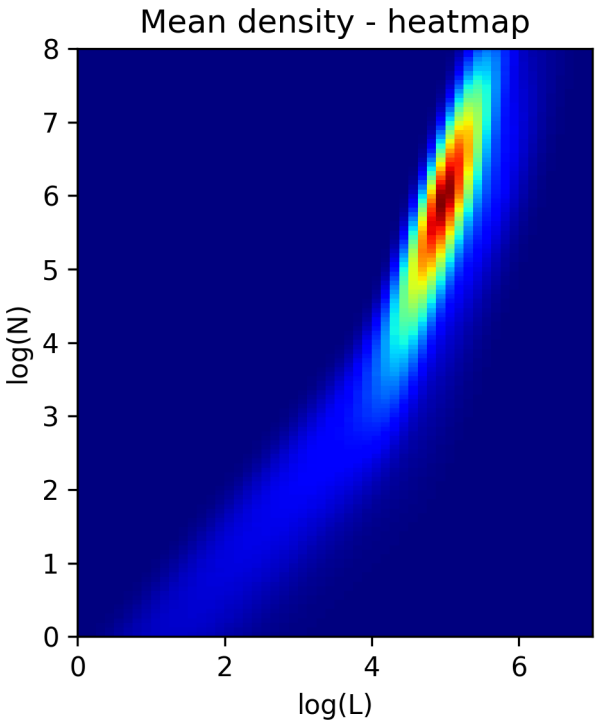

Percent of submodels in model 3 distributed with particular distribution on particular parameter in each of 4 clusters
cluster 1, size=8262
cluster 2, size=6804
cluster 3, size=2754
cluster 4, size=1863
Parameter
log(N)
log(R_*)
log(f_p)
log(n_e)
log(f_i)
log(f_c)
log(f_l)
41% 43% 16%
33% 33% 33%
42% 42% 16%
33% 33% 33%
35% 35% 29%
33% 33% 33%
35% 35% 29%
25% 23% 52%
33% 33% 33%
24% 24% 52%
33% 33% 33%
32% 32% 36%
42% 14% 44%
32% 32% 36%
47% 53% 0%
33% 33% 33%
50% 50% 0%
33% 33% 33%
35% 35% 29%
0% 100% 0%
35% 35% 29%
9% 0% 91%
33% 33% 33%
4% 4% 91%
33% 33% 33%
26% 26% 48%
52% 4% 43%
26% 26% 48%
lognormal
gauss
loglinear
Distribution

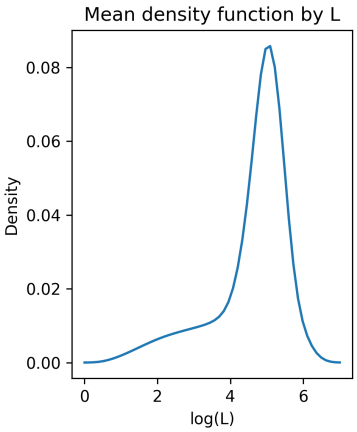

Percentage of distributions
($log(N)$, $log(f_p)$, $log(f_c)$) by (lognormal, gauss, loglinear)
coloured by cluster
3
1
2
4
0
100
200
300
400
500
600
700
(1, 0, 1)
(1, 1, 1)
(0, 0, 1)
(0, 1, 1)
(1, 0, 0)
(1, 0, 2)
(1, 1, 0)
(1, 1, 2)
(0, 0, 0)
(0, 0, 2)
(0, 1, 0)
(0, 1, 2)
(0, 2, 1)
(1, 2, 1)
(2, 0, 1)
(2, 1, 1)
(1, 2, 0)
(1, 2, 2)
(2, 0, 2)
(2, 1, 2)
(0, 2, 0)
(0, 2, 2)
(2, 0, 0)
(2, 1, 0)
(2, 2, 1)
(2, 2, 0)
(2, 2, 2)

