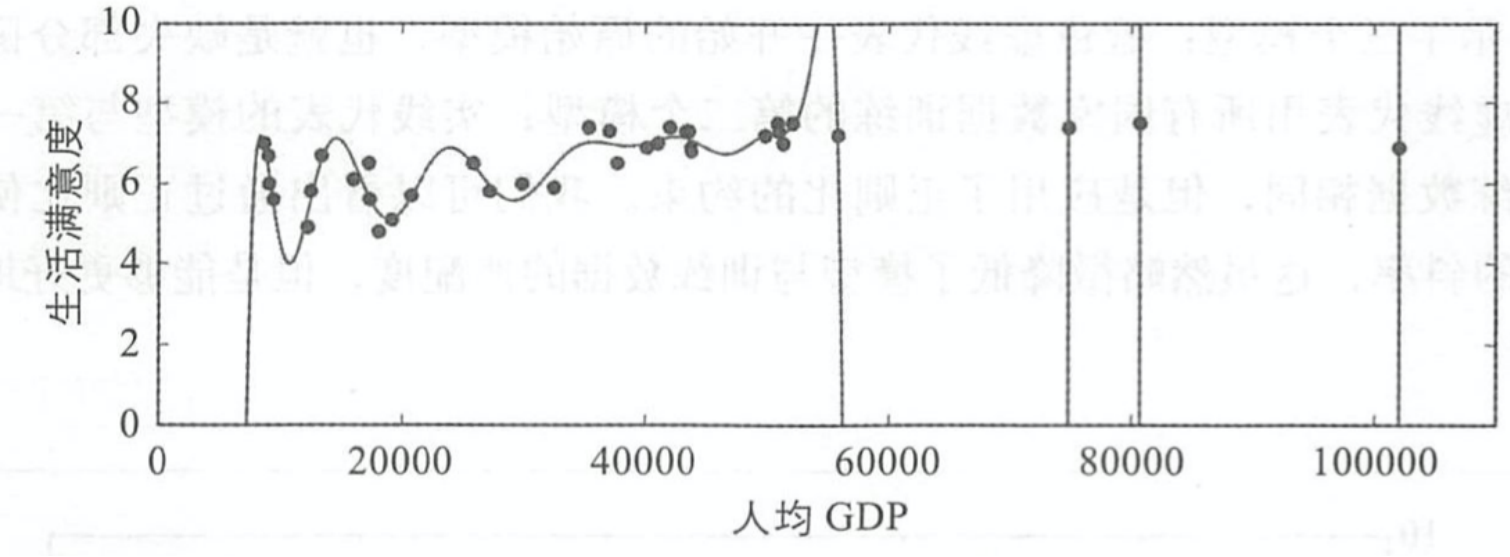
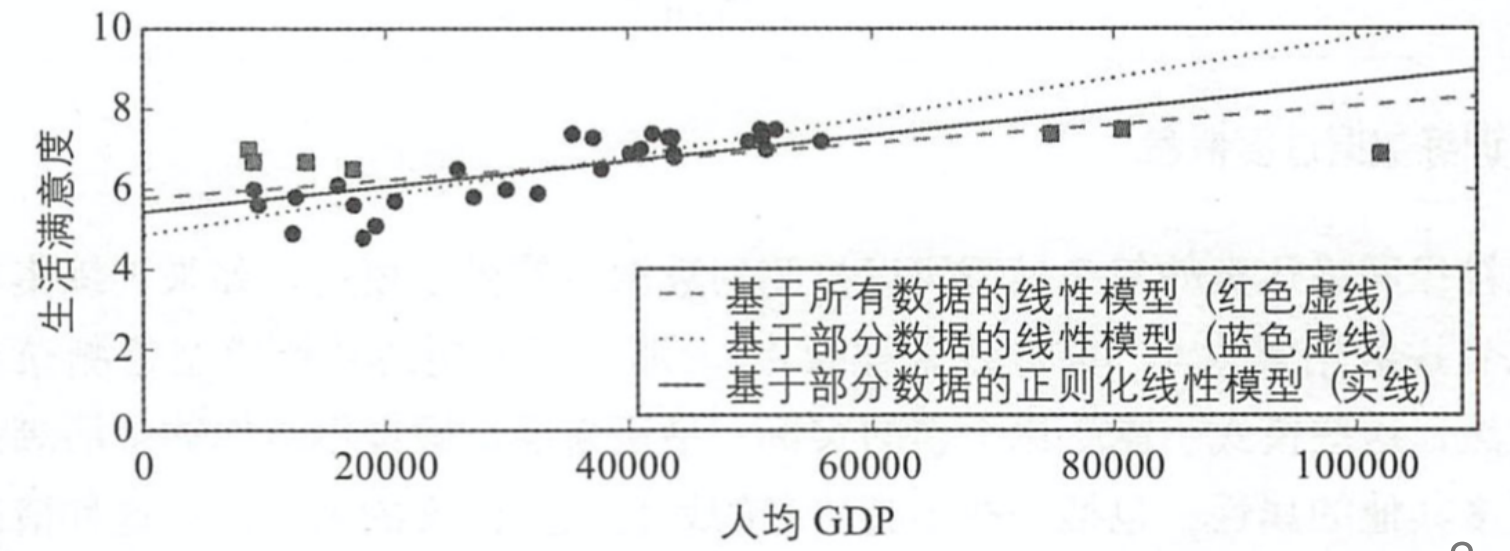
模型正则化

