The Astronomical Journal, 130:84–94, 2005 July

AUTOMATED DETECTION OF CLASSICAL NOVAE WITH NEURAL NETWORKS

S. M. Feeney, V. Belokurov, N. W. Evans, J. An, and P. C. Hewett
Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

M. Bode, M. Darnley, and E. Kerins
Astrophysics Research Institute, Liverpool John Moores University, Twelve Quays House, Egerton Wharf, Birkenhead CH41 1LD, UK

P. Baillon
European Organization for Nuclear Research, CERN, CH-1211 Geneva 23, Switzerland

B. J. Carr
Astronomy Unit, School of Mathematical Sciences, Queen Mary, University of London, Mile End Road, London E1 4NS, UK

S. Paulin-Henriksson
Laboratoire de Physique Corpusculaire et Cosmologie, Collège de France, 11 Place Marcelin Berthelot, F-75231 Paris, France

and

A. Gould
Department of Astronomy, Ohio State University, 140 West 18th Avenue, Columbus, OH 43210
Received 2005 February 24; accepted 2005 April 9

ABSTRACT

The POINT-AGAPE collaboration surveyed M31 with the primary goal of optical detection of microlensing events, yet its data catalog is also a prime source of light curves of variable and transient objects, including classical novae (CNe). A reliable means of identification, combined with a thorough survey of the variable objects in M31, provides an excellent opportunity to locate and study an entire galactic population of CNe. This paper presents a set of 440 neural networks, working in 44 committees, designed specifically to identify fast CNe. The networks are developed using training sets consisting of simulated novae and POINT-AGAPE light curves in a novel variation on K-fold cross validation and use the binned, normalized power spectra of the light curves as input units. The networks successfully identify 9 of the 13 previously identified M31 CNe within their optimal working range (and 11 out of 13 if the network error bars are taken into account). The networks provide a catalogue of 19 new candidate fast CNe, of which four are strongly favored.

Key words: galaxies: individual (M31) — methods: numerical — novae, cataclysmic variables — stars: variables: other

1. INTRODUCTION

One of the greatest advances of modern experimental astrophysics is the automation of photometric surveys, which allows massive amounts of data to be gathered systematically, efficiently, and with the minimum need for human intervention. Such surveys scour large regions of the sky, carefully searching for a wide variety of rare objects and phenomena such as microlensing events (surveys like OGLE, MACHO, and EROS), gamma-ray burst optical counterparts (ROTSE), extrasolar planetary transits (SuperWASP), and near-Earth objects (NEAT). These surveys have provided the scientific community with invaluable information and resulted in many new discoveries, yet they have also left us with a new (and very welcome) problem: how can we sort through the vast data catalogs to reliably filter out objects of interest?

The raw data produced by these surveys are simply collections of the light curves of the objects found in the survey's field of detection. Transient objects hold particular interest for a long list of fields, including cosmology (SNe Ia), single and binary stellar evolution (SNe and cataclysmic variables, respectively), and dark matter studies (microlensing). They are generally rare and have short lifetimes and so must be identified and studied quickly. The sheer size of such data sets means that such transient objects are inevitably present in the catalogs; however, there is still a pressing need to detect objects swiftly and reliably for further study or follow-up. A number of researchers have argued that neural networks may provide a viable solution to this problem (Wozniak et al. 2001; Belokurov et al. 2003, 2004; Brett et al. 2004). Neural networks have already been proven to be useful pattern-recognition tools in astrophysical applications such as galaxy (Lahav et al. 1996) and stellar spectra (Bailer-Jones 1997) classification. They are highly adaptable, easy, and quick to use, but perhaps their most relevant asset in this application is their ability to attach a probability to their classification of an object, thus allowing the user to prioritize their further study.

The contribution of this paper is to provide working neural networks for the detection of classical novae (CNe). These are close interacting binary stars, consisting of a white dwarf primary and a cool red dwarf secondary. The secondary star overflows its Roche lobe and loses mass to the primary. Very occasionally, runaway thermonuclear burning of the degenerate layer of hydrogen accreted by the white dwarf can cause a nova outburst. The nova's brightness rises rapidly to an absolute magnitude of between -6 and -9 before slowly fading back to quiescence. Much remains unknown concerning the abundance and distribution of nova in galaxies because of the lack of systematic surveys. So, there is a need for fully automated, and less subjective, selection of candidate CNe so that more soundly based conclusions concerning the nova rate and distributions can be drawn. Darnley et al. (2004) have already devised one possible systematic algorithm. Here we provide an alternative

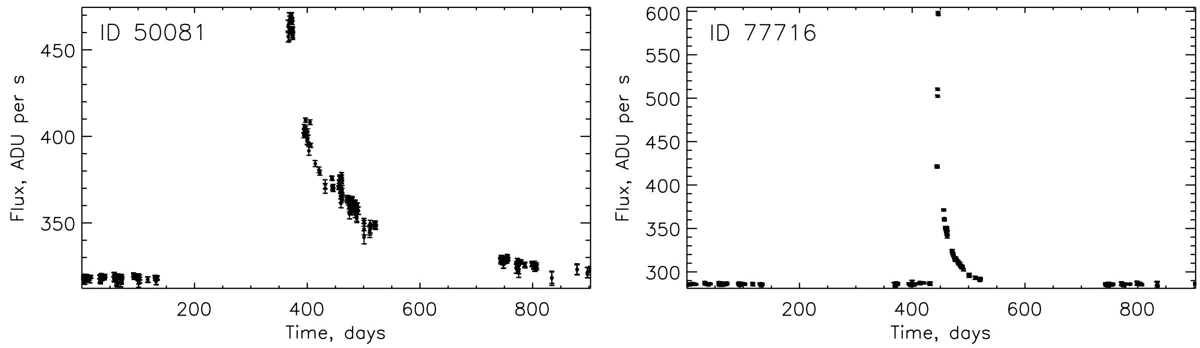

FIG. 1.—*Left*: Light curve of a slow nova in M31, as identified by Darnley et al. (2004; ID: PACN-00-02). Note the decay fluctuations in the declining part of the light curve. *Right*: Light curve of a fast nova in M31, as identified by Darnley et al. (ID: PACN-00-06) and An et al. (2004; ID: 77716).

to the method of Darnley et al. using a novel application of neural networks.

The paper is organized as follows. In § 2, the data set through which we search for CNe light curves is described. This is derived from the POINT-AGAPE microlensing experiment toward M31. Although the primary aim of this experiment is to find microlensing events, the data set of varying light curves is a rich resource for the study of variable stars toward M31 (An et al. 2004). Section 3 discusses the properties of nova light curves and summarizes previous work to find CNe in M31. Next, § 4 provides a short introduction to neural networks for the astronomical user. Section 5 describes the preprocessing and the architecture of neural networks to identify CNe, while § 6 describes the computations. The nova catalog obtained by the networks is presented in § 7.