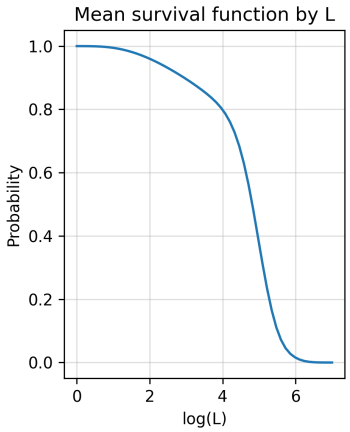
Model III

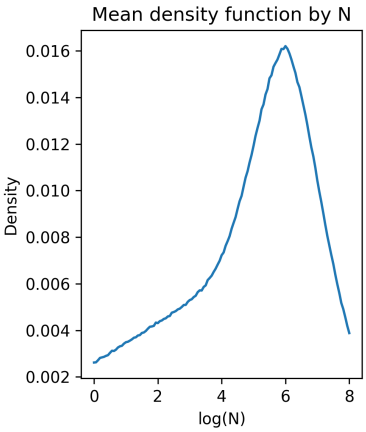
Mean density

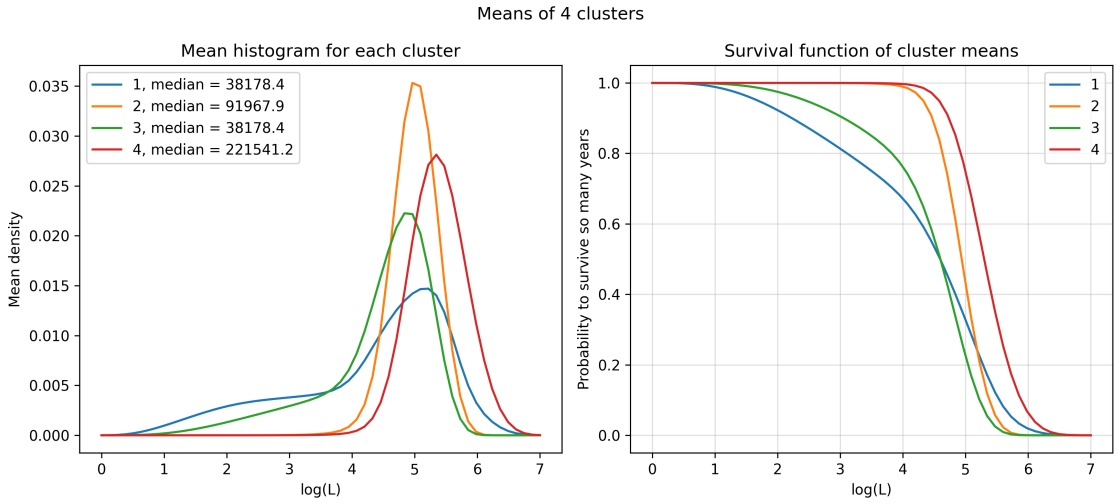


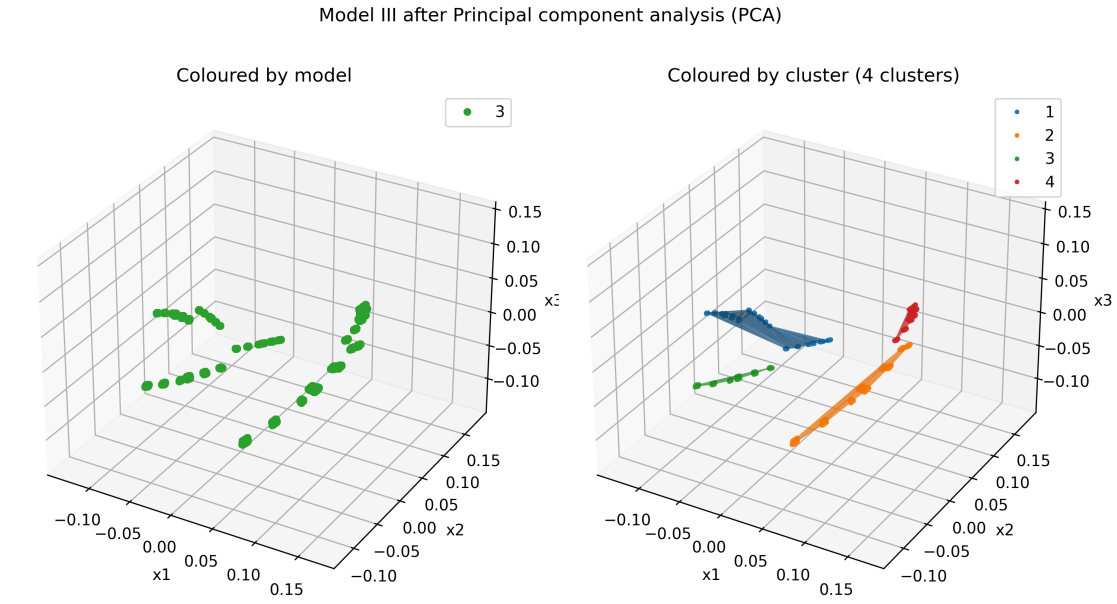

Mean density - heatmap
log(N)
log(L)
0
1
2
3
4
5
6
7
8
0
2
4
6


Mean density function by L
Density
log(L)
0.00
0.02
0.04
0.06
0.08
0
2
4
6


Mean survival function by L
Probability
log(L)
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6

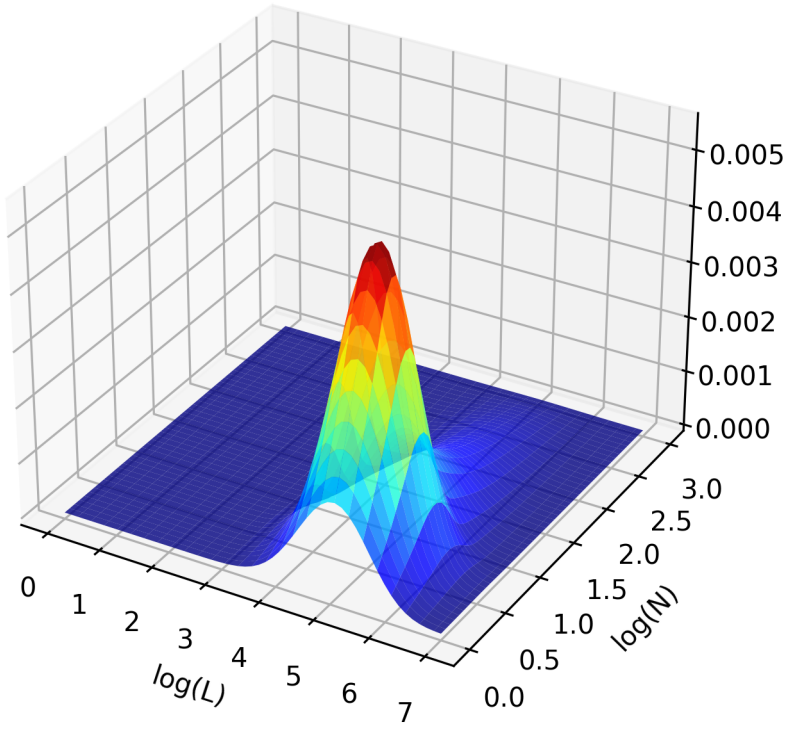

Mean density function by N
Density
log(N)
0.002
0.004
0.006
0.008
0.010
0.012
0.014
0.016
0
2
4
6
8

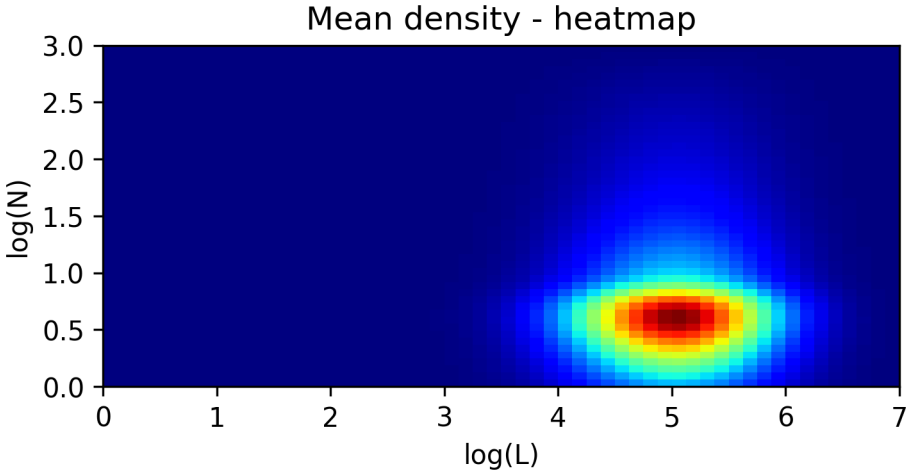


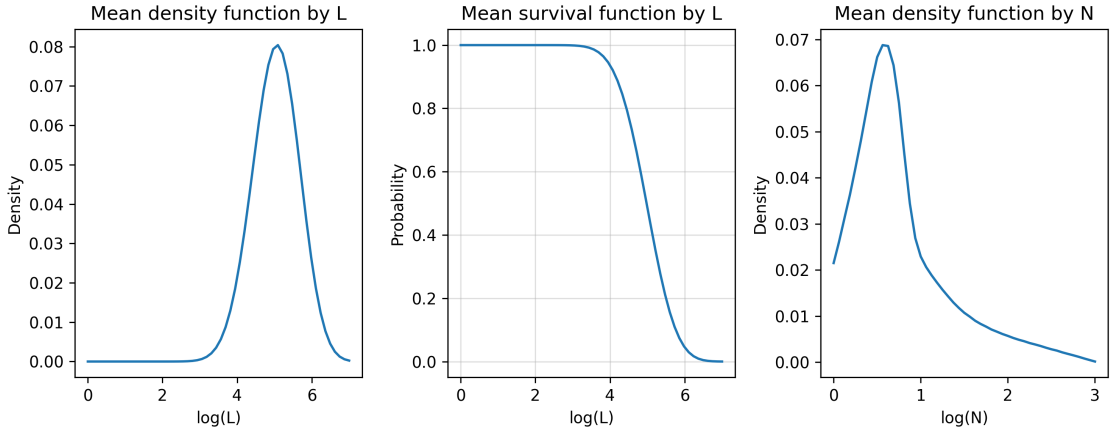


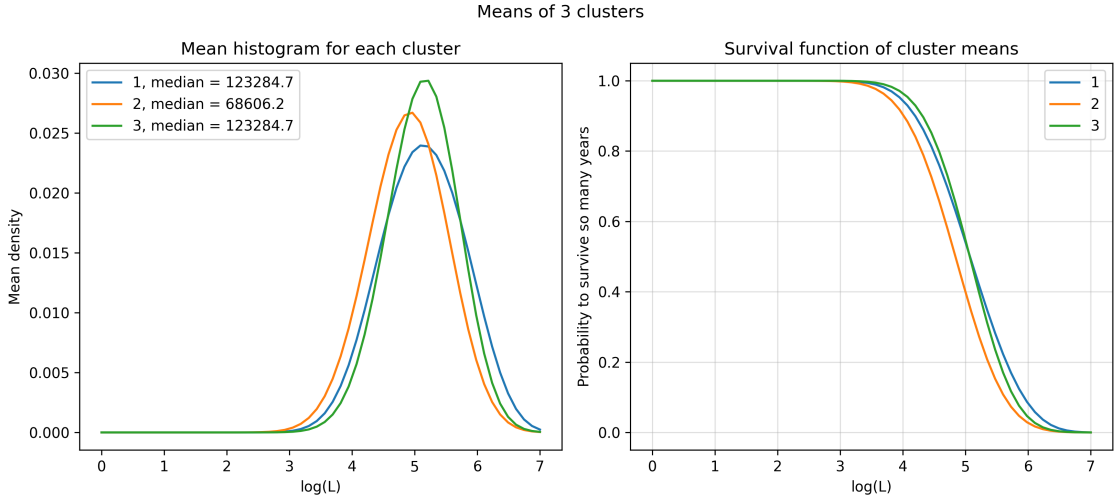
[6]:
```
cluster(model=4, ks=[3])  # model 4
```

