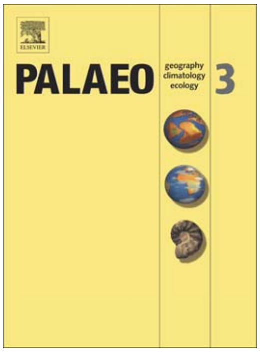


Artificial neural networks reveal a high-resolution climatic signal in leaf physiognomy

Shu-Feng Li, Frédéric M.B. Jacques, Robert A. Spicer, Tao Su, Teresa E.V. Spicer, Jian Yang, Zhe-Kun Zhou

PII: S0031-0182(15)00643-4
DOI: doi: 10.1016/j.palaeo.2015.11.005
Reference: PALAEO 7557

To appear in: *Palaeogeography, Palaeoclimatology, Palaeoecology*

Received date: 7 July 2015
Revised date: 4 November 2015
Accepted date: 5 November 2015

Please cite this article as: Li, Shu-Feng, Jacques, Frédéric M.B., Spicer, Robert A., Su, Tao, Spicer, Teresa E.V., Yang, Jian, Zhou, Zhe-Kun, Artificial neural networks reveal a high-resolution climatic signal in leaf physiognomy, *Palaeogeography, Palaeoclimatology, Palaeoecology* (2015), doi: 10.1016/j.palaeo.2015.11.005

Artificial neural networks reveal a high-resolution climatic signal in leaf physiognomy

Shu-Feng Li[a,e], Frédéric M.B. Jacques[a], Robert A. Spicer[c], Tao Su[a], Teresa E.V. Spicer[d], Jian Yang[d], Zhe-Kun Zhou[a,b*]

[a]Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla 666303, China.

[b]Key Laboratory of Biogeography and Biodiversity, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650204, China.

[c]Environment, Earth and Ecosystems, Centre for Earth, Planetary, Space and Astronomical Research, The Open University, United Kingdom.

[d]State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, the Chinese Academy of Sciences, Beijing 100093, P.R. China.

[e]State Key Laboratory of Paleobiology and Stratigraphy, Nanjing Institute of Geology and Paleontology, Chinese Academy of Sciences, Nanjing 210008, China.

[*]Corresponding author at: Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla 666303, China.

E-mail address: zhouzk@xtbg.ac.cn (Z.-K. Zhou).

Tel: +86-691-8713226

Abstract

The relationship linking leaf physiognomy and climate has long been used in palaeoclimatic reconstructions, but current models lose precision when worldwide data sets are considered because of the broader range of physiognomies that occur under the wider range of climate types represented. Our aim is to improve the predictive power of leaf physiognomy to yield climate signals, and here we explore the use of an algorithm based on the general regression neural network (GRNN), which we refer to as Climate Leaf Analysis with Neural Networks (CLANN). We then test our algorithm on Climate Leaf Analysis Multivariate Program (CLAMP) data sets and digital leaf physiognomy (DLP) data sets, and compare our results with those obtained from other computation methods. We explore the contribution of different physiognomic characters and test fossil sites from North America. The CLANN algorithm introduced here gives high predictive precision for all tested climatic parameters in both data sets. For the CLAMP data set neural network analysis improves the predictive capability as measured by R^2, to 0.86 for MAT on a worldwide basis, compared to 0.71 using the vector-based approach used in the standard analysis. Such a high resolution is attained due to the nonlinearity of the method, but at the cost of being susceptible to 'noise' in the calibration data. Tests show that the predictions are repeatable, and robust to information loss and applicable to fossil leaf data. The CLANN neural network algorithm used here confirms, and better resolves, the global leaf form–climate relationship, opening new approaches to palaeoclimatic reconstruction and understanding the evolution of complex leaf function.

Keywords: artificial neural networks, climate, CLAMP, CLANN, fossil, leaf physiognomy

1. Introduction

With the increasing concern about global climate change, in recent decades there have been new and broader interests in palaeoclimate reconstructions. Palaeobotany has a long tradition of exploiting leaf form to determine past climates (e.g. Bailey and Sinnott, 1915, 1916; Dilcher, 1973; Greenwood, 1993; Jacobs, 1999, 2002; Jacques et al., 2011; Kowalski and Dilcher, 2003; Spicer and Herman, 2010; Srivastava et al., 2012; Su et al., 2013; Wilf, 1997; Wilf et al., 1998; Wing and Wolfe, 1993, 1995). These physiognomic methods have more than one hundred years of history from the first description of the relationship linking the percentage of leaves with entire margins to temperature (Britton and Brown, 1913). Since then both univariate (Wolfe, 1979) and multivariate approaches (Kovach and Spicer, 1996; Jacques et al., 2011; Peppe et al., 2011; Spicer, 2000, 2007; Spicer et al., 2004, 2009; Wolfe, 1990, 1993; Wolfe and Spicer, 1999; Yang et al., 2011, 2015) have been developed to reconstruct temperature, precipitation, and other climatic parameters.

There is a rich literature about the relationship between climate and foliar physiognomy: the percentage of species with entire margined leaves increases with temperature (Wilf, 1997; Wolfe, 1979, 1993), leaf size increases with moisture availability (Givnish, 1987; Peppe et al., 2011), and 'drip tips' are common in warm and humid environments (Leigh, 1975), but common mechanistic links between individual characters and single climate variables across all taxa remain elusive (Jordan, 2011). This is probably because modular genetic control, driven by pleiotropy, influences variation in form under a variable environment, and ultimately leads to natural selection for strongly linked but flexible functional systems (Falconer et al., 1996; Juenger et al., 2005; Rodriguez et al., 2014) and "phenotypic integration," in which functionally related traits covary in complex ways within a given organism (Pigliucci, 2003). Leaves must optimize a variety of ecophysiological functions simultaneously and are developmentally integrated; it seems unlikely then that they would show single-character form-function-environment relationships (Yang et al., 2015).

A practical application of linking physiognomy and climate is the development of tools to retrodict past climate from leaf fossils via some form of function (f):

Climate = f (physiognomic features)

The recent assembly of a large global foliar physiognomic data set (378 sites) demonstrates that in natural woody dicot vegetation an integrated spectrum of leaf form exists across multiple leaf character states and species, and appears more strongly influenced by prevailing climate than biogeographic history. In this data set the co-variation of leaf traits across species suggests strong integration of leaf form (Yang et al., 2015). This work also demonstrates correlations between characters across a wide spectrum of woody dicot taxa despite the inclusion of samples from highly endemic floras. We know, therefore, that there is a relationship linking climate and physiognomy independent of taxonomic composition; however, we have little idea of the form of the function, how complex it is, and its parameters. Using simple relationships to build a complex multivariate function proves to be difficult because we lack information about how the factors interact. Univariate methods, such as leaf margin analysis, reduce the problem to one climatic parameter linked to one physiognomic feature, while digital leaf physiognomy (DLP) first looks at the physiognomic features with the highest explanation power, and then calculates the parameters of the function (Peppe et al., 2011; Royer et al., 2005). Both these approaches ignore, or in the case of DLP try to filter, the integrated nature of leaf form and function. If phenotypic integration results in an overall optimized solution to maximizing photosynthetic return for minimal resource investment, then the assumption that one particular subset of character/climate relationships is more important than another is dangerous when developing a climate proxy that has to be reliable across time and space.

CLAMP does not explicitly filter physiognomic characters but uses a vector-based direct ordination method, Canonical Correspondence Analysis (ter Braak, 1986), to seek physiognomic/climate relationships across 31 leaf characters and a variety of climate variables. Like all previous approaches this uses traditional algebraic methods to compute model parameters. Major climate trends are sought through the cloud of modern natural or naturalized vegetation sites positioned relative to one another based on the leaf physiognomy displayed by at least twenty of their woody dicot component taxa. This cloud of calibration sites form what is known as ‘physiognomic space’. By using observed climate data for each of the vegetation sites climate trends across physiognomic space are determined and expressed as straight-line vectors. These vectors were originally aligned by eye in two-dimensional space (Wolfe, 1993) but subsequently objectively positioned first in two-dimensional space (Kovach and Spicer, 1996) and subsequently in four-dimensional space (Spicer et al., 2003). Higher dimensions carry little additional information for most calibration data sets. Samples with no known climate, such as fossil leaf assemblages, are positioned passively and their position along the vector (the vector score) is used to predict the unknown climate (CLAMP website: http://clamp.ibcas.ac.cn; Kovach and Spicer, 1996; Spicer, 2000; Wolfe and Spicer, 1999).

With small calibration data sets the structure of physiognomic space is relatively simple (Jacques et al., 2011; Spicer, 2000; Stranks and England, 1997), and the vector approach has proved adequate for predicting past climate accurately as measured against other palaeoclimate proxies (Kennedy et al., 2002; Spicer et al., 2003), even accommodating some structural complexity by means of a non-linear regression model for calibrating the vectors. However, with large data sets spanning a diversity of vegetation and climates the ability of the vectors to capture the complexity of physiognomic space and the leaf form-climate relationship degrades (Yang et al., 2015), although the complexity can be visualized using a generalized additive model (Wood, 2011; Yang et al., 2015).

Because of the complexity of the relationship between plants and climate, it is quite likely that non-linear interactions exist among various aspects of the leaf physiognomy-climate relationship. So far, different approaches, such as CLAMP, DLP and other related modified approaches, seek linear trends that may constrain the prediction ability when worldwide data sets are considered. This is because a wider range of physiognomies occurs under the greater diversity of climate types represented as the size and geographic spread of the calibration data set increases. Non-linear relationships should be sought to improve the precision of palaeoclimatic reconstruction from leaf physiognomy.