2. THE LIGHT-CURVE DATA SET

The data used in this paper were gathered by the POINT-AGAPE collaboration working with the Wide Field Camera (WFC) mounted on the 2.5 m Isaac Newton Telescope (INT) on La Palma. The collaboration took images of the Andromeda Galaxy (M31) over the course of three observing seasons (1999–2001), searching for evidence of microlensing events (Aurière et al. 2001; Paulin-Henriksson et al. 2002, 2003; Belokurov et al. 2005). For 1 hr of each observing night, the WFC was used to take images of M31 over two fields to the north and south of M31's central bulge, with each field image formed using the four 4100×2048 CCDs that make up the WFC (see Fig. 1 of An et al. 2004). The raw data produced by the POINT-AGAPE collaboration then consisted of light curves generated from the flux gathered in three passbands by individual pixels in each field image. The passbands used were denoted *g*, *r*, and *i* and are similar to those used by the Sloan Digital Sky Survey. The M31 fields are mainly composed of unresolved stars, and the effects of seeing from epoch to epoch are substantial. In order to build light curves, we use the superpixel method to ensure that the same fraction of flux falls within the window function, irrespective of seeing (Melchior et al. 1999; Ansari et al. 1999; Le Du 2000). This provides superpixel light curves (7×7 pixels in size). Each pixel is $0\farcs33$ on a side, so the 7×7 superpixel is $2\farcs3$ on a side. This matches the typically worst seeing at the INT site, which is about $2''$. The superpixel light curves are then cleaned (for details, see Irwin & Lewis [2001] and An et al. [2004]); a mask of the known CCD defects was constructed, together with regions around all resolved stars detected in the reference frame. After masking, 44,635 variable superpixel *r*-band light curves remained, and this is the catalog through which we search for nova-like light curves.

Although the collaboration produced a very large amount of data and thus greatly increased the chances of discovering new objects, there are two complicating factors that slightly reduce the data's quality and ease of analysis. First, the observations were carried out over the course of three seasons. These seasons correspond to the periods in which M31 was visible from the Northern Hemisphere and mean that the light curves are sampled in runs of ~150 days, with ~200 day gaps (see Fig. 1 for an illustration of the sampling). Three other factors, the limited mounting of the WFC, the limited scheduled observing time on the INT, and the weather, result in the sampled runs consisting of well-sampled periods typically lasting 1–2 weeks, separated by very poorly sampled periods lasting 1–3 weeks. Second, the large distance of M31 means that in most cases single stars are not resolved by the INT. This means that the superpixel light curves almost always consist of flux produced by more than one star, which could result in very exotic light curves, hence limiting our ability to classify objects.

3. CLASSICAL NOVAE IN M31

In CNe, the cool red dwarf secondary overflows its Roche lobe and loses mass to the primary white dwarf. This mass builds up in an accretion disk before falling onto the surface of the white dwarf (see, e.g., Bode & Evans 1989). The main feature of nova light curves is a single outburst,[1] typically increasing the absolute magnitude of the nova to between −6 and −9 before slowly (compared to the initial rise) fading back to the quiescent state. These CN outbursts are caused by the runaway thermonuclear burning of the degenerate layer of hydrogen accreted by the white dwarf. Once a critical amount of hydrogen has been accreted, it begins to burn via the CNO cycle, precipitating thermonuclear runaway and resulting in the ejection of the accreted layer on the white dwarf surface. This explosion and ejection are accompanied by an intense brightening, followed by a gradual decay back to quiescence.

The progress of the nova outburst depends on several parameters, including the mass accretion rate from the secondary and the temperature and mass of the white dwarf (e.g., Prialnik & Kovetz 1995). The outbursts therefore vary from system to system, as shown by the rich variety of CNe light curves in Sterken & Jaschek (1996). However, it is possible to divide novae into speed classes according to the time (t_2) taken to decline by 2 mag from maximum light, the two main classes being fast ($t_2 < 80$ days)

[1] CNe are *required* to have had only one major outburst in historic times. A few CNe in quiescence show smaller outbursts, similar to those in dwarf novae, caused by changes in mass flux through the accretion disk.

and slow ($t_2 > 80$ days) novae (e.g., Payne-Gaposchkin 1957). Fast novae rise rapidly to maximum light, taking 1–2 days, and generally have relatively smooth initial decays with only small fluctuations in their early light curves. Slow novae, on the other hand, can take much longer to reach maximum light and usually have more erratic light-curve decays, with strong fluctuations capable of producing secondary maxima of varying strengths during initial decline. Furthermore, the maximum absolute magnitudes of CNe are correlated to the rate of their decline, which, coupled with their high luminosities, makes CNe potentially important standard candles (Hubble 1929; Cohen 1985). Figure 1 shows the light curves of a slow and fast nova, as previously found in M31.

The light curves of CNe share many features with the light-curve peaks of dwarf novae and recurrent novae. The main distinguishing feature in the light curves of these objects is that dwarf and recurrent novae undergo repeated outbursts. However, the periods between outbursts and the gaps in the POINT-AGAPE sampling could lead to only one peak of a dwarf or recurrent nova light curve being sampled. Hence, we may pick up some stray dwarf or recurrent novae in our final catalog. Dwarf nova outbursts are not detectable in M31. However, they may be present in the POINT-AGAPE catalog as foreground objects, although even this has a very low probability.

Dedicated nova searches of M31 have been carried out ever since Hubble first did so in 1929 (see Table 1 of Darnley et al. [2004] for a list of papers). Very recently, Darnley et al. (2004) and An et al. (2004) have published CNe light curves from the POINT-AGAPE catalog. Darnley et al. (2004) used a pipeline (see Table 3 in their paper) to filter out novae independently of any prior knowledge. This pipeline first selected only objects (defined to be resolved structures with fluxes significantly higher than the local median) present in five consecutive observations, to remove rapid variations. The catalog was further pruned by selecting against periodicity, requiring an adequately sampled primary peak and any secondary peak to be an acceptable size. The remaining candidates were finally required to fit data, rate of decline, color, and color-magnitude criteria before being accepted as nova candidates. An et al. (2004) were primarily interested in the cataloging of the variable stars in the POINT-AGAPE data set. They first constructed a catalog of variable objects by selecting only (suitably cleaned, masked, etc.) superpixel light curves with deviations from their baseline significant enough in size and duration. Novae were then located by looking for variable objects matching (within a 3″ error circle) the positions of novae as published in IAU Circulars. Using these methods, Darnley et al. (2004) gave 20 novae and An et al. (2004) 12 novae light curves, with seven novae common to both papers.

4. AN INFORMAL INTRODUCTION TO NEURAL NETWORKS

This section is intended as a brief introduction to the basics of neural network structure and use as they apply to this paper (for more details, consult Bishop [1995] and MacKay [2003]). Neural networks are pattern recognition tools composed of neurons (or units) arranged in layers. Neurons come in three types: input, hidden, and output. The structure of the networks used in this paper is one layer of input units, one layer of hidden units, and one layer of output units. The neurons in neighboring layers are fully connected with each other, and these connections have assigned to them adaptive weights that are used to calculate the response of a specific neuron to its inputs. The input data are taken as the values of the input units, and the value of each hidden unit is then given by the sum over all connections of the activation value on each input unit, weighted by the weight on the connection. These activation values are calculated using an activation function acting on the value of the unit. The values of the output units are calculated in a similar fashion, except the sum is performed over all connections between the output unit in question and the hidden units. In this paper, the activation function is chosen to be the logistic function, which allows the outputs to be interpreted as a posteriori probabilities.

Before all this can happen, the network must be trained in order to determine the weights. The weights are initially randomized, and the network is presented with a training set, made up of sets of input values (called patterns) for which the desired outputs are known. The outputs produced by the randomly weighted net are compared to the desired values, and the network performance on all patterns is quantified using an error function, namely, the cross entropy error (Bishop 1995; Belokurov et al. 2004). A learning function then uses these errors in conjunction with the values of the hidden units and the hidden-to-output layer weights in order to update the weights and hence reduce the output errors. The errors are also propagated back up to the input-to-hidden layer weights so as to update these weights with the same goal in mind. This whole process, called back-propagation, is carried out a number of times (called epochs) until the desired network performance is reached. With most choices of learning function, it is possible for the network to become overtrained on the training set, with the result that performance on a more general set of inputs is reduced. In this paper, a special learning function (see § 5.4) is used to avoid this problem.

