Hierarchies of Models:
Toward Understanding Planetary Nebulae

Kevin H. Knuth[1], Arsen R. Hajian[2]

[1]*NASA Ames Research Center, Computational Sciences Department, Code IC*
Moffett Field CA 94035 USA
[2]*United States Naval Observatory, Department of Astrometry,*
Washington DC 20392 USA

Abstract. Stars like our sun (initial masses between 0.8 to 8 solar masses) end their lives as swollen red giants surrounded by cool extended atmospheres. The nuclear reactions in their cores create carbon, nitrogen and oxygen, which are transported by convection to the outer envelope of the stellar atmosphere. As the star finally collapses to become a white dwarf, this envelope is expelled from the star to form a planetary nebula (PN) rich in organic molecules. The physics, dynamics, and chemistry of these nebulae are poorly understood and have implications not only for our understanding of the stellar life cycle but also for organic astrochemistry and the creation of prebiotic molecules in interstellar space.

We are working toward generating three-dimensional models of planetary nebulae (PNe), which include the size, orientation, shape, expansion rate and mass distribution of the nebula. Such a reconstruction of a PN is a challenging problem for several reasons. First, the data consist of images obtained over time from the Hubble Space Telescope (HST) and spectra obtained from Kitt Peak National Observatory (KPNO) and Cerro Tololo Inter-American Observatory (CTIO). These images are of course taken from a single viewpoint in space, which amounts to a very challenging tomographic reconstruction. Second, the fact that we have two disparate and orthogonal data types requires that we utilize a method that allows these data to be used together to obtain a solution. To address these first two challenges we employ Bayesian model estimation using a parameterized physical model that incorporates much prior information about the known physics of the PN.

In our previous works we have found that the forward problem of the comprehensive model is extremely time consuming. To address this challenge, we explore the use of a set of hierarchical models, which allow us to estimate increasingly more detailed sets of model parameters. These hierarchical models of increasing complexity are akin to scientific theories of increasing sophistication, with each new model/theory being a refinement of a previous one by either incorporating additional prior information or by introducing a new set of parameters to model an entirely new phenomenon. We apply these models to both a simulated and a real ellipsoidal PN to initially estimate the position, angular size, and orientation of the nebula as a two-dimensional object and use these estimates to later examine its three-dimensional properties. The efficiency/accuracy tradeoffs of the techniques are studied to determine the advantages and disadvantages of employing a set of hierarchical models over a single comprehensive model.

INTRODUCTION

We are only beginning to understand the importance of the later stages of a star's existence. Stars with initial masses between 0.8 and 8 solar masses end their lives as swollen red giants on the asymptotic giant branch (AGB) with degenerate carbon-oxygen cores surrounded by a cool extended outer atmosphere. Convection in the outer atmosphere dredges up elemental carbon and oxygen from the deep interior and brings it to the surface where it is ejected in the stellar winds. As the star ages, the

core eventually runs out of fuel and the star begins to collapse. During this collapse, much of the outer envelope is expelled from the core and detaches from the star forming what is called a planetary nebula (PN) and leaving behind a remnant white dwarf. Despite the wealth of observations the physics and dynamics governing this expulsion of gas are poorly understood making this one of the most mysterious stages of stellar evolution (Maddox, 1995; Bobrowsky et al., 1998).

The carbon and oxygen ejected in the stellar wind and expelled with the PN during the star's collapse are the major sources of carbon and oxygen in the interstellar medium (Henning & Salama, 1998). It is now understood that complex organics, such as polycyclic aromatic hydrocarbons (PAHs) (Allamandola et al., 1985), readily form in these environments (Wooden et al., 1986; Barker et al. 1986). Thus the formation, evolution and environment of PNe have important implications not only for our understanding of the stellar life cycle but also for organic astrochemistry and the creation of prebiotic molecules in interstellar space. In addition, this material will eventually be recycled to form next-generation stars whose properties will depend on its composition.

To better understand the chemical environment of the PN, we need to understand its density distribution as a function of position and velocity. However, without knowledge of the distances to planetary nebulae (PNe), it is impossible to estimate the energies, masses, and volumes involved. This makes knowledge of PN distances one of the major impasses to understanding PN formation and evolution (Terzian, 1993).

