A Non-Linear Model of Nondestructive Estimation of Anthocyanin Content in Grapevine Leaves with Visible/Red-Infrared Hyperspectral[1]

JiangLin Qin[1,2], Donald Rundquist[2], Anatoly Gitelson[2], Zongkun Tan[1],Mark Steele[2]

[1]**Guangxi Institute of Meteorological Disaster-Reducing Research;No.81, Minzu Avenue, Nanning, Guangxi, P.R. China. 530022.**
[2]**Center for Advanced Land Management Information Technologies (CALMIT), University of Nebraska-Lincoln,Lincoln, NE 68583-0973**

Abstract. The anthocyanin(Anth) content in leaves provides valuable information about the physiologocal status of plant. Thus, there is a need for accurate, efficient, practical methodologies to estimate this biochemical parameter. Hyperspectral measurement is a means of quickly and nondestructively assessing leaf Anth in situ. Wet chemical methods has traditionally been used for this purpose. Recently, NIR(near-infrared)/green, red/green, anthocyanin reflectance index(ARI), and a modified anthocyanin refelctance index(MARI) was been used to estimate the anthocyanin content. In this paper, a an artificial-intelligence technique model was introduced to establish the relationship between the anthocyanin content and reflectance of 400-750nm spectum, variation of species and growth stages. The objective of this study was to test the overall performance and accuracy of this new nondestructive techniques for estimating Anth content in grapevine leaves. Although Anth in validation data set was widely variable, the new methods were capable of accurate predicting Anth content in grapevine leaves with a root mean square error below 1.65 mg/m^2,which is lower than that of MARI or ARI [20]. It documents the facts that such an approach is more suitable for developing simple hand-held field instrumentation for accurate nondestructive Anth estimation and for analyzing digital airborne or satellite imagery to assist in making informed decisions vineyard management.

Keywords:Grapes, Hyperspectral, SVM (Support Vector Machine), Anthocyanin.

Corresponding author Jianglin Qin. Tel.: +86 771 5850120; Fax:+86 771 5865594.
E-mail address: zhouli5850120@163.com

2   JiangLin Qin1,2, Donald Rundquist2, Anatoly Gitelson2, Zongkun Tan1,Mark Steele2

1.Introduction / Background

Very often, in grapevine leaves, significant accumulation of anthocyanin is as a result of a number of environmental stresses such as strong light, UV-B irradiation, low temperature, drought, wounding, bacterial and fungal infections, nitrogen and phosphorus deficiencies, certain herbicides and pollutants. It may serves as indicators of grapevine leaf senescence and stress in grapevine, their detection and quantitative assessment can provide important information about response and adaptation of grapevine to environmental stresses.

   Anthocyanins have traditionally been determined through wet-chemical methods, including pigment extraction in a solvent,spectrophotometric determination of absorbance to Anth solution, and conversion from measured absorbance to Anth content[5].However, laboratory procedures are laborious, time-consuming, and destructive to leaves. Anthocyanins, Chlorophylls , carotenoids, and other pigments participate in light absorption in particular bands and can readily be assessed with absorption and reflectance spectroscopy. During past few decades, considerable progress in development of nondestructive methods for estimating physiological state of vegetation has been achieved. But much attention was given to the estimation of Chl and Car content, not much is known about Anth estimation, and lots of them are Linear/or simple nonlinear model. Decades ago, multivariable approach[1] is developed by the USDA, but one of its assumption is the relationships between reflectance and chemical concentration are near-linear. Three bands model[6,7,4] is very popular recently year, but it is very simple non-linear model, and it can not cover the variation in species and in growth stages and even in environmental stresses, only Three bands is far from enough to be used. Very recently, NIR(near-infrared)/green, red/green, anthocyanin reflectance index(ARI, based on refelctances in bands within the green and the red-edge regions[5]), and a modified anthocyanin refelctance index(MARI, based on reflectances in green,red edge, and NIR[5,6]) was been used to estimate the anthocyanin content[20]. However, this research will propose a new approach which cover advantages in lots of methods, include Three bands model and multivariable approach, and e.g, it depends on your research on your problem. It can perfectly model the relationship between Anth content and reflectance of spectra, even environmental factors.

2. Purpose of the Paper

The objectives of the current study was to investigate the performance of a reflectance-based nondestructive technique to estimate Anth in grape leaves that may contain anthocyanin , chlorophylls and carotenoids, and that may contain the scattering of leaf surface. And specifically to (1) identify the optimal position

of spectral bands and their width as they relate to Anth content in grape;(2) identify the position of spectral bands as they relate to chlorophyll, carotenoids content so as to remove the afects of this two pigment in estimating Anth in grape leaves;(3)select near-infrared (NIR) band so as to remove the scattering of the grape leaf internal structure and leaf surface;(4) develop a quantitative model that accurately estimates Anth content in grapevine leaves fully employ hyperspectral reflectance at selected narrow portions of the visible spectrum (400-750 nm); (5) evaluate the performance of the new techniue.

3. Material and Methods

3.1 Selected Grape Cultivars

Five grape cultivars were investigated as our research: DeChaunac, Edelweiss, Lacrosse, St.Croix, Stpepin. The red cultivars included DeChaunac and Saint Croix, while the whites were Edelweiss and Lacrosse. DeChaunac, a red French-American hybrid developed by Albert Seibel and introduced into Canada in 1946, is a hybrid of Seibel 5163 and Seibel 793. The vine is vigorous and more disease resistant than other French hybrids, and is cold hardy to temperatures of -26° C (-15° F).   Saint Croix, a red Vitis riparia cross between ES-283 and ES-193, was introduced by Elmer Swenson in 1981. Saint Croix is a vigorous vine with known resistance to black rot and is cold hardy to temperatures of -35° C (-32° F). Edelweiss, a white Vitis labrusca cross between Minnesota 78 and Ontario cultivars, was introduced by Wisconsin grape breeder Elmer Swenson in 1980. It is known as a vigorous vine resistant to foliage diseases and is cold hardy to temperatures of -34° C (-30° F). Lacrosse is a white winegrape originated by by Elmer Swenson in Osceola, Wisconsin and introduced in 1983. It is a hybrid of (Minn 78 $\times$ Seibel 1000) $\times$ Seyval, with a medium sized berry, medium cluster and is very winter hardy. It produces a high quality, fruity, non-labrusca wine and the medium vigor vine has good tolerance to 2,4-D and resistance to disease. Stpepin.