The purpose of this work is not to present an alternative palaeoclimate proxy to those currently in use, but to explore a different way of revealing the information content of physiognomic space. In this study, we explore a new non-linear approach to approximate the function linking climate and physiognomy. The general regression neural network (GRNN) is a type of artificial neural network (ANN) that can approximate to both linear and nonlinear regressions (Specht, 1991). The GRNN is particularly advantageous with sparse data in a real-time environment, because the regression surface is instantly defined everywhere (Specht, 1991). As such the GRNN is a useful technique to investigate the climate and physiognomy relationship. We tested GRNN on two different physiognomy data sets and compared our results with those obtained from other computational methods. We also tested the GRNN using different physiognomic characters and fossil sites from North America.

2. Material and methods

2.1. Leaf physiognomy and climatic data sets

Two data sets were used in this study. The CLAMP global data set (Yang et al., 2015; the CLAMP website: http://clamp.ibcas.ac.cn) and the DLP data set (Peppe et al., 2011). Both data sets have a similar structure: a physiognomic data set that encapsulates leaf characteristics for each sampling site, and a meteorological data set describing the climate data for the same sites.

The CLAMP global data set used here is made up of 378 sites worldwide. The meteorological data usually consists of 11 parameters retrieved from a gridded data set (New et al., 2002; Spicer et al., 2009). The physiognomic data consist of a string of 31 characters describing leaf physiognomy across at least 20 taxa for each of those sites.

The DLP data set consists of 92 sites around the world. The meteorological data is made up of 10 parameters retrieved from WorldClim (Hijmans et al., 2005). The physiognomic data consist of 28 characters. Because the CLAMP data set is larger than the DLP data set, and thus potentially more complex, we chose to use the CLAMP data set in detailed tests of GRNN.

2.2. CLANN algorithm

We developed an algorithm based on GRNN. The predicted value (target) $\widehat{Y}$ to input vector *X* in the GRNN is computed by the equation (Specht, 1991):

$$\widehat{Y}\,(X) = \frac{\sum_{i=1}^{n} Y_i \exp\left(-\frac{D_i^2}{2\sigma^2}\right)}{\sum_{i=1}^{n} \exp\left(-\frac{D_i^2}{2\sigma^2}\right)} \qquad (1)$$

Where *i* is number of hidden nodes (samples) *i* = 1, 2, 3, ..., *n*. The optimal value of σ, which here denotes the spread, can be determined by cross-validation (Specht, 1991). D_i^2 is the Euclidian distance between the prediction site X_i and each known sites *X*, which is given by the following equation:

$$D_i = \sqrt{(X - X_i)^T\,(X - X_i)} \qquad (2)$$

We were able to simulate a GRNN (Fig. 1; Supplementary Algorithm S1, S2) using an algorithm we call CLANN (Climate Leaf Analysis with Neural Networks). The physiognomic data set corresponds to the input, whereas the meteorological data set corresponds to the target. The active sites were used as a training data set for the network. Because variables with large magnitudes are combined with those with small magnitudes, the former can mask the effect of the latter due to the larger weights associated with them (Sandhya, 2006), so it is crucial to normalize data prior to ANNs training process (Sola and Sevilla, 1997). All parameters of the active physiognomic data set were normalized so that their minimum and maximum values ranged between -1 and +1. The physiognomic parameters of the passive sites were transformed using the same function. The transformed active physiognomic data set was used for the weights of the input layer (Fig. 1). The meteorological data of the active data set were used as the weight of the pattern layer (Fig. 1).

Fig. 1

2.3. Performance

2.3.1. Cross-validation

Because ANNs are prone to over-training (Plumb et al., 2005), we used a repeated k-folds cross-validation process to test the precision of the approach (Fig. 1). All sites were randomly separated to 10 folds using the 10-fold validation method in R. One of these folds was selected for test data and the other 9 folds for calibration data. The meteorological parameters of the test data were reconstructed using the trained model. This 10-fold validation was repeated 10 times. The mean values of predictions for all sites were compared to the observed meteorological data.

2.3.2. Diagnostic values

Two diagnostic values were calculated to measure the performance of the model: 1) the R-squared (R^2) between the predicted and observed values, 2) the standard deviation of the residuals (SD).

2.3.3. Spread selection

The spread is the only parameter that can be adjusted in a GRNN. A lower spread will give relatively higher weights to active sites near the passive site. A higher spread will tend to give more similar weights over all the data set. In other words, a lower spread increases the influence of local sites while a high spread gives results that tend towards the mean of the data set. Different spreads were tested for our model; tested values were from 0.1 to 2.00 with an increment of 0.01. The R^2 between the observed values and the values predicted under the 10-fold cross-validation procedure was used as a measure of performance of the model. For each climatic parameter, we selected the spread that gave the highest R^2.

2.4. Significance of the model

To test if there is a real climatic signal retrieved by the CLANN algorithm from leaf physiognomy, we composed an artificial physiognomy file consisting of random numbers, and compared our results from the CLAMP data set with that from this random data set. The random physiognomic data set was built using the function ‘runif’ in R, following the rules of scoring for CLAMP (Wolfe, 1993). For example in CLAMP scoring the sum score of the characters 'no teeth', 'rounded teeth' and 'acute teeth' is 100 and the total score of the three leaf base characters is 100. The random numbers were constructed in such a way that such restrictions in the scoring regime (see the CLAMP website, http://clamp.ibcas.ac.cn, for details) were adhered to. The algorithm used is given in supplementary Algorithm S3. The cross-validation procedure was used with this random data set as if it were the physiognomic data set along with the meteorological data set used for CLAMP. The spread was adjusted for this data set. R^2 values of predicted versus observed values for the random data set and the R^2 values of predicted versus observed values for the CLAMP data set were compared to test the validity of the model.

2.5. Comparison with previous methods

To calculate the performance of CLAMP under a cross-validation process, we used the R package ‘Vegan’ (Oksanen, 2015). The models followed the usual calculation method (Jacques et al., 2011; Yang et al. 2011).

The DLP calculations for the cross-validation procedure were carried out using the software R. The multiple regression models for DLP data were built according to Peppe et al. (2011). The regression parameters were selected using stepwise model selection by AIC in R package ‘MASS’ (Ripley et al., 2015).

2.6. Contribution of each character

The physiognomic characters can be grouped in seven classes: lobed (character #1), leaf margin (characters #2 to 7), leaf size (characters #8 to 16), leaf apex (characters #17 to 20), leaf base (characters #21 to 23), length-to-width ratio (L:W, characters #24 to 28), and leaf shape (characters #29 to 31).

The influence of each character class was analyzed in two ways.

(1) a new input data set created by excluding a class from the physiognomic data set. The cross-validation procedure is carried out using this new input data set. This is done for all classes turn-by-turn.

(2) a new input data set created using only one feature class of the physiognomic data set. The cross-validation procedure is carried out using this new input data set. This is done for all classes turn-by-turn.

Spreads are adjusted for each newly designed data set by training the new data sets respectively using the CLANN algorithm (see supplementary Algorithm S2).

2.7. Application to fossil assemblages

Canonical correspondence analyses (CCAs) were carried out to understand the differences between regions and to visualize where the fossil sites were located in relation to calibrated physiognomic space (physiognomic space defined by modern vegetation sites growing within a known climate regime). A CCA was carried out on the physiognomic data of the modern CLAMP data set and another on the 82 fossil sites from Eocene to Pliocene in North America presented in Yang et al. (2011). Both CCAs were made using the R package ‘Vegan’ (Oksanen, 2015).

Palaeoclimates were reconstructed from 82 Eocene to Pliocene fossil sites in North America (36 Paleogene fossil sites and 46 Neogene fossil sites), to test the parity between CLAMP and CLANN when reconstructing palaeoclimates. A pared t-test was performed with IBM SPSS Statistics software (version 20, IBM Corporation, Somers, NY, USA) to explore the differences in the reconstructed palaeoclimates between CLAMP and CLANN. We also drew box plots to show how large these differences were among the 11 reconstructed climate parameters.

3. Results

3.1. Performance

The best spreads, R^2, and P of the CLANN algorithm for the CLAMP and DLP data set are shown in Table 1.

Table 1

For the CLAMP data set, the predictive power as measured by R^2 ranged between 0.42 and 0.86 for all climatic parameters. The relationship between predicted and observed values for all climatic parameters is highly significant (see *P* values). The relationships between observed and predicted values for MAT and GSP are indicated on Fig. 2. Comparison of CLAMP and CLANN predictions for all of the 11 climate parameters are shown in the Supplementary Figs. S1-S11. The CLANN algorithm gives high predictive power across all 11 parameters.

Fig. 2

For the DLP data set, the predictive power is low for MAP, with an R^2 of only 0.21. However, the relationship between observed and predicted values is highly significant for both MAP and MAT (Table 1).

3.2. Significance

The R^2 and *P* between the predicted and observed values for the CLAMP physiognomic data set and a random physiognomic data set are shown in Table 1. All R^2 values for the random data set are very low. The relationship between the predicted and observed values for the random data set is not significant for all climatic parameters. These results show that the CLANN algorithm gives higher predictive power than CLAMP for the real CLAMP data set, but presents very poor predictions for the random data set. This shows that the CLANN algorithm retrieves climate information from the CLAMP data set and does not impose patterns where none exist.