More recently, detection of the expansion parallax has been demonstrated to be an important distance estimation technique. It requires dividing the Doppler expansion velocity of the PN, obtained from long-slit spectroscopy, by the angular expansion rate of the nebula, measured by comparing two images separated by a time baseline of several years. Two epochs of images of PNe were obtained from the Very Large Array (VLA) with a time baseline of about 6 years, and have resulted in increasingly

NGC 3242
Balick, Hajian,
Terzian, Perinotto, Patriarchi
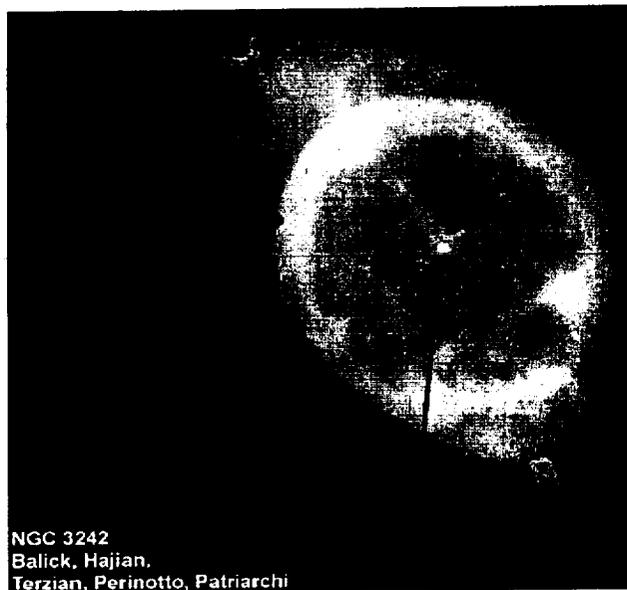
FIGURE 1. A Hubble Space Telescope (HST) image of NGC 3242 (Balick, Hajian, Terzian, Perinotto, Patriarchi) illustrating the structure of an ellipsoidal planetary nebula.

reliable distance estimates to 7 nebulae (Hajian et al., 1993; Hajian & Terzian 1995, 1996). However, successfully application of th s technique requires that one relate the radial Doppler expansion rate to the observed tangential expansion. This is straightforward for spherical nebulae, but for the most part distances to PNe with complex morphologies remain inaccessible. More recently using images from the Hubble Space Telescope (HST), distance estimates to 5 more nebulae have been obtained. Using two techniques, the magnification method and the gradient method, Palen et al. (2002) resolved distances to 3 PNe and put bounds on another. Reed et al. (1999) estimated the distance to a complex nebula (NGC 6543) by identifying bright features and relying on a on a heuristic model of the structure of the nebula derived from ground-based images and detailed long-slit spectroscopy (Miranda & Solf, 1992). This work emphasized the utility of the model-based approach to reconciling the measured radial expansion velocities to the observed tangential angular motions.

To accommodate complex PNe, we have adopted the approach of utilizing an analytic model of the nebular morphology, which takes into account the physics of ionization equilibrium and parameters describing the density distribution of the nebular gas, the dimensions of the nebula, its expansion rate, and its distance from earth. Bayesian estimation of the model parameter values is then performed using data consisting of images from the Wide Field Planetary Camera (WFPC2) on the HST and long-slit spectra from the 4m telescopes at Kitt Peak National Observatory (KPNO) and Cerro Tololo Interamerican Observatory (CTIO). In our preliminary work (Hajian & Knuth, 2001) we have demonstrated feasibility of this approach by adopting a model describing the ionization boundary of a PN based on an assumed prolate ellipsoidal shell (PES) of gas – the ionization-bounded PES model (IBPES) (Aaquist & Kwok, 1996; Zhang & Kwok, 1998). One of the difficulties we have encountered is the fact that the forward computations of the complete IBPES model are extremely time consuming. For this reason we have been investigating the utility of adopting a hierarchical set of models, where each successive model captures a new feature of the nebula neglected by the previous model.

A HIERARCHICAL SET OF MODELS

The inspiration of utilizing a finite hierarchical set of models comes in part from the process of scientific advancement itself where each new theory, viewed as a model of a given physical object or process, must explain the phenomena explained by the previous theories in addition to describing previously unexplainable phenomena. The apparent utility of such a process is rooted in fact that hierarchical organization is a very efficient means to constructing a system of great complexity. In this application we consider a series of three models approaching the uniform ellipsoidal shell model (UES) of an ellipsoidal PN, which describes the PN as an ellipsoidal shell of gas of uniform density.