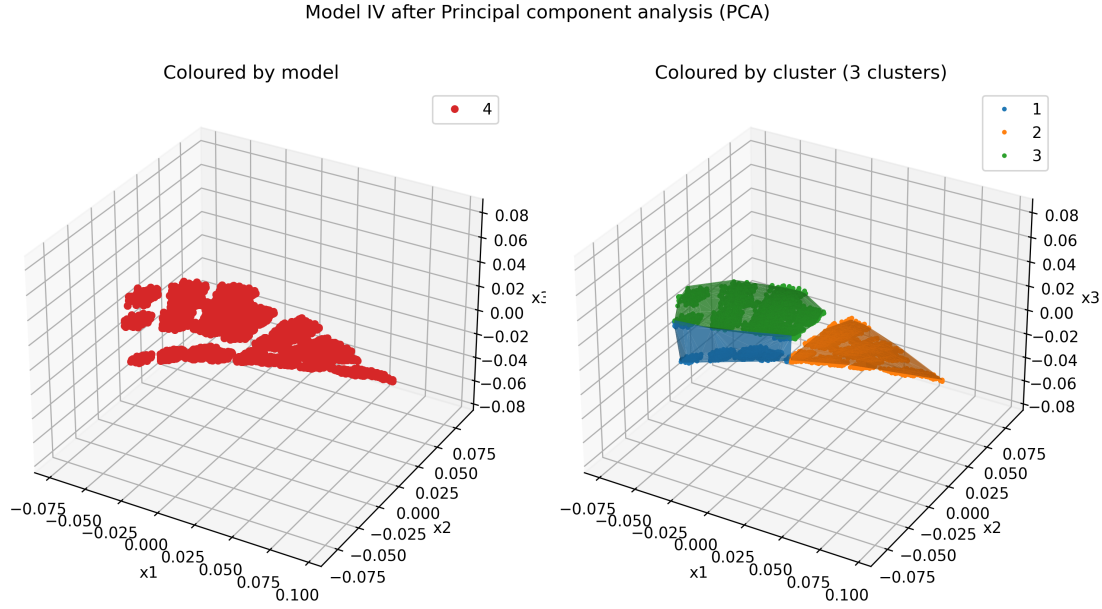
Model IV

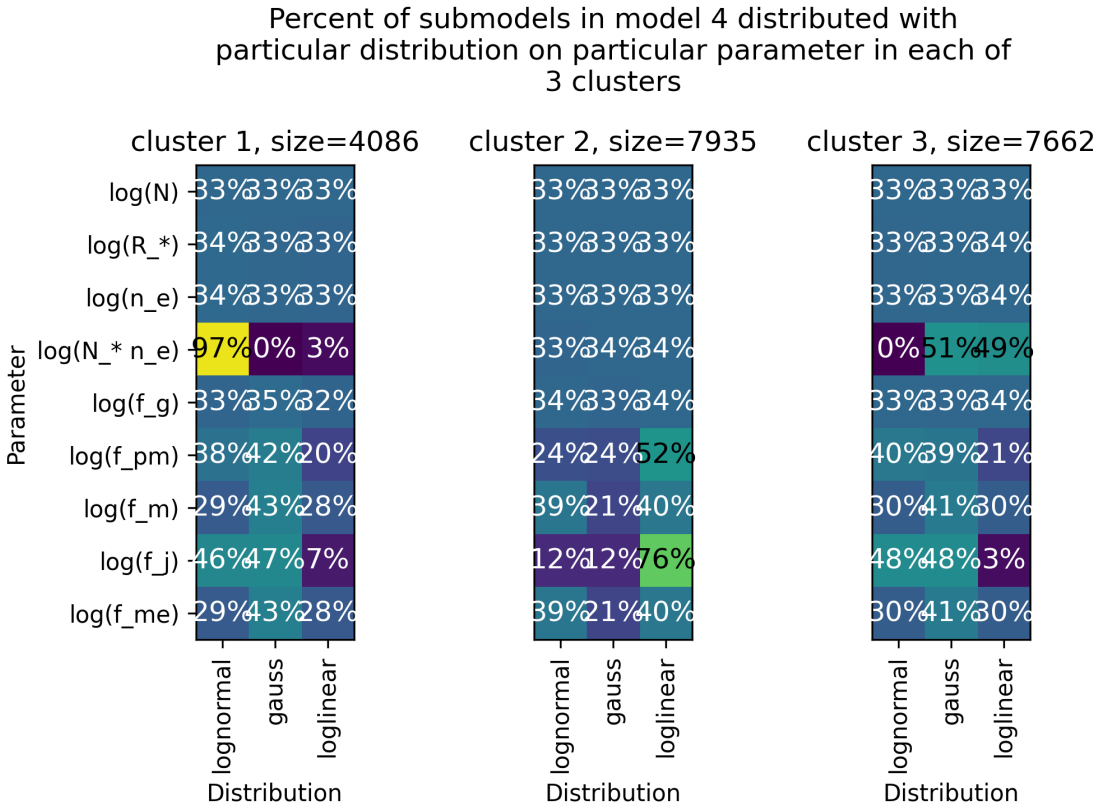

Mean density
0.005
0.004
0.003
0.002
0.001
0.000
0
1
2
3
4
5
6
7
log(L)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
log(N)

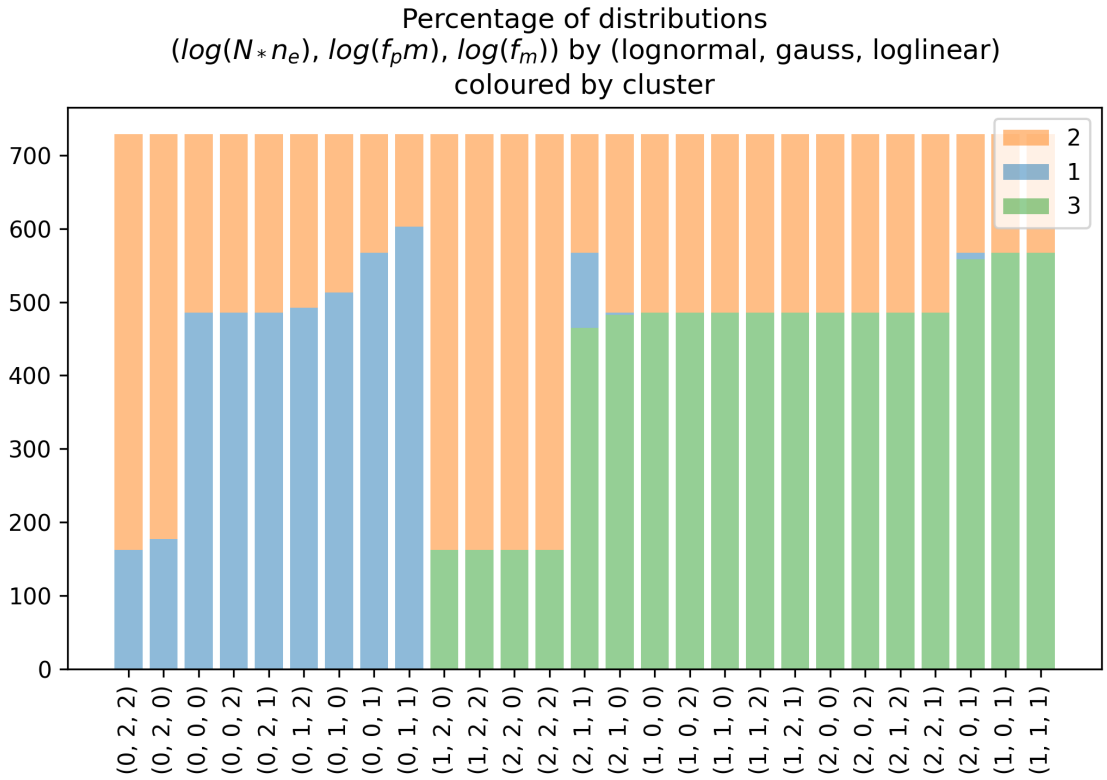

Mean density - heatmap
3.0
2.5
2.0
1.5
1.0
0.5
0.0
log(N)
0
1
2
3
4
5
6
7
log(L)