3.3. Comparison with previous methods

Diagnostic values of CLANN and CLAMP, including R^2 and standard deviations, are given in Table 2. Whatever the climatic parameter considered and whatever the diagnostics used, CLANN always gives better precision than CLAMP. Supplementary Figs S1-11 compare the relationships between observed and predicted values for CLAMP and CLANN. For 3-DRY and RH, there are greatly improved predictions by CLANN compared to CLAMP. For 3-DRY and RH, the R^2 of CLAMP predictions are around 0.16 and 0.30 respectively, whereas CLANN gives much better predictive power. Note that these values are different from those for the whole data set because they refer to the mean statistics of the 10-fold cross validation subsamples of the full data set (Fig. 1). Figs. S1-11 show that sample points are less dispersed using the CLANN model than for CLAMP, and that the model regression line is closer to the y=x line for CLANN than for CLAMP.

Table 2

Diagnostic values of CLANN and DLP are also given in Table 2. For all climatic parameters, except GDD and GSDD, CLANN always gives better diagnostic values than DLP. For GDD and GSDD, the diagnostic values are quite similar between the DLP and CLANN predictions.

3.4. Contribution of each character

The R^2 for each climatic parameter when a character class is lost is shown on Fig. 3. Whatever the class of character excluded and whatever the climatic parameter considered, the R^2 is similar to the one obtained with all characters present.

Fig .3

The R^2 for each climatic parameter when only one character class is included is shown on Fig. 4. In most instances, the R^2 is clearly different from zero. For temperature-related parameters, the features concerning the teeth give the highest R^2. For precipitation-related parameters, leaf size characters typically have the highest predictive power. No character class alone reaches the R^2 levels obtained with all characters.

Fig. 4

3.5. Comparison between CLAMP and CLANN for fossil sites

CCA axes 1 v 2 and CCA axes 1 v 3 (Fig. 5) show the distribution of 82 North American Paleogene and Neogene fossil sites (black open circles), within the cloud of modern sites (coloured symbols) that define physiognomic space for the calibration data set. All the 82 fossil sites are located within modern physiognomic space. This implies that the palaeoclimate for all of these fossil sites can be reconstructed using CLANN.

Fig. 5

Based on the results of palaeoclimates reconstructed using CLAMP and CLANN (Supplementary Table S1, S2 respectively), we explored the difference between these two methods using the paired t-test. Table S3 shows that there were significant differences in six palaeoclimate parameters reconstructed by CLANN and CLAMP, implying that there were differences between these two models when reconstructing palaeoclimates. For these six palaeoclimate parameters, CLANN generates higher prediction values for WMMT, LGS, GSP, and 3-WET, while it gives lower values for CMMT and RH. For 3-DRY, although CLANN greatly improved predictive precision compared to CLAMP, the t-test result shows no significant difference in the mean values between these two methods.

The box plots show that CLANN produces obviously higher median values for MAT, WMMT and LGS, while it gives distinctly lower median values for CMMT, GSP and MMGSP. For 3-DRY, the box plots show only a small difference in the median value; while CLANN gives a larger range of reconstructed climate values. The results also show that CLANN predictions have larger ranges than those of CLAMP for all the moisture-related climate parameters.

4. Discussion

4.1. An improved palaeoclimatic model

A skeptic could argue that our results are an artifact of the algorithm, and that there is no real climatic signal to be retrieved from leaf physiognomy. To test this, we compared our results to results obtained from an artificial physiognomy file consisting of random numbers. When using this random input data set, all R^2 values for the relationships between predicted and observed values were below 0.01 ($P > 0.3$ for all tested climate variables; see Table 1). Whereas with the observed leaf physiognomy data set (Table 2), the smallest R^2 we get is 0.42 (see Table1; for 3-WET, which refers to precipitation during the three consecutive wettest months). Therefore, we conclude that the CLANN algorithm is capable of revealing structure in the data that is present in the real observations but absent in the random artificial data.

In this study, two indices were used to test the performance of the new CLANN method. For all climatic parameters, CLANN performs better than CLAMP (Table 2). CLANN also performs better than DLP for eight parameters, while there are no significant differences between CLANN and DLP performance for GDD and GSDD parameters (Table 2). However, where the same climate parameter is used in both CLAMP and DLP, CLAMP exhibits the greater predictive precision.

Among the methods of palaeoclimatic reconstruction based on leaf physiognomy, CLANN is the one that performs best based on R^2 metrics. For example, based on the CLAMP data set, CLANN gives an R^2 of 0.86 for MAT, which means that CLANN can explain 86% of the MAT variability based on the leaf physiognomy variability. A perfect model should give 100%. Our model is not perfect; we need to investigate possible sources of errors. One source of error is the modelling itself, which does not perfectly fit all data. Other sources of error concern imperfections in the data sets: the actual climatic parameters experienced by the vegetation may be different from those given by the gridded data set; the leaf sampling process may have missed one or two species and even if the scoring process is well defined, there can be occasional scoring errors. All sources of error contribute to the overall model error and methodological improvements cannot rectify error within the data sets. Within these constraints it appears that the CLANN model is closer to the minimal error compared to CLAMP and DLP models.

4.2 Advantages and limits

Previous palaeoclimate reconstructions methods based on leaf physiognomy relied on a general relationship (trends) between physiognomic features and climate variables across whole data sets. The CLANN method is based on a GRNN; therefore, it allows local adaptations of the model to the data (Specht, 1991). However, it raises the question of the importance of the local adaptations over the general relationship. It is not an easy question to answer: a too generalized model overlooks local adaptations that might be important. For example particular combinations of leaf features, exhibited within large global foliar physiognomic data sets, might be important for detecting marked seasonal variations in water availability and thus a monsoon signal (Jacques et al., 2011) and these may be missed if a too generalized model is used. However, too strong an importance given to local adaptations diminishes the power of the model when used on new sites. In GRNN, the local importance is defined by the spread: a high spread favours a more general relationship; a low spread favors local adaptations. For a very high spread, the model gives the same value (the mean of the calibration data set) for all sites. For a very low spread, the model gives the exact value for all calibrating points, and 0 for all other possible points. Using a repeated k-fold cross-validation approach, we were able to tune the best spreads for our data set and avoid over-fitting. If the CLANN method were to be used on another data set, new spread values would have to be tuned and even then a low spread would make predictions vulnerable to the characteristics of individual calibration sites. In situations where species diversity is low and/or sampling or scoring contains errors, or the gridded calibration climate data do not reflect well the local conditions experienced by the leaves (e.g. in topographically complex mountainous terrain), the position of a calibration site in physiognomic space may be anomalous and lead to erroneous CLANN predictions for unknown (fossil) sites. This will give rise to 'noise' and a large predictive range offered by CLANN. In CLAMP the influence of such poor calibration sites is low, but the cost is a reduction in precision.

A limit of CLANN is that it cannot be used for sites that fall outside the range of its calibrating data set. This limitation is true for any arithmetic model: the error of the model increases sharply outside the calibrating range. Even if the models can be extrapolated beyond the calibration range, the cost in increasing errors is hidden. In the case of CLANN, when a site is far from the range of the calibrating data set, the results will be 0: this gives a good control for such sites. However, we recommend that users always test if a fossil site is included in the range of calibration. CLANN by itself does not provide a visualization of physiognomic space, or the relationship of a fossil site to calibrated physiognomic space, but this can be done using CCA. The CCA results (Fig. 5) show that all of the fossil sites of North America were located within the physiognomic space occupied by the modern global data set, and indicate that the CLANN model derived from this calibration is appropriate for exploring the palaeoclimates represented by the fossil sites.

4.3. Applicability on different data sets

We tested our computation methods on two data sets that were built with different scoring strategies: the CLAMP data set and the DLP data set. CLANN works on both data sets, which indicates that it may also perform well on other data sets. If scoring improvements are proposed, or new scoring methods developed, we suggest that CLANN represents one of the best computation methods to explore their properties.

4.4. Complex multivariate relationships

The relationship between leaf physiognomy and climatic parameters is complex and multivariate. Ecologists are interested in leaf physiognomic function in relation to the environment, including climate. For palaeoclimatic reconstruction, a strong relationship is more important than fully understanding functional mechanisms. In contrast to linear models, the GRNN used in CLANN allows the influence of a feature to vary inside physiognomic space. This is one of the reasons why CLANN offers higher precision than other methods: it models a higher complexity of the relationships than those methods.

Previous studies on CLAMP show that the climatic signal of one parameter is not only present in one leaf feature, but that all features can contribute to this signal (Spicer et al., 2011; Yang et al., 2015). However some authors suggest that some characters correlate only weakly with climate in the present day (Peppe et al., 2010). This characteristic is an important issue in palaeobotany, where some leaf features cannot be scored because of taphonomic loss and poor preservation in the fossil, or correlations may change over time. This leads to a question: is the reconstruction method robust to character loss? To test this, 1) each character class was excluded one by one from the analysis (Fig. 3); and 2) only one character class was included in each analysis (Fig. 4). These results show that a climatic signal is retrieved for all climatic parameters whatever the analysis carried out; and there is only a small difference in the R^2 value when some character information is lost. Therefore, our results show that the climatic signal is encoded in all character classes and that the CLANN algorithm, like CLAMP, is robust against character loss. Our results also show that in the calibration data some traits make a higher contribution to the prediction of temperature and precipitation related parameters (e.g. teeth and leaf size contribute greatly to temperature and precipitation predictions respectively) than others, although there is strong integration of leaf form as evidenced by the covariation of leaf traits across species (Yang et al., 2015). Does this mean that some leaf characters are more 'important' than others for palaeoclimate and the rest ignored? Based on the observation that in some parts of the modern world such as New Zealand even the proportion of toothed to non-toothed leaves shows no correlation with mean annual temperature (Stranks and England, 1997; Kennedy et al., 2014; Yang et al., 2015) clearly the 'importance' of a character is not universal between geographic regions and is equally unlikely to be static through time. The purpose of our character removal test is to explore the sensitivity of CLANN to character loss, not to argue that one suite of characters is more important than another for retrieving climate signals.