The purpose of the first model is to perform a relatively trivial task - discover the center of the PN in the image. The second model is designed to discover the extent,

eccentricity and orientation of the PN. Finally the third model works to estimate the thickness of the ellipsoidal shell. Each of these models treats the image of the nebula as a two-dimensional object, which drastically minimizes the computational burden imposed by working with a three-dimensional model. As these models approach the three-dimensional UES model they grow in complexity with increasing numbers of parameters. Several of these parameters are of course nuisance parameters of relevance only to that specific model and necessary only to enable one to perform the forward computations of creating an image of the nebula from hypothesized model parameter values. As the models grow in complexity, the forward computations become more time consuming. However, as some of the parameters have been well-estimated by the previous models, both the dimension and the volume of the hypothesis space to be searched grows smaller relative to the total hypothesis space of the current model thus reducing the effort needed to approach the solution.

Methodology

The parameters for each of the three models to be presented were estimated by maximizing the posterior probability found simply by assigning a Gaussian likelihood and uniform priors. To enable comparison of the models rather than the techniques used to find an optimal solution, gradient ascent was used in each case to locate the maximum a posteriori (MAP) solution. Stopping criteria were defined so that if the change in each of the parameter values from the previous iteration to the present were less than a predefined threshold the iterations would terminate. The thresholds typically became more stringent for the more advanced models. This is because highly refined estimates obtained from a primitive model do not necessarily correspond to higher probable solutions for a more advanced model.

Discovering the Center

Discovering the center of the PN is a straightforward task. Many quick-and-dirty solutions present themselves, with perhaps the most obvious being the calculation of the center of mass of the intensity of the image This can typically place the center to within several pixels in a 500x500 image. However, several confounding effects can limit the accuracy of this estimate. First, the entire image is not used in the analysis. The central star and its diffraction spikes are masked out so that those pixels are not used. Asymmetric placement of the mask with respect to the center of the nebula can dramatically affect estimation of the center of mass. In addition, by not masking the central star and diffraction spikes similar problems can occur as these high intensity pixels are rarely symmetric. Furthermore, it is not assured that the star is situated in the center of the nebula. A second problem is that the illumination of the nebula may not be symmetric, and third the nebula itself might not be symmetric. As we are currently focusing our efforts on well-defined ellipsoidal PNe, these two issues are less relevant than the first.
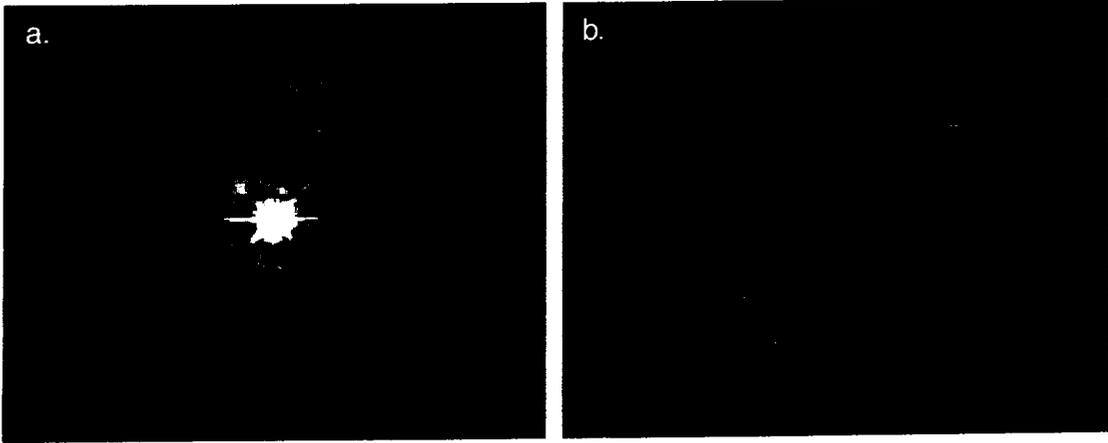
FIGURE 2. a. The planetary nebula IC 418 (Sahai, Trauger, Hajian, Terzian, Balick, Bond, Panagia, Hubble Heritage Team). b. The masked image ready for analysis. Note that the regions outside the nebula are not masked, as they are as important for determining the extent of the nebula as the image of the nebula itself.