Mean density function by L
Density
log(L)
Mean survival function by L
Probability
log(L)
Mean density function by N
Density
log(N)


Means of 3 clusters
Mean histogram for each cluster
1, median = 123284.7
2, median = 68606.2
3, median = 123284.7
Mean density
log(L)
Survival function of cluster means
Probability to survive so many years
1
2
3
log(L)

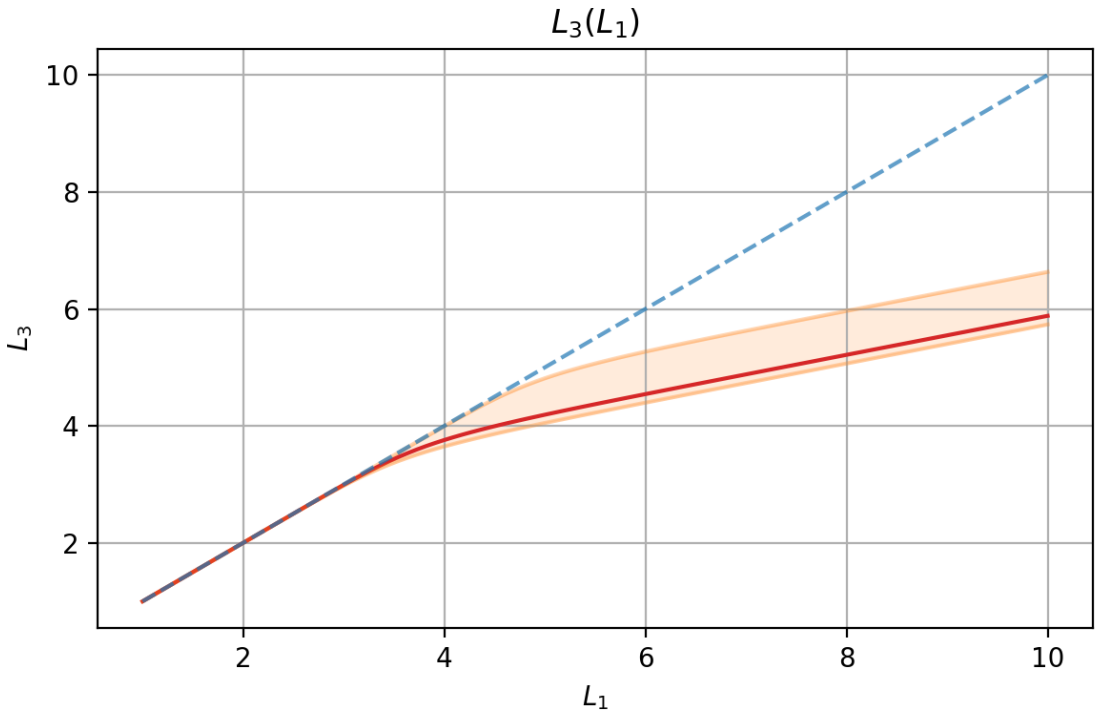

Model IV after Principal component analysis (PCA)
Coloured by model
4
Coloured by cluster (3 clusters)
1
2
3
x1
x2
x3


Percent of submodels in model 4 distributed with particular distribution on particular parameter in each of 3 clusters
cluster 1, size=4086
cluster 2, size=7935
cluster 3, size=7662
Parameter
log(N)
log(R_*)
log(n_e)
log(N_* n_e)
log(f_g)
log(f_pm)
log(f_m)
log(f_j)
log(f_me)
33% 33% 33%
34% 33% 33%
34% 33% 33%
97% 0% 3%
33% 35% 32%
38% 42% 20%
29% 43% 28%
46% 47% 7%
29% 43% 28%
33% 33% 33%
33% 33% 33%
33% 33% 33%
33% 34% 34%
34% 33% 34%
24% 24% 52%
39% 21% 40%
12% 12% 76%
39% 21% 40%
33% 33% 33%
33% 33% 34%
33% 33% 34%
0% 51% 49%
33% 33% 34%
40% 39% 21%
30% 41% 30%
48% 48% 3%
30% 41% 30%
lognormal
gauss
loglinear
Distribution

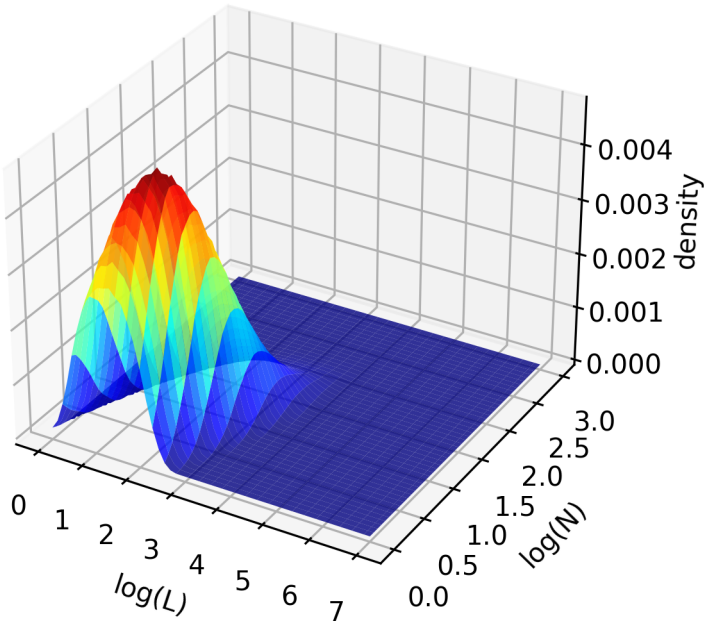


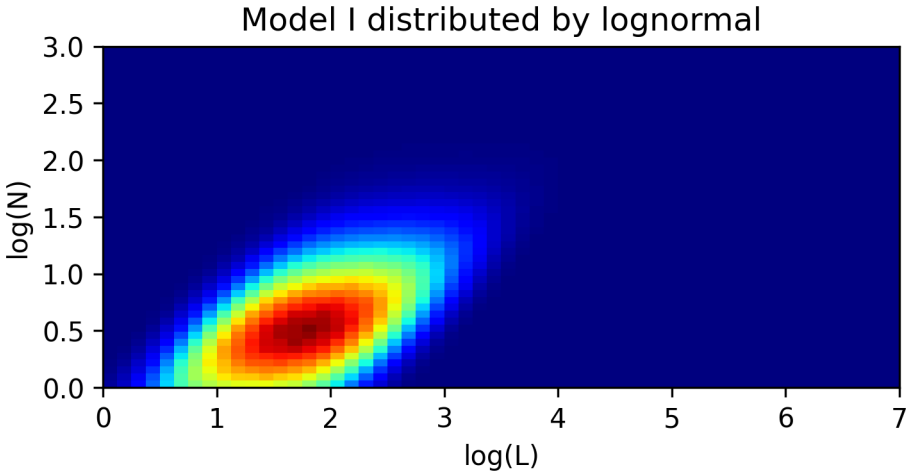
[7]:
```
def polyDeltaL(logL1):
    minP, meanP, maxP = -2, np.log10(1.8), np.log10(5)
    L1 = 10 ** logL1
    a1 = 10 ** (9 - np.array([minP, meanP, maxP]) - np.log10(np.pi * 4))
    a1 = [np.roots([1, 3 * L1, 3 * L1 ** 2 + a, L1 ** 3]) for a in a1]
    return [- min(a, key=lambda x: np.abs(x.imag)).real for a in a1]


xi = np.linspace(1, 10, 100)
xi3 = np.array([xi, xi, xi]).T
h = np.array([polyDeltaL(x) for x in xi])
ca = [("tab:orange", 0.3), ("tab:red", 1), ("tab:orange", 0.3)]

for ylab, ydat in [("L_3", np.log10(np.abs(10 ** xi3 - h)))]:
    plt.figure(figsize=(6, 4), dpi=200, tight_layout=True)
    plt.fill_between(xi, ydat[:, 0], ydat[:, 2],
                     color=ca[0][0], alpha=ca[0][1] / 2)
    for i in range(3):
        plt.plot(xi, ydat[:, i], color=ca[i][0], alpha=ca[i][1])
    if "L_1" not in ylab:
        plt.plot(xi, xi, "--", alpha=0.7)
    plt.title(f"${ylab}(L_1)$")
```