3.2 Field Sampling of Leaves

144 leaves were sampled during seven fields campaigns undertaken on: June 26,2006 to June 28,2007. Individually leaves were selected based upon the variety they were and how healthy they looked. The selected leaves were cut from the canopy, immediately sealed in a plastic bag with a small amount of water, and paced in a cooler with ice.

   After fields sampling was completed, the leaves were transported to the lab, and areas of homogeneous pigmentation on each were identified and delineated with a permanent marker. In order to facilitate record-keeping, high resolution

4   JiangLin Qin1,2, Donald Rundquist2, Anatoly Gitelson2, Zongkun Tan1,Mark Steele2

digital photographs were acquired for each group of sampled leaves using an olympus C-175 4.0 megapixel camera.

Fig. 1. Leaves, as photographed by means of the digital camera, were selected to provide samples of the various cultivars. Measurement areas were delineated on the leaves themselves. These Saint Croix leaves were sampled on July 10, 2007.

3.3 Pigment Extraction

Analytical extraction of anthocyanins is problematic since the pigment in solution absorbs at 525 nm, the same band used for analytical chlorophyll extraction. To separate absorption by the two pigments, a process of pheophytinization is conducted. The process is performed by acidifying the neutral solution to convert chlorophylls to pheophytins, which while in solution, will absorb at 654 nm and 648 nm. In a neutral solution, anthocyanins are colorless. The acidification causes the red anthocyanin absorption feature to develop. This, in addition to the conversion of chlorophylls to pheophytins, facilitates analytical extraction of anthocyanin

  For each leaf sample, two to three 1-cm diameter discs were cut from the marked area. Pigments contents were calculated in a two step procedure. The punches were weighed and ground in 100% methanol using a mortar and pestle. The leaf tissue was ground until the pulp turned white in color and all pigments were suspended in the solution. The resulting homogenate was centrifuged in test tubes for 6 minutes. Absorption spectra of the solution were recorded using a cary spectrophotometer. The cary spectrophotometer was configured to measure absorption of the sample at 1 nm intervals between 400 nm and 800 nm.

chlorophylls a and b and carotenoid contents were calculated from the spectra using coefficients described by Lichtenthaler [15]. Once chlorophyll absorption spectra were collected, chlorophylls were converted to pheophytins by adding HCL to each sample to produce a 99.9% methanol 0.1% HCL solution. Absorption spectra of the converted solution were recorded using the cary spectrophotometer and absorption by anthocyanins was calculated using coefficients described by Lichtenthaler [15].

Laboratory analytical anthocyanin of 144 leaves yielded an adequate range of pignment concentrations, ranging from 1.34 mg/m^2 to 64.74 mg/ m^2 with a mean of 11.71 mg/ m^2. The range of Anthocyanin in this study was similar to that observed by Gamon and Surfus[3] .

3.4 Measuring Leaf Reflectance

Spectral-reflectance measurements were collected, at leaf level, for each of the five grape cultivars noted above using "leaf clip" with 2.3 mm diameter bifurcated fiber optic attached to both an ocean optics UBS2000 radiometer and to an ocean optics LS-1 tungsten halogen light source. The ocean optics UBS2000 radiometer uses a charged coupled device(CCD) to measure radiance with a spectral resolution of 1.5 nm across 2024 individual spectral channels ranging from 350nm to 1010nm in wavelength. The instrument has a 12 bit radiometric resolution; thus it records levels of reflectance ranging from 0 to 4095. The an ocean optics LS-1 light source uses a regulated power supply and a tungsten halogen filament bulb burning at 3100K to output a steady beam of light with a spectral range between 260 and 2500nm. In this study, the light source was turned on at least 15 minutes prior to scanning to allow the bulb and filament to warm and stabilize.

The plastic leaf clip, used to position the ocean optics fiber against individual grape leaves, consisted of a black polyvinyl chloride(PVC) attachment and a bifurcated glass fiber optic with a diameter of 2.3 mm. This fiber optic is transmissive between 400 and 1000 nm and was positioned at a 60° angle relative to the leaf surface in order to minimize the spectral range of the instrument, was held against the abaxial side of the leaf to reduce extraneous reflectance from the reflect light being transmitted through the leaf.

The ocean optics radiometer was calibrated prior to each data-collection session using a Labsphere spectral reference panel(North Sutton, NH) with a nominal reflectance of 99% between 250 and 2500nm. First, the reflectance panel was held tightly against the optic, and a spetral scans were taken prior to each data-collection session. The sensor was operated using the CALMIT Data Acquistion Program, which uses a single calibration scan collected at the time nearest to the acquisition of target scans to compute reflectance. To insure that a calibration scan used for reflectance calculation was accurate, at least three additional reflectance-panel scans were recorded for each dataset. The fourth calibration scan was plotted and compared to the plots of the previous three, it was used in subsequent processing. The reflectance spectrum was calculated as a ratio of leaf radiance of the calibration standard at wavelength .

6   JiangLin Qin1,2, Donald Rundquist2, Anatoly Gitelson2, Zongkun Tan1,Mark Steele2

A total of 6 reflectance measurements were acquired within the marked homogenous area of each leaf using the ocean optics UBS2000 radiometer. Care was taken to distribute the location of the spectral scans throughout the entire marked area superimposed on the leaf in order to acquire an accurate representation of greenness. No data were taken at leaf edges where stain was present due to ink from the darker. The average of the six scans per sample was calculated in order to establish a single representative reflectance spectrum per leaf. Each spectral profile was numbered to correspond with the leaf from which it was acquired.