For this reason, we adopted a simple two-dimensional circular Gaussian distribution as a model of the two-dimensional image of the nebular intensity.

$$G(x, y) = I_o \, Exp\left[-\frac{(x - xo)^2 + (y - yo)^2}{2\sigma^2} \right],$$ (1)

where I_o is the overall intensity parameter, σ is the overall extent of the PN, and (xo, yo) are the coordinates of the center of the nebula in the image. While the fall-off of the PN intensity is not Gaussian, the symmetry of the nebula and the symmetry of the Gaussian work in concert to allow adequate estimation of the PN center. In practice this technique was acceptable, however it was found that the Gaussian distribution could shift to try to hide some of its mass in masked out areas of the image. This was especially noticeable for nebulae asymmetrically situated in the image so that the edge of the nebula was close to the edge of the image. In this case, it was found that the estimate of the center could be off by a few pixels.

As this is the first stage, we did not work to develop a more sophisticated model for center estimation, although such a model will probably be useful to find the centers of more complex non-ellipsoidal PNe. Rather, the center estimates are refined by the next model, which is designed to better describe the intensity distribution.

In summary, four parameters are estimated by the Gaussian model (Gauss): the center (xo, yo), the general extent σ, and the overall intensity I_o.

Discovering the Extent, Eccentricity and Orientation

To determine the extent, eccentricity and orientation of the PNe, we must adopt a more realistic model. To first-order the ellipsoidal PNe look to be ellipsoidal patches of light, for this reason we utilized a two-dimensional sigmoidal hat function defined by

$$S(x,y) = I_o \left(1 - \frac{1}{1 + Exp\left[-a(r(x,y)-1)\right]} \right) \qquad (2)$$

where

$$r(x,y) = \left(C_{xx}(x-xo)^2 + 2C_{xy}(x-xo)(y-yo) + C_{yy}(y-yo)^2\right)^{\frac{1}{2}} \qquad (3)$$

and

$$C_{xx} = \frac{\cos^2\theta}{\sigma_x^2} + \frac{\sin^2\theta}{\sigma_y^2}$$

$$C_{xy} = (\sigma_y^{-2} - \sigma_x^{-2})\sin\theta\,\cos\theta \qquad (4)$$

$$C_{yy} = \frac{\sin^2\theta}{\sigma_x^2} + \frac{\cos^2\theta}{\sigma_y^2}$$

where I_o is the overall intensity parameter, a is the intensity falloff at the edge of the PN, σ_x and σ_y are extents of the PN along the minor and major axes, θ is the orientation of the PN in the image and (xo, yo) are the coordinates of its center. Thus three new parameters are estimated by the sigmoidal hat model (SigHat), and in addition the three old parameters are refined.

Figure 3a shows the intensity profile of SigHat characterized by its relative uniform intensity across the nebula with a continuously differentiable falloff. The falloff region allows the model to accommodate variability in location of the outer edge of the PN in addition to aiding the gradient ascent method used to locate the optimal solution. Given initial estimates of the PN center and general extent, the algorithm was able to identify these parameters with relative ease.
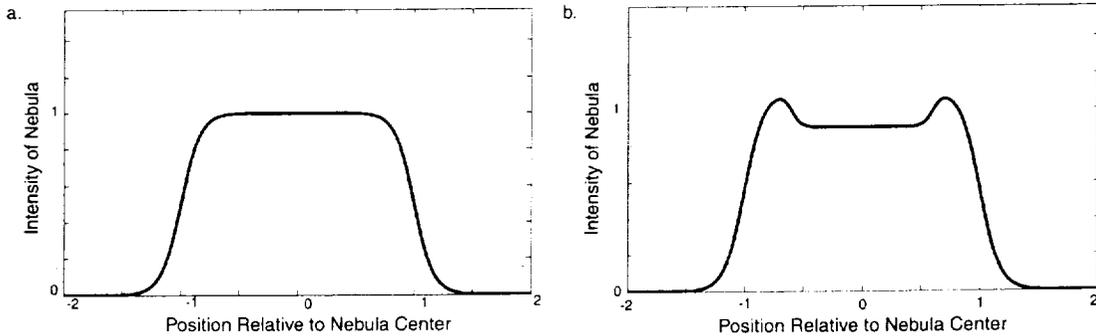
FIGURE 3. a. Intensity profile of the sigmoid hat function (2) used to estimate extent, eccentricity and orientation. b. Intensity profile of the dual sigmoid hat function (5) used to estimate the thickness of the gaseous shell.