The process behind training neural networks is the minimization of the error function (as applied to the training set) with respect to the adaptive weights within the network. This error function may not have just a global minimum in the multidimensional weight space but could have a number of local minima instead. In any case, networks trained using the exact same training set for the same number of epochs but using different initial weights (and therefore different starting points in this space) will converge to slightly different final weights. In the case of multiple minima, this means that networks can follow different error minimization paths into entirely separate minima, some of which might classify the general set (as opposed to the training set) much better than others. We can turn this fact to our advantage by using network committees (see Bishop [1995], §§ 9.6 and 10.7), produced by training groups of networks on the same training set but with initial weights randomly chosen from a range of values. These networks therefore sample a region (rather than a point) of the weight space around the error function minimum/a and hence produce a range of results when classifying the final test set. The results can then be averaged out over the committee to take account of a whole range of network "opinions," making sure poor quality networks stuck in high-error minima do not overly affect the results.

5. NETWORK PREPARATION

5.1. *The Training Set*

The ideal training set should contain examples of all forms of stellar variability we expect the networks to encounter, along with as many examples of nova light curves as possible. The usual process is to build the training set from a comprehensive selection of example nova and variable star light curves taken from existing data catalogs. We do not do this for two reasons. First, there are not enough well-sampled nova light curves in the

g, r, and i bands in the standard catalogs for our purposes. Therefore, we are obliged to simulate nova light curves from templates. Second, all of the other forms of variability needed for the training set are already present in the POINT-AGAPE catalog, and we can therefore use the data set itself to provide the nonnova examples required to build the training set, using a variation on a technique called K-fold cross validation (see below and Bishop 1995, § 9.8.1).

In K-fold cross validation, the data set is first partitioned into K separate segments. A network is then trained using a training set containing all of the data from $K-1$ segments before being tested on the remaining segment. This process is then repeated, each time choosing a different segment to be left out of the training set, until all K choices for the omitted segment have been covered. The test errors are then averaged out over all K results to create a much more robust estimate of the network performance, hence providing one of the two main advantages of using this technique. The second advantage is that all of the examples in the data set are used in both training and testing, in effect creating a large training set without the need for any "external" data. The major disadvantages are that the training process must be repeated K times, and that some or all of the training sets will contain nova-type light curves present in the catalog falsely identified as nonnova objects. We therefore use a new variation on the technique, training networks using just one data segment before testing the networks on the remaining $K-1$ segments. We believe this is advantageous, as it reduces both processing time and the risk of training set contamination while still retaining the benefits of normal K-fold cross validation.

The final form for the training set is 1000 simulated nova light curves, assigned desired output probabilities of 1, and 1000 randomly chosen POINT-AGAPE light curves, with desired output probability 0. The decision to use exactly 1000 POINT-AGAPE light curves is a compromise: 1000 POINT-AGAPE light curves should include a sufficient cross section of the forms of variability while greatly reducing individual training times and keeping the number of falsely classified nova examples down to $O(1)$ per training set.[2] The main drawback to using 1000 POINT-AGAPE light curves is that the training process must be repeated $\sim$40 times and is therefore quite slow. The number of nova examples is chosen to overwhelm any falsely classified novae and also to create networks biased toward producing false positives rather than false negatives. Overrepresenting the novae (as compared to their natural frequency) in the training set increases the prior probability of finding a nova in the set, and hence training using such sets produces networks that are much more likely to misclassify nonnovae as novae than vice versa (see § 6.2). This is exactly the trend required considering that we are trying to locate a very rare phenomenon. Of course, the drawback to permitting more false positives than false negatives is that an additional algorithm may be needed after the neural network search to root out the contaminants.

5.2. *Nova Templates*

Six novae identified by An et al. (2004) (see Table 1) are chosen as templates. They are selected as having well-sampled peaks with intermediate decay timescales; their half-widths at 5% of maximum light (an indication of the total length of the decay) are all in the range 40–100 days. Three other novae (An et al. IDs 26277, 78668, and 83479) were also originally included as templates, but their inclusion reduced the consistency (in terms of both decay timescale and shape) of the simulated portion of the training set and resulted in poor final network performance. Note that, because of the limited timescales covered by the templates and the differences in the light curves of CNe of different speed classes, we expect our networks to suffer when asked to classify novae with much longer or shorter timescales. The template light curves are fitted using a model function consisting of a flat background, a steep linear rise, and a function $f(t)$ of the form shown below to match the decay:

TABLE 1

An et al. (2004) and Darnley et al. (2004) Identification Numbers of the Template Novae, along with Estimates of Decay Time

An et al. ID	Darnley et al. ID	Half-Width at 5% of Max. Light (days)
25851	PACN-99-05	64.8
26021	PACN-00-04	59.8
26946	Not present	72.7
77324	PACN-01-06	99.6
77716	PACN-00-06	45.5
83835	Not present	42.0

$$f(t) = A_1 \exp\left[\frac{-(t-t_m)}{\tau_1}\right] + A_2 \exp\left[\frac{-(t-t_m)}{\tau_2}\right] + B, \quad (1)$$

where A_i are the relative sizes of the exponentials, t_m is the time of maximum light, τ_i are the exponential decay timescales, and B is the value of the background flux. Figure 2 shows an example of such a model function.

To create the 1000 simulated novae, we repeat the following procedure. First, a random template is selected, and its peak is shifted randomly in time within the time limits of the POINT-AGAPE measurements. A POINT-AGAPE light curve is then chosen at random from the catalog, and its sampling times are used to sample the newly shifted model function. At this point, we require that there are at least 10 sampling times present in the first 30 days after the peak time of the shifted model, to ensure that enough of a signal is present.[3] A small amount of Gaussian noise is then added to the sampled, shifted model in order to create simulated novae light curves that are as similar in form as possible to the original novae (see Fig. 2).

5.3. *Preprocessing and Network Inputs*

The computational power required to use a network grows quickly with each added input. It is therefore usual to preprocess the light curves, that is, to extract a small number of features from the data to use as inputs. In this paper, we reduce each light curve to its power spectrum before binning and suitably normalizing both the individual power spectra *and* the training set as a whole (Belokurov et al. 2003, 2004).

The first reason for reducing the data to their power spectra is that the features that distinguish the nova-type light curves from the other forms of variability—i.e., the event timescales, the singular nature of the eruptions, and the shape of the nova

[2] There are $\sim$40,000 light curves in the catalog, with $O(20)$ true nova examples present. Hence, choosing 1000 POINT-AGAPE examples per training set gives $\sim$0.5 false nova-type light curves per set.

[3] Without this requirement, many of the simulated nova light curves have very small peaks (or none at all). There was, therefore, a large constituent group of the training set whose light curves were dominated by the random Gaussian fluctuations we added, and so the networks simply learned to recognize these light curves instead of the nova-like light curves.

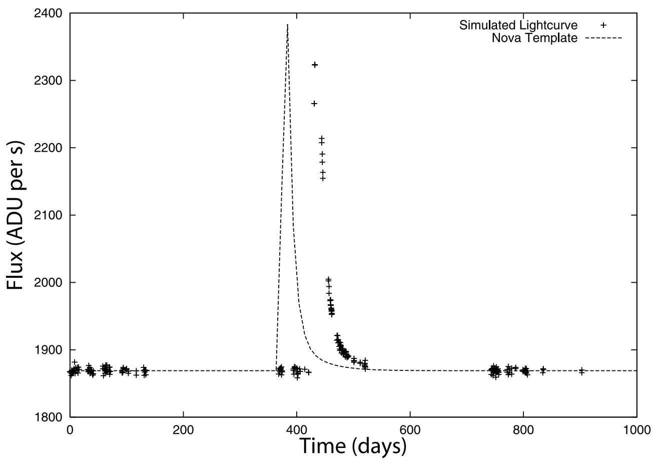

FIG. 2.—Model function for template nova 26021, along with simulated light curve.