4. The Procedure for Developing the Quantitative Model

4.1 To create variable set for the new model.

4.1.1 Confirming sensitive spectrum and/or factors.

The relationship between the measured anthocyanin content and spectral reflectance at visible wavelengths (400-750 nm) is a complex non-linear form. The first task in developing an appropriate quantitative model to summarize that relationship is to determine which specific wavelengths in the visible spectrum have the greatest sensitivity for detecting subtle changes in anthocyanin content. The sensitivity calculation first makes use of the standard error in the regression of two datasets. In our case, the first dataset was the anthocyanin content associated with the sampled leaves corresponding to each of the five cultivars at every one of 4 growth stages. Each data-collection campaign in the field was considered as a separate growth stage (5 total field campaigns during 2006 and 2007). For the year of 2006, the first stage represents May 26, 2006,19 samples for Edelweiss and 25 samples for StCroix; the second stage represents June 1 and 9,2006, 10 samples for StCroix, Edelweiss, stpepin; the fourth stage represents September 19,2006,22 samples for stcroix. For the year of 2007, the second stage represents June 5, 2007, the third stage represents June 28,2007, every stage includes 6 samples for DeChaunac, Edelweiss, Lacrosse, StCroixper.The anthocyanin content datasets for the five cultivars were regressed against visible spectral reflectance, with the latter interpreted for our study as the wavelength region between 400 and 750 nm. The purpose of the many regression calculations was to select the most useful portions of the visible range for estimating Anthocyanin content.

As an illustration of the procedure, consider Fig.2 where the data collected on June 28, 2007 are shown. Notice that the x-axis is the visible spectrum as defined above. The y-axis represents the standard error of the respective calculations associated with all 5 field campaigns. The solid black line represents the result for all samples combined (4 grape cultivars times 6 scans per leaf) for one growth stage. The red-line represents the series of standard errors for the

DeChaunac grape cultivar, green represents Edelweiss, turquoise represents
Lacrosse, and purple represents Saint Croix . The lowest points along the
respective lines indicate the spectral regions or wavelengths where the greatest
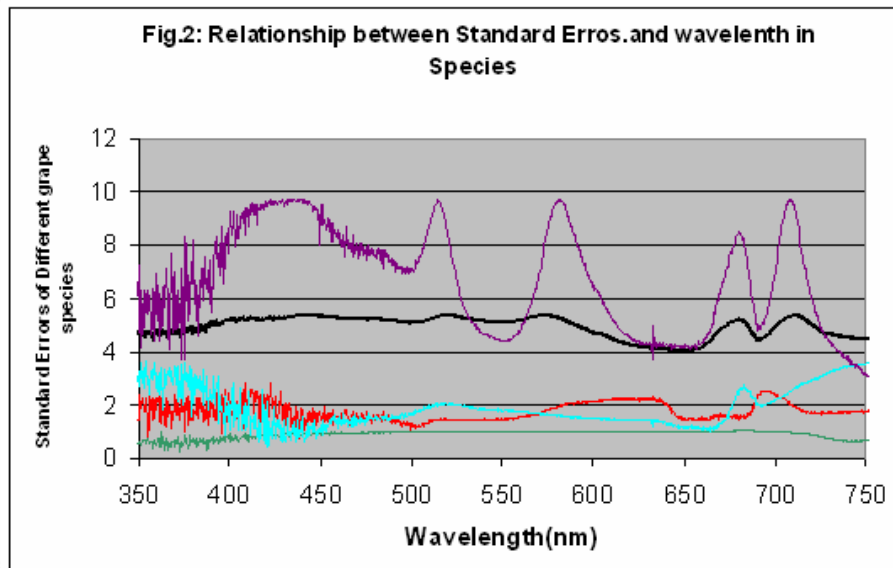sensitivity occurs.

Fig. 2. The graphic illustrates the variation in the sensitivity of the various spectral
regions to changes in Anth content. The solid black line represents the result for all
samples combined (4 grape cultivars 6 times scans per leaf) for one growth stage. The
red-line represents the series of standard errors for the DeChaunac grape cultivar, green
represents Edelweiss, turquoise represents Lacrosse, and purple represents Saint Croix .
The data shown here are from leaves sampled on June 28, 2007.

The next step in the procedure was to calculate the correlations (R) and
coefficients of determination (R^2) between anthocyanin and reflectance for those
locations along the lines with the lowest standard errors. This was done merely
to confirm the selection of wavelengths that will be used in the final model. The
result of this latter calculation was the selection of fifteen spectral regions for use
in the modeling procedure.   The fifteen selections include the following
(wavelengths in nm): 400.3 to 404.8;450.22 to 453.19;494.61 to 499.38;533.65 to
536.18; 550.62 to 553.14;641.13 to 643.55;676.2 to 679.95;682.33 to 686.07;689.8
to 692.51; 696.56 to 699.93; 716.71 to 719.71;724 to 726;721.04 to 722.71;739.94
to 742.57;748 to 750.

Next, the reflectances for the individual spectral ranges noted above are
integrated by means of averaging these reflectance over the wavelength ranges
(e.g., 400.3 to 404.8 nm, which contains band centers at 400.3;400.68;401.06;
401.43; 401.81; 402.19; 402.57; 402.95; 403.32; 403.7; 404.08; i.e., *eleven*
individual channels).   The averages became input variables to the developing

model, along with two other factors: 1) grapevine cultivar (each cultivar was labeled, and that label, expressed numerically); and 2) growth stage of the vine (i.e., each stage was labeled, and that label, expressed numerically). The procedure was repeated for the other 16 data-acquisition dates. The quantitative model was developed by using the seventeen factors to estimate anthocyanin content.

The sensitive channel is also variation between grape species: From the Fig. 2, we find very apparent variation in sensitivity between grape species. St.Croix is much more variable, there are many lower points on its sensitive curve, they are about at 499, 554, 625, 661, 693,750 nm. Normally situation, DeChaunac and Edelweiss is much more sensitive, and Lacrosse is lower sensitive, but it is not always right, we still need to confirm by caculating coefficient r^2-value. The lower points is also different between grape species, for example, there is a lower point at 693.52 nm for St.Croix, but that similar lower point for Lacrosse is changed to point 686.75 nm, and for DeChaunac and Edelweiss, there is similar lower point may be changed to 721.71 nm and 737.96 nm. Another example, there is a lower spectral fields about in 625-661 nm for St.Croix; but it is changed to spectral fields 652.88 to 686.41 nm for DeChaunac; and it is changed to 532.9 to 640.43 nm for Lacrosse; and there is not apparent similar lower spectral fields for Edelweiss.