Discovering the Thickness

The effect of imaging a three-dimensional ellipsoidal shell of gas is to produce an ellipsoidal patch surrounded by a ring of higher intensity. To capture the thickness of the shell without resorting to an expensive three-dimensional model, we model the

image as the difference of two sigmoidal hat functions with the thickness of the shell being estimated as the difference in extent of the two functions

$$T(x,y) = I_+ S_+(x,y) - I_- S_-(x,y) \qquad (5)$$

where $S_+(x,y)$ and $S_-(x,y)$ are the sigmoidal hat functions in (2), expect each has its own falloff parameter a_+, a_- and the extents are related by the thickness ratio Δ

$$\sigma_{x-} = \Delta \cdot \sigma_{x+}$$
$$\sigma_{y-} = \Delta \cdot \sigma_{y+} \qquad (6)$$

We call this model the dual sigmoidal hat model (SigHat2). A typical profile is shown in Figure 3b.

At this point the center, orientation, and extent parameters were taken to be well-estimated and the focus was on determining the thickness ratio Δ and estimating the nuisance parameters I_+, I_-, a_+, and a_-. During the course of our investigation, we found that the estimation of I_+, I_- proved to be rather difficult with either highly oscillatory steps or very slow convergence. Investigation of the landscape of the hypothesis space proved to be quite informative; as it was found that the MAP solution was a top peak of a long narrow ridge. This finding led us to employ a transformation from the parameters I_+, I_- to

$$I_a = I_+ + I_-$$
$$I_b = I_+ - I_- \qquad (7)$$

so that

$$T(x,y) = \frac{1}{2}(I_a + I_b) S_+(x,y) - \frac{1}{2}(I_a - I_b) S_-(x,y). \qquad (8)$$

With this reparameterization, the hypothesis space is transformed so that the highly probable regions are not as long and narrow. This was found to aid convergence eliminating the oscillatory steps and allowing the solution to converge more quickly to the higher probability regions. SigHat2 estimates only five parameters, the nuisance parameters I_a, I_b, a_+, a_-, and the thickness Δ.

PERFORMANCE

There are three aspects important to determining the degree to which performance has been improved by taking this hierarchical approach. First, it is expected that the speed at which optimal estimates can be obtained would be increased. Second, we might expect that the increase in speed comes at the cost of accuracy, however this accuracy could presumably be regained by applying the ultimate model for a minimal number of additional iterations. Third, by employing a set of hierarchical models we can rule out regions of the hypothesis space that are irrelevant and avoid the difficulties of local maxima. This aspect is extremely important in complex estimation tasks where the hypothesis space may be riddled with local maxima. Due to the high-

dimensionality of the spaces involved, the ex stence, number and location of these local maxima is almost impossible to demons rate explicitly. However, we expect that the set of models applied hierarchically will result in fewer occurrences of non-optimal solutions than the ultimate model applied alone.

Evaluation Methodology

The same method to obtain an optimal estimate, gradient ascent, was used for each model to assure that the utility of the models themselves were being compared rather than the optimization technique. All code was written and executed in Matlab 6.1 Release 12.1 and run on the same machine (Dell Dimension 8200, Windows 2000, Pentium 4, 1.9 GHz, 512K RAM).

