PUBLICATIONS OF THE ASTRONOMICAL SOCIETY OF THE PACIFIC, **122**:839–852, 2010 July

Inference of Surface Parameters from Near-Infrared Spectra of Crystalline H_2O Ice with Neural Learning

LILI ZHANG,[1,2] ERZSÉBET MERÉNYI,[2] WILLIAM M. GRUNDY,[3] AND ELIOT F. YOUNG[4]

Received 2010 January 27; accepted 2010 May 21; published 2010 June 22

ABSTRACT. The near-infrared spectra of icy volatiles collected from planetary surfaces can be used to infer surface parameters, which in turn may depend on the recent geologic history. The high dimensionality and complexity of the spectral data, the subtle differences between the spectra, and the highly nonlinear interplay between surface parameters make it often difficult to accurately derive these surface parameters. We use a neural machine, with a Self-Organizing Map (SOM) as its hidden layer, to infer the latent physical parameters, temperature and grain size from near-infrared spectra of crystalline H_2O ice. The output layer of the SOM-hybrid machine is customarily trained with only the output from the SOM winner. We show that this scheme prevents simultaneous achievement of high prediction accuracies for both parameters. We propose an innovative neural architecture we call Conjoined Twins that allows multiple (k) SOM winners to participate in the training of the output layer and in which the customization of k can be limited automatically to a small range. With this novel machine we achieve scientifically useful accuracies, $83.0 \pm 2.7\%$ and $100.0 \pm 0.0\%$, for temperature and grain size, respectively, from simulated noiseless spectra. We also show that the performance of the neural model is robust under various noisy conditions. A primary application of this prediction capability is planned for spectra returned from the Pluto-Charon system by New Horizons.

1. REMOTE EXPLORATION OF PLANETARY SURFACES FROM SPECTRA—APPLICATION TO THE NEW HORIZONS MISSION

Surface parameters, such as grain size and temperature, often have a wide range of implications on the formation history of landforms and atmospheres of solar system objects. For icy planetary bodies (e.g., Kuiper Belt objects and major satellites around giant planets), grain size can be tied to an object's recent volatile transport history and thermal processes of crystallization or radiative disruption on its surface (Clark et al. 1983; Hansen & McCord 2004). In the case of ice temperatures, the vapor pressures of many ices are steep exponential functions of their temperatures. Hence, temperature can be useful as a constraint on an object's atmospheric column abundance (Tryka et al. 1993). However, the available knowledge of these surface parameters is fairly limited. For example, the constraints on Pluto's surface pressure are indirect. Pluto's N_2 ice is in vapor pressure equilibrium with the N_2 atmosphere, buffered at a temperature of 40 ± 2 K during the early 1990s, as estimated from a weak spectral feature of N_2 ice at 2.147 μm (Tryka et al. 1994). This N_2 ice temperature is likely to change seasonally, leading to large changes in atmospheric pressure (Elliot et al. 2003; Sicardy et al. 2003; Spencer et al. 1997).

NASA's New Horizons space mission, which is a one-way journey to the Kuiper Belt and beyond, is expected to investigate the icy surfaces of remote planetary bodies such as Pluto, Charon, Nix, and Hydra (Young et al. 2008). The onboard infrared imaging spectrometer (Reuter et al. 2008) will map the surfaces of Pluto and Charon in 2015 at wavelengths from 1.25 to 2.5 μm. From these acquired near-infrared (NIR) ice spectra, the physical parameters that influence the reflectance values across all wavelengths can potentially be inferred (Grundy & Schmitt1998; Grundy et al. 2002).

However, accurate inference of these parameters is nontrivial. The complicated manifestation of and interplay between the physical parameters underlying the measured spectra, the often subtle variation in the spectral shapes in response to significant changes in parameters, and the high dimensionality of the spectra pose considerable challenges. Previous efforts to infer surface parameters have not yet produced satisfactory results. A simple and purely empirical method was proposed by Fink & Larson (1975) for retrieving H_2O ice temperatures from reflectance spectra. They developed a calibration curve of a feature at 6056 cm^{-1} (~1.65 μm) and used it to determine ice temperatures for the Galilean satellites Europa, Ganymede, and the rings of Saturn. Their method was limited to objects that display H_2O ice absorptions. Another competing method to infer H_2O ice temperatures was discussed by Grundy et al. (1999),

[1] Rice Quantum Institute, Rice University, Houston, TX 77005; llzhang@rice.edu.

[2] Department of Electrical & Computer Engineering, Rice University, Houston, TX 77005.

[3] Lowell Observatory, Flagstaff, AZ 86001.

[4] Space Studies Department, Southwest Research Institute, Boulder, CO 80302.

involving construction of a suite of models with various free parameters, fitting them to the spectra and also fitting them to different segments of the spectra separately, and comparing the resulting collection of best-fit temperatures. If most of the models and most of the spectral segments agreed on the temperature, that temperature was accepted as likely correct. When different models gave systematically different temperatures, the results were taken as probably meaningless. Such techniques worked well for some applications. However, when the surface conditions were more complicated than the model could address, such as in the presence of mixed phases or complex grain size distribution, this method produced poorer results.

We approach the challenge of parameter inference with neural learning, specifically with a hybrid supervised neural architecture containing a Self-Organizing Map (SOM) (§ 2.3). As we will see, this approach yields high, scientifically useful accuracies. The work presented in this article has three necessary components for the development of a comprehensive pattern recognition system for inferring surface parameters from NIR spectra:

1. Investigation of the approach with noiseless spectra of crystalline H_2O ice (§ 3);
2. Analysis of the robustness of the neural models to different levels of noise (§ 4);
3. Classification of spectra of different ice species (§ 5).

We start with H_2O ice because it is one of the most commonly found materials in the solar system, for which a great deal of relevant data and experience have been accumulated. We focus on crystalline (as opposed to amorphous) H_2O ice because observations of Charon are consistent with crystalline ice (Brown & Calvin 2000) and because amorphous ice is a less interesting problem, with virtually no sensitivity to temperature in its spectrum. In follow-up work we will investigate the inference capability of the neural approach under increasing levels of realism such as extension to other types of ices (CO_2, CO, NH_3) and to mixtures of ices. These efforts can ultimately lead to the capability of accurately inferring surface conditions from ice spectra such as those returned by New Horizons.

2. NEURAL NETWORK MODELING FOR INFERENCE OF SURFACE PARAMETERS

2.1. Methodology and Training Data

We need a large number of labeled spectra to train the neural models. However, real spectra of Pluto and Charon with sufficient resolution in spatial and other aspects are scarce. Training with real data sets does not necessarily produce a model that spans all plausible ranges of physical parameters. In such situations, and if available, realistic synthetic data or laboratory spectra can be used for training. For example, Gilmore et al. (2004) developed a carbonate identifier by training a back propagation neural network with a large number of laboratory spectra of carbonate and noncarbonate minerals. The resulting autonomous system was successful in various simulated Martian scenarios. Ramsey et al. (2002) presented a Bayesian approach for mineral identification from NIR reflectance spectra. The recognition rate in the experiments with laboratory spectra of a variety of solid and powdered rock samples exceeded that of human experts. Similarly, we use synthetic spectra for the training of our neural machines to fit models to the entire NIR spectral range (as opposed to piecewise models). We then validate the inference capability of the trained models with synthetic test spectra. Reliably performing models will be deployed to infer unknown parameters from spectra taken from real planetary surfaces.

The synthetic spectra are generated on a parameter grid through a radiative transfer code (Grundy 1995) based on the Hapke model, which is the most common way to describe the interaction of a solid particulate surface with incident sunlight (Hapke 1993; Cruikshank et al. 1998). The ice optical constants used in the Hapke model are also synthetically generated with a model that fits to laboratory spectra with 17 temperature-dependent Gaussians (Grundy & Schmitt 1998). The parameter grid has 126 temperatures with 2 K spacing between 20 and 270 K, and 9 grain sizes logarithmically spaced from 0.0003 to 3.0 cm. This set of parameters covers a meaningful range of surface conditions for the Pluto-Charon system, at sufficient resolution for scientific study. The spectral resolution, 230 band passes in the NIR range (1–2.5 μm), is close to the resolution of the sensor used on the New Horizons spacecraft (Reuter et al. 2008).

2.2. Challenges in Inferring Latent Variables

From a computational standpoint, temperature and grain size can be called *latent variables*, because they are not directly observed, but are rather inferred from directly measured *observable variables* (the reflectance values). They have a *global* influence on the spectral shapes, as opposed to ions or molecular compounds, which cause *local* absorptions (limited to some wavelengths) and therefore may be determined from individual absorption bands. Temperature and grain size, however, are hard to infer from piecewise spectral segments, because they have complicated and intertwined effects on the reflectance values. As seen from the sample spectra in Figure 1, both parameters deepen the absorptions, e.g., at 1.3 μm and 1.65 μm, but the change in the band depths is a nonmonotonic function of grain size. Hence, we build models to infer the two parameters from the entire NIR spectra, which potentially have more information than single absorption bands. In addition, temperature has a much smaller effect on the spectral brightness than grain size. This results in many crossovers between spectra with different temperatures (Fig. 1, top); the subtle changes in spectral shapes caused by temperature makes the differentiation between temperatures difficult. A sensitive classifier is needed to distinguish between these spectra such that accurate inference can be made.