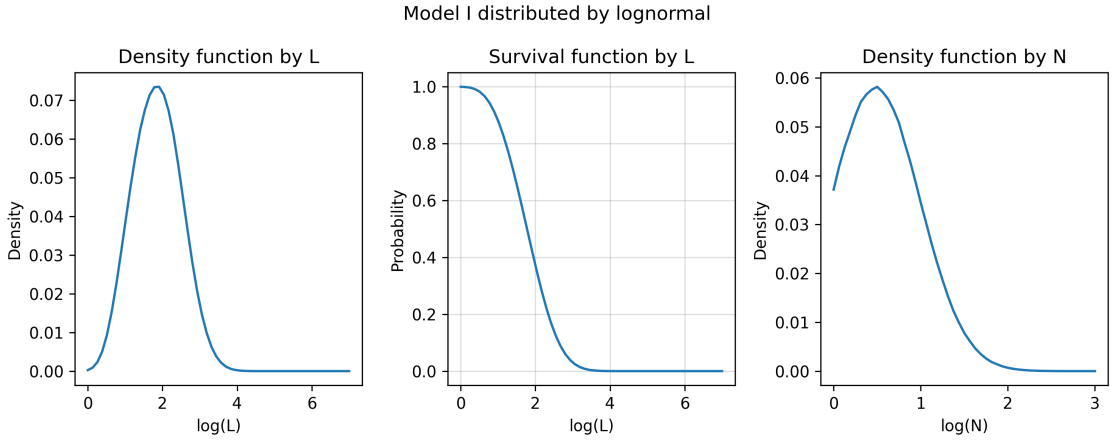
```
    plt.xlabel("$L_1$")
    plt.ylabel(f"${ylab}$")
    plt.grid()
    plt.savefig(f"out/model3-model1.png")
    plt.show()
```

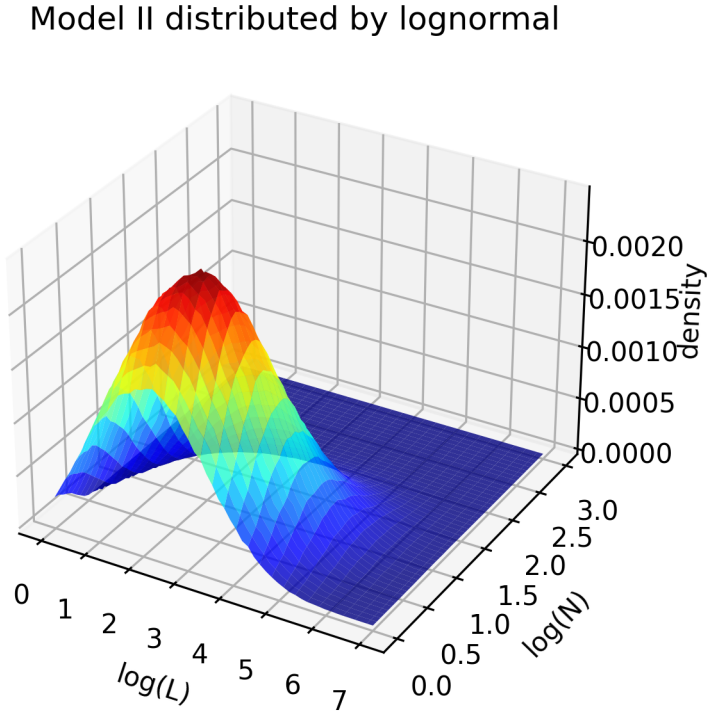


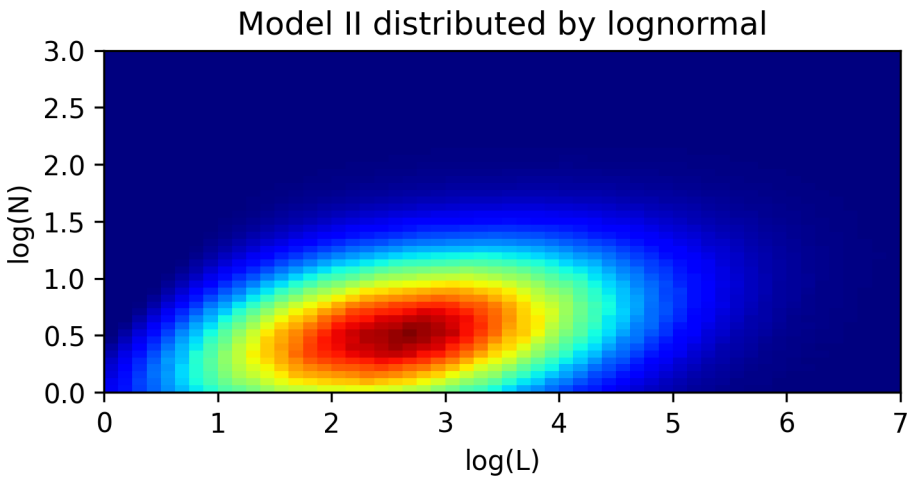
[8]:
```
from plot3D_L import *
draw_histograms3D(model=1, distribution=tuple(0 for _ in range(6)),␣
 ↪supermodel=1)
draw_histograms3D(model=2, distribution=tuple(0 for _ in range(2)),␣
 ↪supermodel=1)
draw_histograms3D(model=3, distribution=tuple(1 for _ in range(6)),␣
 ↪supermodel=1)
draw_histograms3D(model=4, distribution=tuple(0 for _ in range(8)),␣
 ↪supermodel=1)
draw_histograms_N3D()
plt.show()
```

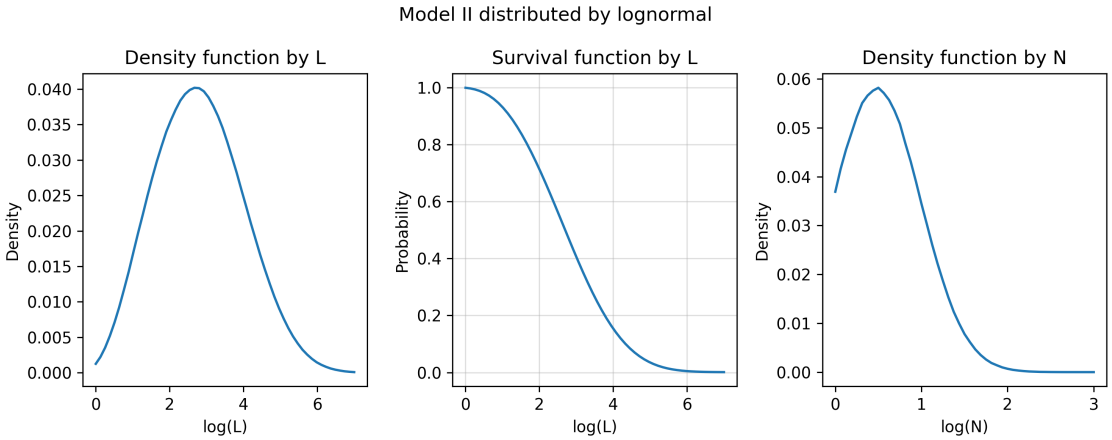

Model I distributed by lognormal
0.004
0.003
0.002
0.001
0.000
density
0
1
2
3
4
5
6
7
log(L)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
log(N)