The models were tested on four synthetic PN images (350 x 400 pixels) constructed using the UES model. Figure 1a shows one such synthetic data set (Case 1). Figures 1b, c, and d show the three results from the models Gauss, SigHat and SigHat2 respectively. Note that Gauss has located the center of the PN and its general extent. SigHat has effectively captured its eccentricity, orientation and the extent of the projections of its major and minor axes. Finally SigHat2 has made an estimate of the thickness of the gaseous shell. This estimate however is not as well defined as the others due to fact that the meaning of the shell thickness in the UES model is qualitatively different than the thickness in the SigHat2 model. One can look at progressing from SigHat2 to UES as a paradign shift, which will ultimately result in a much better description of the bright ring in the image.
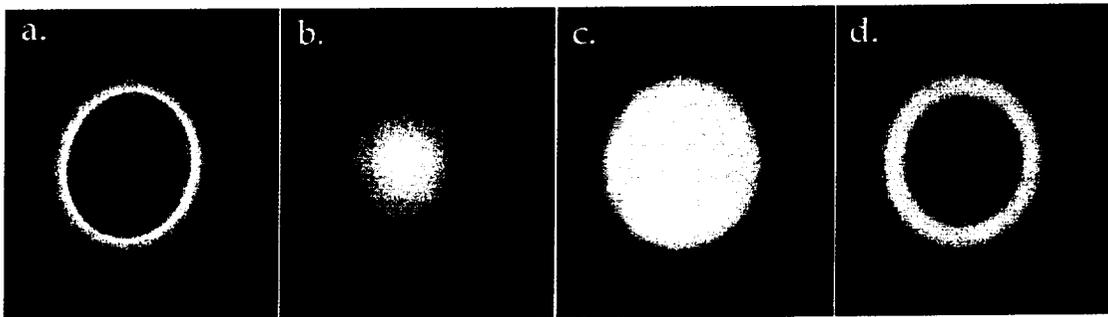
FIGURE 5. a. Synthetic image of PN made from parameterized UES model. b. Gaussian model used to discover center of the PN. c. Sigmoid hat model capturing extent, eccentricity and orientation. d. Dual sigmoid hat model estimating the thickness of the nebular shell. Note that as the dual sigmoid hat model and the UES model implement thickness differently the match cannot be perfect.

Rates of Convergence

As expected the amount of time required to complete an iteration of the gradient ascent step varied from one model to the next Gauss required an average of 6.76 s/iteration, whereas SigHat required an average of 14.74 s/iteration, and SigHat2 required an average of 12.85 s/iteration. Although the SigHat2 is more complex than SigHat, fewer parameters are being updated, as the center position, extent, eccentricity, and orientation are assumed to be well estimated and are held constant.

In contrast, one step of the UES model used to generate the synthetic images requires on the order of one half hour of time under identical circumstances for a single iteration depending on the spatial extent of the PN in the image.

TABLE 1. Iterations Required for Convergence				
Trial	Gauss	SigHat	SigHat2	SigHat Alone
1	20	14	16	42
2	21	21	17	39
3	24	50	7	X
4	36	36	13	61
Avg Iters	25.67	23.67	15.33	47.33
Avg Time	173.83 s	350.33 s	197.62 s	699.51 s

Table 1 shows the number of iterations required for each model to sufficiently converge for the four cases considered. The model SigHat was started using as initial conditions those estimated by Gauss, and similarly for SigHat2, which followed SigHat. In addition, we tested SigHat alone without the aid of Gauss to determine whether the hierarchical progression actually improved the rate of convergence. Case 3 proved to be difficult due to the object's small size in the image and the specific combination of its orientation and eccentricity. We found that SigHat alone was unable to obtain a solution. For this reason the averages at the bottom of the table reflect only the three cases where all algorithms were successful. In each case SigHat took longer to converge when applied alone than when it was preceded by Gauss, with an average of 699.51s as compared to 524.16s for the sum of Gauss and SigHat.

Goodness of Fit

The hierarchical application of the models also improved the accuracy of the estimates as can be seen in Table 2 which shows the goodness of fit measured by $-\log(likelihood)$, where smaller numbers correlate with higher probability solutions. Note that comparisons across trials are meaningless as the $\log(likelihood)$ is not normalized and is dependent on the extent of the object in the image. This is evident in case 3 where the fit was relatively poor and the object's extent was small with respect to the dimension of the image. Most important is the comparison between the results for SigHat and SigHat Alone. In all three cases, the goodness of fit for SigHat run alone was worse than that for SigHat when preceded by Gauss. This demonstrates that not only is it faster to apply the models hierarchically, but the results obtained better describe the data.

Throughout the course of these experiments it was found that local maxima do exist in the hypothesis space and that the models can become stuck. This was even more problematic when applied to real images. For example, the SigHat model with its limited extent can easily become attached to the high intensity regions in the shells of

TABLE 2. Goodness of Fit as measured by: - log(*likelihood*)				
Case	Gauss	SigHat	SigHat2	SigHat Alone
1	5029	1868	751	2332
2	7024	2055	1137	2790
3	1485	205	423	X
4	4343	317	244	340