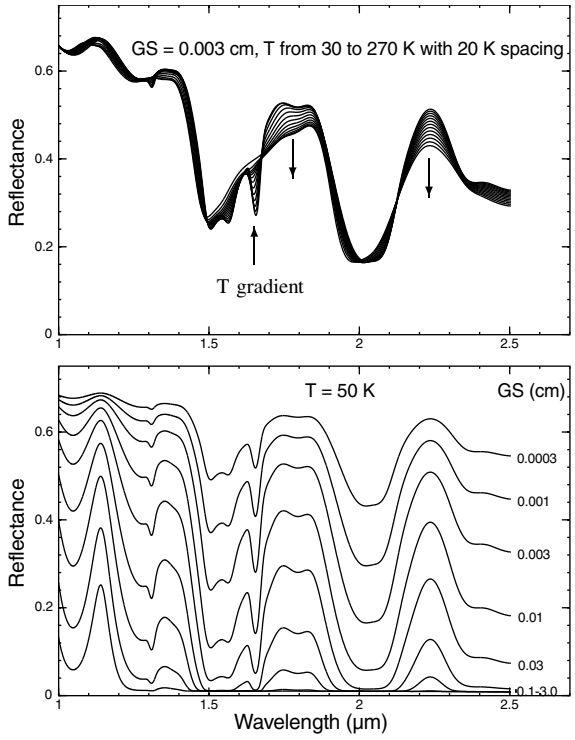

FIG. 1.—Sample synthetic spectra of crystalline H_2O ice. *Top:* variation in the spectral shape as a function of temperature (T), for one fixed grain size (GS), 0.003 cm. *Bottom:* variation in the spectral shape as a function of grain size, at 50 K.

2.3. The SOM-Hybrid Architecture for Supervised Learning

The neural architecture we use is a two-layer fully connected feedforward network, with a two-dimensional (2D) SOM as the hidden layer (Fig. 2). The 230-neuron input buffer takes a 230-band spectrum $\mathbf{x}(=[x_1, x_2, \ldots, x_{230}]^T)$ as an input vector in each learning step. Each neuron i in the SOM lattice A, of N neurons, is connected to all inputs via a 230-element weight vector $\mathbf{w}_i$ (the ith row vector in a $N \times 230$ weight matrix $\mathbf{W}$ that

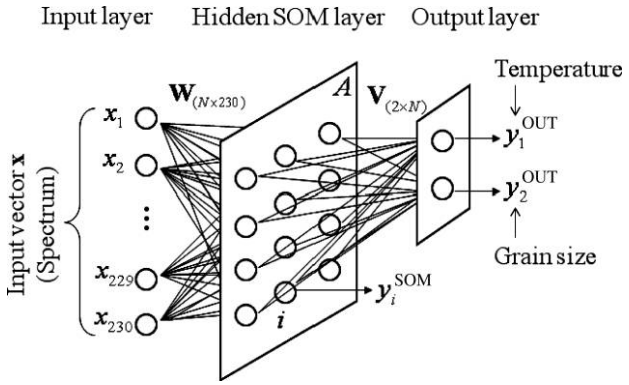

FIG. 2.—SOM-hybrid neural architecture.

connects the input buffer and the hidden layer). The output layer, connected to the SOM by a $2 \times N$ weight matrix $\mathbf{V}$, has two neurons that compute y_1^{OUT} and y_2^{OUT}, the inferred values of the physical parameters. This machine performs a topologically ordered mapping of the input spectra on the SOM lattice in a first, unsupervised learning phase, during which the output layer is idle. The resulting map reflects the similarity relationships of the high-dimensional input data, thus clusters can be identified. For general background on SOMs we refer to Kohonen (2001). We use a lesser-known variant, the Conscience algorithm (DeSieno 1988), of which a brief summary is given later in this section. Upon convergence of this phase, i.e., when the SOM weights $\mathbf{W}$ no longer change significantly, the output layer is turned on, to perform the second, supervised, learning according to equations (1) and (2). The outputs of the output layer are computed as

$$y_p^{\mathrm{OUT}} = \sum_{i \in A} v_{pi} y_i^{\mathrm{SOM}} \qquad p = 1, 2. \tag{1}$$

v_{pi} is the element in the pth row and ith column of the weight matrix $\mathbf{V}$. Temperature and grain size are learned through adjustment of v_{pi} to minimize the total squared error of the outputs according to the so-called delta rule (Widrow & Smith 1964):

$$\Delta v_{pi} = \alpha y_i^{\mathrm{SOM}} (Y_p^{\mathrm{OUT}} - y_p^{\mathrm{OUT}}) \tag{2}$$

where Y_p^{OUT} is the desired output, α is a learning rate. During the supervised training of the output layer, the SOM can continue its unsupervised learning with a small learning rate, for fine-tuning of the SOM weights $\mathbf{W}$.

This construction has been successful in a number of planetary applications. Howell et al. (1994) revised asteroid taxonomy by analyzing the clusters identified from an SOM and making the classification scheme more self-consistent. Another example is the accurate classification of a large number of clay-bearing soils with subtle spectral differences due to different clay species, for landslide hazard study from AVIRIS imagery (Rudd & Merényi 2005). The capability of this neural architecture to finely discriminate spectral shapes is a result of using an SOM in the hidden layer. We use the same architecture to target the parameter inference problem in this study.

SOMs, in general, are adaptive vector quantizers that place quantization prototypes (SOM weights $\mathbf{w}_i$, $i \in A$) in the data space such that they best represent the manifold distribution. (To distinguish the SOM weights $\mathbf{W}$ and the weights of the output layer $\mathbf{V}$, we call the SOM weights prototypes.) At the same time, the prototypes are organized (indexed) on the rigid low-dimensional lattice A according to their similarities. This allows capture of clusters in high-dimensional space from the 2D lattice without reduction of dimensionality. We use the Conscience variant (DeSieno 1988) of the original Kohonen SOM because it can achieve maximum entropy mapping. This means that

approximately equal numbers of data samples (spectra) will be represented by the SOM prototypes, providing the best possible approximation of the data probability density function (*pdf*) with the given number of prototypes. Maximum entropy mapping (encoding of input data by vector quantization prototypes) is optimal for information transfer. The Kohonen SOM has been shown suboptimal in this sense. The following describes the Conscience algorithm. The prototypes $\mathbf{w}_i$ ($i \in A$) are updated through many iterations (learning steps) described through equations (3)–(6). An input vector $\mathbf{x}$ is first randomly selected. An SOM winner neuron (or best matching unit, BMU) c is found such that

$$\|\mathbf{w}_c - \mathbf{x}\|^2 - b_c \leq \|\mathbf{w}_j - \mathbf{x}\|^2 - b_j \quad \forall\ j \in A \tag{3}$$

with a bias b_j applied to each neuron j. The bias b_j is computed from the winning frequency p_j, of neuron j, as

$$b_j = \gamma(t) \times ((N \times p_j) - 1). \tag{4}$$

p_j is updated as

$$p_j^{\text{new}} = p_j^{\text{old}} + \beta(\delta_{jc} - p_j^{\text{old}}), \tag{5}$$

where δ_{jc} is the Kronecker delta, β and γ are user-specified parameters. $\gamma = 0$ makes the Conscience algorithm equivalent to the original Kohonen algorithm. After winner selection, prototypes $\mathbf{w}_j$ are updated as

$$\mathbf{w}_j^{\text{new}} = \mathbf{w}_j^{\text{old}} + \alpha(t) h_{c,j}(t)(\mathbf{x} - \mathbf{w}_j^{\text{old}}) \tag{6}$$

where $\alpha(t)$ is a time-decreasing learning rate. $h_{c,j}(t)$ is a neighborhood function that defines the activation of a SOM neuron j by the BMU c to learn from the same input $\mathbf{x}$. For the Kohonen SOM, the neighborhood is commonly a Gaussian kernel centered over the BMU, which needs to be large at the beginning and diminish with time, to help avoid distortions in the map. The Conscience variant, however, enables the use of a fixed and small neighborhood size, such as an 8-neighbor square configuration in this case, which lightens the computational burden compared to the Kohonen SOM.