Then, by detail analysis the datum and figures of every growth stage, we also find significantly different sensitivity between grape species. But unfortunately, if we only pay attention to the rough-black line, e.g if we ignore the situation of the difference between grape species , you shouldn't get those findings, the result will became worse.

The sensitivity of spectral channels is changed and/or shifted according to growth-stage. From growth stage of June 5,2007 to that of June 28,2007, the shape of all the sensitive curve is changed a lot. Taken St.Croix and DeChaunac as example, the similar lower points at about 550 and 680 nm is still retained, but the exact point is changed and corresponding sensitivity is also changed. And the shape of all curve is changed. While in June 5,2007, the St.Croix is very sensitive at point 550nm, but DeChaunac is not very sensitive at 550 nm. However, while the growth stage move to June 28, 2007, the most sensitive channel of St.Croix and DeChaunac is changed to point at 693.86 and 506 nm separately.

Normally, the sensitivity of separate species is much more high than that of no considering species difference situation: Detail analysis of the datum and graphs of all growth stage shows that in the most situation and/or growth stages, the sensitive curve of separate species is usually under the curve of the rough-black one. It indicates that modeling on separate species will be much more effective, or we should use some kind of non-linear technique and taking consider of species and growth stage difference to get more effective model.

4.1.2 Interaction of anthocyanin, chlorophylls, carotenoids and other foliar chemistry in absorption and Reflectnce.

Absorption feature are broadened by multiple scattering and often interfere with one another. Organic compounds also absorb in similar wavebands, so that a wavelength is never uniquely related to a chemical. For instance, both of Anth, Chl and Car is absorb at about 550,680 nm, it makes the problem very intractable. From Fig.3a&3b, we can see that in some cases, while the Anth content is very lower, but Chl and/or Car maybe very high. For example, Edelweiss in June 5, 2007, Ant content is 2.36 mg/m^2, but Chl and/or Car content is 454.5 and 97.6 mg/ m^2 separately. However, the same grape species Edelweiss in May 26, 2006, Ant content is the similar 2.364 mg/ m^2, but Chl and/or Car content is 24.49 and 11.22 mg/m^2 separately. So the similar Anth content, but Chl and Car content is variable a lot, from 454.5 and 24.49 mg/m^2 for Chl and from 24.49 to 11.22 mg/m^2 for Car. As for sensitive channel of at 550.62 to 553.14 and 676.2 to 679.95 nm, the reflectance is 11.78 and 3.2 for Edelweiss in June 5, 2007 individually, and is 23.57 and   5.19   for Edelweiss in May 26,2006 separately, it is drastically variable from 11.78 to 23.57 for band of 550.62 to 553.14 nm and from 3.2 to 5.19 for band of 676.2 to 679.95 nm. So, in case of Anth lower content, the reflectance of spectrum maybe drastically variable, and so as for the absorption. The relationship between them is a very complex non-linear form, non-linear adjustment is necessary.
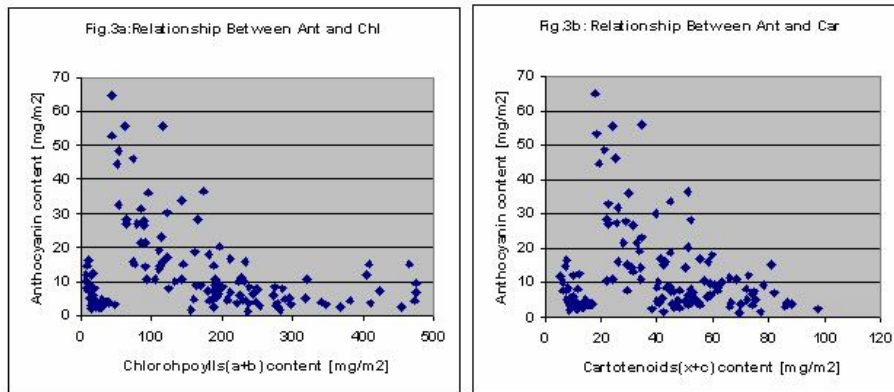

Fig. 3. The graphic illustrates that the reflectance and absorption of the Anth content affects by Chl and Car content. Both of Figures illustrates that in lower Anth content, the Chl and Car variates sharply, especially for the Anth content ≤15mg/ m^2

After researching above in detail, we confirm following sensitive spectrum as part of our modeling factors. There are totally fifteen spectral fields (see above), each spectral field takes the average value of the nearest several channels. Some of them are Chl or Car sensitive spectrum. For instance, 550.62 to 553.14,641.13 to 643.55, 689.8 to 692.51,696.56 to 699.93.So that the affection of Chl and/or Car can be compensated in the new model.

4.1.3 Introduce of the Secondary Index.

We introduce Secondary Index, such as MARI to improve accuracy of the model. Here is the format of MARI

$$MARI_{\lambda NIR} = [1/\rho_{\lambda Green} - 1/\rho_{\lambda Re}] \times \rho_{\lambda NIR} \tag{1}$$

Here we select λ_{Green} as 550.62 to 553.14 nm spectrum, λ_{Re} as 696.56 to 699.93 nm spetrum; and λ_{NIR} as 721.04 to722.71,739.94 to 742.57,748 to 750 nm separately. So we get three MARI indices, label as $MARI_{721}$, $MARI_{739}$, $MARI_{750}$ separately. We found that there is variation on MARI between grapevine species and growth stage.

We also select $1/\rho_{\lambda Green}$ and $1/\rho_{\lambda Re}$ and NIR700/Visible676.2 as the secondary index, totally six factors [6,7,4].