模型拥有的自由度越低，就越不容易过度拟合数据。

过拟合

（Overfitting）



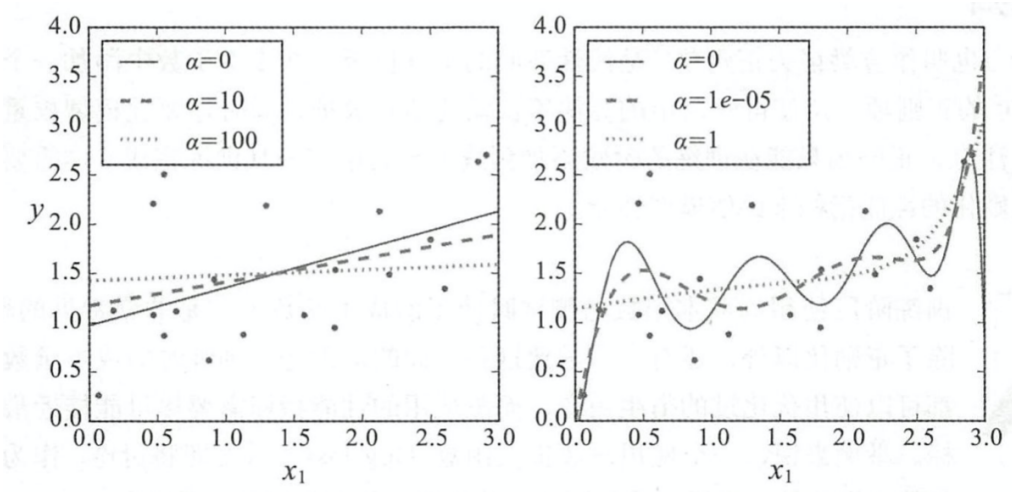


岭回归（Ridge Regression）

$$J(\theta) = \mathrm{MSE}(\theta) + \alpha \frac{1}{2} \sum_{i=1}^{n} \theta_i^2$$

- 超参数 α 控制模型的正则化程度。
- 数据标准化?

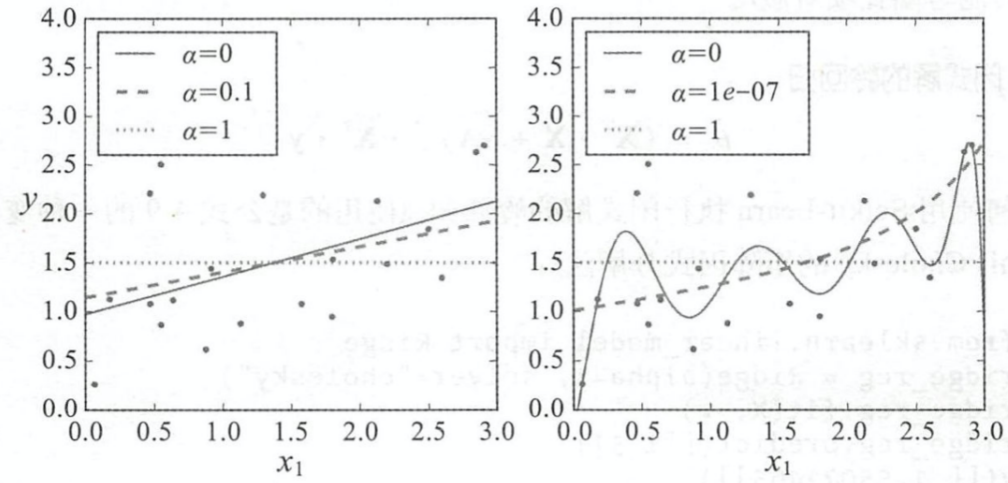
注意：正则化只能在训练的时候添加到成本函数中，一旦训练完成，你需要使用未经正则化的性能指标来评估模型性能。