In the supervised learning phase, the customary way to pass information from the SOM to the output layer is the winner-takes-all (WTA) mode, in which only the BMU c sends a nonzero signal, 1, to the weighted sum in equation (1), all other SOM neurons send 0. Hence, in the WTA mode, this weighted sum is reduced to one term:

$$y_p^{\text{OUT}} = v_{pc} \qquad p = 1, 2. \tag{7}$$

In § 3.2 we will introduce an alternative way by which multiple SOM winners can contribute to the training of the output layer.

3. LEARNING TEMPERATURE AND GRAIN SIZE FROM NOISELESS SPECTRA

We assess the neural modeling for noiseless spectra. as a necessary initial step. The performance we achieve on noiseless data will serve as a benchmark in the noise sensitivity analysis we give in § 4.

In the unsupervised training phase of the SOM, we use all 1134 synthetic spectra (126 temperatures, 9 grain sizes) of crystalline H_2O ice. In the supervised training phase, we conduct tenfold jackknifing (cross-validation) to assess the performance of the trained predictive models. In each jack-knife run, 1134 spectra are randomly split with a 1:9 ratio into a test and a training set. The prediction results are averages of 10 jack-knife runs.

3.1. Relationship Between the Parameters as Expressed in the SOM

Since the unsupervised learning phase is important for fine discrimination of the spectral shapes in subsequent supervised learning, it is useful to examine how the converged SOM reflects the manifold structure and, specifically, what can be seen in terms of the influence by temperature and grain size. We visualize the distance (in the data space) between the prototypes that belong to two adjacent SOM neurons as a "fence" between the respective cells (Fig. 3, left). The gray scale intensity of the "fence" is proportional to the this distance. With this visualization, we can find out how similar (or dissimilar) the SOM prototypes are. We can see that the SOM is clearly separated into grain size clusters, typically by double-fenced black corridors, such as the diagonal one that separates the dark blue and the magenta clusters from the light blue and the yellow clusters. Since each prototype is the average of all spectra that map to it, we can see the variation in the spectral shapes caused by different grain sizes across the grain size clusters (Fig. 4). The prototypes within each grain size cluster are organized with respect to temperature. The temperature-dependent features in the prototypes change in an orderly fashion from one end of the cluster to another (from top to bottom, left to right, and in other directions). Figure 3, right, illustrates this for the 0.003 cm (yellow) grain size group. The prototypes learned from spectra with low temperatures have a strong absorption at 1.65 μm (in red boxes). This feature gradually disappears toward the right for high temperatures (in red circles). The observed dominant effect of grain size on the SOM clustering can be explained by the grain size dominance on the reflectance spectra. The difference in the spectral brightness is substantial between two grain size categories (Fig. 1, bottom). In contrast, the changes in temperature cause negligible changes in brightness, but they cause band centers to shift, as well as significant changes in relative band depth (Fig. 1, top). Therefore, the structure of this data manifold as suggested by the clustering in the SOM agrees with the spectral properties we know.

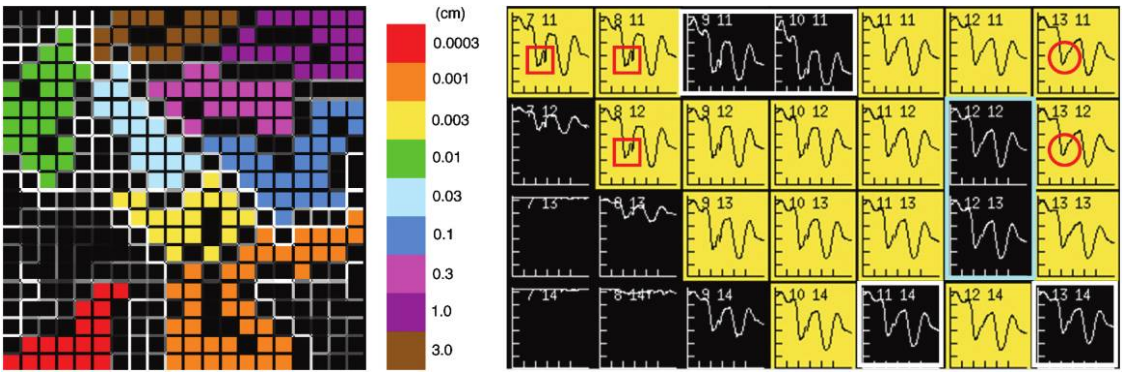

FIG. 3.—*Left:* A 20 × 20 SOM trained with synthetic spectra of crystalline H_2O ice. Grid cells represent SOM neurons. In the SOM, we only color the neurons that represent spectra of H_2O ice. The colors indicate the known grain sizes as keyed at right. The "fences" between adjacent cells have gray scale intensities proportional to the Euclidean distances (in feature space) between the prototypes of the respective neurons. White is large distance. The unlabeled (*black*) cells, such as those between the *red* and the *green* clusters, mostly indicate prototypes of spectra of other ices, N_2, CH_4 etc. This information is not shown here. Some black cells—typically in the narrow corridors between grain size groups, e.g., between the blue and the yellow clusters—are prototypes with no data mapped to them. Whether a prototype has data mapped to it is not shown in this representation. *Right:* An example of how spectra are organized within a grain size group according to temperatures. Here, part of the *yellow* group is shown with prototypes plotted in the SOM cells. An orderly change in the prototype shapes from left to right can be observed in response to increasing temperatures. The *red boxes* and *circles* exemplify differences in temperature-dependent absorption features at low and high temperatures, respectively. The *light blue* and *white boxes* indicate the empty prototypes of this grain size group, inside and at the boundaries, respectively.

Closer inspection of the SOM clustering reveals that, without exception, all input spectra mapped to any prototype within a grain size cluster have the same grain size label (not shown in Fig. 3, left). Thus, perfect learning of grain size can be easily achieved in the WTA mode, as in equation (7). In contrast, with approximately 25–30 prototypes in a grain size cluster, for the representation of 126 different temperatures, each prototype is forced to form an average (a mixture) of spectra across different temperatures. In the WTA mode, the output temperature value for any input spectrum that maps to a prototype is trained to approximate the average temperature represented by that prototype. This severely limits the resolution of the inferred temperature values. However, we can observe that, in general, each spectrum with a specific temperature also contributes to the spectral mixtures in several neighboring prototypes. This suggests that the inclusion of neighboring prototypes of the BMU in the weighted sum sent to the output layer may help better reconstruct a specific temperature. The question is which and how many neighbors to include. We propose a principled approach next, which also allows a natural customization for the prediction of different parameters.

3.2. *k*-Winners-Take-All and Choices of *k*

By the SOM formula, the response of any SOM neuron (SOM output) to an input vector is indicative of the similarity between the respective prototype and the input vector. In the most frequent implementations, the SOM output is either proportional to the inner product of the input vector and the SOM prototype (Kohonen 2001), or inversely proportional to the Euclidean distance between them. In the WTA mode, binary thresholding is applied to the SOM responses, assigning 1 to the BMU c and 0 to the rest as

$$y_i^{\mathrm{SOM}} = \begin{cases} 1 & i = c \\ 0 & i \neq c \end{cases}. \tag{8}$$

In this work, we allow k SOM winners, indexed as i_q ($q = 1, \ldots, k$) in the SOM lattice, to be simultaneously active with nonzero output values (k-winners-take-all or kWTA). y_i^{SOM} is inversely proportional to the distance d_i between the prototype $\mathbf{w}_i$ and the input vector $\mathbf{x}$:

$$y_i^{\mathrm{SOM}} = \begin{cases} 1 & i = i_1 (= c) \\ \frac{d_1}{d_1 + d_i} & i = i_2, i_3, \ldots, i_k. \\ 0 & i \neq i_1, i_2, \ldots, i_k. \end{cases} \tag{9}$$

Then these y_i^{SOM} are normalized as

$$y_{i_q}^{\mathrm{SOM}} = \frac{y_{i_q}^{\mathrm{SOM}}}{\sum_{q=1}^{k} y_{i_q}^{\mathrm{SOM}}}. \tag{10}$$

With k winners, the weighted sums in the output layer have k terms:

$$y_p^{\mathrm{OUT}} = \sum_{q=1}^{k} v_{pi_q} y_{i_q}^{\mathrm{SOM}} \qquad p = 1, 2. \tag{11}$$

Obviously WTA is a special case of kWTA with $k = 1$. NeuralWare's implementation in Neural Works Professional II/Plus (NeuralWare 2003) is a special case of this, for $k = 3$. In practice, k can be a much smaller number than N, because the outputs of most of the SOM neurons are insignificant for a given