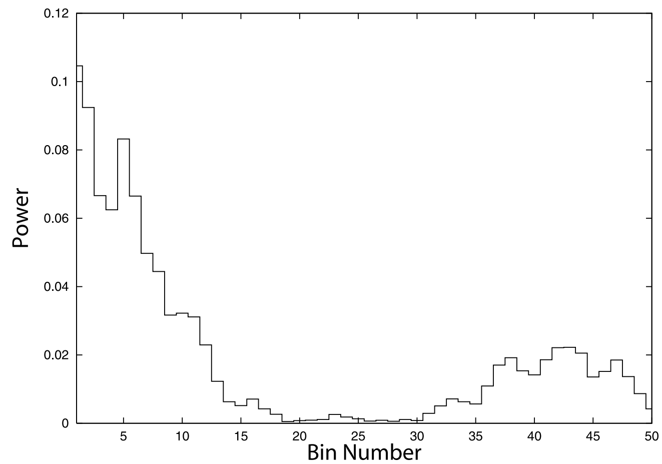

FIG. 3.—Binned, normalized power spectrum (prior to full training set normalization) of a nova light curve. Higher bin numbers correspond to higher frequencies.

peaks—all manifest themselves in the power spectrum. To see this, consider a simplified nova eruption as a top-hat function of width w. The Fourier transform of a top-hat function of width w in positive frequency space is (half) a sinc function, with a central peak of half-width π/w. Hence, we expect that the power spectra of our actual nova eruptions with decay timescales τ to be distortions of sinc functions with central peaks of widths of the order of π/τ. From this line of reasoning, we expect the almost singular nature of the nova eruptions to make their power spectra sinclike, with the individual timescales affecting the widths of the sinc peaks and the shapes of the outbursts distorting the power spectra as a whole. Some evidence for this can be found in Figure 3, which shows that the nova power spectra do indeed resemble sinc functions with roughly correct peak widths.

A further reason for choosing the power spectrum is that the features we wish to select *against*, such as periodicity or random variations, should also manifest themselves in the power spectra of the nonnova objects. The power spectrum is also invariant under time translation of the initial light curve. Furthermore, the power spectrum is easily binned, which allows for the reduction in dimensionality to produce practical networks, although care *must* be taken to ensure that too much information is not lost. Because of the uneven time sampling of the POINT-AGAPE light curves, we used the Lomb-Scargle periodogram (Press et al. 1992) to calculate the power spectra. The power spectra are determined in the frequency range 0–0.3 day^{-1}, as this range of values contains a significant number of CNe power spectrum features. The power spectra are all binned into 50 constant width bins, as this results in a manageable number of network inputs but still retains the resolution of the original power spectra.

The next preprocessing technique is to normalize each individual binned power spectrum. This has two positive effects: first, it ensures that all of the inputs are consistently drawn from within the same range (from zero to one), and second, it reduces the chances of the networks classifying two differently shaped power spectra simply because they contain a similar size peak. Normalizing the individual light curves helps the networks classify objects by the shapes of their power spectra, rather than the size of any peaks the power spectra contain. An example of a binned normalized nova power spectrum as it appears at this stage of preprocessing is shown in Figure 3.

The last preprocessing technique is to shift the first input of each pattern in the training set by the mean of all the first inputs and then scale it by dividing by the standard deviation of all of the first inputs. This is repeated for each input, so that *all* of the networks' inputs are not only drawn from the same range but also have comparable magnitudes, which forces the networks into classifying the set using *all* of the inputs provided. As an illustrative example, prior to the introduction of this technique, the nova power spectrum typically has low-frequency bin powers a factor of 10^2 greater than its high-frequency bin powers (see Fig. 3). Now we would consider a 10% variation in the power in any bin to be equally important, but a 10% variation in a high-frequency bin would appear to the networks to be much less important than a 10% variation in a low-frequency bin. By scaling the inputs as described, the networks classify using the relative, and not absolute, sizes of bin power variations between different objects.

5.4. *Network Architecture*

The networks used in this paper are all created using the Stuttgart Neural Network Simulator[4] and are made up of one layer of 50 input units, one layer of 24 hidden units, and one layer consisting of one output unit. (The reasons behind this choice are given shortly.) The units in the hidden layer are fully connected to both the input and output layers, and the value of the output unit gives the a posteriori probability that the subject light curve is a nova, given the weights and the inputs calculated for the subject. Our networks use as a learning function resilient back-propagation with adaptive weight decay (RpropMAP). Particularly high adaptive weights correspond to very strong pattern recognition and therefore tend to suggest overfitting of the training set. During the training process, RpropMAP therefore automatically allows the highest weights to decay intelligently so as to keep the network as generally applicable as possible. Hence, when using RpropMAP, there is no need for the validation process required by other learning functions. (A much fuller explanation can be found in Bishop [1995], §§ 9 and 10.)

The last choice to make is the number of units. Choosing the number of input and output units is straightforward: these numbers are simply determined by the number of inputs (in our case, 50) and outputs (in our case, one) that the networks receive and produce, respectively. However, in tasks such as this,

[4] See http://www-ra.informatik.uni-tuebingen.de/SNNS.

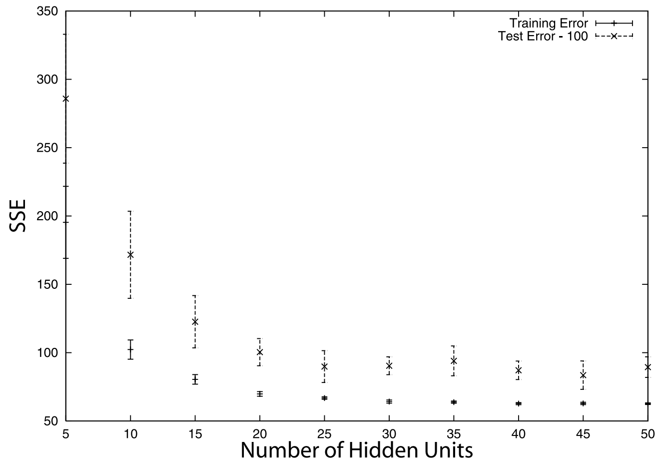

FIG. 4.—Errors in training and test sets for a range of numbers of hidden units. (SSE stands for the sum of the squared errors of all outputs.) Note that test set errors have been shifted down by 100 to aid comparison.

it is impossible to choose the required number of hidden units N_H theoretically. Instead, N_H must be determined experimentally, by examining the behavior of the errors produced by networks of differing N_H in classifying the training set and a new test set (same form as the training set but totally new light curves).

We expect both the training and test errors to decrease at first with increasing N_H. Low-N_H networks are very simple, so increasing the number of hidden units increases the network's complexity and hence ability to map the decision boundary between the classes of object. However, for some values of N_H, we expect the behavior of the training and test errors to diverge, with the training error continuing to decrease but the test error either leveling off or beginning to rise. This differing behavior occurs because the networks have become complex enough to start to overtrain on the training set. The final number of hidden units is therefore chosen to be the value of N_H at which the training and test error behaviors diverge, as this gives the best general network performance.

A plot of the mean errors produced by our networks in classifying the training and test sets against N_H is shown in Figure 4. These results are produced by training committees of 10 networks for each value of N_H, with each network given initial weights drawn randomly from the range -3 to 3. The networks are trained for 1000 epochs, after which the final errors in classifying the training set are recorded. The trained networks are then each tested using the same test set. The training and test errors are finally averaged out over each committee, thereby providing mean values to represent more reliably the performance of the different size networks. The standard deviations are also computed to give some idea of the mean error spread.