套索回归（Least Absolute Shrinkage and Selection Operator Regression, Lasso)

$$J(\theta) = \mathrm{MSE}(\theta) + \alpha \sum_{i=1}^{n} |\theta_i|$$

- Lasso 回归最重要特点是倾向于完全消除最不重要 特征的权重。



弹性网络

$$J(\theta) = \mathrm{MSE}(\theta) + r\alpha \sum_{i=1}^{n} |\theta_i| + \frac{1-r}{2}\alpha \sum_{i=1}^{n} \theta_i^2$$

- 弹性网络是岭回归与 Lasso 回归的中间地带。
- 一般而言，弹性网络优于 Lasso 回归，因为当特征数量超过训练实例数量，又或者是几个特征强相关时，Lasso 回归的表现会不稳定。

Regularization for PWA

arXiv:1505.05133v1

The decay amplitudes are constructed

$$|\mathscr{M}(\vec{x}, m_{abc})|^2 = \sum_M \left| \sum_i a_i e^{i\phi_i} A_i^M(\vec{x}) \right|^2 + |a_{nr}|^2$$

J^P	m_X	Γ_X	M	j	l	Isobar mass	Isobar width	Isobar daughters
1^+	1.00	0.30	0	1	0	0.75	0.10	bc
2^-	1.45	0.25	0	2	0	1.10	0.15	bc
2^-	1.45	0.25	0	1	1	0.60	0.10	ab, ac
2^+	1.10	0.15	1	1	2	0.60	0.10	ab, ac
1^+	1.25	0.25	1	1	0	0.75	0.10	bc
non-resonant								abc