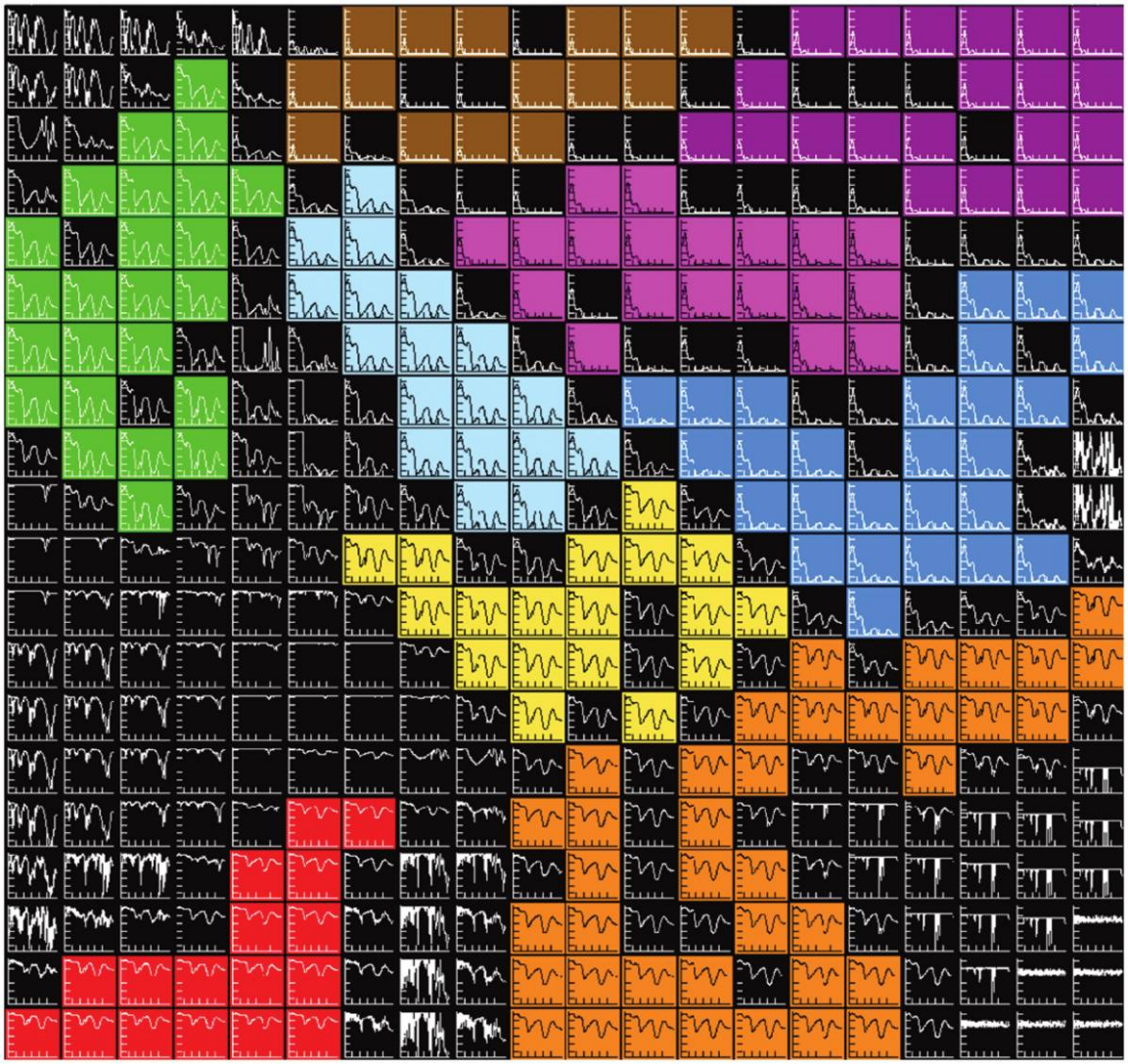

FIG. 4.—Learned prototypes plotted in their respective cells in the same SOM as in Fig. 3, *left*. The "fences" between adjacent SOM cells are turned off for clarity. An orderly change in the temperature-dependent features in the prototypes can be observed from one end of each grain size cluster to another. This can be seen in more detail for the yellow grain size group in Fig. 3, right.

input. We will use the term "interpolation off" or "noninterpolating mode" for $k = 1$ (or WTA), and use "interpolation on" or "interpolating mode" for $k > 1$ (or kWTA).

In order to find the right k for the inference of temperature and grain size, it is helpful to examine the neighboring relationships of the Voronoi cells that result from the segmentation of the data space (Voronoi tessellation) with respect to the SOM prototypes. A Voronoi cell comprises the data points represented by the prototype that serves as the Voronoi centroid. A derivative of the Voronoi tessellation is the Delaunay triangulation, in which two Voronoi centroids are connected by an edge if their cells share a boundary. Figure 5, left, provides an illustration of a Delaunay triangulation, for an artificial 2D data set "exclamation mark." This data set consists of two disconnected parts, in the shape of an ellipse and a circle, respectively. The gap between the two parts cannot be detected from the Delaunay triangulation. For a better representation of discontinuities in data manifolds, Martinetz & Schulten (1994) proposed the induced Delaunay triangulation. They showed that under certain circumstances, the edges of the induced Delaunay triangulation could be effectively constructed through so-called Hebbian learning (Hebb 1949) one form of which is SOM learning. A *connection* (an edge in the induced Delaunay triangulation) is defined between two Voronoi cells (as Voronoi neighbors) if their centroids form a pair of BMU and second BMU for at least one data point. In Figure 5, right, the induced Delaunay graph for the "exclamation mark" data is constructed through their method. The separation between the upper and lower parts of

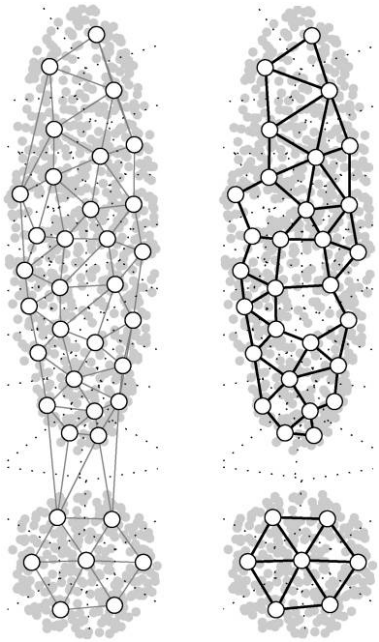

FIG. 5.—Illustration of Voronoi tessellation, Delaunay triangulation and induced Delaunay triangulation with a 2D "exclamation mark" data set, which has uniformly distributed data points (*gray dots*) in two disconnected regions. *Open circles* represent the 36 SOM prototypes that learned this data, and serve as Voronoi centroids. Voronoi cells are delineated by dotted lines. *Left:* The Delaunay triangulation (*thin solid lines*) does not help separate the two disconnected parts in the data set. *Right:* The induced Delaunay triangulation (*thick solid lines*) highlights the discontinuity.

the "exclamation mark" is made obvious by the disconnection in the graph. In a similar manner, we now further define *second-tier connections* between two Voronoi centroids that form a pair of BMU and third BMU for at least one data point. We will call the connections defined by Martinetz & Schulten (1994) *first-tier connections*. Visualization of these first-tier and second-tier connections, and possibly others, can help understand the connectedness of the manifold structure. For data dimensionality larger than 3, we cannot show the connections in the data space, but we can visualize them on the SOM, as in Figure 6 for the 230-dimensional ice spectra. We find that almost all first-tier as well as second-tier connections are between immediate lattice neighbors, with connection length $= 1$ (city block distance between the SOM cells at the two ends of a connection) (Villmann et al. 1997). The small number of first-tier connections with length > 1 are called *violating connections*, which indicate that prototypes neighboring in the data space are not indexed as neighbors in the SOM lattice. From an inspection of the ranks of the first-tier connections, we can determine that all violating connections for this data are *local* violations (Taşdemir & Merényi 2009), which reflect the smearing of temperature representation across neighboring prototypes, as discussed in § 3.1.