The first feature to note in Figure 4 is that the test error values are all significantly larger than their corresponding training errors. This is because there are likely to be numerous light curves in the test set of which there are no similar examples in the training set, because of the random selection of the POINT-AGAPE light curves included in each set. This increases the risks of misclassification. The most important information to take from the plot is the behavior of the errors. For small values of N_H, the behavior of both the training and test errors is very similar, as expected. For N_H between 20 and 25, however, the behavior of the two errors begins to differ: the training error continues decreasing asymptotically, whereas the test error levels out within its error bars. We therefore use 24 hidden units in the networks to produce our final results.

TABLE 2

IDENTIFICATION NUMBERS OF THE M31 NOVAE, AS IDENTIFIED BY DARNLEY ET AL. (2004), AN ET AL. (2004), OR BOTH, MISSING FROM CLEANED DATA SET

An et al. ID	Darnley et al. ID
10889	PACN-99-01[a]
28862	PACN-99-02[a]
82483	PACN-99-03[a]
93392	PACN-99-04[a]
49835	PACN-99-07[a]
26946[b]	PACN-00-01[b]
83835[b]	PACN-01-02[b]
26277[c]	...
26285[c]	...
78668[c]	...
79136[c]	...

NOTE.—Ellipses indicate no ID available.
[a] Nova identified by Darnley et al. (2004).
[b] Nova identified by both Darnley et al. (2004) and An et al. (2004).
[c] Nova identified by An et al. (2004).

6. PRODUCTION OF FINAL RESULTS

6.1. *The Network Probabilities*

Forty-four committees consisting of 10 networks, each with 50 input units, 24 hidden units, and one output unit, are created with random initial weights. These networks are trained using training sets as described in § 5.1 for 1000 epochs, taking care to record the POINT-AGAPE light curves used and the 50 input means and standard deviations (as described at the end of § 5.3) for each training set. The trained networks are then used to classify two data sets, which are preprocessed in the same fashion as the training set but are normalized using the input means and standard deviations specific to each committee. The first data set is the cleaned POINT-AGAPE catalog, and the second consists of all of the novae identified by An et al. (2004) and Darnley et al. (2004) missing from the catalog, as listed in Table 2. The initial form of the results is therefore a set of 440 probabilities for each POINT-AGAPE object and each previously identified nova. Each object's results are first averaged out over the 10 networks in each committee, producing 44 committee probabilities and errors for each object, before these values are averaged over the committees. The POINT-AGAPE objects' probabilities and errors are averaged out over only those committees in whose training sets they did not feature, whereas the previously identified novae's values are averaged over all 44 committees.

6.2. *Decision Boundary Determination*

The final task is to set the decision boundary for classification, that is, to determine the probability value an object must exceed in order to be classified as a CN. This requires the network's performance to be quantified in terms of numbers of false positives (POINT-AGAPE objects with probabilities greater than that of the decision boundary) and negatives (simulated novae with probabilities less than that of the decision boundary) for a range of decision boundary choices. The decision boundary is chosen so as to optimize the rates at which these false classifications occur.

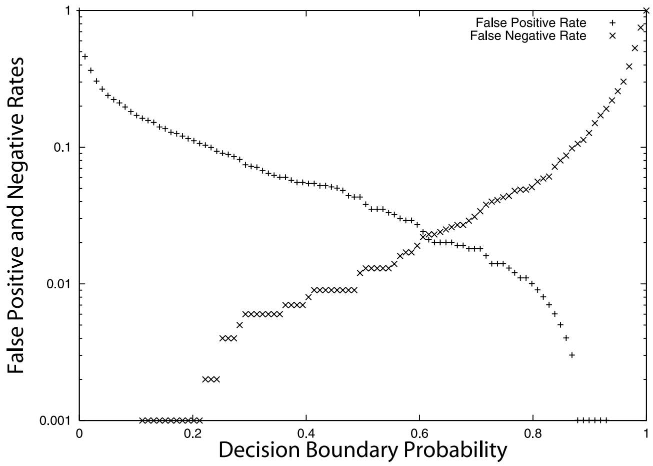

FIG. 5.—Rates of false positive and negative classifications for a range of decision boundary probability values.

First, a test set is produced using 1000 new simulated novae and the POINT-AGAPE light curves used to train one network committee. The test set was then preprocessed and classified by the other committees, and the results were averaged out in the same way as the POINT-AGAPE catalog results in § 6.1. The decision boundary probability $p_{\rm db}$ is set at different probability values between 0 and 1 and the numbers of false positives $N_{\rm fp}$(test) and negatives $N_{\rm fn}$(test) for the test set determined. The results are plotted in Figure 5.

If, in our catalog, the number of nonnova objects $N^{\rm tot}_{\rm var}$(cat) were approximately equal to the number of novae $N^{\rm tot}_{\rm nov}$(cat), the standard procedure would be to choose the decision boundary such that the rates of false positives and negatives are equal. In actuality, however, we expect $N^{\rm tot}_{\rm var}({\rm cat}) \sim 44{,}600$ and $N^{\rm tot}_{\rm nov}({\rm cat}) \sim 20$. This means that if the decision boundary were chosen to be the point at which the rates $r_{\rm fp}$(test) and $r_{\rm fn}$(test) were equal (i.e., $r_{\rm fp}({\rm test}) = {\rm r_{fn}}({\rm test}) \sim 0.025$), then the number of expected false positives is $\sim$1100, much bigger than the number of true novae expected. The choice of decision boundary must therefore be taken to minimize the number of false positives while ensuring that most novae are still detected. Accordingly, the decision boundary probability is fixed to be 0.95. At this value, $r_{\rm fn}({\rm test}) \sim 0.2$ from Figure 5, so we expect 20% of the true novae to be missed. No value for $r_{\rm fp}$(test) is available for this $p_{\rm db}$ (probably because the test set was too small to contain any POINT-AGAPE objects with outputs as high as 0.95); however, an upper bound on the value can be found by taking the last nonzero value, which is $\sim$0.001. For this decision boundary, we therefore expect <45 false positives, a much more manageable number comparable to the total number of true novae expected.

7. THE NOVA CATALOG

The nova catalog comprises 47 objects classified by the networks as having probabilities greater than 0.95 of being novae and is made up of nine previously identified novae (discussed in § 7.1), 19 new nova candidates, and 19 probable contaminants (all discussed in § 7.2).

7.1. *Previously Identified Novae*

The average probabilities produced for the 25 CNe previously identified by An et al. (2004) and Darnley et al. (2004) are shown in Table 3. Also included in this table are two decay timescales: the half-widths of the peaks at $1/e$ and 5% of maximum light (t_e and $t_{5\%}$, respectively), chosen to give an indication of the timescale of the initial (t_e) and overall ($t_{5\%}$) decay. The networks trained in this paper correctly identify nine of the novae (using the criterion from § 6.2), with three further novae falling within their probability errors' distance of the classification cutoff. A plot of the probabilities assigned to the 25 novae against their 5% timescales is shown in Figure 6. Examination of this plot indicates two main trends in the data. The first trend is that the novae that are classified with higher probabilities also have much smaller probability errors than the misclassified novae. The poorly classified (probabilities of 0.7 and lower) novae in particular are therefore classified much better by some networks than others, which suggests that their power spectra are being confused. The confusion could be because the power spectra of these objects are similar to those of POINT-AGAPE objects present in only some networks' training sets or because the networks have never seen this form of nova before.