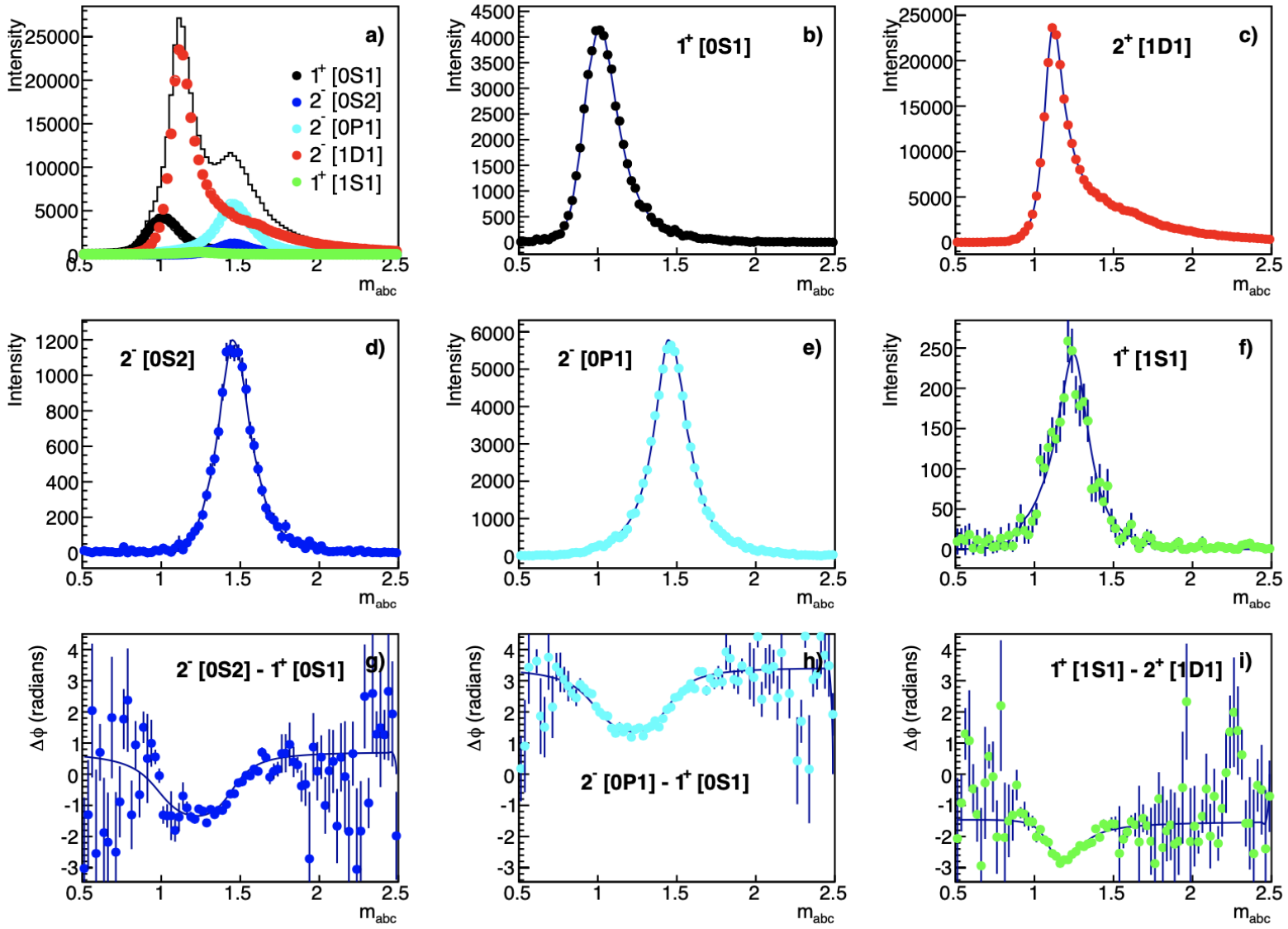


Figure 2. Results of extended maximum likelihood fit with the true p.d.f. from Table 1: a) - f) intensities for each resonant amplitude in the true p.d.f. labeled $J^P[Mlj]$ and g) - i) phase differences between each resonant wave and the corresponding reference wave with the same M. Closed symbols represent the measured intensities and phase differences, while the true model values are indicated by the solid curves.

