

A. Effects of the additional ppd feature layer

To test whether adding the angular resolution (*ppd*) to the latent code can model the non-linear effects correctly, we compare the performance of the architecture with or without the ppd feature layer. We denote the architectures as PU-CNN-Vanilla or PU-CNN-PPD for the case without or with the ppd feature layer. The means and standard errors of likelihoods for 5-fold cross validation are shown in Table 2. From Table 2 we can observe that introducing the additional ppd feature layer improves the performance. The reason for choosing the middle of the neural network for concatenating the ppd feature layer is two fold: Firstly, concatenating the ppd feature layer in the middle will only introduce a 2X2 feature layer, which will not increase the number of parameters to fit greatly. Secondly, there are three deconvolutional layers with ReLU activations that can provide enough capacity to model the non-linear effects of the angular resolution.

Method	Likelihood
PU-CNN-Vanilla	0.8556 ± 0.015
PU-CNN-PPD	0.8772 ± 0.016

Table 2. Effects of the additional ppd feature layer.