4.5. Regional differences

It has been suggested that several regions have a different leaf physiognomy-climate relationship (Jacques et al., 2011; Little et al., 2010; Peppe et al., 2011; Stranks and England, 1997; Su et al., 2010). These regions include Australia and New Zealand (Stranks and England, 1997). For New Zealand, the predictive 3-DRY climate values are displaced away from the observed values (Fig. S8), indicating that 3-DRY cannot be reconstructed confidently but this is easily explained because there is no proper dry season in New Zealand. Extreme cold sites from Siberia are outliers in CLAMP (Spicer et al., 2004), especially for temperature parameters (e.g. MAT and CMMT). With CLANN, there are no outliers (Fig. S1, S3). From these results we conclude that the approach we explore here is valid worldwide. Because this new method correctly reconstructs modern climate from tropical to cold regions, it is also likely to reconstruct well the diversity of past climates.

Regional constraints have long been recognized for univariate physiognomic methods (reviewed in Steart et al., 2010). When the leaf form/climate relationship is examined using multiple leaf characters climate dominates over phylogeny in determining this relationship, but the structure of physiognomic space can be complex and that this complexity varies among climate parameters (Yang et al., 2015). Because CLANN has the ability to adapt to localized trends within physiognomic space, regional variations in the relationship between leaf form and climate, particularly variations in the mix of characters that result in leaf optimization for local climatic regimes, is captured by CLANN.

Nevertheless, it is clear that a non-random phylogenetic signal is present in both leaf traits and the distribution of plants: (1) some families have only leaves with entire margin (such as Magnoliaceae), whereas others have only toothed leaves (such as Betulaceae); (2) the distribution of plants is not independent of their taxonomy, some families are exclusively tropical, whereas others are mostly represented in cold or temperate regions. An exaptive scenario has been suggested by some authors to explain the distribution of toothed-margin species in cold regions (Little et al., 2010), but our results show that each class of character contains some climatic signal for all climatic parameters (Fig. 4). Exaptive scenarios to explain the distribution of all the studied leaf features along the gradient of all studied climatic parameters therefore seem highly improbable. An adaptive scenario, where leaf features are selected by the climate, is thus more parsimonious. Therefore, as in Yang et al. (2015), we suggest that the climatic signal present in leaf physiognomy (especially for temperatures) is mostly independent of phylogeny.

4.6. CLANN as a palaeoclimate proxy

The 82 Paleogene and Neogene fossil sites analyzed here demonstrate differences between CLAMP and CLANN. The box plots show that CLANN predictions may cover larger ranges of climate space than CLAMP. Are these greater ranges a reflection of reality or are they an artifact? To answer this question it is necessary to compare the methodological differences between CLAMP and CLANN. In CLAMP the standard statistical engine is Canonical Correspondence Analysis, the outcome of which is the summary of climate trends across physiognomic space represented by linear vectors. Such trends inevitably compromise precision because they cannot accommodate complexities in physiognomic space (Yang et al., 2015). In CLANN the derived climate signal is weighted towards the climate experienced by those calibration sites that have the most similar physiognomies to that of the unknown (fossil) site. The derived signal is thus in large part dependent on the number of calibration sites, and their properties, that are used to obtain the climate of the unknown site. Which calibration sites are used to derive the prediction is determined by the CLANN 'spread' parameter. Potentially this approach, like the local multivariate regression approach (Stranks and England, 1997) and its multiple regression surfaces derivative (Yang et al., 2015), can yield more precise climate predictions because it better accommodates the complexities of physiognomic space. However, the fewer (spread-determined) calibration samples used to derive climate predictions in CLANN, compared to the full data set that is used in CLAMP, renders it highly sensitive to variation among the calibration sites due to proximity to the boundaries of physiognomic space, poor sampling, poor scoring, or uncertainties in the modern climate data. This meteorological 'noise' is most pronounced in topographically complex mountainous regions (the location of many of the calibration sites because this is where natural vegetation survives today) and particularly in precipitation data (Spicer et al., 2009). It is perhaps to be expected then that the ranges in CLANN predictions for fossil sites are uniformly higher than those for CLAMP, most notably in the precipitation variables (Fig. 6). In future this climatic noise may be reduced by gridded data that takes into account meso-scale meteorological processes as well as aspect, but at the moment meteorological noise is a major limitation on increasing both the precision and accuracy of multivariate foliar physiognomic palaeoclimate proxies.

Acknowledgements

This study was supported by the National Basic Research Program of China (No. 2012CB821901), National Natural Science Foundation of China (No. 41372035), the Foundation of the State Key Laboratory of Paleobiology and Stratigraphy, Nanjing Institute of Geology and Paleontology, Chinese Academy of Sciences (No. 153107), as well as the CAS 135 program (XTBG-F01). This work is part of the NECLIME (Neogene Climate Evolution of Eurasia) network.

Supplementary content

References

Bailey, I.W., Sinnott, E.W., 1915. A botanical index of Cretaceous and Tertiary climates. Science 41, 831-834.

Bailey, I.W., Sinnott, E.W., 1916. The climatic distribution of certain types of angiosperm leaves. American Journal of Botany 3, 24-39.

Britton, N.L., Brown, A., 1913. An illustrated flora of the northern United States, Canada and the British possessions: from Newfoundland to the parallel of the southern boundary of Virginia, and from the Atlantic Ocean westward to the 102d meridian. Scribner.

Falconer, D.S., Mackay, T.F., Frankham, R., 1996. Introduction to quantitative genetics, fourth ed. Trends in Genetics 12, 280.

Falconer, D.S., Mackay, T.F., 1996. Introduction to quantitative genetics. Addison Wesley Longman, Harlow, Essex, UK.

Givnish, T.J., 1987. Comparative studies of leaf form: assessing the relative roles of selective pressures and phylogenetic constraints. New Phytolologist 106, 131-160.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25, 1965-1978.

Jacobs, B.F., 1999. Estimation of rainfall variables from leaf characters in tropical Africa. Palaeogeography, Palaeoclimatology, Palaeoecology 145, 231-250.

Jacobs, B.F., 2002. Estimation of low-latitude paleoclimates using fossil angiosperm leaves: examples from the Miocene Tugen Hills, Kenya. Paleobiology 28, 399-421.

Jacques, F.M.B., Su, T., Spicer, R.A., Xing, Y.W., Huang, Y.J., Wang, W.M., Zhou, Z.K., 2011. Leaf physiognomy and climate: Are monsoon systems different? Global and Planetary Change 76, 56-62.

Juenger, T., Pérez ‐ Pérez, J.M., Bernal, S., Micol, J.L., 2005. Quantitative trait loci mapping of floral and leaf morphology traits in Arabidopsis thaliana: evidence for modular genetic architecture. Evolution and Development 7, 259-271.

Kennedy, E.M., Spicer, R.A., Rees, P.M., 2002. Quantitative palaeoclimate estimates from Late Cretaceous and Paleocene leaf floras in the northwest of the South Island, New Zealand. Palaeogeography, Palaeoclimatology, Palaeoecology 184, 321-345.

Kennedy, E.M., Arens, N.C., Reichgelt, T., Spicer, R.A., Spicer, T.E.V., Stranks, L., Yang, J., 2014. Deriving temperature estimates from Southern Hemisphere leaves. Palaeogeography, Palaeoclimatology, Palaeoecology 412, 80-90.

Kovach, W.L., Spicer, R.A., 1996. Canonical correspondence analysis of leaf physiognomy: a contribution to the development of a new palaeoclimatological tool. Palaeoclimates 2, 125-138.

Kowalski, E.A., Dilcher, D.L., 2003. Warmer paleotemperatures for terrestrial ecosystems. Proceedings of the National academy of Sciences 100, 167-170.

Leigh, E.G., 1975. Structure and climate in tropical rain forest. Annual Review of Ecology and Systematics 6, 67-86.

Little, S.A., Kembel, S.W., Wilf, P., 2010. Paleotemperature proxies from leaf fossils reinterpreted in light of evolutionary history. PLoS One, 5, e15161.

Oksanen, J., 2015. Multivariate analysis of ecological communities in R: vegan tutorial (http://vegan.r-forge.r-project.org/), pp. 1-43.

Peppe, D.J., Royer, D.L., Cariglino, B., Oliver, S.Y., Newman, S., Leight, E., Enikolopov, G., Fernandez-Burgos, M., Herrera, F., Adams, J.M., Correa, E., Currano, E.D., Erickson, J.M., Hinojosa, L.F., Hoganson, J.W., Iglesias, A., Jaramillo, C.A., Johnson, K.R., Jordan, G.J., Kraft, N.J., Lovelock, E.C., Lusk, C.H., Niinemets, U., Penuelas, J., Rapson, G., Wing, S.L., Wright, I.J., 2011. Sensitivity of leaf size and shape to climate: global patterns and paleoclimatic applications. New Phytolologist 190, 724-39.

Peppe, D.J., Royer, D.L., Wilf, P., Kowalski, E.A., 2010. Quantification of large uncertainties in fossil leaf paleoaltimetry. Tectonics 29, TC3015.

Pigliucci, M., 2003. Phenotypic integration: studying the ecology and evolution of complex phenotypes. Ecology Letters 6, 265-272.

Plumb, A.P., Rowe, R.C., York, P., Brown, M., 2005. Optimisation of the predictive ability of artificial neural network (ANN) models: a comparison of three ANN programs and four classes of training algorithm. European Journal of Pharmaceutical Sciences 25, 395-405.