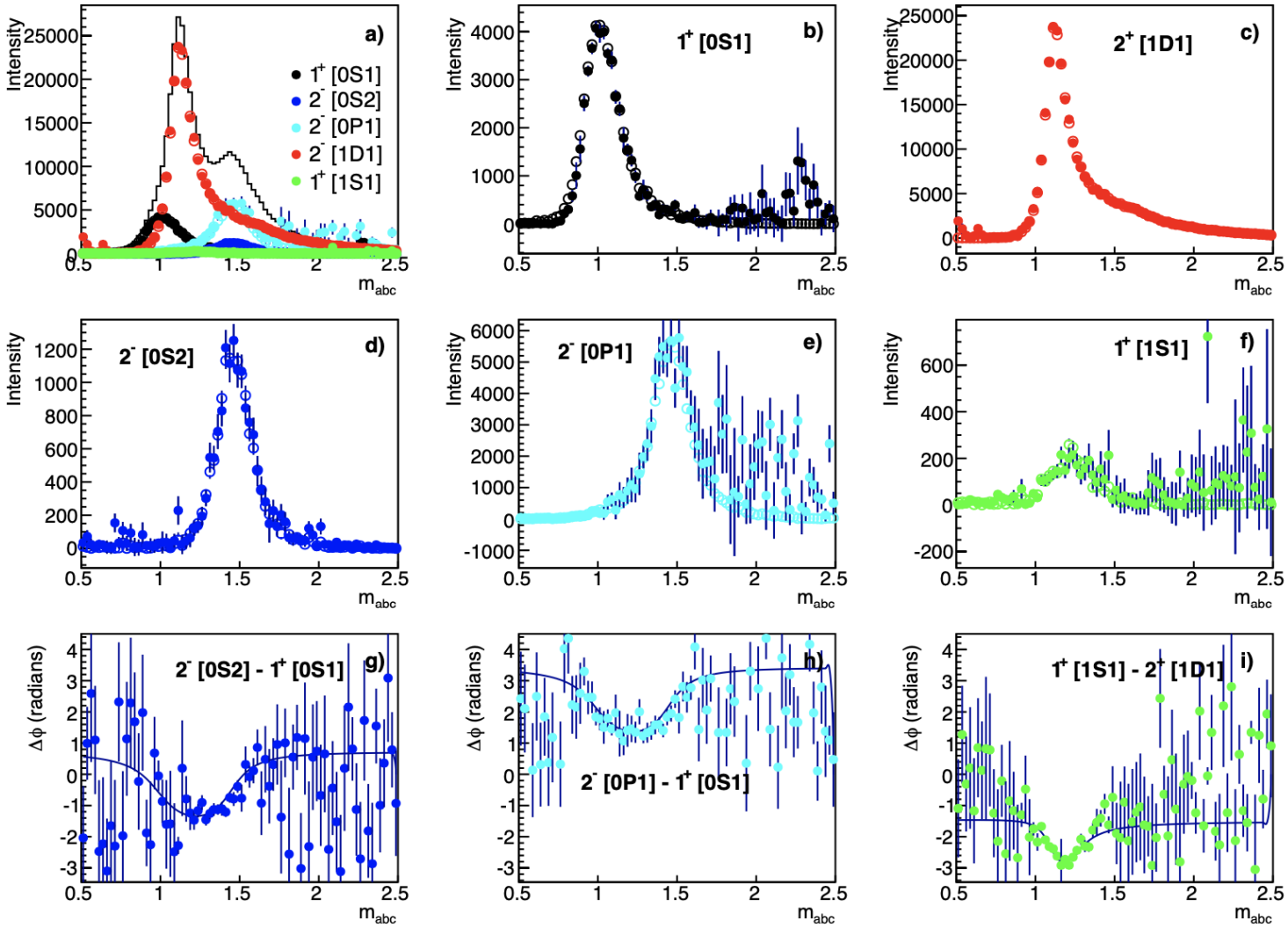


Figure 3. Results of extended maximum likelihood fit with the full p.d.f., including the additional 35 extraneous amplitudes, and no LASSO regularization (*i.e.* $\lambda = 0$): a) - f) intensities for each resonant amplitude in the true p.d.f. labeled $J^P[Mlj]$ and g) - i) phase differences between each resonant wave and the corresponding reference wave with the same M. Closed symbols represent the measured intensities and phase differences for the full p.d.f., open symbols represent the intensities determined from the true p.d.f. fit in Fig. 2, and the curves indicate the true model values.

Lasso for PWA

Applying the LASSO to this analysis involves augmenting the minimization quantity as follows

$$-2\log\mathscr{L} + \lambda\left[\sum_{i,M}\sqrt{\int |a_i e^{i\phi_i} A_i^M(\vec{x})|^2 d\vec{x}} + \sqrt{\int |a_{nr}|^2 \, d\vec{x}}\right]$$