Having demonstrated the soundness of the map, we can further elaborate on what information the connections in Figure 6 convey. All first-tier and second-tier connections are confined within their respective grain size clusters. As a result, both second and third BMUs either have the same grain size label as the BMU, or have no label (with no spectra mapped to them). This confirms that the Voronoi neighbors of the BMU (connected in the SOM) cannot help the prediction of grain size because their inclusion in the weighted sums, equation (11), merely brings in redundant information. However, as suggested in § 3.1, the prediction of temperature can benefit from the Voronoi neighbors. The question is how many neighbors are sufficient. By definition, the number of first-tier connections of a given prototype gives the number of Voronoi neighbors for that prototype. The SOM prototypes representing spectra of H_2O ice have at most three Voronoi neighbors (Fig. 6, top), thus we may use up to 3 Voronoi neighbors to help improve the prediction of temperature. Further, we can quantify the importance of each Voronoi neighbor with its *connection strength*, which is defined as the number of data points that select this pair of prototypes as

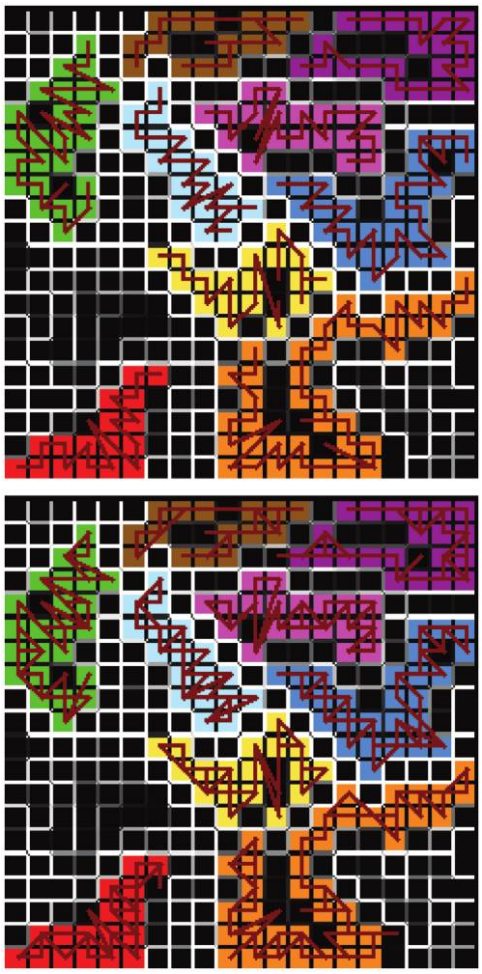

FIG. 6.—*Top:* First-tier connections between SOM prototypes (edges of the induced Delaunay triangulation) drawn as *maroon line segments* connecting cells in the same SOM as in Fig. 3, *left*. *Bottom:* Second-tier connections.

the BMU and the second BMU (Taşdemir & Merényi 2009). The Voronoi neighbors of a given prototype can be ranked according to their respective connection strengths, and average connection strengths can be computed for each ranking across all prototypes (Table 1). This statistics shows that there is no connection to fourth-ranking neighbors (i.e., any prototype has a maximum of three neighbors) and the third-ranking neighbors have negligible connections. This justifies $k = 3$ for the prediction of temperature for this data set.

3.3. *k*WTA Versus WTA in Supervised Learning

In this study, we do the supervised training for temperature and grain size in both the interpolating and the noninterpolating modes. The results are shown as correlations between predicted and true values, in Figure 7. Since both physical parameters have large ranges, we quantify the prediction accuracy as the percentage of test spectra for which the true parameter value was predicted with less than 5% relative error. The accuracies shown in Table 2, top left block, confirm that, as expected, the interpolating mode with $k = 3$ helps improve the prediction of temperature significantly, from 76.4% to 83.0%. The relatively poor results for the prediction of temperature occur mostly at the extreme values, ~20 K and ~270 K (Fig. 7, left block), due to limited availability of synthetic training spectra with optical constants in these ranges. With this boundary effect excluded from the calculation of the prediction accuracy, temperatures are predicted with 91.8% accuracy (Table 3) for test spectra within the [50 K, 240 K] range. In contrast, the prediction accuracy of grain size decreases considerably in the interpolating mode. Experiments that naturally come to mind for regaining the high accuracy for grain size, such as increasing the grid resolution of the training data, increasing the SOM size, or increasing both, are not helpful in this case, as seen from Table 2. For details we refer to Zhang et al. (2009). Although it is possible that with an SOM even larger than 40×40, or with many more training steps, we may be able to achieve the same accuracies as in Table 2, top left block, the extra computational cost required makes that solution undesirable for practical purposes. Discussion of the computational cost is given in § 3.4.2.

TABLE 1

STATISTICS OF CONNECTIONS TO VORONOI NEIGHBORS, FROM THE HIGHEST TO THE LOWEST RANKING, ACROSS SOM PROTOTYPES THAT REPRESENT H_2O ICE SPECTRA

Rank of neighbors	1	2	3	4
Number of connections	209	184	7	0
Average connection strength	7.6	3.7	1.1	0

3.4. Conjoined Twins–A Neural Architecture for Inference of Two Latent Variables

3.4.1. Novel Architecture

From the results described in § 3.3, we find that the opposing preferences of temperature and grain size for the interpolation modes cannot be satisfied simultaneously. However, we can encode this duality in a novel way into a neural architecture we call Conjoined Twins (Zhang et al. 2009). It has the same structure as in Figure 2, but with "twin heads," two copies of the output layer (Fig. 8). Both "heads" share the same "body" of knowledge in the SOM, but use it in customized ways. Head #1 pulls the SOM output only from the BMU (noninterpolating mode) for the training of the output layer (eq. [7]). This head becomes a grain size specialist, achieving perfect prediction for grain size. The prediction of temperature from its second output neuron is discarded. Similarly, head #2 specializes in temperature by drawing the outputs from the first three BMUs and forming a three-term weighted sum according to equation (11) (interpolating mode with $k = 3$). The grain size prediction from this head is discarded. Importantly, the customization of k for both heads is determined from the statistics of the connections between the SOM prototypes by straightforward computation, as discussed in § 3.2. With the customization of k in the Conjoined Twins we obtain high prediction accuracies for both parameters.

3.4.2. Computational Cost

The increase in computational cost with an additional "head" is relatively small compared to increasing the size of the SOM, for two reasons. First, the training of the SOM is typically longer (takes more training steps) than the training of the output layer. Second, the cost of each training step of the SOM is much larger than the cost of a training step in the output layer. With N SOM prototypes and dimension D, it takes $(3D + 6)N$ operations for the SOM to calculate the distances between an incoming input vector and all the prototypes, for winner selection (eqs. [3]–[5]), and $4DN$ operations for updating the SOM prototypes $\mathbf{w}_j$ (eq. [7]). For a 20×20 SOM ($N = 400$) that learns 230-dimensional data ($D = 230$), it takes 646,400 ($=278{,}400 + 368{,}000$) operations to learn from one input vector. For a 40×40 SOM ($N = 1600$), the number of operations is 2,585,600 (4 times larger) for one learning step. In contrast, adding a "twin head" carries a small overhead. In the noninterpolating mode (WTA), one training step has $5N$ operations: N for setting SOM outputs (eq. [8]), $2N - 1$ for calculating y_p^{OUT} in the output layer (eq. [1]), and $2N + 1$ for updating v_{pi} (eq. [2]). In the interpolating mode (kWTA), one step costs $3k + 5N - 2$ operations, where k is the customized number of SOM winners to be used by this "head." This includes $k + N - 1$ operations for setting the SOM outputs (eq. [9]), $2k - 1$ operations for normalizing the SOM outputs (eq. [10]), $2N - 1$ operations