Table 1.   r ^{2}-value between Anthocyanin content Vs part of Indices

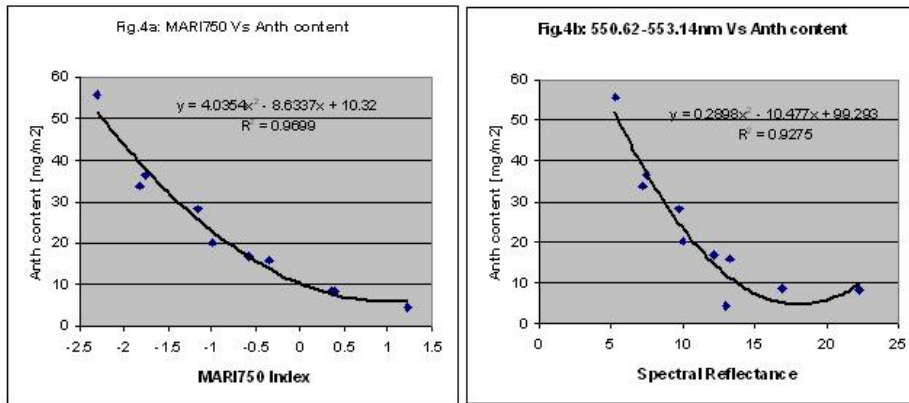| Index	MARI750	550.62 to553.14	696.56 to699.93	748 to750	$1/\rho_{550.62\ to\ 553.14}$	$1/\rho_{696.56\ to\ 699.93}$
r ^{2}	0.658	0.2506	0.0155	0.5275	0.61	0.0693

Fig. 4. The graphic illustrates that in some case, 3 band model/or MARI can improve the Linear feature between Anth content and original data space $\mathfrak{R}m$ in local fields, e.g. sub-dimension. It can makes a asymptote form of Fig.4b to be a near linear form of Fig.4a.It also makes that the scattering points in Fig.4b shrinks closed to the curve in Fig.4a. Especially in lower Anth content points. The data shown here are from leaves sampled on June 9,2006 of Stpepin.

From Fig.4a&4b which the data is taken from June 9,2006 of Stpepin, we can see that by introducing 3 band model/or MARI as index, the Linear feature between Anth content and original data space $\mathfrak{R}^m$ in local fields has been improved ; in lower value points of Anth content, the points become much more closed to the curve, e.g. shrunk to the curve(Fig.4a), which shows that Errors will

become more little. Those MARI indices will make that the learning procedure converges much faster.

By introducing six secondary indices, we have got improvement on Er_i (See Eq.(5)) percentage Error, e.g. Er_i divides by its measured Anth content. After introducing secondary Indices, the average value of top ten samples of most high Er_i percentage Error has been decreased, from 221.77% to 80.73% in our learning and predict-testing procedure, and the sample number of the most high Er_i Error percentage (here it is 7 samples which Er_i Error percentage 100% in our learning and testing procedure) has also been decreased to 3 samples. Those improvements will be very important in lower value point of Anth content. We can infer that the new model cover advantages in many methods, include MARI, ARI,3 bands model, because we can introduce as many indices as possible into the new model. And by learning procedure, the new model optimizes its accuracy and ability of its generalization or ability of prediction.

We can infer that the new model cover advantages in many methods, include MARI, ARI,3 bands model, because we can introduce as many indices as possible into the new model. And by learning procedure, the new model optimizes its accuracy and ability of its generalization or ability of prediction.

4.2 SVM Approach to Establishing Non-Linear Model

Support Vector Machine is a kind of high-degree non-linear method [25, 10,11], it also possesses good robustness and lots of another good feature, those characteristics make it become one of the most prominent machine learning technique for high dimension sparse data. The principle of the technique is summarized below.

Taking measured Anthocyanin content Y_i and its corresponding factors $<x_{i1},x_{i2},.....,x_{im}>^T$, or we can note factor vector as X_i , it is belong to $\Re^m$, here m = 25 represents factors number, i≤n represents sample number, as a pair dataset, normally it is called as sample.

The learning task is defined a fixed array X of n points(X_1, X_2,...., X_n). Each data point has a desired Y_i. For training, the learner receives the labels Y_i for a random subset Y_1,

$$Y_1 \subset [1, 2, …, n] \qquad\qquad | Y_1 | = l \qquad\qquad (2)$$

Of l<n (training) data points. Here $| Y_1 |$ represents the number of element of subset Y_l. The goal of the learner is to predict the labels Y_i of the remaining (testing) points in X as accurately as possible.

On another aspect, we can say that the principle of SVM is based on the hypothesis that all samples possess the similar features, so we can establish model from its subdatasets(learning) that can predict the remaining(testing) subdatasets. In our case, anthocyanin , Chlorophoylls and Cartotenoids reflect and absorpt at fixed channel of light spetrum in the same grape species and similar growth stage, their reflectance and absorption also follow some rules, so we can infer that every sample possesses the similar feature. Our training goal is

to extract the rule, e.g. the relationship of anthocyanin content and factors, which is hidden behind subdatasets (learning), so that we can we can predict the remaining(testing) subdatasets. In our research, we randomly sort all samples, and then cut about two third as the training datasets (90 samples), remaining points as the testing datasets (54 samples).

We use SVM (*SVM^{Light}V6.01*) programmer to train the training datasets, and get the model, then use the testing datasets to test the model. Next we optimizes its accuracy and ability of its generalization or ability of prediction as likelihood as possible. Finally we had gotten best fit model, the function between Y_i and $<x_{i1}, x_{i2}, \ldots, x_{im}>^T$

$$Y = \omega \cdot \phi(X) + b \qquad (3)$$

Here ω is Weigh Vectors, b is Deviation Vector. And $\phi(X)$ is project function from $\Re^m$ space to $\Re$ space and can be replaced by a kernel function, here we select Gauss kernel that is expressed by

$$k(x_{ik}, x_{ij}) = \exp(- \parallel x_{ik} - x_{ij} \parallel^2 / (2 \delta^2)) \qquad (4)$$

Here δ is the parameter of rbf kernel.

4.3 Modeling effect from our training datasets

The effect of the model that we get from training dataset is shown in the Fig. 5a&5b. By the procedure of the training, we have established a kind of project from $\Re^m$ space to $\Re$ space. While in $\Re^m$ space, the points of original datasets of factors are existed as a curve fields, now they are shown in a linear form, and existed in a narrow channel. The strength of SVM model for the measured Anth content and predicted Anth content   was very significant, its r^2-value of 0.97289 (calculated according format).