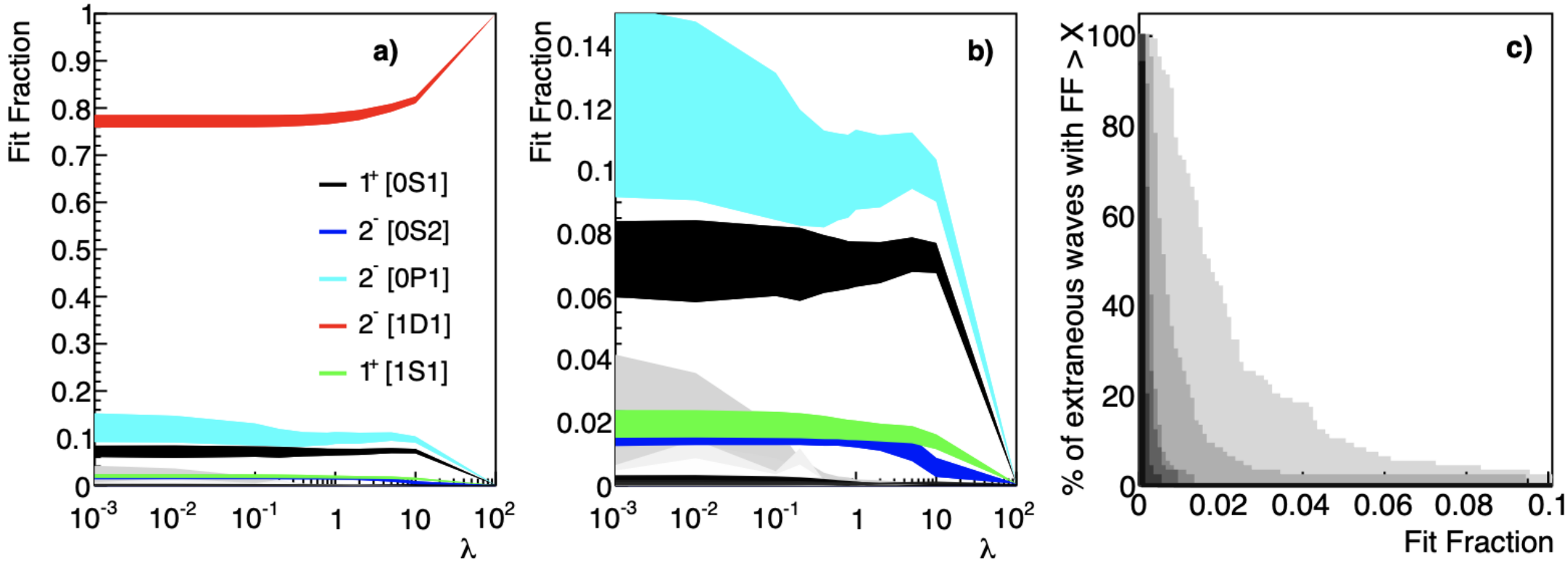


Figure 5. Results of fitting an ensemble of 100 data sets in a single mass bin with $m_{abc} = 1.25$. a) Fit fractions for each of the resonant amplitudes in the full p.d.f. as a function of λ, where the width of each band represents the standard deviation obtained from the 100 data samples in the ensemble. b) is the same as a), but with a zoomed y-axis to better display the growing contributions of the extraneous resonant amplitudes as $\lambda \to 0$. c) Percentage of extraneous resonant amplitudes with fit fraction $> X$ for $\lambda = 0, 0.1, 0.4, 0.8, 2, 10$ (increasing from light grey to black); this demonstrates the reduction in extraneous amplitudes contributing to the fit as λ is increased.

Information entropy

Akaike and Bayesian information criteria

$$\mathrm{AIC}(\lambda) = -2\log\mathscr{L} + 2r, \quad \mathrm{BIC}(\lambda) = -2\log\mathscr{L} + r\log n$$

In this case, we have chosen *r* to be the number of resonant amplitudes with a fit fraction larger than 10^{-3}.

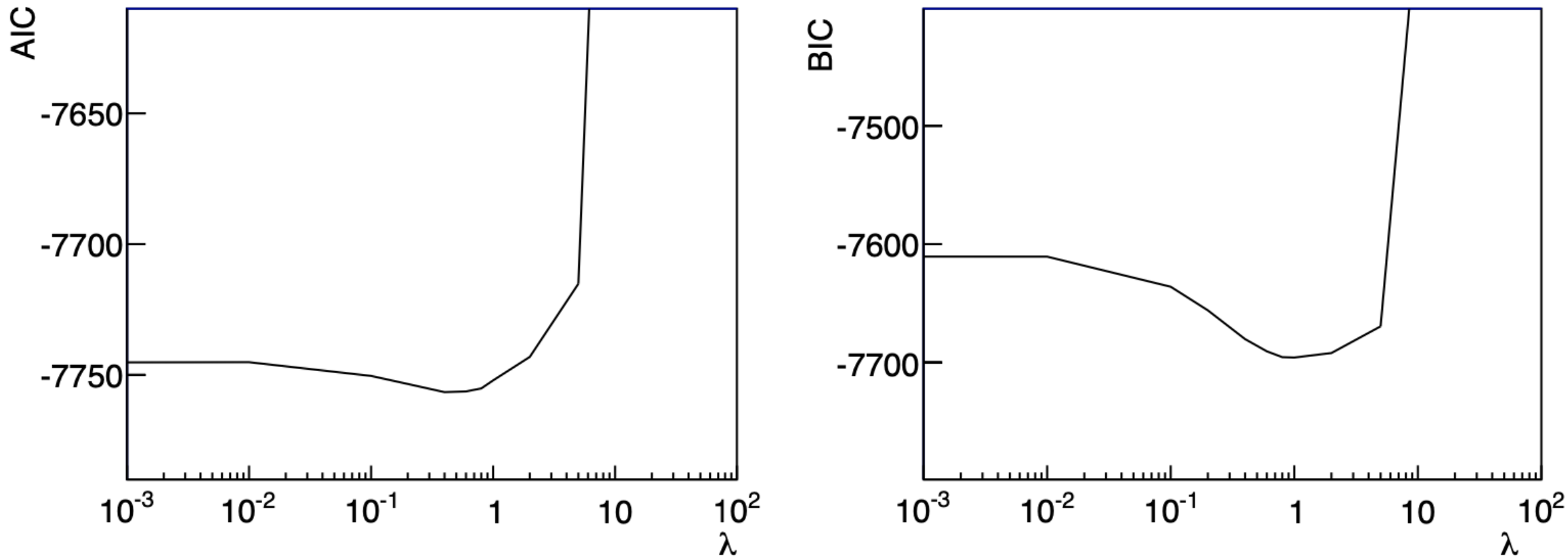


Figure 6. Average a) AIC and b) BIC as a function of λ for the ensemble of 100 independent samples with $m_{abc} = 1.25$.