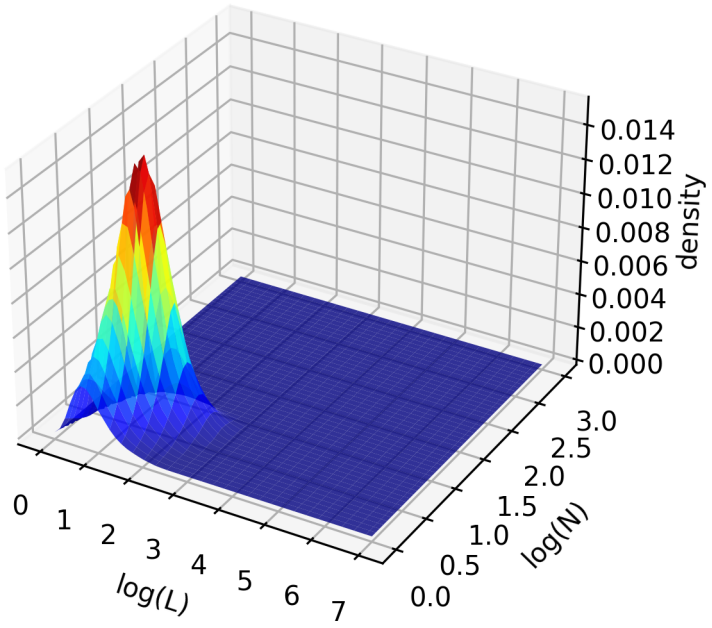

Model I distributed by lognormal
3.0
2.5
2.0
1.5
1.0
0.5
0.0
log(N)
0
1
2
3
4
5
6
7
log(L)

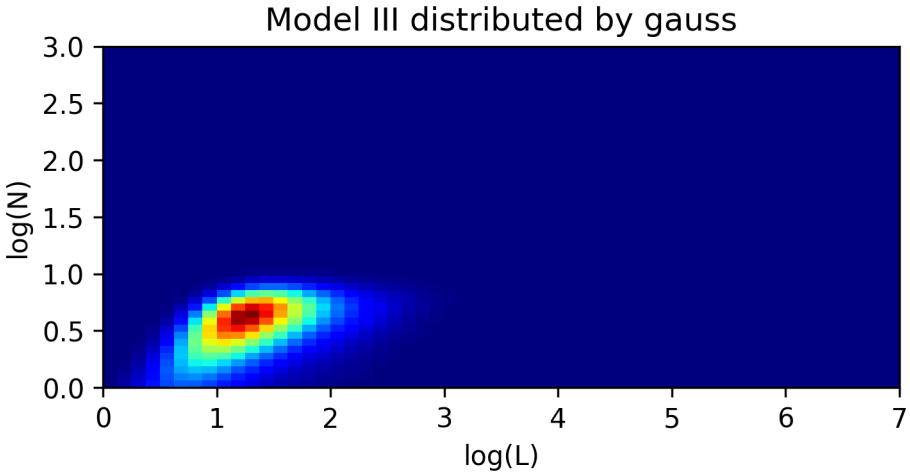

Model I distributed by lognormal
Density function by L
Density
log(L)
Survival function by L
Probability
log(L)
Density function by N
Density
log(N)

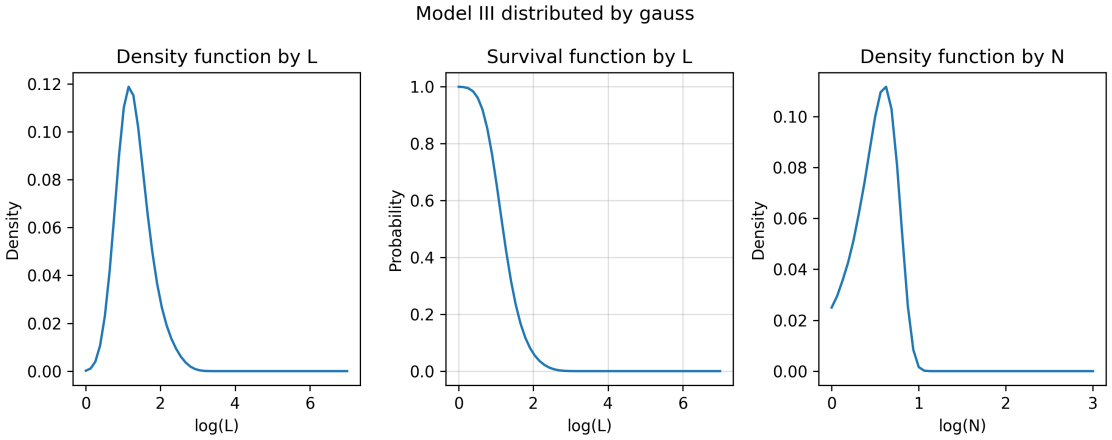

Model II distributed by lognormal
density
log(L)
log(N)

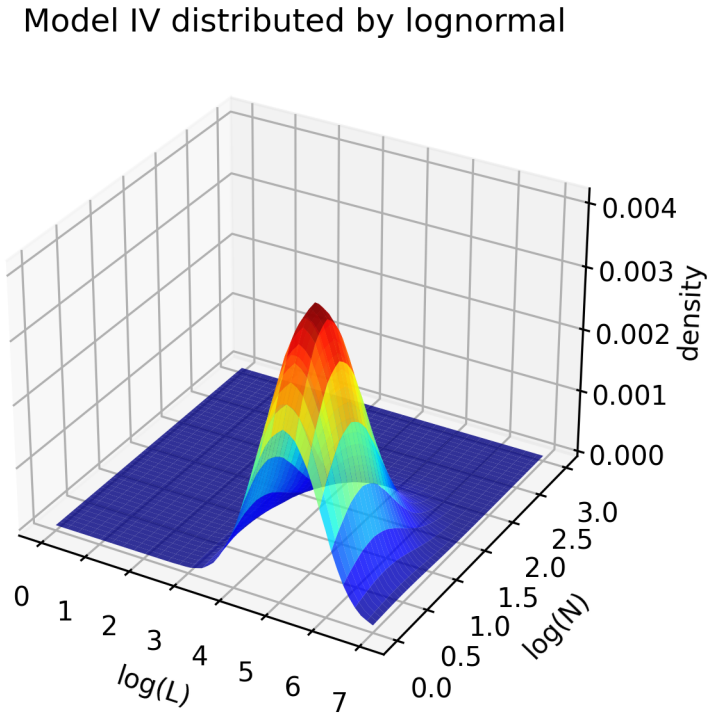

Model II distributed by lognormal
log(N)
log(L)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
1
2
3
4
5
6
7

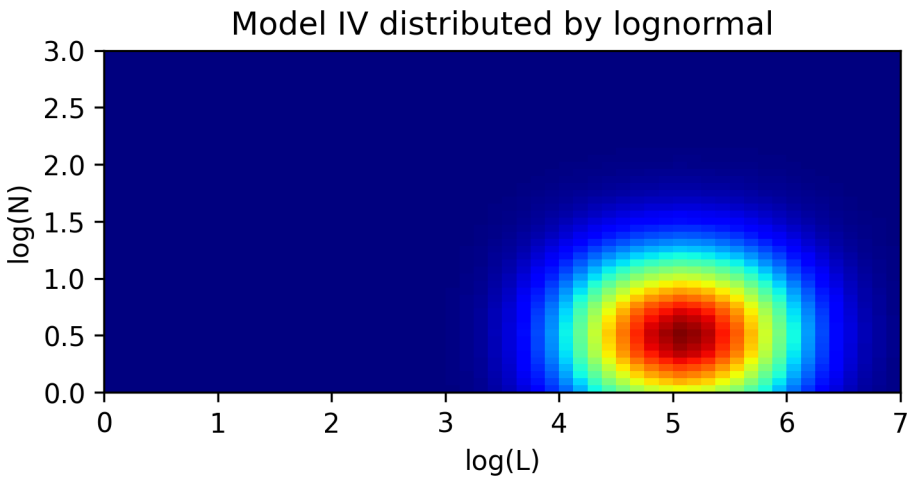

Model II distributed by lognormal
Density function by L
Density
log(L)
Survival function by L
Probability
log(L)
Density function by N
Density
log(N)