Fig.5b shows Error distribution of the model, Average Error is 1.3823 mg/m^2, e.g. 13.32%. From the Fig.5b, we also find that the Error is not apparently increased with the increasing of Ant content. This characteristic shows that the effect of the model is better than normally statistical method. Here is the format of calculating Error, it is recorded as Er_i shortly.

$$Er_i = abs(Y_{\_predicted\ i} - Y_{\_original\ i}). \qquad (5)$$

Here i=1,2,…,n. From Fig.5b, we can find that almost all Er_i is under about 1.69 mg/m^2 line. r^2-value of 0.252 shows that the Er_i is not apparently relative to the Anth   content   increase.   This   phenomenon   shows   that   almost   all   useful information is transformed into SVM model, the Noise is "pure Noise".

To estimate how sensitive is each of predicted Anth content evaluated in this study, to change in reality Anth content(e.g.,Measured Anth content in Fig.5a), the Noise Equivalent Anth content(NE$_{\Delta Anth}$) was calculated[28].

$$NE_{\Delta Anth} = RMSE(\text{predicted Anth vs. Anth})/[d(\text{predicted Anth})/d(\text{Anth})]. \qquad (6)$$

Here RMSE (predicted Anth vs.Anth) is root mean square error(RMSE) of the relatioanship between predicted Anth and Anth content in reality, and d(predicted Anth) /d(Anth) is the first derivative of predicted Anth with respect to Anth. Here d(predicted Anth) /d(Anth) ≈1.
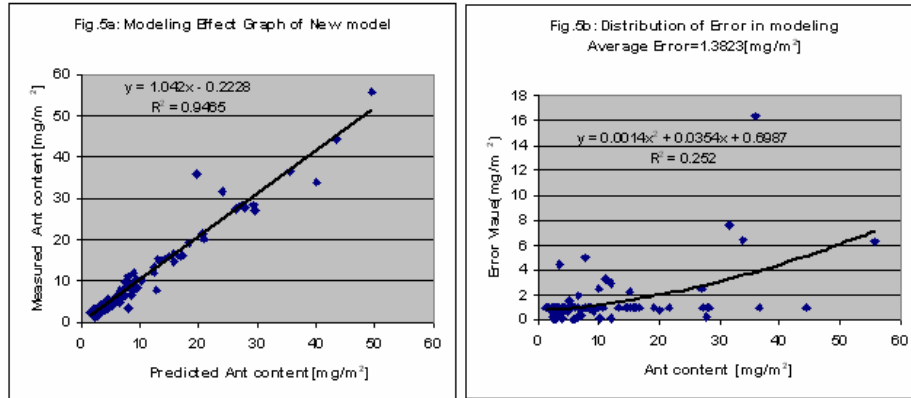

Fig. 5. The graphic illustrates that the effect of the new model. Fig. 5a shows that almost all(except one) point have shrunk into two parallel line. Fig.5b shows that the Noise is not apparently increased with the increasing of Ant content. And almost all points are below under 2 mg/m², it illustrates that Eri of lots of points is lower than 2 mg/m²

Fig.5b shows that Average Error is 1.3823 mg/m². From the format (6), we also find that the Noise Equivalent is 1.65 mg/m², it is far lower than that of ARI and MARI([20] Mark R.Steele,2009, we use the same data), it is 2.93 and 2.23 mg/m² separately. From Fig.5b, we can find that lots of Error is under about 1.7 mg/m² line, and lots of it near zero line. r^2 of 0.252 shows that the Error is not apparently relative to the Chl content increase. This phenomenon shows that almost all useful information is transformed into SVM model, the Noise is "pure Noise".

4.3 Predict-testing effect of our Model by the testing datasets

Fig.6a&6b also show the good performance of the SVM model. The strength of prediction for Anth content of the predict-testing datasets and Measured Anth content is also very significant, its r^2-value of 0.96176 (calculated according format), it is similar as that of the modeling.

The Error of the predict-testing datasets in Fig.5b also shows that the nice performance of SVM. Its Average Error is 3.618 mg/m²,e.g 25.99%, the Er_i is increased a little with the increase of the Anth content.

We can conclude that the performance of predicting ability of the SVM is very well.
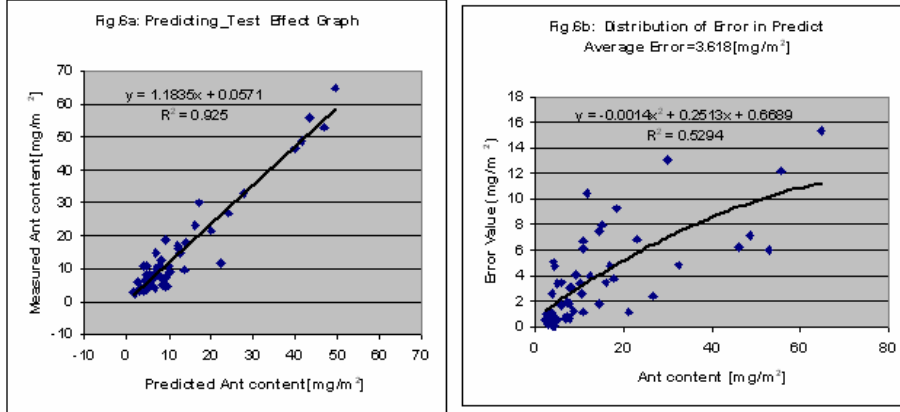
Fig. 6. The graphic illustrates the predicting performance. Fig.6a shows that all scattering points are existed between two parallel narrow lines, and its r^2 of 0.96176 (calculated according format) is similar as that of the modeling in Fig.5a. In Fig.6b, its Average Error is 3.618 mg/m^2,e.g 25.99%, the Er_i is increased a little with the increase of the Anth content.