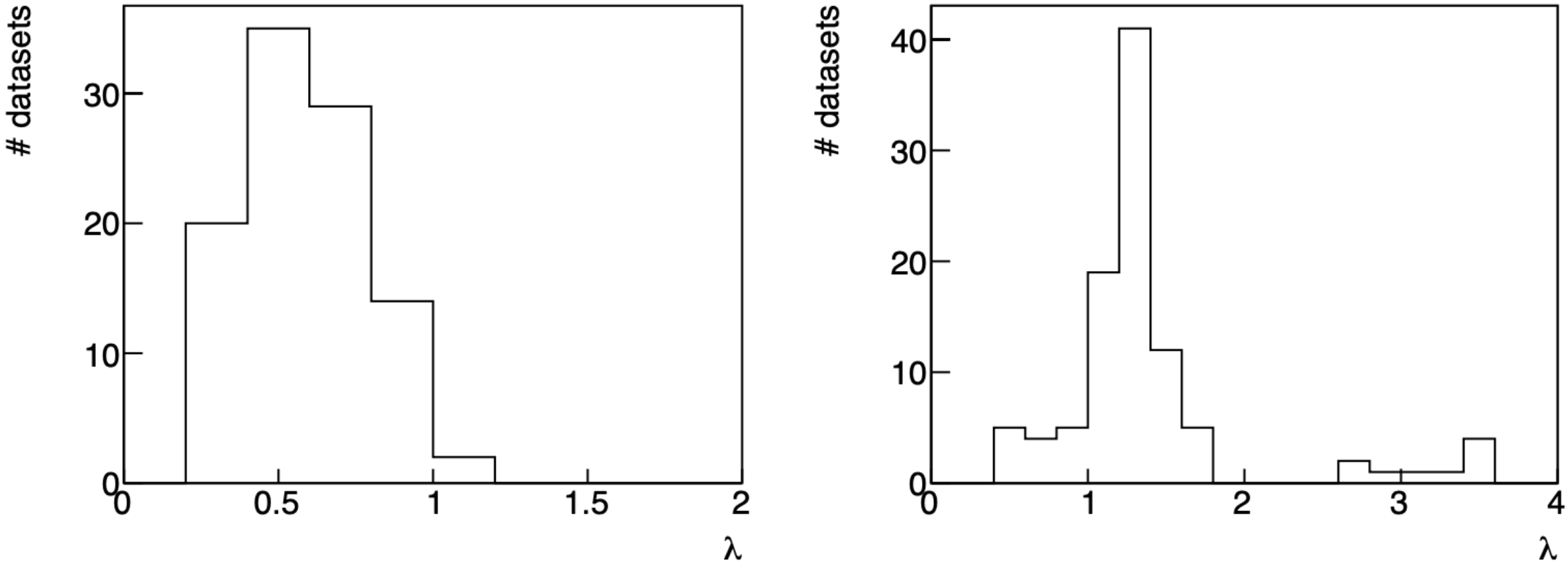


Figure 7. Distribution of λ values which minimize a) AIC and b) BIC for the ensemble of 100 independent samples with $m_{abc} = 1.25$.