Ripley, B., Venables, B., Bates, D.M., Hornik, K., Gebhardt, A., Firth, D., 2015. Package 'MASS' (http://cran.r-project.org/web/packages/MASS/index.html), pp. 1-164.

Rodriguez, R.E., Debernardi, J.M., Palatnik, J.F., 2014. Morphogenesis of simple leaves: regulation of leaf size and shape. Wiley Interdisciplinary Reviews: Developmental Biology 3, 41-57.

Royer, D.L., Wilf, P., Janesko, D.A., Kowalski, E.A., Dilcher, D.L., 2005. Correlations of climate and plant ecology to leaf size and shape: potential proxies for the fossil record. American Journal of Botany 92, 1141-1151.

Sandhya, S., 2006. Neural Networks for Applied Science and Engineering. Auerbach publications, New York, pp. 253.

Sola, J., Sevilla, J., 1997. Importance of input data normalization for the application of neural networks to complex industrial problems. Nuclear Science, IEEE Transactions on 44, 1464-1468.

Specht, D.F., 1991. A general regression neural network. Neural Networks, IEEE Transactions on 2, 568-576.

Spicer, R.A., 2000. Leaf Physiognomy and Climate Change, in: Culver. S.J. and Rawson, P. (Eds.), Biotic Response to Global change: the Last 145 Million Years. Cambridge University Press, Cambridge, pp. 244-264.

Spicer, R.A., Harris, N.B.W., Widdowson, M., B., A., Herman, Guo, S., Valdes, P.J., Wolfek, J.A., Kelley, S.P., 2003. Constant elevation of southern Tibet over the past 15 million years. Nature 421, 622–624.

Spicer, R.A., 2007. Recent and Future Developments of CLAMP: Building on the Lagacy of Jack A. Wolfe. Cour. Forsch. -Inst. Senckenberg 258, 109-118.

Spicer, R.A., Herman, A.B., 2010. The Late Cretaceous environment of the Arctic: A quantitative reassessment based on plant fossils. Palaeogeography, Palaeoclimatology, Palaeoecology 295, 423-442.

Spicer, R.A., Herman, A.B., Kennedy, E.M., 2004. Foliar physiognomic record of climatic conditions during dormancy: Climate Leaf Analysis Multivariate Program (CLAMP) and the cold month mean temperature. The Journal of Geology 112, 685-702.

Spicer, R.A., Bera, S., De Bera, S., Spicer, T.E.V., Srivastava, G., Mehrotra, R., Mehrotra, N., Yang, J., 2011. Why do foliar physiognomic climate estimates sometimes differ from those observed? Insights from taphonomic information loss and a CLAMP case study from the Ganges Delta. Palaeogeography, Palaeoclimatology, Palaeoecology 302, 381-395.

Spicer, R.A., Valdes, P.J., Spicer, T.E.V., Craggs, H.J., Srivastava, G., Mehrotra, R.C., Yang, J., 2009. New developments in CLAMP: Calibration using global gridded meteorological data. Palaeogeography, Palaeoclimatology, Palaeoecology 283, 91-98.

Srivastava, G., Spicer, R.A., Spicer, T.E.V., Yang, J., Kumar, M., Mehrotra, R., Mehrotra, N., 2012. Megaflora and palaeoclimate of a Late Oligocene tropical delta, Makum Coalfield, Assam: Evidence for the early development of the South Asia Monsoon. Palaeogeography, Palaeoclimatology, Palaeoecology 342-343, 130-142.

Steart, D.C., Spicer, R.A., Bamford, M.K., 2010. Is Southern Africa different? An investigation of the relationship between leaf physiognomy and climate in southern

African mesic vegetation. Review of Palaeobotany and Palynology 162, 607–620.

Stranks, L., England, P., 1997. The use of a resemblance function in the measurement of climatic parameters from the physiognomy of woody dicotyledons. Palaeogeography Palaeoclimatology Palaeoecology 131, 15-28.

Su, T., Jacques, F.M.B., Spicer, R.A., Liu, Y.S., Huang, Y.J., Xing, Y.W., Zhou, Z.K., 2013. Post-Pliocene establishment of the present monsoonal climate in SW China: evidence from the late Pliocene Longmen megaflora. Clim. Past 9, 1911-1920

Su, T., Xing, Y.W., Liu, Y.S., Jacques, F.M.B., Chen, W.Y., Huang, Y.J., Zhou, Z.K., 2010. Leaf Margin Analysis: A New Equation from Humid to Mesic Forests in China. Palaios 25, 234-238.

ter Braak, C.J.F., 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 67, 1167-1179.

Tokar, A.S., Johnson, P.A., 1999. Rainfall-runoff modeling using artificial neural networks. Journal of Hydrologic Engineering 4, 232-239.

Wilf, P., 1997. When are leaves good thermometers? A new case for leaf margin analysis. Paleobiology 23, 373-390.

Wilf, P., Wing, S.L., Greenwood, D.R., Greenwood, C.L., 1998. Using fossil leaves as paleoprecipitation indicators: an Eocene example. Geology 26, 203-206.

Wing, S.L., Greenwood, D.R., 1993. Fossils and fossil climate: the case for equable continental interiors in the Eocene. Philosophical Transactions of the Royal Society of London B: Biological Sciences 341, 243-252.

Wolfe, J.A., 1979. Temperature parameters of humid to mesic forests of Eastern Asia and relation to forests of other regions of the Northern Hemisphere and Australasia: analysis of temperature data from more than 400 stations in Eastern

Asia. United States Geological Survey Professional Paper 1106, 1-37.

Wolfe, J.A., 1990. Palaeobotanical evidence for a marked temperature increase following the Cretaceous/Tertiary boundary. Nature 343, 153-156.

Wolfe, J.A., 1993. A method of obtaining climatic parameters from leaf assemblages. Annual Review of Earth and Planetary Sciences 23, 119-142.

Wolfe, J.A., Spicer, R.A., 1999. Fossil leaf character states: multivariate analysis, in: Jones, T.P., Rowe, N.P. (Eds.), Fossil plants and spores: modern techniques. Geological Society, London, pp. 233-239.

Yang, J., Spicer, R.A., Spicer, T.E.V., Li, C.S., 2011. 'CLAMP Online': a new web-based palaeoclimate tool and its application to the terrestrial Paleogene and Neogene of North America. Palaeobiodiversity and Palaeoenvironments 91, 163-183.

Yang J., Spicer R.A., Spicer T.E.V., Arens N.C., Jacques F.M.B., Su T., Kennedy E.M., Herman A.B., Steart D.C., Srivastava G., Mehrotra R.C., Valdes P.J., Mehrotra N.C., Zhou Z.K., Lai J.S., 2015. Leaf Form-Climate Relationships on the Global Stage: An Ensemble of Characters. Global Ecology and Biogeography 10, 1113-1125.

Figure Captions

Fig. 1 Schematic illustration of the CLANN algorithm. (1) The entire dataset of 378 sites were randomly separated to 10 folds using 10-fold validation method in R. One of these folds was selected for test data and the other 9 folds for calibration data. This 10-fold validation was repeated 10 times. (2). Two hundred spreads (from 0.01 to 2, in steps of 0.01) were selected for model training. For each model the calibration was optimized with 20×10 iterative cycles (epochs) (3). Next, calibration data were randomly separated to 10 folds using 10-fold validation method in R. One of these folds was selected for validation data and the other 9 folds for training data. (4). GRNN training. (5) The diagnostic values were calculated by comparing the predicted values and the real values. (6). The models were calibrated based on the optimal spread values. (7). Once the models were calibrated they were used to predict climate values for test data. The climate parameters of 378 sites were subsequently predicted using all the calibrated models 10 iterative cycles (epochs). (8) The mean of all predicted climate parameters of 10 iterative cycles were calculated as a measure of the performance of the model.

Fig. 2 Climate Leaf Analysis with Neural Networks (CLANN) predictions for mean annual temperature (MAT) and growing season precipitation (GSP). The black line represents a 1:1 correspondence between the observed and predicted values, the red line is the regression representing the CLANN-derived relationship.

Fig. 3 Spider diagram showing the effect of character loss tested by the Climate Leaf Analysis with Neural Networks (CLANN). The radii represent the R-squared (R^2)

between the predicted and observed values for the complete physiognomic dataset (All), and other seven new datasets created by excluding a class of leaf characters from the physiognomic dataset each time. The physiognomic characters are grouped in seven classes: lobing, teeth, size, apex, base, length-to-width ratio (L:W), and shape. Abbreviations for the climate parameters: MAT, mean annual temperature; WMMT, mean temperature of the warmest month; CMMT, mean temperature of the coldest month; LGS, length of the growing season; GSP, growing season precipitation; MMGSP, mean monthly growing season precipitation; 3-WET, precipitation of the three consecutive wettest months; 3-DRY, precipitation of the three consecutive driest months; RH, annual mean relative humidity; SH, annual mean specific humidity; ENT, enthalpy.

Fig. 4 Spider diagram showing the climatic signal coded by each feature class tested by the Climate Leaf Analysis with Neural Networks (CLANN). The radii represent the R-squared (R^2) between the predicted and observed values for the complete physiognomic dataset (All), and other seven new datasets created by using just a single class of leaf characters from the physiognomic dataset each time. The physiognomic characters are grouped in seven classes: lobing, teeth, size, apex, base, length-to-width ratio (L:W), and shape. Abbreviations are as in Fig. 3.