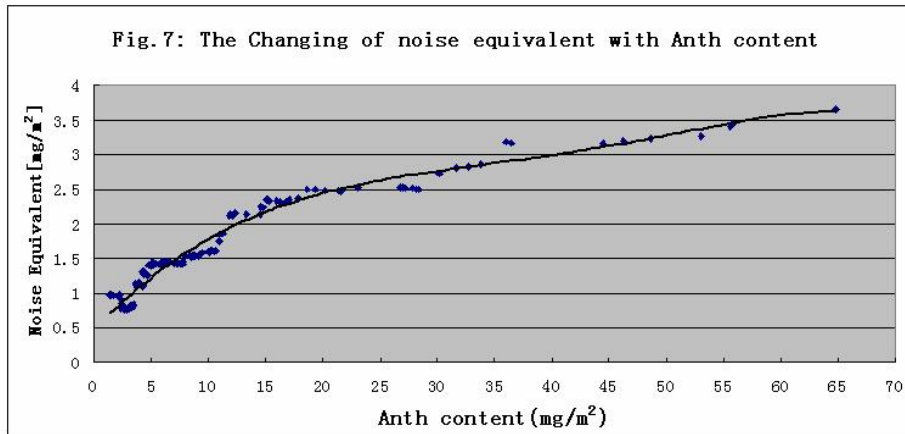## 4.4 Comparison with the existing methods

Fig. 7. Noise equivalent of SVM model was plotted versus Anth content throughout the range of Anth from 1.34 mg/m^2 to 64.74mg/m^2.

To further evaluate the accuracy of the new model, we calculated noise equivalent values($NE_{\Delta Anth}$) for SVM and plotted against measured Anth content(Fig.6). At very lower Anth content(while Anth≤5 mg/m^2), the $NE_{\Delta Anth}$ was very low, then the $NE_{\Delta Anth}$ increased sharply to 1.44 mg/m^2 ,but all $NE_{\Delta Anth}$ value is still lower than 1.44mg/m^2, and with the average $NE_{\Delta Anth}$ 1.0 mg/m^2 .

From Anth $=5$ mg/m^2 to 10 mg/m^2, the NE$_{\Delta Anth}$ value increases very little, but there is a fastly increasing from Anth $=10$ mg/m^2 to 12 mg/m^2. And then the NE$_{\Delta Anth}$ value increases slowly with Anth content. and it is lower than 3.5 mg/m^2. While Anth$\leq$15 mg/m^2, the whole average NE$_{\Delta Anth}$ is 1.0 mg/m^2; While Anth$\geq$45 mg/m^2, the average NE$_{\Delta Anth}$ is 3.4mg/m^2, the value is a higher than that of Anth$\leq$15mg/m^2. For the whole range from 1.34 mg/m^2 to 64.74 mg/ m^2, NE$_{\Delta Anth}$ of SVM has average NE$_{\Delta Anth}$ = 1.65 mg/m^2, which is lower than that of MARI=2.23 mg/ m^2 and ARI=2.93 mg/ m^2 in [20].

The poor performance of the normally statistics methods is due to that it cannot describe the complex relationship of the Anth content and its reflectance, and the variation in species and growth stage. They cannot fully employ the wealth information of the hyperspectral

4.5 A simple description about the theoretical principle over procedures

The learning procedure of SVM guarantees structural risk minimization of the model, and optimizes its accuracy, so SVM usually exhibit good generalization performance and high accuracy. The procedure in 4.2 item can ensure evenly distributed samples, so that the learning subset of samples can cover the $\Re^m$ space much more perfectly, and finally to ensure ability of generalization is strong.

5. Result and Discussion

Red/green[3],Modified ACI(e.g,$\rho_{\lambda NIR}/\rho_{\lambda 530}$)[27],ARI[5],MARI[5,6] are some traditional methods to estimate pigment content in leaves nondestructively. The authors of [20] used those indices to estimate Anth content in grape leaves with the same data of this study. They found that Anth prediction by Red/green was poorest with determination coefficient($r^2 \approx 0.76$) and highest RMSE (7.63 mg/ m^2);Anth prediction by ARI was much more accurate that the Red/green($r^2 \approx 0.96$,RMSE<2.93 mg/ m^2); and MARI was the most accurate in Anth prediction($r^2 \approx 0.97$,RMSE<2.23 mg/ m^2).

In this study, we found that both of SVM and MARI/ARI has a coefficient of determination $r^2 \approx 0.96$ for the whole Anth ranging.The Noise equivalents of Anth estimation by both the MARI and ARI has shown in [20]. This study uses the same data, but NE of SVM has average NE = 1.65 mg/m^2 ,for the whole range from 1.34 mg/m^2 to 64.74 mg/ m^2, which is lower than that of MARI=2.23 mg/ m^2 and ARI=2.93 mg/ m^2 in [20],e.g. the result has been improved a lots.

For the Anth$\leq$15 mg/m^2, the NE$_{\triangle Anth}$ of MARI or ARI was about 2.23 to 2.93 mg/m^2 , but NE$_{\triangle Anth}$ of SVM was 1.00mg/m^2,thus SVM was much more accurate (less noisy) than both MARI and ARI, it is very sensitive for early detection of plant environmental stress.

From Fig.5b&Fig.6b, we can see that that the Average Error is not apparently relative to the Anth content increase,even if Anth content is beyond 45 mg/m^2,it is great interest for detection of heavily plant environmental stress.

6. Conclusion

By introducing SVM, we had developed a more accurate quantitative model to like leaf Anth content in grape with spectral reflectance in visible/Red-Infrared region. The fifteen important spectral regions in visible/Red-Infrared region were identified: 400.3 to 404.8;450.22 to 453.19;494.61 to 499.38;533.65 to 536.18; 550.62 to 553.14;641.13 to 643.55;676.2 to 679.95;682.33 to 686.07;689.8 to 692.51; 696.56 to 699.93; 716.71 to 719.71;724 to 726;721.04 to 722.71;739.94 to 742.57;748 to 750. are most responsive to changes in leaf Anth content in grape. We also found that the senisitive channel is variation between grape species and growth-stages.
    Outperformance of the new methods is partially due to that it removed the affect of anthocyanin, carotenoids and it removed the affect by the scattering of the grape leaf internal structure and leaf surface.
    The new methods was found to be capable of accurately estimating Anth contents across the whole Anth range , thus the new methods is likely to allow accurate Anth determination in *Vitis vinifera* vines, especially can used for quantitative assessment of the early stage of plant environmental stress.