The second trend is that novae with $t_{5\%}$ in the range 30–140 days are generally classified much better than those outside the range, apart from three exceptions in the range (IDs 10739, 14026, and 50100; specifically marked in Fig. 6) and one outside (ID 50081). On closer inspection of the exceptions within the range, reasons for their misclassification become apparent. The light curve of nova 50100 has a very significant second peak and even some evidence for a third, as well as a confusing bump in the later part of the light curve. We therefore do not expect to classify this object well. Light curve 14026 actually has a much slower decay than is indicated by its $t_{5\%}$ value (the reason behind this being its poorly sampled decay) and so should be located further right in the plot. Its light curve also features a second bump in the early stages of its decay. Light curve 10739 at first appears to be ideal for our networks, but its peak is poorly sampled near maximum. This seems to hinder the Lomb periodogram, as nova 10739's power spectrum contains large amounts of high-frequency noise. These objects, therefore, should either really not be found in this region of the plot or possess features that make them differ from the template nova light curves our networks are trained to recognize.

Discarding these objects, the networks correctly identify 8 out of the 13 novae found in the preferred $t_{5\%}$ range ($\sim$62% efficiency). Allowing for the error bars on the network outputs, a further three novae fall above the decision boundary ($\sim$92% efficiency). Therefore, the networks can be reliably used to recognize typical novae with timescales in the range 30 days $< t_{5\%} \lesssim 140$ days but not outside this range. Note that this range is actually slightly larger than the range of timescales used in the template light curves (i.e., 40 days $< t_{5\%} <$ 100 days), as the networks can generalize to some extent. The rapid falloff of the network's response for novae with $t_{5\%}$ much greater than 100 days is to be expected, as slow CNe are much more likely than fast CNe to have decay fluctuations and secondary peaks and hence be significantly different to the template novae. The low-$t_{5\%}$ falloff of the network's response is also expected, as it corresponds to the power spectrum range becoming too small to fit in the main features of the novae's sinclike power spectra (see § 5.3). These falloffs mean that in order to recognize novae with $t_{5\%}$ values outside of the preferred range, we will have to alter the preprocessing techniques.

Additionally, one further nova outside the preferred timescale range is detected. We note that the positive classification of light curve 50081 is highly inconsistent with the results for other slow novae. Its light curve is well sampled and clearly belongs to a very slow nova, and yet its power spectrum appears

TABLE 3

Probability Values Assigned to Novae Previously Located by Darnley et al. (2004), An et al. (2004), or Both

Object's An et al. ID	Object's Darnley et al. ID	Half-Width at $1/e$ of Max. Light (days)	Half-Width at 5% of Max. Light (days)	Averaged Network Response
10889	PACN-99-01[a]	10.8	99.3	0.863 ± 0.061
28862	PACN-99-02[a]	55.9	291.1	0.253 ± 0.106
82483	PACN-99-03[a]	13.2	38.3	0.849 ± 0.073
93392	PACN-99-04[a]	24.3	267.6	0.423 ± 0.124
25851[b]	PACN-99-05[b]	7.8	64.8	0.977 ± 0.012[c]
10739	PACN-99-06[a]	13.2	54.0	0.611 ± 0.152
49835	PACN-99-07[a]	28.6	178.5	0.710 ± 0.160
26946[b]	PACN-00-01[b]	19.2	72.7	0.961 ± 0.017[c]
50081	PACN-00-02[a]	72.4	433.9	0.977 ± 0.008[c]
24225	PACN-00-03[a]	13.6	87.8	0.963 ± 0.016[c]
26021[b]	PACN-00-04[b]	26.8	59.8	0.901 ± 0.066[d]
50100	PACN-00-05[a]	37.3	108.8	0.482 ± 0.170
77716[b]	PACN-00-06[b]	10.6	45.5	0.984 ± 0.008[c]
87092	PACN-00-07[a]	25.3	135.5	0.976 ± 0.011[c]
81539[b]	PACN-01-01[b]	79.5	159.7	0.394 ± 0.156
83835[b]	PACN-01-02[b]	7.7	42.0	0.993 ± 0.003[c]
14026	PACN-01-03[a]	60.3	103.0	0.690 ± 0.153
82840	PACN-01-04[a]	18.1	81.7	0.917 ± 0.056[d]
1881	PACN-01-05[a]	23.9	94.7	0.985 ± 0.008[c]
77324[b]	PACN-01-06[b]	27.6	99.6	0.986 ± 0.008[c]
26277[e]	...	13.5	93.3	0.887 ± 0.064[d]
26285[e]	...	5.0	5.1	0.625 ± 0.183
78668[e]	...	10.1	336.5	0.258 ± 0.117
79136[e]	...	0 (1-point peak)	0	0.022 ± 0.007
83479[e]	...	3.8	16.0	0.788 ± 0.068

Note.—Ellipses indicate no ID available.
[a] Nova identified by Darnley et al. (2004).
[b] Nova identified by both Darnley et al. (2004) and An et al. (2004).
[c] Object definitely classified as a nova by our networks.
[d] Object just misclassified (i.e., whose probabilities *plus errors* overlap the decision boundary).
[e] Nova identified by An et al. (2004).

to be recognizable to the neural networks. We currently have no explanation as to why the networks should pick it up, as nothing comparable to it appears in the training set, but as it appears in a region where little response is expected, it is more of an added bonus than a troubling anomaly.[5]

[5] A duplicate of 50081, namely, 50153, is also detected. However, it is removed from the list of new nova candidates by human intervention.

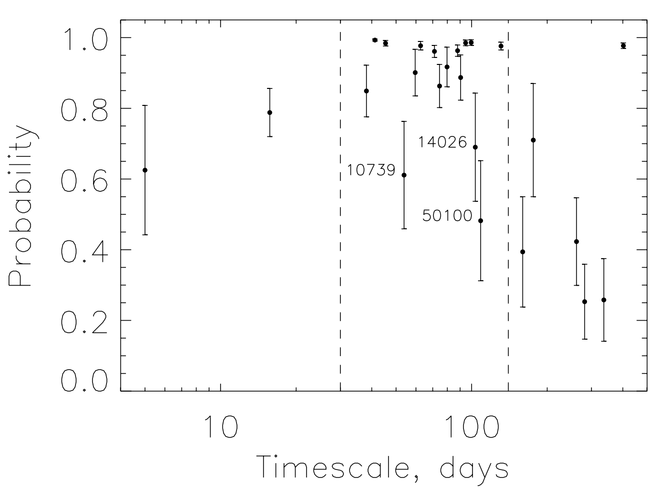

Fig. 6.—Nova probability vs. 5% timescale for the 25 previously identified novae. The region within which the highest nova sensitivity is reached is indicated with dashed lines.

7.2. *New Nova Candidates*

The nova catalog also contains 19 light curves, which, on inspection, are either recognizable as novae or exhibit some nova characteristics and hence can be classified as candidates for newly discovered novae. The IDs, locations, and probabilities of these 19 candidates are listed in Table 4, while their light curves are displayed in Figure 7.

The candidates can be roughly separated into four groups according to which nova features they exhibit. The first group consists of 1430, 42808, 50177, and 74935. Their light curves contain most or all of the desired features and hence make excellent nova candidates. The second group has light curves in which only the first few measurements of a rise toward a peak are present (IDs 58826, 66538, 73732, 80951, and 92933), with no sampling of the decay. The third group has light curves with samples present that suggest some form of decay from a peak but no measurements of the rise or peak itself (IDs 2973, 86283, 88205, 89701, 93095, and 95935). The fourth group has light curves that feature prominent, sharp peaks but not much clear evidence for the characteristic nova rise or decay (IDs 6251, 39995, 42075, and 86234) and which could therefore be very fast novae or simply instrumental defects. It is difficult to say for certain that objects in these three groups are novae without more data. The locations of the 19 candidates in Table 4, together with the nine candidates in Table 3, are shown in Figure 8, superposed on the optical isophotes of M31. These are all the candidates with a network probability >0.95.