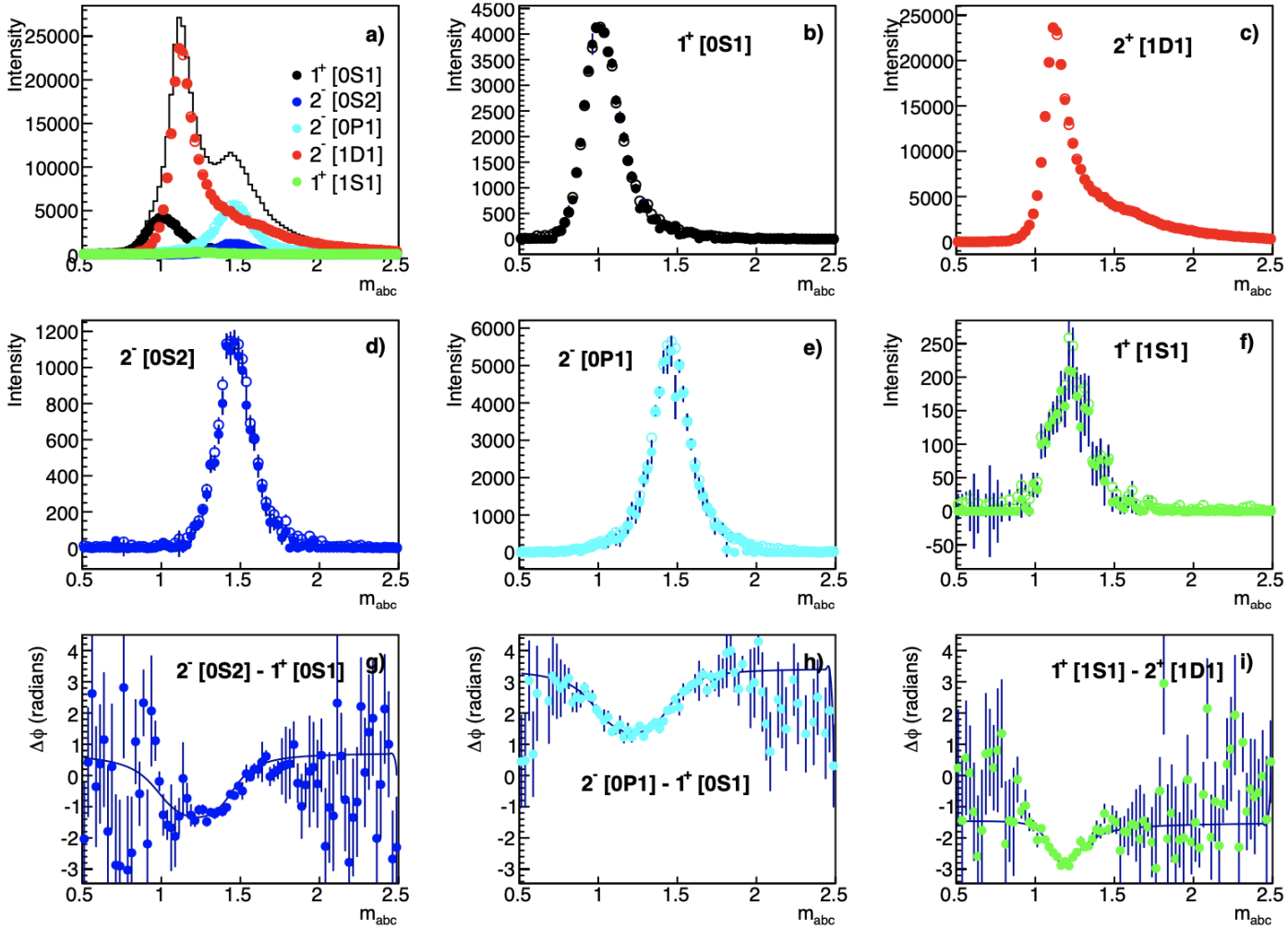


Figure 8. Results of extended maximum likelihood fit with the full p.d.f., including the additional 35 extraneous amplitudes, and using BIC to select the optimal λ for each m_{abc} bin: a) - f) intensities for each resonant amplitudes in the true p.d.f. labeled $J^P[Mlj]$ and g) - i) phase differences between each resonant wave and the corresponding reference wave with the same M. Closed symbols represent the measured intensities and phase differences for the full p.d.f. using the LASSO, open symbols represent the intensities determined from the true p.d.f. fit in Fig. 2, and the curves indicate the true model values. The agreement between the fit results using the true model p.d.f. *vs* the full model using BIC-LASSO is good enough that is difficult to see both results in these plots since they overlap.

$J^P[Mlj]$	Generated	True p.d.f.	No LASSO	BIC-LASSO
$1^+[0S1]$	41.74	41.78 ± 0.76	40.37 ± 1.73	41.83 ± 0.86
$2^-[0S2]$	10.55	10.12 ± 0.52	10.58 ± 0.96	10.11 ± 0.71
$2^-[0P1]$	49.21	48.67 ± 0.91	53.40 ± 5.87	48.04 ± 2.62
$2^-[1D1]$	86.58	86.64 ± 0.40	86.47 ± 1.01	86.5 ± 0.77
$1^+[1S1]$	1.77	1.90 ± 0.23	1.62 ± 0.42	1.60 ± 0.40

Table 2. Mean and standard deviation of the measured fit fractions for Model I.