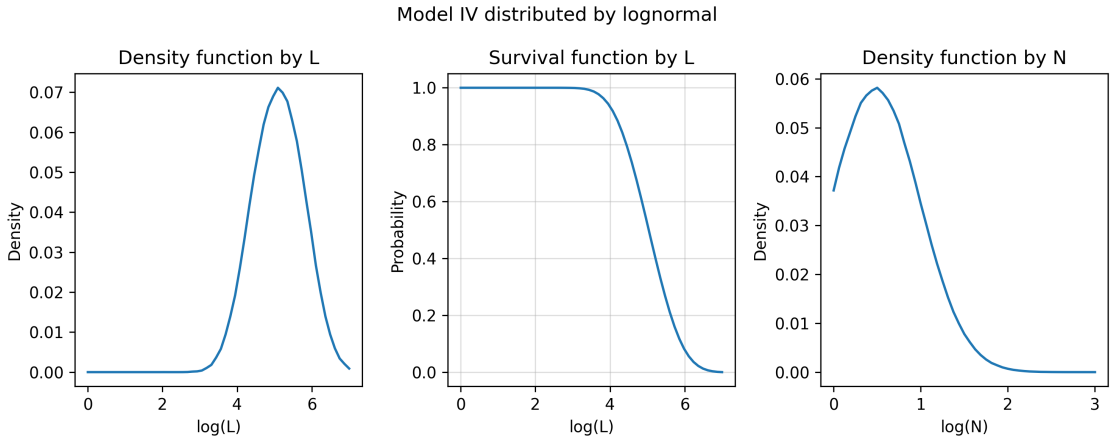

Model III distributed by gauss
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
density
0
1
2
3
4
5
6
7
log(L)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
log(N)

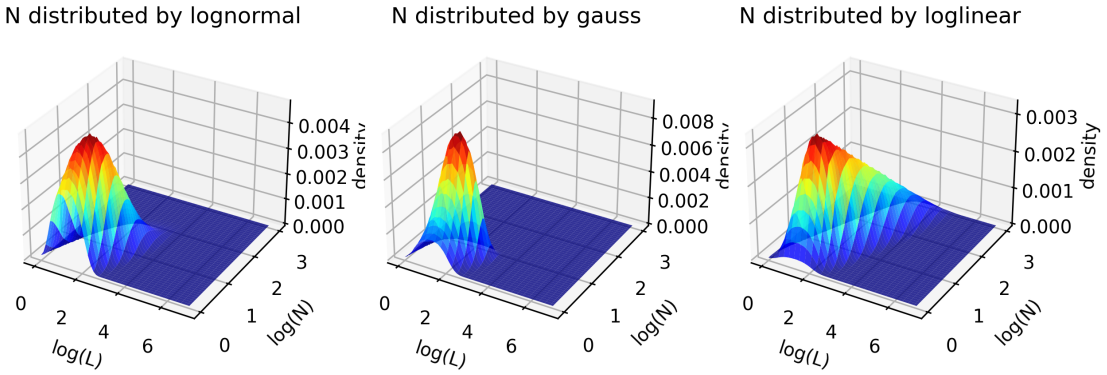

Model III distributed by gauss
3.0
2.5
2.0
1.5
1.0
0.5
0.0
log(N)
0
1
2
3
4
5
6
7
log(L)


Model III distributed by gauss
Density function by L
Survival function by L
Density function by N
Density
Probability
log(L)
log(L)
log(N)

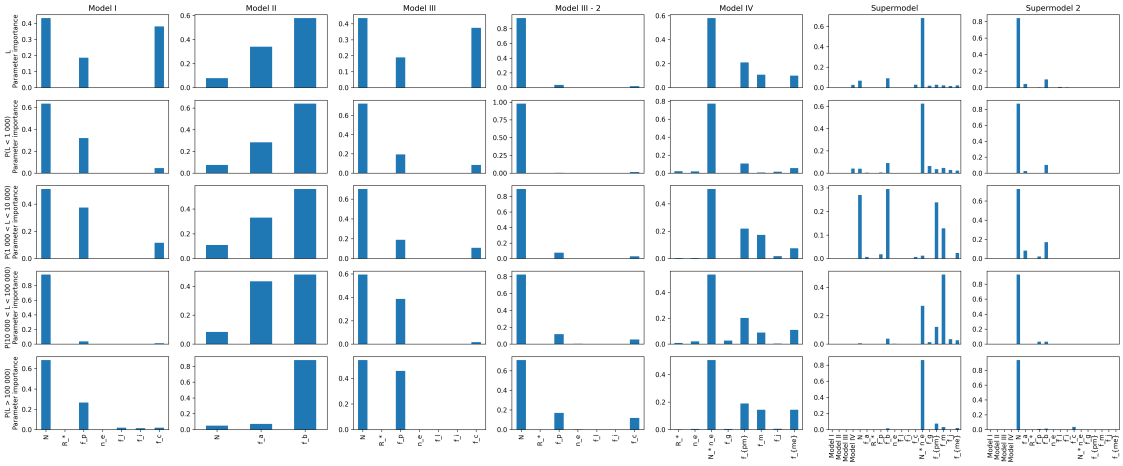

Model IV distributed by lognormal
density
log(L)
log(N)


Model IV distributed by lognormal
log(N)
log(L)


Model IV distributed by lognormal
Density function by L
Density
log(L)
Survival function by L
Probability
log(L)
Density function by N
Density
log(N)


N distributed by lognormal
N distributed by gauss
N distributed by loglinear
log(L)
log(N)
density