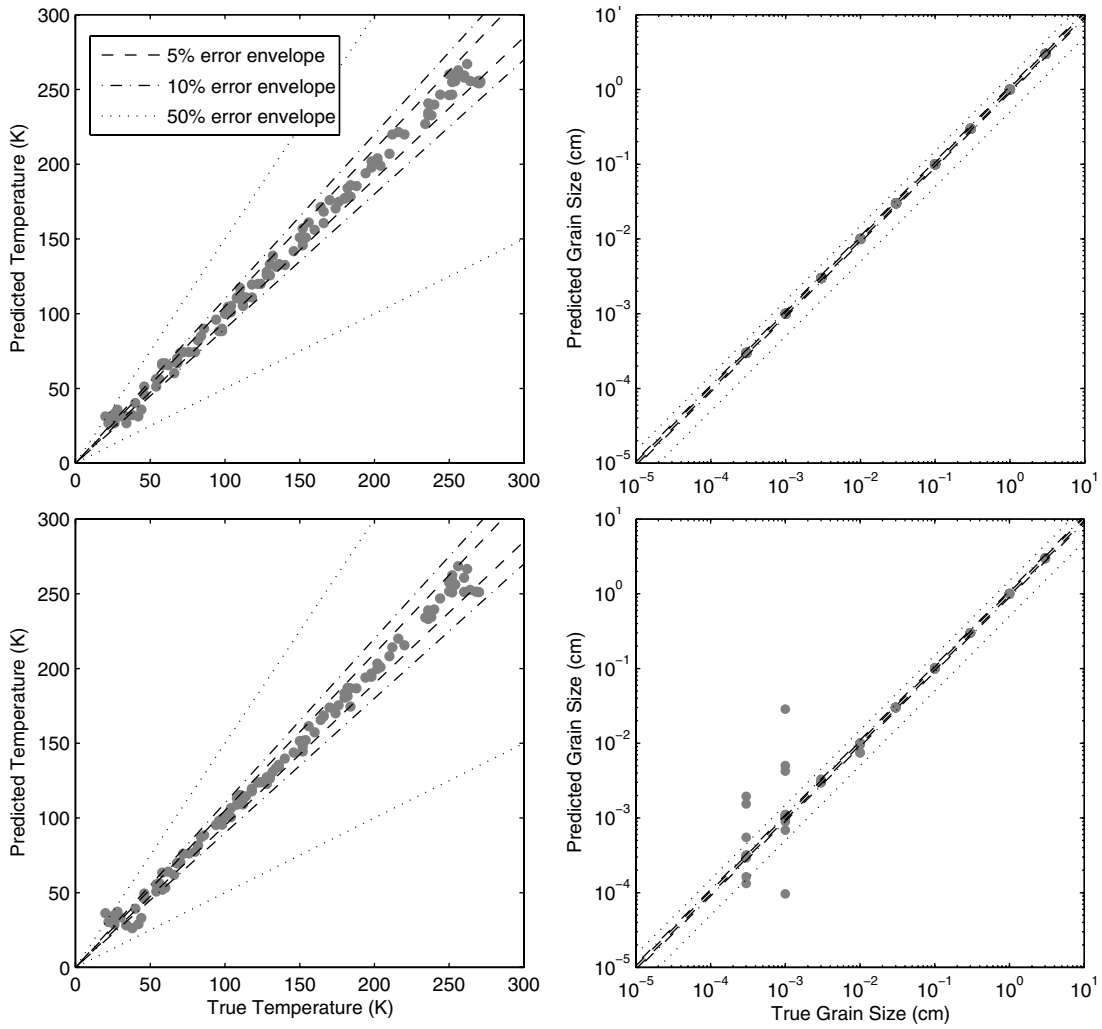

FIG. 7.—Correlation of predicted and true values of temperature (*left block*) and grain size (*right block*) (Zhang et al. 2009). Data are shown as gray dots. *Top:* Results obtained with the SOM in the noninterpolating mode. *Bottom:* Results obtained with the SOM in the interpolating mode. The *dashed, dash-dotted,* and *dotted lines* indicate 5%, 10%, and 50% error envelopes, respectively. The prediction of temperature is better in interpolating mode, while the prediction of grain size is better in noninterpolating mode (Table 2).

for calculating $y_p^{\rm OUT}$ in the output layer, and $2N+1$ for updating v_{pi}. As discussed in § 3.2, k is typically a small number ($\ll N$). For a 20×20 SOM, 2000 additional operations are needed for WTA, or 2007 for kWTA ($k = 3$). Hence, the extra computational cost of adding a "twin head" is negligible, and independent of the data dimensionality. This makes the Conjoined Twins approach especially suitable for high-dimensional data.

TABLE 2

PREDICTION ACCURACIES OF GRAIN SIZE (GS) AND TEMPERATURE (T) FOR TWO SEPARATE DATA SETS

		Data with 9 Grain Sizes		Data with 81 Grain Sizes	
Interpolation mode		Noninterpolating	Interpolating	Noninterpolating	Interpolating
20×20 SOM	GS	$100.0\% \pm 0.0\%$	$76.4\% \pm 4.4\%$	$74.1\% \pm 1.5\%$	$78.8\% \pm 1.7\%$
	T	$76.2\% \pm 2.6\%$	$83.0\% \pm 2.7\%$	$31.9\% \pm 1.2\%$	$52.5\% \pm 1.8\%$
40×40 SOM	GS	-	-	$97.8\% \pm 0.4\%$	$54.5 \pm 1.2\%$
	T	-	-	$60.3\% \pm 1.2\%$	$77.9\% \pm 1.0\%$

NOTE.—The two data sets contain 9 and 81 grain sizes, respectively, with 20×20 and 40×40 SOMs, each in the interpolating ($k = 3$) and the noninterpolating modes (Zhang et al. 2009). Results are averages of 10 jack-knife runs.

TABLE 3

PREDICTION ACCURACIES OF GRAIN SIZE (GS) AND TEMPERATURE (T) WITH THE DATA SET CONTAINING 9 GRAIN SIZES AND THE 20×20 SOM

T (K)		Noninterpolating	Interpolating ($k = 3$)
[20,270]	GS	$100.0\% \pm 0.0\%$	$76.4\% \pm 4.4\%$
	T	$76.2\% \pm 2.6\%$	$83.0\% \pm 2.7\%$
[50,240]	GS	$100.0\% \pm 0.0\%$	$73.7\% \pm 4.7\%$
	T	$82.3\% \pm 3.7\%$	$91.8\% \pm 1.2\%$

NOTE.—Prediction accuracies are calculated for the whole data set with T$\in$[20 K, 270 K] and for the subset of data with T$\in$[50 K, 240 K], respectively. Results are averages of 10 jack-knife runs.

We can achieve 100% accuracy for the prediction of grain size in the interpolating mode ($k = 3$) by running the supervised phase for ~2 million steps, more than twice as long as with the Conjoined Twins (750,000 steps). This means that the inclusion of 3 SOM winners can produce good prediction for both parameters, but more computational time is required by this machine with single interpolation mode to perform as well as the Conjoined Twins for the inference of grain size. For a 20×20 SOM ($N = 400$) and 230-dimensional data ($D = 230$), it takes a total of $\sim 2.1 \times 10^{11}$ ($=(276{,}000 + 2000 + 2007) \times 750{,}000$) operations for the Conjoined Twins to learn both parameters well. The machine with single interpolation mode, $k = 3$, needs $\sim 5.6 \times 10^{11}$ operations ($=(276{,}000 + 2007 + 2007) \times 2{,}000{,}000$) to achieve similar results. Thus, the Conjoined Twins machine is a more economical solution to this parameter inference problem.

4. NOISE SENSITIVITY ANALYSIS

As our eventual goal is to infer surface parameters from real spectra acquired in space missions under various noise conditions, we conduct a noise sensitivity analysis to evaluate the robustness of our neural models. To address the noise conditions that are common for spectral measurements, we add noise to the noiseless data set used in the previous section (1134 spectra), producing noisy versions of the data with seven different signal-to-noise ratio (S/N) levels, $\mathrm{S/N} = 256$, 128, 64, 32, 16, 8, and 4. Since, in general, noise can be additive (due to sky background, dark current, and read noise, for example) and multiplicative (due to the Poisson shot noise of photons from the object itself), our noise generation routine lets the user specify two parameters: the amplitude of the generated noise vector (specified as a target S/N) and the fraction of the noise that is additive as opposed to multiplicative. Obviously, for bright sources, the multiplicative noise will dominate, and vice versa for faint sources. In all of the noise vectors generated for the purposes of this study, we assume that the additive and multiplicative noise sources are roughly equal. The two noise sources

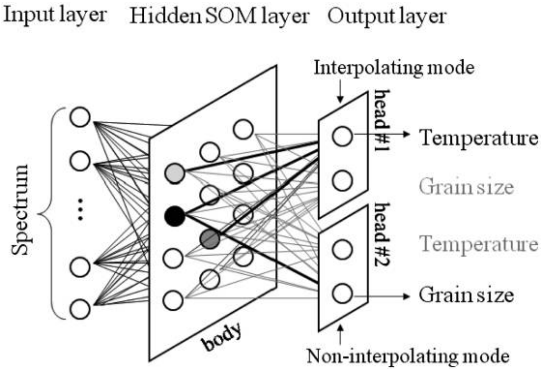

FIG. 8.—Conjoined Twins architecture. One "head" of the Conjoined Twins works in noninterpolating mode, using the output from the BMU (*black* neuron in the SOM) to predict grain size. The other "head," working in interpolating mode, uses the first, second, and third BMUs (*black, dark,* and *light gray* neurons) to predict temperature.

are assumed to be uncorrelated and are therefore added in quadrature by $\sigma_{\mathrm{total}} = (\sigma^2_{\mathrm{add}} + \sigma^2_{\mathrm{mult}})^{1/2}$, where σ represents the standard deviation of the noise. Because multiplicative noise depends on the spectral values themselves and additive noise does not, we choose a typical spectral value (the median of the spectrum, in fact) at which the multiplicative and additive noise components have equal standard deviations. We generate two batches of noisy data sets, one batch with 1 noisy version, the other with 10 noisy versions for each noiseless spectrum. We refer to these as NoisyData1 ($7 \times 1134 = 7938$ spectra) and NoisyData10 ($7 \times 10 \times 1134 = 79{,}380$ spectra).

The noise sensitivity analysis consists of two parts. First, we compare the models trained on the NoisyData10 data set with five S/N levels (inf, 256, 128, 64, and 32), as shown in Table 4. The inference capabilities of the resulting models are tested on data with eight different S/N levels (inf, 256, 128, 64, 32, 16, 8, and 4). For each case, we do 3 threefold jack-knife runs. The training set for each case comprises $\frac{2}{3}$ of the spectra with the training S/N, selected randomly. The remaining $\frac{1}{3}$ of the spectra, together with the spectra with other SNR levels, make up the corresponding test set. Second, we investigate the influence of the size of the noisy training set on the prediction accuracy, by comparing the models trained with the NoisyData1 and the NoisyData10 data sets.