Fig. 5 Canonical correspondence analysis (CCA) showing that the 82 fossil sites used as CLANN test samples all fall within the physiognomic space defined by the modern dataset. CCA axes 1 v 2 and CCA axes 1 v 3 shows the distribution of 82 North American Paleogene and Neogene fossil floras (black colour open circles), within the cloud of modern sites (coloured symbols) that define physiognomic space for the calibration dataset.

Fig. 6 Box plots showing the differences of 12 reconstructed climate parameters for the 82 fossil sites using CLAMP and CLANN. Abbreviations for the climate parameters are as in Fig. 3.

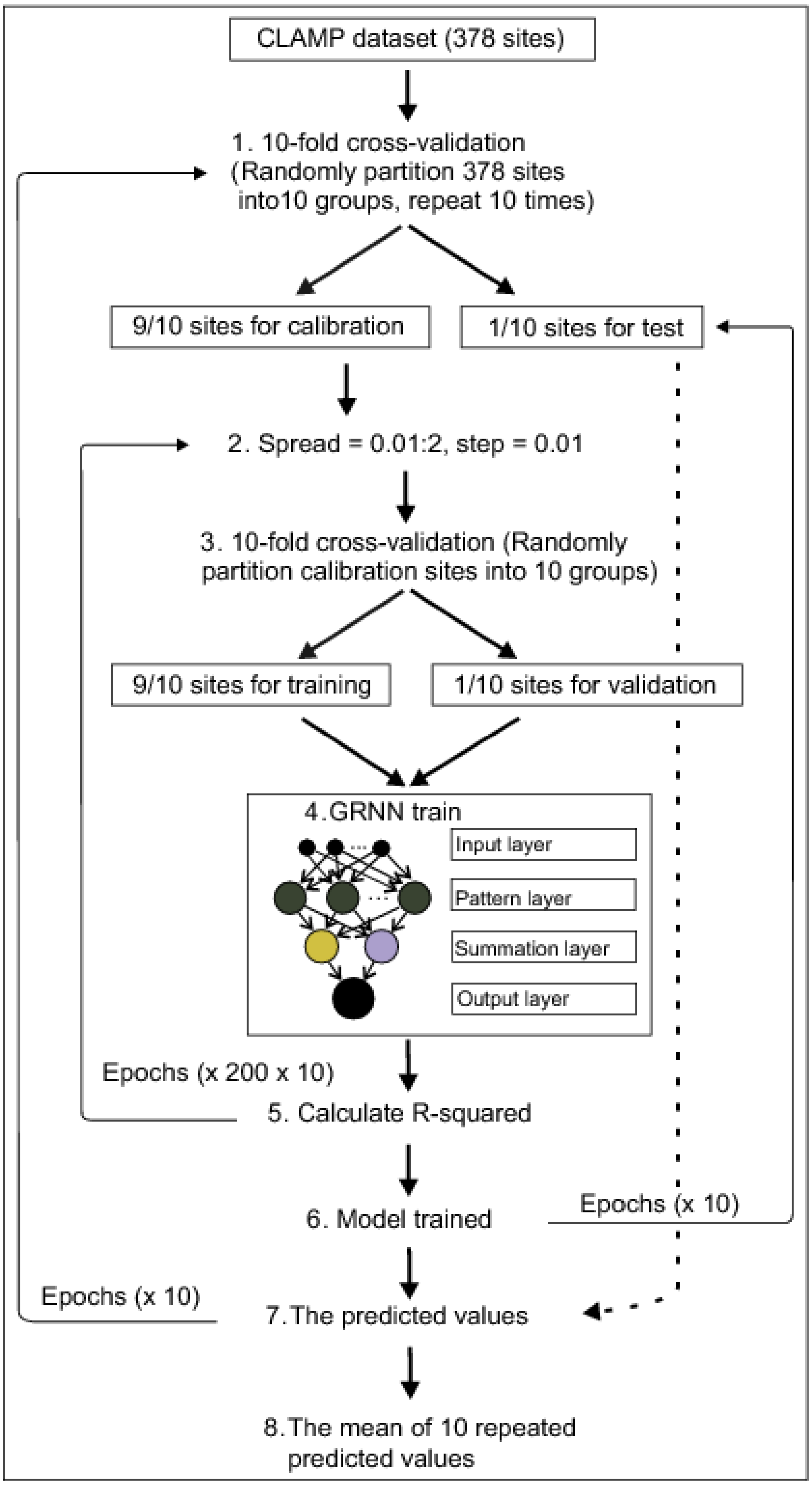

CLAMP dataset (378 sites)
1. 10-fold cross-validation (Randomly partition 378 sites into10 groups, repeat 10 times)
9/10 sites for calibration
1/10 sites for test
2. Spread = 0.01:2, step = 0.01
3. 10-fold cross-validation (Randomly partition calibration sites into 10 groups)
9/10 sites for training
1/10 sites for validation
4.GRNN train
Input layer
Pattern layer
Summation layer
Output layer
Epochs (x 200 x 10)
5. Calculate R-squared
6. Model trained
Epochs (x 10)
Epochs (x 10)
7.The predicted values
8.The mean of 10 repeated predicted values

Figure 1

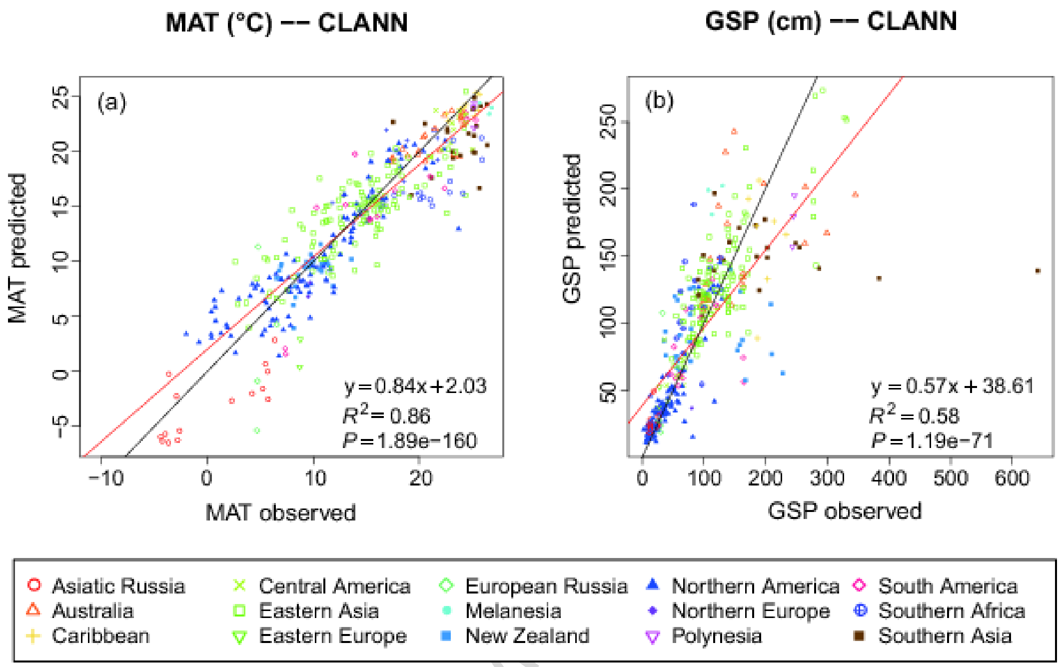

MAT (°C) -- CLANN
GSP (cm) -- CLANN
(a)
(b)
MAT predicted
MAT observed
GSP predicted
GSP observed
-5
0
5
10
15
20
25
-10
0
10
20
50
100
150
200
250
0
100
200
300
400
500
600
y=0.84x+2.03
R^2= 0.86
P=1.89e-160
y = 0.57x + 38.61
R^2 = 0.58
P = 1.19e-71
Asiatic Russia
Australia
Caribbean
Central America
Eastern Asia
Eastern Europe
European Russia
Melanesia
New Zealand
Northern America
Northern Europe
Polynesia
South America
Southern Africa
Southern Asia