[9]:
```
modeli_napake()
```

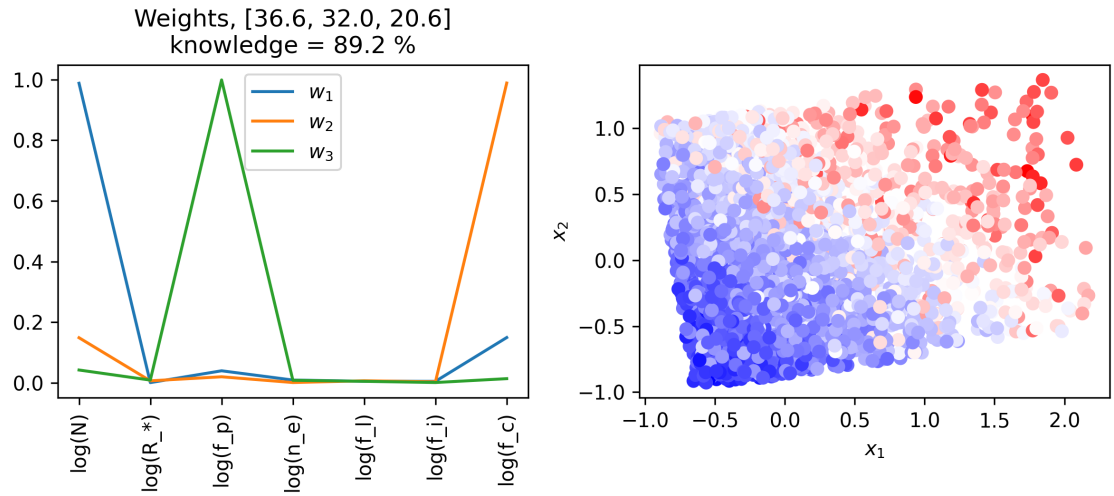


[10]:
```
data = pd.read_csv('collectedData/meti_tabela_csv.csv', index_col=0)
parameters = list(data.columns)[1:-1]
parameters_nomath = [p.replace("_pm", "_{pm}").replace("_me", "_{me}") if "("␣
 ↪in p else p.replace("_", " ") for p in parameters]
dataTotal = np.load('collectedData/meti_parameters.npy')
labelsTotal = np.load('collectedData/meti_labels.npy')
data = dataTotal[dataTotal[:, 0] == 1, 1:]
labels = labelsTotal[dataTotal[:, 0] == 1]
data2 = dataTotal[dataTotal[:, 0] == 2, 1:]
labels2 = labelsTotal[dataTotal[:, 0] == 2]
columns = [[4, 6, 7, 9, 10, 11, 12], [4, 5, 8], [4, 6, 7, 9, 10, 11, 12], [6,␣
 ↪9, 13, 14, 15, 16, 17, 18]]
podatki = [(data[data[:, 0] == 1, :][:, columns[0]], [parameters[c] for c in␣
 ↪columns[0]], labels[data[:, 0] == 1], 'Model I')]
podatki += [(data[data[:, 1] == 1, :][:, columns[1]], [parameters[c] for c in␣
 ↪columns[1]], labels[data[:, 1] == 1], 'Model II')]
podatki += [(data[data[:, 2] == 1, :][:, columns[2]], [parameters[c] for c in␣
 ↪columns[2]], labels[data[:, 2] == 1], 'Model III')]
podatki += [(data2[data2[:, 2] == 1, :][:, columns[2]], [parameters[c] for c in␣
 ↪columns[2]], labels2[data2[:, 2] == 1], 'Model III - 2')]
podatki += [(data[data[:, 3] == 1, :][:, columns[3]], [parameters[c] for c in␣
 ↪columns[3]], labels[data[:, 3] == 1], 'Model IV')]
```

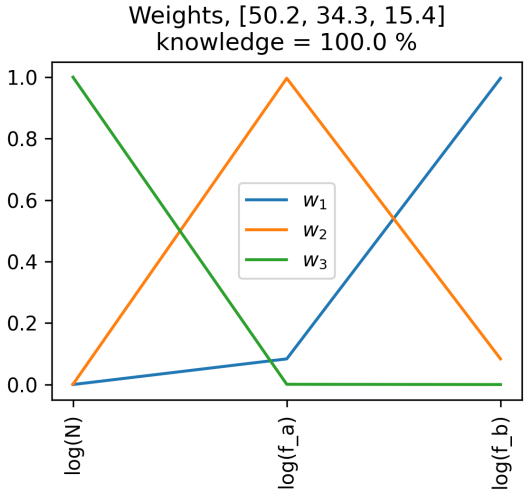
[11]:
```
for d, p, l, m in podatki:
    pca = PCA().fit(d)
    data = pca.transform(d)
    plt.figure(figsize=(8, 4), dpi=300, tight_layout=True), plt.suptitle(m)
    plt.subplot(121)
```

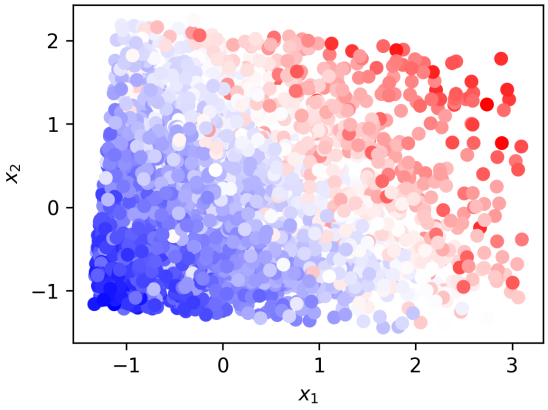
```
    plt.plot(np.abs(pca.components_.T[:, 0]), label="$w_1$")
    plt.plot(np.abs(pca.components_.T[:, 1]), label="$w_2$")
    plt.plot(np.abs(pca.components_.T[:, 2]), label="$w_3$")
    plt.xticks(list(range(len(p))), p, rotation=90)
    plt.title(f"Weights, {list(np.round(pca.explained_variance_[:3] / np.
 ↪sum(pca.explained_variance_) * 100, 1))}"
              f"\nknowledge = {(np.sum(pca.explained_variance_[:3]) / np.
 ↪sum(pca.explained_variance_)) * 100:.1f} %")
    plt.legend(loc="best")
    ax = plt.subplot(122)
    ax.scatter(data[:, 0], data[:, 1], c=l, cmap="bwr")
    plt.xlabel("$x_1$"), plt.ylabel("$x_2$")
plt.show()
```

Model I

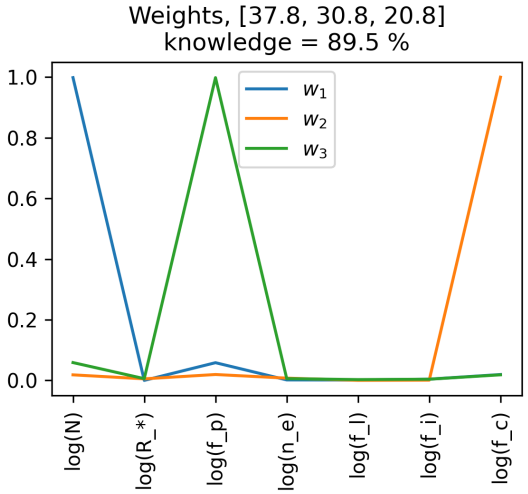


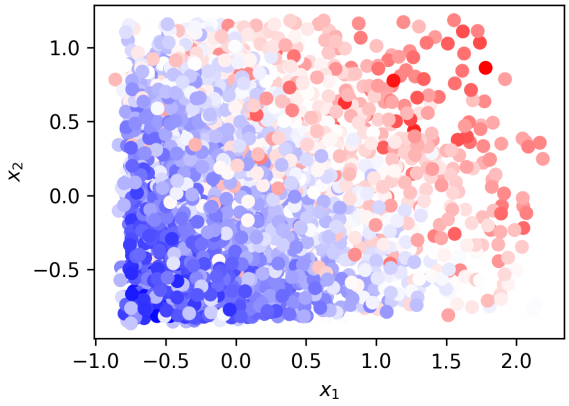
Model II



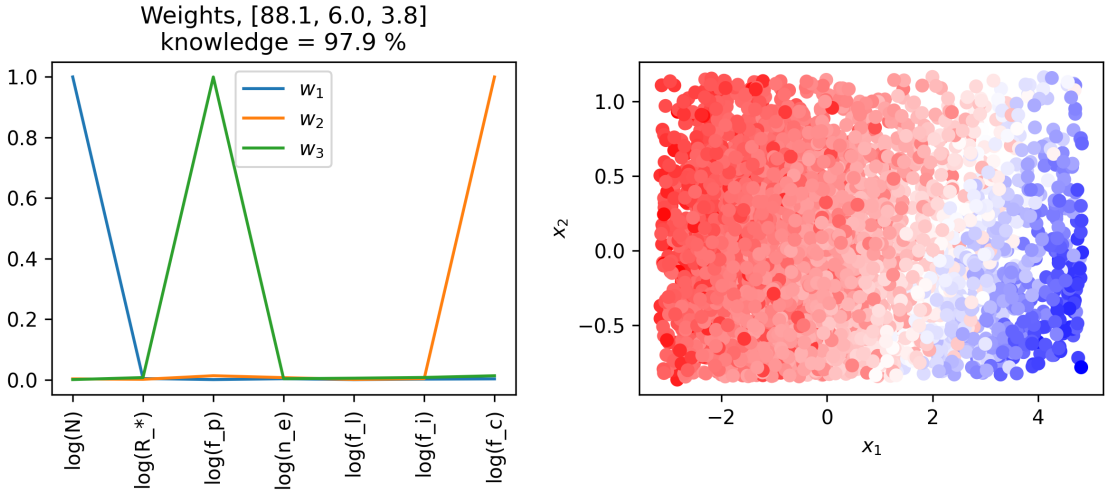


Model III





Model III - 2



Model IV