In both experiments, we reuse the SOM that learned with the noiseless data and train the "twin heads" with the noisy data in the supervised learning phase. A rationale for reusing the SOM is that we expect to train the SOMs in our models mostly with synthetic data, thus, we have no limitation in using noiseless data. In addition, the use of the same SOM (trained with noiseless spectra) across all cases helps separate the effect of training the "twin heads" with different noisy data sets from the effect of training the SOM with noisy data. A follow-up task should be to assess what noise levels make significant difference in training the SOM.

TABLE 4

PREDICTION ACCURACIES FOR TEMPERATURE (T) AND GRAIN SIZE (GS) TABULATED FOR DIFFERENT S/N LEVELS OF THE TRAINING AND TEST DATA (INFINITY, INF)

		T Prediction Accuracy (%)					GS Prediction Accuracy (%)				
Training S/N		inf	256	128	64	32	inf	256	128	64	32
Test S/N	4	45.5	46.0	46.1[a]	46.0	46.1[a]	87.4	87.7	88.5	89.2[a]	87.8
	8	55.6	56.0	56.3	56.4[a]	55.9	94.5	94.7	95.3	95.9[a]	93.9
	16	64.7	65.8	66.0[a]	65.9	65.6	97.4[a]	97.4[a]	97.1	97.4[a]	93.9
	32	72.7	73.7	74.0[a]	73.0	71.5	98.8[a]	98.5	97.9	97.5	93.6
	64	78.4	79.3[a]	79.3[a]	77.0	74.3	99.3[a]	98.8	98.2	97.5	93.3
	128	82.0	82.4[a]	81.7	78.1	74.9	99.8[a]	99.0	97.9	97.2	93.1
	256	83.5[a]	83.3	81.8	78.2	74.6	99.9[a]	99.0	97.9	97.2	92.8
	inf	83.0[a]	82.5	80.8	76.1	72.3	100.0[a]	99.0	97.8	97.1	93.0

NOTE.—Each prediction accuracy is an average of 3 jack-knife runs. Variances of all prediction accuracies are less than 0.6 for T, and less than 0.3 for GS (not shown here).

[a] Maximum prediction accuracies for test sets with different S/N levels.

The best results in Table 4 show that the difference between the S/N levels of the training and the test data is important for the prediction of temperature. The training set with the highest S/N, namely the noiseless set, does not always generate the best predictions. For instance, from the noisy data with $S/N = 64$, temperature is best inferred with the model trained on data with $S/N = 128$. This makes sense because training with noisy data is similar to training with a larger variety of noiseless training samples that have the same variance as the noisy data. This helps the prediction from noisier samples which have an even larger variance than the training data. Table 4 suggests that the training sets with S/N 2–8 times as high as the S/N of the test data produce the highest accuracies. In contrast, for grain size, the noiseless training set produces the best prediction accuracy for all test sets with $S/N \geq 16$. The markedly lower accuracies produced by the models resulting from noisy training data can be explained by noise-induced blurring of the boundaries between grain size groups. Two noisy spectra with different grain sizes can map to the same SOM prototype at the boundary of two clusters (such as the ones in the white boxes in Fig. 3, right). This causes confusion during the training of the grain size specialist "head". For the test sets with SNR levels 4 and 8, the best results are produced with the models with $S/N = 64$ of the training data. However, for these two test sets, the advantage of the best models over others is small ($\leq 2.0\%$ increase in accuracy), thus may not be conclusive.

Tables 5 and 6 show the prediction accuracies achieved with the NoisyData10 and the NoisyData1 data sets, respectively. Their difference, in Table 7, indicates the improvements in prediction accuracies caused by the larger sizes of the training sets. For the prediction of temperature, the improvement in accuracy is prominent when the test data set has an S/N level at least twice as high as the S/N level of the training data. When the S/N level of the training set is 8 times as large as the S/N level of the test data, the advantage of using 10 noisy versions for training vanishes. For the prediction of grain size, however, the tendency is consistent. The results with NoisyData10 are always better than with NoisyData1. One general conclusion from Table 7, for both parameters, is that, in most cases the noisier the training set the greater improvement in accuracy can be achieved with more (in this case 10 times more) noisy training spectra.

The results described here demonstrate good consistency in the performance of the neural models under a wide range of noisy conditions. The statistics in Tables 4–7 will help choose the most suitable model for inference of temperature and grain size from real spectra when noise estimate for real data is available.

5. CLASSIFICATION OF ICE SPECIES

In addition to crystalline H_2O ice, a variety of ice types occur in the Pluto-Charon system. In this section we show the capability of the SOM-hybrid neural machine for distinguishing

TABLE 5

PREDICTION ACCURACIES PRODUCED WITH THE NOISYDATA10 DATA SET, CONTAINING 10 NOISY VERSIONS FOR EACH NOISELESS SPECTRUM

		T Accuracy (%)			GS Accuracy (%)		
Training S/N		128	64	32	128	64	32
Test S/N	4	46.1	46.0	46.1	88.5	89.2	87.8
	8	56.3	56.4	55.9	95.3	95.9	93.9
	16	66.0	65.9	65.6	97.1	97.4	93.9
	32	74.0	73.0	71.5	97.9	97.5	93.6
	64	79.3	77.0	74.3	98.2	97.5	93.3
	128	81.7	78.1	74.9	97.9	97.2	93.1
	256	81.8	78.2	74.6	97.9	97.2	92.8
	inf	80.8	76.1	72.3	97.8	97.1	93.0

NOTE.—This table shows a subset of the results in Table 4, for easy comparison with Table 6.

TABLE 6

PREDICTION ACCURACIES PRODUCED WITH THE NOISYDATA1 DATA SET, CONTAINING ONE NOISY VERSION FOR EACH NOISELESS SPECTRUM

		T Accuracy (%)			GS Accuracy (%)		
Training S/N		128	64	32	128	64	32
Test S/N	4	46.6	46.3	46.0	88.5	87.8	87.0
	8	56.7	57.0	56.9	94.4	92.4	90.9
	16	66.7	65.3	64.3	94.8	93.6	89.4
	32	73.2	71.7	-	97.3	96.5	-
	64	78.9	-	70.3	97.5	-	90.4
	128	-	76.2	70.5	-	94.9	89.0
	256	79.3	75.7	70.5	96.8	95.7	89.4
	inf	77.1	71.9	67.4	96.6	95.3	89.4

NOTE.—Entries are missing when the test and the training S/N levels coincide, because in those cases the sole noisy version is pulled into the training set and the test set is empty.

between N_2, CH_4, H_2O, CO_2, NH_3, and CH_3OH ices. We use the same neural architecture as shown in Figure 2, but this time we perform classification with it, instead of regression. For this, we encode the class labels of the 6 ice types as six-dimensional unit vectors at the output layer, which now has 6 neurons. The target output vector for an input spectrum of ice type i ($i = 1, 2, \ldots, 6$) is written as $[e_1, e_2, e_3, e_4, e_5, e_6]^T$ ($e_j = \delta_{ij}, j = 1, 2, \ldots, 6$), where δ_{ij} is the Kronecker delta. The class (ice type) of a spectrum is predicted by the component with the largest value in the output vector, produced by the trained model.

In this experiment, we use a relatively small data set that was not systematically generated on a regular grid with respect to temperature and grain size. The number of spectra available for each ice type is shown in parenthesis in Table 8. The resolution of temperature for the first three ice types is ~10 K, but within different temperature ranges due to the availability of the related optical constants. The spectra of these three ice types have nine different grain sizes in the 0.0001–3.0 cm range. The spectra of the remaining three ice types in Table 8 have only one temperature but 12 different grain sizes in the 0.0001–0.4 cm range. This mixed availability of parameter resolutions might be the case in reality, therefore assessment of classification performance with these data is of value.