The difference images of all 19 candidates have been examined, and the point-spread functions (PSFs) constructed. If

TABLE 4

The IDs, Right Ascension and Declination, Network Probabilities, Decay Timescales (Where Possible), and Nova Features of the 19 New Nova Candidates

Object ID	R.A. (J2000.0)	Decl. (J2000.0)	Half-Width at $1/e$ of Max. Light (days)	Half-Width at 5% of Max. Light (days)	Network Probability	Nova Features
1430..............	00 44 36.564	41 27 24.159	47.7	91.6	0.966 ± 0.016	M
2973..............	00 44 30.446	41 18 13.510	9.6	∼25	0.959 ± 0.041	D, F
6251..............	00 44 05.978	41 22 19.384	11.0	17.0	0.975 ± 0.012	P, F
39995............	00 44 33.854	41 37 27.775	5.0	5.8	0.971 ± 0.018	P, F
42075............	00 44 29.306	41 35 42.513	1.5	15.1	0.975 ± 0.012	P, F
42808............	00 44 22.572	41 29 50.579	31.9	89.5	0.955 ± 0.020	M
50177............	00 43 15.816	41 29 12.045	29.3	64.8	0.955 ± 0.011	M
58826............	00 42 05.645	41 02 49.409	No decay	No decay	0.986 ± 0.005	R
66538............	00 41 46.976	40 45 28.867	No decay	No decay	0.984 ± 0.005	R
73732............	00 43 33.160	41 06 44.146	No decay	No decay	0.954 ± 0.020	R
74935............	00 43 18.538	41 09 48.496	36.7	84.4	0.954 ± 0.009	M
80951............	00 42 31.148	41 14 25.462	No decay	No decay	0.961 ± 0.017	R
86234............	00 43 41.720	41 01 04.352	No decay	No decay	0.989 ± 0.014	P, F
86283............	00 43 44.186	41 01 48.806	6.0	6.0	0.968 ± 0.016	D, F
88205............	00 43 27.490	40 57 16.057	∼3	∼10	0.980 ± 0.009	D, F
89701............	00 43 12.483	40 54 05.979	3.0	107.8	0.953 ± 0.026	D, F
92933............	00 42 48.081	40 57 20.327	No decay	No decay	0.954 ± 0.023	R
93095............	00 42 44.604	40 57 04.521	3.6	52.1	0.979 ± 0.008	D, F
95935............	00 42 24.705	41 02 12.578	∼3	∼10	0.977 ± 0.015	D, F

Note.—Units of right ascension are hours, minutes, and seconds, and units of declination are degrees, arcminutes, and arcseconds. For nova features, M = most or all, R = rise only, D = decay only, P = peak only, and F = possible fake, as judged by examination of the image frames.

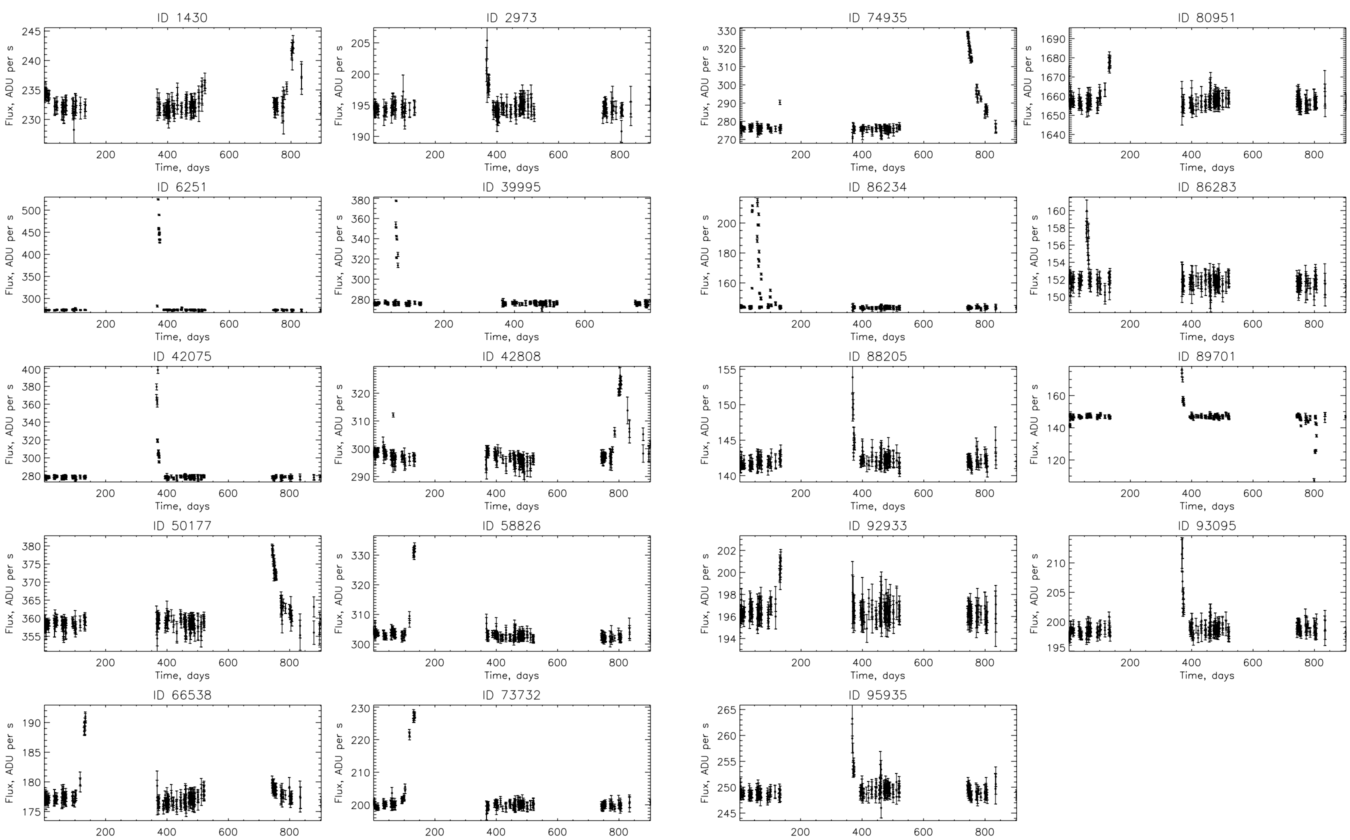

Fig. 7.—Light curves of the 19 new nova candidates. The r-band flux in ADU s^{-1} is plotted against time in JD $-$ 2, 451, 392.5. The four strong candidates are 1430, 42808, 50177, and 74935.