Figure 2

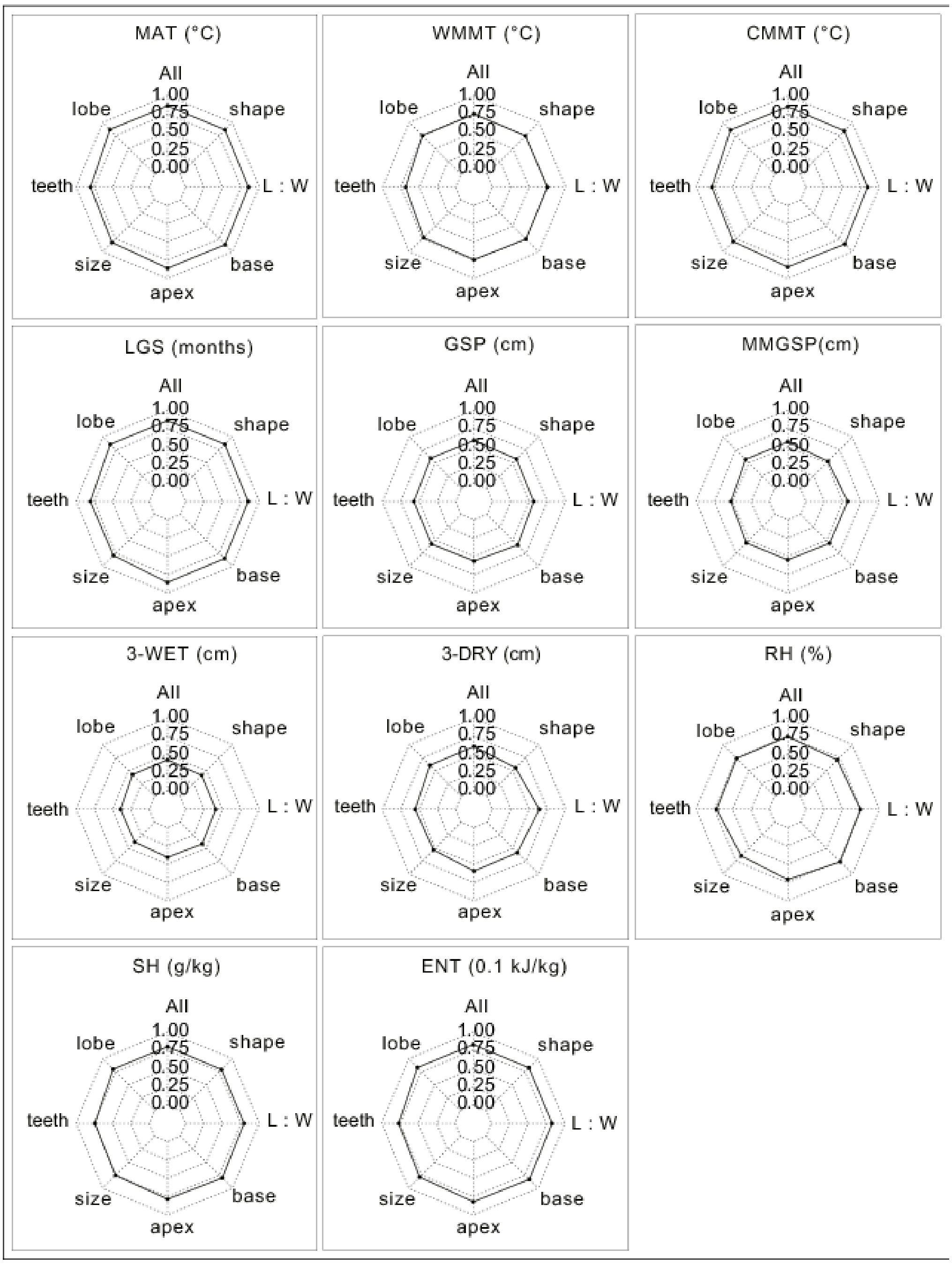

MAT (°C)
WMMT (°C)
CMMT (°C)
LGS (months)
GSP (cm)
MMGSP(cm)
3-WET (cm)
3-DRY (cm)
RH (%)
SH (g/kg)
ENT (0.1 kJ/kg)
All
shape
L : W
base
apex
size
teeth
lobe
1.00
0.75
0.50
0.25
0.00

Figure 3

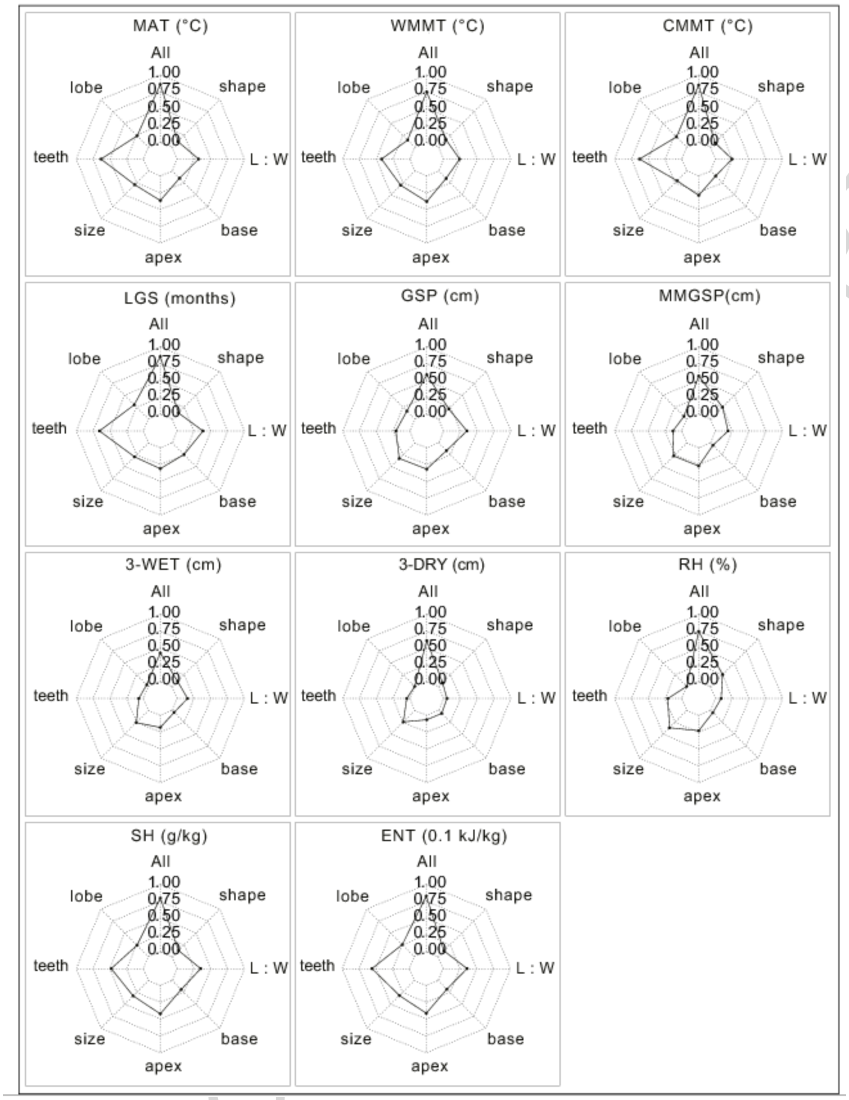

MAT (°C)
WMMT (°C)
CMMT (°C)
LGS (months)
GSP (cm)
MMGSP(cm)
3-WET (cm)
3-DRY (cm)
RH (%)
SH (g/kg)
ENT (0.1 kJ/kg)
All
shape
L : W
base
apex
size
teeth
lobe
1.00
0.75
0.50
0.25
0.00

Figure 4

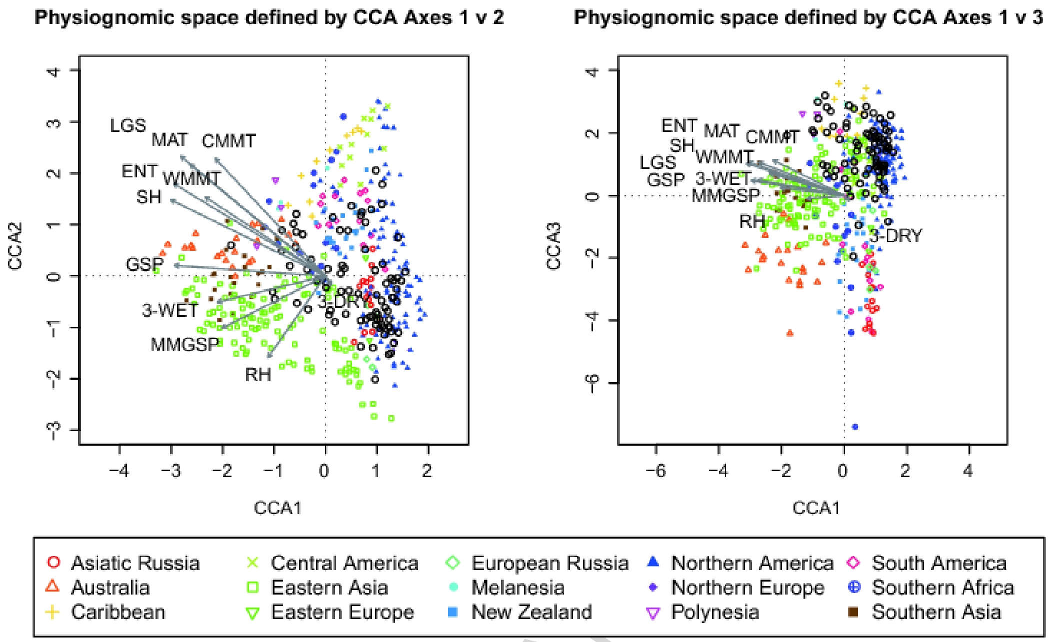

Physiognomic space defined by CCA Axes 1 v 2
Physiognomic space defined by CCA Axes 1 v 3
CCA2
CCA1
CCA3
CCA1
LGS
MAT
CMMT
ENT
WMMT
SH
GSP
3-WET
MMGSP
RH
3-DRY
ENT
MAT
CMMT
SH
WMMT
LGS
GSP
3-WET
MMGSP
RH
3-DRY
4
3
2
1
0
−1
−2
−3
−4
−3
−2
−1
0
1
2
4
2
0
−2
−4
−6
−6
−4
−2
0
2
4
○ Asiatic Russia
△ Australia
+ Caribbean
× Central America
□ Eastern Asia
▽ Eastern Europe
◇ European Russia
● Melanesia
■ New Zealand
▲ Northern America
◆ Northern Europe
▽ Polynesia
◇ South America
⊕ Southern Africa
■ Southern Asia