TABLE 7

DIFFERENCE OF TABLE 5 AND TABLE 6, SHOWING THE IMPROVEMENT IN PREDICTION ACCURACIES WITH NOISYDATA10 COMPARED TO NOISYDATA1

		Improvement in T Accuracy (%)			Improvement in GS Accuracy (%)		
Training S/N		128	64	32	128	64	32
Test S/N	4	−0.5	−0.3	0.1	0	1.4	0.8
	8	−0.4	−0.6	−1.0	0.9	3.5	3.0
	16	−0.7	0.6	1.3	2.3	3.8	4.5
	32	0.8	1.3	-	0.6	1.0	-
	64	0.4	-	4.0	0.7	-	2.9
	128	-	1.9	4.4	-	2.3	4.1
	256	2.5	2.5	4.1	1.1	1.5	3.4
	inf	3.7	4.2	4.9	1.2	1.8	3.6

TABLE 8

CLASSIFICATION ACCURACIES FOR 6 ICE TYPES ACHIEVED BY THE SOM-HYBRID CLASSIFIER IN THE INTERPOLATING AND THE NONINTERPOLATING MODES

Ice Type (number of spectra available)	Classification Accuracy (%)	
	Noninterpolating	Interpolating
N_2 (63)	98.3 ± 4.1	100.0 ± 0.0
CH_4 (63)	91.7 ± 9.8	100.0 ± 0.0
H_2O (234)	93.1 ± 4.5	100.0 ± 0.0
CO_2 (12)	83.3 ± 25.8	100.0 ± 0.0
NH_3 (12)	75.0 ± 27.4	100.0 ± 0.0
CH_3OH (12)	91.7 ± 20.4	100.0 ± 0.0
Overall Accuracy (%)	92.8 ± 2.5	100.0 ± 0.0
Average Accuracy (%)	88.9 ± 6.0	100.0 ± 0.0

NOTE.—Classification results are averages of 6 jack-knife runs. The overall and average accuracies are the correctly classified test spectra, and the average of class accuracies, respectively.

Similarly to the regression of parameters, training of this classifier consists of the unsupervised and supervised learning phases. In the unsupervised phase, we use all 396 spectra to train the SOM. After the SOM converged, we perform 6 sixfold jack-knife runs for the supervised learning. The data are split in a 1:5 stratified random sampling for each jack-knife run, to ensure samples from all classes in every fold.

As seen from Table 8, the neural model trained in the interpolating mode outperforms the model trained in the noninterpolating mode. The significantly lower accuracies obtained with the noninterpolating mode is not a result of obscurity in the boundaries between different ice clusters in the SOM, but a result of the small training data set. In this SOM (not shown here), clusters representing 6 different ice types separate cleanly, similarly to the separation between grain size clusters shown in Figure 3, left. However, due to the small number of the spectra used in the supervised training, only a portion of the SOM neurons fire during the training of the output layer. Many weights that connect the output layer and the SOM do not have a chance to be trained in the noninterpolating mode. Whenever a test spectrum activates an SOM neuron that never won in the supervised training phase, the model produces a meaningless result. To remedy this, we switch to the interpolating mode with $k = 3$ so that each incoming spectrum stimulates 3 SOM winners. This potentially involves the updating of more weights in the training of the output layer. The 100% classification accuracies, in Table 8, show that the discrimination of pure ice species from NIR spectra will not be problematic. This encourages us to pursue

this approach for more complicated surface composition conditions.

6. DISCUSSION AND CONCLUSION

This article proposes an innovative neural learning architecture, Conjoined Twins, that enables effective inference of two latent physical parameters, temperature and grain size, from NIR synthetic spectra. We achieve perfect, $100.0\% \pm 0.0\%$ prediction accuracy for grain size and $83.0\% \pm 2.7\%$ for temperature. If we exclude the problematic end regions of the temperatures where adequate training data are unavailable, the prediction accuracy for temperature in the remaining 50–240 K range is $91.8\% \pm 1.2\%$ (Table 3). For Charon, and for N_2-ice–free regions of Pluto, diurnal and latitudinal temperature variations of tens of K are expected, with temperatures in the 40–70 K range. Our neural model needs to be able to retrieve temperatures with less than ~3 K error in order to be able to resolve these temperature differences and map the thermal inertia across these surfaces. Since 3 K represents ~5% error in this temperature range, it is reasonable to assume such performance for 80%–86% of the measured spectra according to Table 2, top left block. The obtained knowledge can in turn assist in unraveling the processes undergoing in the interiors and in the atmospheres of these planetary bodies. Our neural model also shows robustness under various noise conditions. The real spectral measurements returned by New Horizons are expected to have an S/N of ~50. From the noise sensitivity analysis reported in Table 4, we see that the prediction accuracies for grain size from spectra with $S/N = 32$ and $S/N = 64$ are nearly perfect (~99%), and temperature predictions within $<5\%$ error are produced for 74%–79% of the spectra, which is still useful for science.

In this study, the observed interplay between the two parameters and the neighboring relationships between the Voronoi cells of SOM prototypes help us justify the choices of k, the number of SOM winners to use in the supervised learning. k has been determined as 1 for the prediction of grain size, and 3 for the prediction of temperature, for this particular data set and with this 20×20 SOM. For other data, or for a different size of the SOM, examination of the SOM's knowledge prior to the supervised learning phase is necessary for correct customization of k. An upper limit of k can be found fully automatically from the statistics of the data distribution and the SOM's knowledge, such as in Table 1. However, determination of the optimum k for the individual latent variables needs interactive examination of the SOM. For example, the use of less than 3 SOM winners for temperature prediction was ruled out as it yielded poorer results than those obtained with $k = 3$. One desirable follow-up task is to replace this interactive finding of the best k value below the upper limit with a principled automated approach.

The sampling of training data plays an important role for the development of inference capability. One possibility to improve the prediction of temperature is to generate training spectra on a nonlinear temperature grid with higher resolution at the boundary temperature ranges where poor predictions are observed in this study with linear temperature grid (Fig. 7, left block).

In follow-up work we plan to test our approach for the inference of temperature and grain size from spectra of other ices. If we can achieve similar success as with crystalline H_2O ice, we will attempt to gradually incorporate more realistic factors, such as viewing and illumination geometry, macroscopic roughness of the surface, crystallinity, abundances of various ice types in areal mixtures and, further, in intimate mixtures. The gradually increasing level of realism will present escalating levels of difficulty in the neural modeling, which may require additional innovations. Step-by-step extensions to the Conjoined Twins presented in this article may help lead to models suitable for inference of surface conditions from real NIR planetary spectra returned by space missions.

This work was partially supported by grants NNG05GA63G and NNG05GA94G from the Applied Information Systems Research Program, NASA, Science Mission Directorate. William M. Grundy gratefully acknowledges partial support from NASA Planetary Geology & Geophysics grant NNG04G172G to Lowell Observatory.

REFERENCES

Brown, M. E., & Calvin, W. M. 2000, Science, 287, 107
Clark, R. N., et al. 1983, Icarus, 56, 233
Cruikshank, D. P., et al. 1998, Solar System Ices (Dordrecht: Kluwer), 579
DeSieno, D. 1988, in IEEE Intl Conference on Neural Networks, 1, 117
Elliot, J. L., et al. 2003, Nature, 424:165
Fink, U., & Larson, H. P. 1975, Icarus, 24, 411
Gilmore, M. S., et al. 2004, Icarus, 172, 641
Grundy, W. M. 1995, Ph.D. thesis, Univ. Arizona
Grundy, W. M., & Schmitt, B. 1998, J. Geophys. Res., 103, 25809
Grundy, W. M., et al. 1999, Icarus, 142, 536
Grundy, W. M., et al. 2002, Icarus, 155, 486
Hansen, G. B., & McCord, T. B. 2004J. Geophys. Res., 109, E01012
Hebb, B. 1993, Theory of Reflectance and Emittance Spectroscopy (New York: Cambridge Univ. Press)
Hebb, D. O. 1949, The Organization of Behavior (New York: Wiley)
Howell, E. S., et al. 1994, J. Geophys. Res., 99, 10847
Kohonen, T. 2001, Self-Organizing Maps, (3rd ed., Berlin: Springer-Verlag)
Martinetz, T., & Schulten, K. 1994, Neural Networks, 7:507
NeuralWare 2003, Neural Computing–A Technology Handbook for Professional II/PLUS and NeuralWorks Explorer (Pittsburgh: NeuralWare Inc.)

Ramsey, J., et al. 2002, Data Min. Knowl. Discov., 6, 277
Reuter, D. C., et al. 2008, Space Sci. Rev., 140:129
Rudd, L., & Merényi, E. 2005, in Proc. 14th AVIRIS Earth Science and Applications Workshop, ed. R. Green (Pasadena, CA: JPL)
Sicardy, B., et al. 2003, Nature, 424, 168
Spencer, J. R., et al. 1997, in Pluto and Charon ed. S. A. Stern, & D. J. Tholen (Tucson: Univ. Arizona Press), 435
Taşdemir, K., & Merényi, E. 2009, IEEE Trans. Neural Networks, 20, 549
Tryka, K. A., et al. 1993, Science, 261, 751
Tryka, K. A., et al. 1994, Icarus, 112, 513
Villmann, T., et al. 1997, IEEE Trans. Neural Networks, 8, 256
Widrow, B., & Smith, F. W. 1964, Computer and Information Sciences (COINS) Symp. Proc. (Washington, DC: Spartan Books), 288
Young, L. A., et al. 2008, Space Sci. Rev., 140, 93
Zhang, L., et al. 2009, in Proc. 7th WSOM Advances in Self-Organizing Maps, ed. J. Príncipe, & R. Miikkulainen, (LNCS 5629; Berlin: Springer), 362