PNe that possess sufficient inclination to produce the effect. For example consider the high intensity region in the limb of IC418 near the top edge of the picture in Figure 6a). SigHat can become trapped covering this high-intensity region. Local maxima are especially a problem for SigHat2, which can hide in a dark region outside the PN by making itself invisible, i.e. equating I_+ and I_- while minimizing the shell thickness. Another interesting hiding behavior was observed with the SigHat model, which could settle inside the central masked region of Figure 6a. We have found that this misbehavior is avoided by first capturing the center and general extent with Gauss. Figure 6 below shows the results of applying the hierarchy of models to IC418.
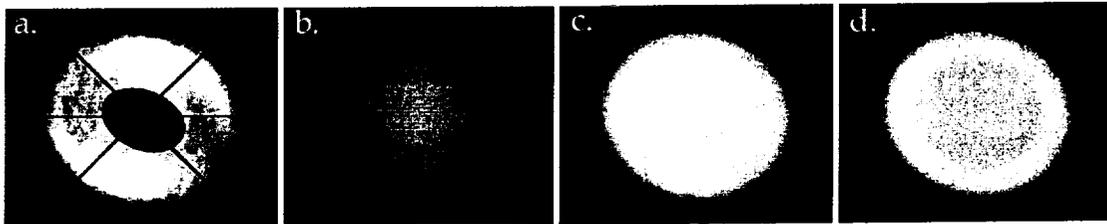
FIGURE 6. a. IC418 masked for analysis. b. Gauss is used to discover the center and general extent of the object. c. SigHat captures its extent, eccentricity and orientation. d. Finally SigHat2 estimates the thickness of the nebular shell. This estimate is difficult as the intensity of IC418 apparently varies as a function of latitude, however this is most likely due to the inclination of the PN - a feature not captured by SigHat2. The thickness estimate obtained nevertheless places us in the correct region of parameter space, which will facilitate more sophisticated analyses.

Estimates of Parameters

The models were quite capable of estimating the parameters to accuracies much greater than what is needed to aid the higher order models. Table 3 shows the evolution of the parameter estimates for Case 2. Note that the values of most of the parameters are frozen for SigHat2. All estimates are within acceptable ranges of error (less than 5%), especially as they are only being used to obtain ballpark estimates for use with higher-order three-dimensional models. The larger errors in the extent and the shell thickness are due to the different ways in which the models use these parameters to create the images. That is, these parameters quantify very different concepts and hence are not perfectly reconcilable.

TABLE 3. Evolution of Parameter Estimates				
Gauss	SigHat	SigHat2	True Values	Percent Error
$xo = 169.778$	169.965	169.96?	170	0.02%
$yo = 212.492$	209.806	209.80?	210	0.09%
$\sigma = 99.331$	$\sigma_x = 117.467$	117.46?	120	2.11%
	$\sigma_y = 173.117$	173.11?	180.53	4.10%
	$\theta = 0.2509$	0.2509	0.25	0.36%
		$\Delta = 0.6?1$	0.66	1.67%

As expected we found that the orientation was quite difficult to detect as the projected image of the object became more circular, either due to the eccentricity of the object or its inclination toward or away from the viewer. However, an elliptical nebula does not quite have an elliptical high-intensity ring when the object is inclined. The approximate eccentricity of the central region is typically higher than that of the outer edge of the nebula, as can be seen in IC418 in the region of the higher intensity regions of the projected shell. For this reason, it is probably wise to continue to estimate the orientation in SigHat2 as the shape of the darker inner region of the nebula provides more information about the orientation than the bright outer regions.

DISCUSSION

The idea of using a hierarchy of models to understand a physical system is based on the observation that present scientific theories are built on a framework of earlier theories. Each new layer of this framework must explain a new phenomenological aspect of the system in addition to everything that was explained by previous theories. There are of course fits and starts as a paradigm shift may qualitatively change the direction taken by this hierarchical progression. Yet even in such cases, the old theories are quantitatively sufficient to describe the phenomena that they were designed to model. Hierarchical organization is well known to be an efficient means to generating complex systems and, as it is a useful technique in theory building, we have chosen to examine its usefulness in efficient parameter estimation. The particular hierarchical succession of models employed in this work was chosen to successively estimate larger and larger numbers of parameters approaching the uniform ellipsoidal shell model of a PN.

We found that not only are the results obtained using a hierarchical set of models more accurate, but they are also obtained more quickly. We expect that as we progress to the UES model and then the IBPES model the observed speed-up and accuracy increase will become even more significant as these models represent the PN as a three-dimensional object, which requires a substantial increase in computational effort. Furthermore, by hierarchically applying a set of models, which better and better describe the object, we minimize the possibility that the estimate may converge to a locally optimal solution.