Figure 5

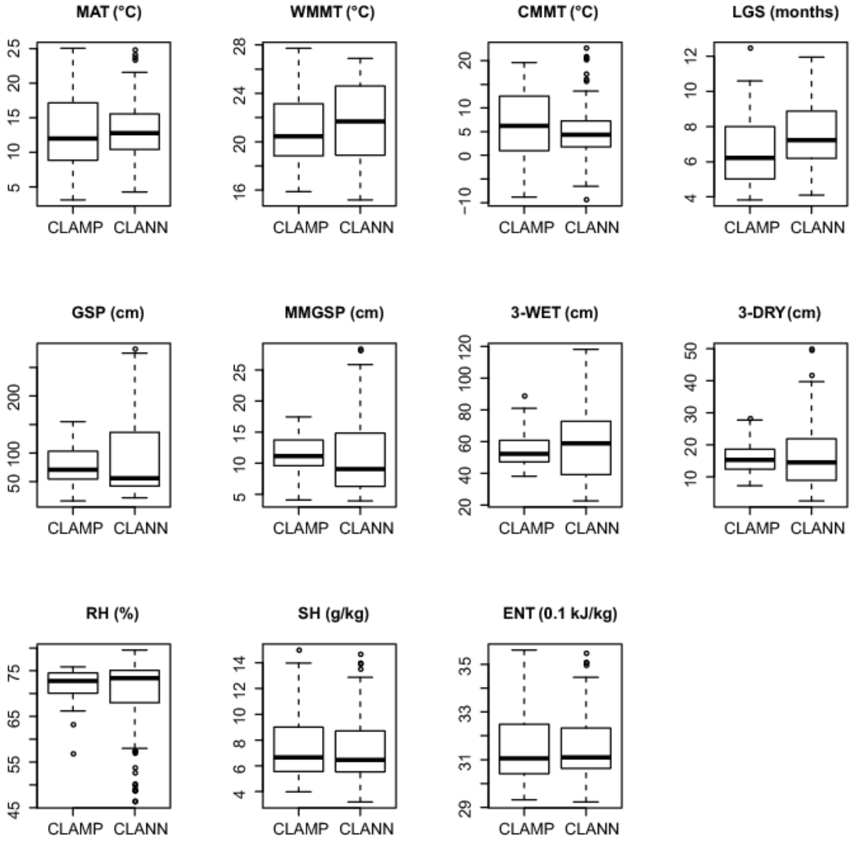

MAT (°C)
WMMT (°C)
CMMT (°C)
LGS (months)
GSP (cm)
MMGSP (cm)
3-WET (cm)
3-DRY (cm)
RH (%)
SH (g/kg)
ENT (0.1 kJ/kg)
CLAMP
CLANN
5
10
15
20
25
16
20
24
28
−10
0
5
10
20
4
6
8
10
12
50
100
200
25
30
40
60
80
100
120
45
55
65
75
14
29
31
33
35

Figure 6

Table Captions

Table 1. Results of the Climate Leaf Analysis with Neural Networks (CLANN) algorithm applied to the Climate Leaf Analysis Multivariate Program (CLAMP) PhysgGlobal378 dataset with a corresponding high resolution gridded climate data available from the CLAMP website (http://clamp.ibcas.ac.cn), a random physiognomic dataset constructed using CLAMP scoring protocols, and the digital leaf physiognomy (DLP) datasets of Peppe et al. (2011). Abbreviations: MAT, mean annual temperature; WMMT, mean temperature of the warmest month; CMMT, mean temperature of the coldest month; LGS, length of the growing season; GSP, growing season precipitation; MMGSP, mean monthly growing season precipitation; 3-WET, precipitation of the three consecutive wettest months; 3-DRY, precipitation of the three consecutive driest months; RH, annual mean relative humidity; SH, annual mean specific humidity; ENT, enthalpy; MART, Mean annual range in temperature; GSMT, growing season mean temperature; GDD, Growing degree days; GSDD, Growing season degree days; GSL, Growing season length; MAP, Mean annual precipitation.

Table 2. Comparison of model statistics for the Climate Leaf Analysis Multivariate Program (CLAMP), and the digital leaf physiognomy (DLP) with the newly introduced Climate Leaf Analysis with Neural Networks (CLANN). In terms of the R-squared (R^2) and the standard deviation (SD), CLANN has a better precision than CLAMP for all studied climatic parameters from CLAMP dataset. For the DLP dataset, both methods give similar precision for GDD and GSDD, but CLANN has a better precision for all other parameters than DLP. Abbreviations: MAT, mean annual

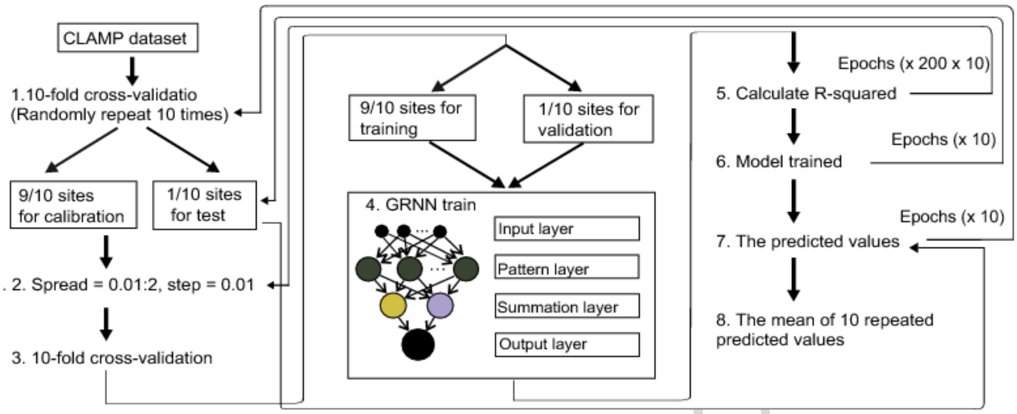
temperature; WMMT, mean temperature of the warmest month; CMMT, mean temperature of the coldest month; LGS, length of the growing season; GSP, growing season precipitation; MMGSP, mean monthly growing season precipitation; 3-WET, precipitation of the three consecutive wettest months; 3-DRY, precipitation of the three consecutive driest months; RH, annual mean relative humidity; SH, annual mean specific humidity; ENT, enthalpy. Note that the CLAMP statistics are slightly different from those given in Yang et al. (2015) because those given here are calculated on the basis of the same random 90% subsets of the full data set as used in CLANN for training, whereas those in Yang et al. (2015) are based on the full data set.

Table 1

CLAMP dataset				**Random dataset**		**DLP dataset**			
Parameter	Spread	R^2	P	R^2	P	Parameter	Spread	R^2	P
MAT (°C)	0.52	0.86	1.89×10^{-160}	0.0018	0.976	MAT (°C)	0.50	0.66	1.40×10^{-21}
WMMT (°C)	0.55	0.75	1.87×10^{-113}	0.0024	1.141	WMMT (°C)	0.45	0.38	1.04×10^{-09}
CMMT (°C)	0.49	0.85	5.69×10^{-152}	0.0022	0.636	CMMT (°C)	0.48	0.73	1.15×10^{-26}
LGS (months)	0.56	0.85	7.93×10^{-156}	0.0027	0.931	MART (°C)	0.50	0.63	1.49×10^{-18}
GSP (cm)	0.57	0.58	1.19×10^{-71}	0.0018	0.732	GSMT (°C)	0.47	0.53	2.59×10^{-15}
MMGSP (cm)	0.58	0.56	7.83×10^{-67}	0.0020	0.808	GDD (days)	0.50	0.63	1.08×10^{-19}
3-WET (cm)	0.56	0.42	3.63×10^{-44}	0.0037	0.354	GSDD (days)	0.49	0.73	1.29×10^{-26}
3-DRY (cm)	0.52	0.60	2.52×10^{-75}	0.0098	1.575	GSL (days)	0.39	0.39	7.18×10^{-08}
RH (%)	0.49	0.74	3.92×10^{-110}	0.0030	0.810	GSP (cm)	0.39	0.38	2.04×10^{-09}
SH (g/kg)	0.56	0.80	7.42×10^{-130}	0.0026	0.800	MAP (cm)	0.57	0.21	4.73×10^{-05}
ENT (0.1 kJ/kg)	0.55	0.83	3.77×10^{-105}	0.0023	0.828				

Table 2

CLAMP dataset					DLP dataset				
Parameter	CLAMP method		CLANN method		Parameter	DLP method		CLANN method	
	R^2	SD	R^2	SD		R^2	SD	R^2	SD
MAT(°C)	0.71	4.10	0.86	2.85	MAT (°C)	0.54	4.15	0.66	2.72
WMMT(°C)	0.41	3.97	0.75	2.59	WMMT (°C)	0.27	3.51	0.38	2.73
CMMT(°C)	0.62	6.92	0.85	4.40	CMMT (°C)	0.63	6.12	0.73	3.40
LGS(months)	0.66	1.92	0.85	1.26	MART (°C)	0.38	5.85	0.63	3.32
GSP(cm)	0.41	56.96	0.58	47.86	GSMT (°C)	0.46	3.84	0.53	3.06
MMGSP(cm)	0.28	6.11	0.56	4.77	GDD (days)	0.66	865.77	0.63	872.54
3-WET(cm)	0.20	33.55	0.42	28.56	GSDD (days)	0.77	598.15	0.73	653.99
3-DRY(cm)	0.16	13.40	0.60	9.23	GSL (days)	0.09	132.39	0.39	52.31
RH(%)	0.30	9.68	0.74	5.89	GSP (cm)	0.16	162.21	0.38	56.36
SH(g/kg)	0.67	1.99	0.80	1.58	MAP (cm)	0.05	92.42	0.21	77.97
ENT(0.1 kJ/kg)	0.72	1.07	0.83	0.84					



Graphical abstract

Highlights:

- We developed an artificial neural network to test leaf form/climate relationships.
- The new algorithm (CLANN) reveals a high-resolution climatic signal in leaf form.
- CLANN predictions are repeatable, robust to information loss, and precise.
- The new method is applicable to fossil leaf data and could form a new climate proxy.