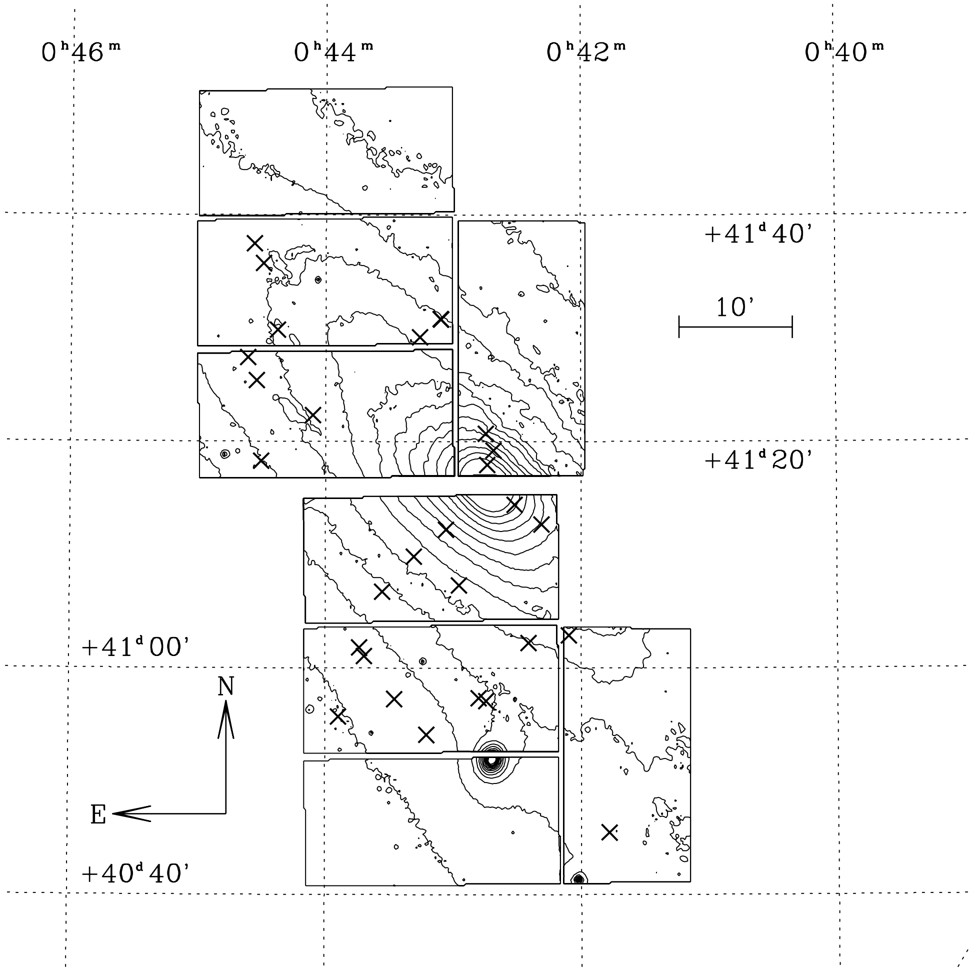

Fig. 8.—Locations of the 19 candidates in Table 4 plus the nine previously identified novae from Table 3. This is the entire sample of candidates with a network probability >0.95.

the PSF is not roundish with a size controlled by the seeing, then this suggests that the candidates may be fakes. Performing this test yields the result that perhaps 10 of the candidates are spurious (2973, 6251, 39995, 42075, 86234, 86283, 88205, 89701, 93095, and 95935). Reassuringly, we note that none of these are classified as having most nova features in Table 4. Finally, the nova catalog also contains 19 contaminants that appear to be true variable objects and are primarily made up of the light curves of superpixels covering periodic stars such as Miras and Cepheids, although many light curves exhibit some other superposed form of variability.

8. CONCLUSIONS

This paper has presented working neural networks for the identification of fast CNe. The use of *K*-fold cross validation and the choice of preprocessing techniques (i.e., reducing the light curve to a suitably binned and normalized power spectrum) has produced a set of neural networks capable of detecting the fast CNe present in the POINT-AGAPE survey. This conclusion is borne out by the consistently high nova probabilities assigned to the previously identified novae with 30 days $< t_{5\%} \lesssim 140$ days, the detection of four strong new nova candidates in the POINT-AGAPE catalog, and a further 15 possible candidates. This adds further weight to the claims by a number of authors (Wozniak et al. 2001; Belokurov et al. 2003, 2004; Brett et al. 2004) that neural networks offer a promising solution to the problem of light-curve identification in massive variability surveys.

The variation of *K*-fold cross validation used in this paper is new and particularly well adapted to the search for rare objects in a large data set. Usually, in *K*-fold cross validation, the data set is first partitioned into *K* separate sets. A network is then trained using a training set containing all of the data from $K - 1$ segments and tested on the remaining data. Our variation on this technique is to train the networks using just one POINT-AGAPE data segment before testing the networks on the remaining $K - 1$ segments. This is beneficial, as the processing time is substantially reduced. In many circumstances, there would be a risk of training set contamination using this variation on *K*-fold cross validation. However, CNe are very scarce in the POINT-AGAPE data set. So, the POINT-AGAPE light curves themselves can be used for the nonnova examples in the training set with little risk of contamination. The nova examples produced in the training set must be produced with templates. This method, therefore, can be used to find any rare light curves in a massive variability survey, provided suitable templates exist.

Nonetheless, the networks cannot be used in their current form to obtain a nova rate for M31. Very fast novae are missing because the POINT-AGAPE sampling rate is just not good enough to detect them. As demonstrated by Figure 6, the networks also do not detect enough slow, bumpy novae. Furthermore, these novae are more often than not assigned high probabilities, yet these probabilities fall below the classification cutoff because the networks produce too many false positives. The difficulty here is that artificial templates for slow novae are harder to construct, as they exhibit a greater morphology in the declining part of the curve. The best way to overcome this is to use known examples of slow CNe as part of the training set. Unfortunately, there are very few such light curves available in the *g*, *r*, and *i* passbands of the POINT-AGAPE survey. This, however, may become possible in the future using transformed colors. The extension of the networks to slow novae may also require modifications to the preprocessing technique, as the power spectra of slow novae are different (less sinclike) from those of fast novae.

Finally, it is worth mentioning that the limiting factor for detection of fast novae is actually the temporal sampling of the POINT-AGAPE data set. As fast CNe are the brightest CNe, they are still easy to detect even against the bright bulge of M31. Although we have not carried out a full efficiency analysis, it is clear that the networks successfully detect the CN types on which the system was trained, up to the limit imposed by the temporal sampling.

V. B. and E. J. K. are supported by the Particle Physics and Astronomy Research Council of the United Kingdom, while J. A. is supported by the Leverhulme Trust. Work by A. G. is supported by NSF grant 02-01266.

REFERENCES

An, J. H., et al. 2004, MNRAS, 351, 1071
Ansari, R., et al. 1999, A&A, 344, L49
Aurière, M., et al. 2001, ApJ, 553, L137
Bailer-Jones, C. A. L. 1997, PASP, 109, 932
Belokurov, V., Evans, N. W., & Le Du, Y. 2003, MNRAS, 341, 1373
———. 2004, MNRAS, 352, 233
Belokurov, V., et al. 2005, MNRAS, 357, 17
Bishop, C. M. 1995, Neural Networks for Pattern Recognition (New York: Oxford Univ. Press)
Bode, M., & Evans, A. 1989, Classical Novae (London: Wiley)
Brett, D., West, R. G., & Wheatley P. J. 2004, MNRAS, 353, 369
Cohen, J. G. 1985, ApJ, 292, 90
Darnley, M. J., et al. 2004, MNRAS, 353, 571
Hubble, E. 1929, ApJ, 69, 103
Irwin, M., & Lewis, J. 2001, NewA Rev., 45, 105
Lahav, O., Naim, A., Sodré, L., Jr., & Storrie-Lombardi, M. C. 1996, MNRAS, 283, 207
Le Du, Y. 2000, Ph.D. thesis, Université Pierre et Marie Curie - Paris VI
MacKay, D. J. C. 2003, Information Theory, Inference and Learning Algorithms (Cambridge: Cambridge Univ. Press)
Melchior, A.-L., et al. 1999, A&AS, 134, 377
Paulin-Henriksson, S., et al. 2002, ApJ, 576, L121
———. 2003, A&A, 405, 15
Payne-Gaposchkin, C. 1957, The Galactic Novae (Amsterdam: North Holland)
Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. 1992, Numerical Recipes in C: The Art of Scientific Computing (Cambridge: Cambridge Univ. Press)
Prialnik, D., & Kovetz, A. 1995, ApJ, 445, 789
Sterken, C., & Jaschek, C. 1996, Light Curves of Variable Stars (Cambridge: Cambridge Univ. Press)
Wozniak, P. R., et al. 2001, BAAS, 33, 1495