Acknowlegements

Reference

[1] Curran,P.J.,1989. Remote Sensing of Foliar Chemistry. Remote Sens. Environ. 30:271-278.
[2] Dugald C.C., and L.B.Christopher. The Ecophoysiology of Foliar Anthocyanin. The Botanical Review. 69(2):149-161.
[3] Gamon, J., and J. Surfus. 1999. Assessing leaf pigment content and activity with a reflectometer. New Phytol. 143:105-117.

[4] Giorgio Dall'Olmo ,A.G.Anatoly, C.R.Donald,2003. Towards a unified approach for remote estimation of chlorophyll-a in both terrestrial vegetation and turbid productive waters. Geophsical Research Letters, 30(18),1938, doi:10.1029/2003GL018065.

[5] Gitelson,A.A., Mark N. Merzlyak and Olga B. Chivkunova,2001. Optical Properties and Nondestrstructive Estimeation of Anthocyanin Content in Plant Leaves.Photochemistry and Photobiology,74(1):38-45.

[6] Gitelson,A.A., Galina P. Keydan, and Mark N.Merzlyyak,2006. Three-band model for nonvasive estimation of chlorophyll, carotenoids, and anthocyyanin content in higher plant leaves. Geophsical Research Letters, VOL.33,L11402, doi:10.1029/2006 GL026457.

[7] Gitelson ,A.A., Andrés Viña, Veróna Ciganda, C.R.Donald, and J.A.Timothy,2005. Remote estimation of canopy chlorophyll content in crops. Geophsical Research Letters, VOL.32,L10843, doi:10.1029/2005 GL022688.

[8] Grant,L.,1987. Review article. Diffuse and Specular Characteristics of Leaf Reflectance .Remote Sensing of Environment.22:309-322.

[9] Joachims, T. ,1999a. Estimating the generalization performance of a SVM efficiently (LS VIIIReport 25). Universität Dortmund, Germany.

[10] Joachims, T. ,1999b. Making large-scale SVM learning practical. In B. Schölkopf, C. Burges, &A. Smola (Eds.), Advances in kernel methods - Support vector learning. Cambridge, MA: MIT Press.

[11] Joachims,T.,2002.Learning to Classify Text using Support Vector Machines Methods, Theory, and Algorithms. Dissertation, Kluwer.

[12] Kearns, M., & Ron, D. ,1997. Algorithmic stability and sanity-check bounds for leave-one-out crossvalidation. Proceedings of the Tenth Conference on Computational Learning Theory (pp. 152{162). New York: ACM Press

[13] Kevin S.G., R.M.Kenneth, H.S.Richard, and J.G.Jessica, 2000. Functional role of anthocyanins in the leaves of Quintinia serrata A. Cunn..Journal of Experimental Botany.51(347).1107-1115.

[14] Klinkenberg,R.,and T. Joachims,2000. Detecting Concept Drift with Support Vector Machines. Proceedings of the Seventeenth International Conference on Machine Learning (ICML), Morgan Kaufmann.

[15] Lichtenthaler, H.K. 1987. Chlorophyll and carotenoids: Pigments of photosynthetic biomembranes. Meth. Enzymol. 148:331-382.

[16] Pirie, A. and M.G.Mullins,1976. Changes in Anthocyanin Phenolics Content of Grapevine Leaf and Fruit Treated with Sucrose, Nitrate, and Abscisc Acid. Plant Physiol. 58:468-472.

[17] Qin, J.L.,2001. Preliminary Inquiry On Technical Support System for the Precision-Farming with Chinese Characteristics. Transaction Of the Chinese Society Of Agricultral Engineering .17(3):1-6.

[18] Joachims,T.,2002, Optimizing Search Engines Using Clickthrough Data, Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD), ACM.

[19] Mark R.Steel, Anatoly A.Gitelson,Donal C.Rundquist,2008.Research Note Nodestructive Estimation of Leaf Cholorophyll Content in Grapes.American Journal of Enology and Viticulture,59(3):299-305

[20] Mark R. Steele, and A.A.Gitelson,Donald C.R.,2009.Research note nondestructive estimation of anthocyanin content in grapevine leaves. American Journal of Enology and Viticulture,60:87-92.

[21] Samuel O.N., and K.S. Gould,2003.Anthocyanins in leaves: light attenuators or antioxidant. Functional Plant Biology.30:865-873.

[22] Taylor S.Feild, David W.Lee, and N.Michele Holbrook. Why Leaves Turn Red in Autumn. The Role of Anthocyanins in Sensescing Leaves of Red-Osier Dogwood, Plant Physiology,October 2001,Vol.127,pp.566-574.

[23] Tsochantaridis, I., T.Hofmann, T.Joachims, and Y.Altun,2004. Support vector machine learning for interdependent and structured output spaces. ICML.

[24] Tsochantaridis, I., T.Joachims, T.Hofmann, and Y.Altun,2005. Large margin methods for structured and interdependent output variables. JMLR, 6, 1453–1484.

[25] Vapnik, V. ,1998. Statistical learning theory. Chichester, UK: Wiley.

[26] Wahba, G. ,1999. Support vector machines, reproducing kernel hilbert spaces, and randomized gacv.In B. Schölkopf, C. Burges, & A. Smola (Eds.), Advances in kernel methods - Support vector learning.Cambridge, MA: MIT Press.

[27] van den Berg, A.K., and T.D. Perkins. 2005. Nondestructive estimation of anthocyanin content in autumn sugar maple leaves. HortScience 40:685-686.

[28] Viña, A, and A.A. Gitelson. 2005. New developments in the remote estimation of the fraction of absorbed photosynthetically active radiation in crops. Geophys. Res. Lett. 32:L17403.

[29] Wantanachaturaporn, P., K.A.Manoj, and P.K.Varshney,2008. Multisource Classification Using Support Vector Machines: An Empirical Comparison with Decision Tree and Neural Network Classifiers.Photogrammetric Engineering& Remote Sensing. February 2008.74(2):239-246.

[30] Yu, C.N., T.Joachims, R.Elber, and J.Pillardy, 2007.Support vector training of protein alignment models.RECOMB.