

بررسی اثر روش‌های مختلف پیش‌پردازش داده‌های طیفی محلول‌های قندی در ناحیه زیرقرمز نزدیک بر مدل تخمین غلظت ساکارز

محمدعرفان بهرامی^۱، مسعود هنرور^۲، کیوان انصاری^{۳*}، بهاره جمشیدی^۴

۱- دانشجوی دکتری مهندسی علوم و صنایع غذایی، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات تهران، تهران، ایران، کدپستی: ۱۴۷۷۸۹۳۸۵۵

۲- دانشیار، گروه علوم و صنایع غذایی، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات تهران، تهران، ایران، کدپستی: ۱۴۷۷۸۹۳۸۵۵

۳- استادیار، گروه پژوهشی نمایش رنگ و پردازش تصاویر رنگی، پژوهشگاه رنگ، تهران، ایران، صندوق پستی: ۱۶۷۶۵-۶۵۴

۴- استادیار، موسسه تحقیقات فنی و مهندسی کشاورزی، سازمان تحقیقات، آموزش و ترویج کشاورزی، کرج، ایران، صندوق پستی: ۳۱۵۸۵-۸۴۵

تاریخ دریافت: ۱۳۹۷/۸/۷ تاریخ پذیرش: ۱۳۹۷/۹/۲۶ در دسترس به صورت الکترونیکی از: ۱۳۹۸/۶/۴

چکیده

در سال‌های اخیر استفاده از روش‌های طیف‌سنجی زیرقرمز نزدیک برای ایجاد یک مدل تخمین خصوصیات فیزیکی و شیمیایی ترکیبات در صنایع غذایی گسترش یافته است. در این تحقیق نسبت به بررسی تاثیر برخی از روش‌های پیش‌پردازش داده‌های طیفی همچون میانگین‌گیری متحرک، تصحیح پراکنش افزایشنده، هموارسازی سائیتزکی-گولای، توزیع نرمال استاندارد، مشتق اول و مشتق دوم بر مدل تخمین میزان ساکارز موجود در شربت‌های کارخانه تولید شکر اقدام شد. نتایج نشان داد که روش‌های پیش‌پردازش توزیع نرمال استاندارد و مشتق‌گیری دوم می‌توانند بهبود موثری در تخمین توسط مدل تدوین شده داشته باشند.

واژه‌های کلیدی: پیش‌پردازش طیفی، طیف‌سنجی زیر قرمز نزدیک، ساکاروز، شربت‌های قندی.

Investigating the Effect of Various Pre-Processing Methods of the Spectral Data of Sugar Syrups on the Model Estimation of Sucrose Concentration in Near Infrared Region

M. E. Bahrami¹, M. Honarvar¹, K. Ansari², B. Jamshidi³

1- Department of Food Science and Technology, Islamic Azad University, Science and Research Branch of Tehran, P. O. Code: 1477893855, Tehran, Iran

2- Department of Color Imaging and Color Image Processing, Institute for Color Science and Technology, P. O. Box: 16765-654, Tehran, Iran

3- Agricultural Engineering Research Institute, Agricultural Research Education and Extension Organization (AREEO), P. O. Box: 31585-845, Karaj, Iran

Received: 29-10-2018

Accepted: 17-12-2018

Available online: 26-08-2019

Abstract

Using Near-infrared spectroscopy methods to create the physical and chemical properties estimation model of material has been developed and considered by the food industries in recent years. The present study was conducted to investigate the statistical effects of various pre-processing techniques such as Moving Average (MA), Multiplicative Scatter Correction (MSC), Savitzky-Golay (SG), Standard Normal Variate (SNV), First and Second derivative on the Near-Infrared spectral data of the sugar syrups on sucrose concentration estimation model. The results showed that the SNV and the Second derivative techniques could have an effective improvement in the estimation model. *J. Color Sci. Tech.* 13(2019), 99-106©. Institute for Color Science and Technology.

Keywords: Spectral pre-processing, Near infrared spectroscopy, Sucrose, Sugar syrups.

۱- مقدمه

غلظت یک ماده مورد نظر از طریق طیف NIR به دست آمده بکار گرفته می‌شود، از مهم‌ترین و پرکاربردترین روش تجزیه و تحلیل داده چند متغیره محسوب می‌گردد که از آن می‌توان به منظور تجزیه و تحلیل نمونه‌های پیچیده نیز استفاده نمود [۶]. برای چنین روشی، ابتدا غلظت ماده مورد نظر در نمونه‌های آزمایشی با استفاده از روش‌های استاندارد و به کمک دستگاه‌های اندازه‌گیری دقیق ارزیابی و تعیین می‌شود. سپس از نمونه‌ای آزمایشی در محدوده طیفی زیرقرمز نزدیک، طیف‌گیری می‌گردد. آنگاه از روش رگرسیون چند متغیره برای تعیین یک مدل کالیبراسیون مناسب که به خوبی نیز از نظر آماری معنادار باشد، جهت ارتباط میان غلظت ماده مورد نظر در ترکیب نمونه‌های آزمایشی با داده‌های طیفی استفاده می‌شود. بدین ترتیب این مدل باید بتواند برای پیش‌بینی غلظت نمونه‌های مجهول مورد استفاده قرار گیرد [۷].

اما به جهت بهبود نتایج، لازم است که پیش پردازشی بر روی داده‌های طیفی خام اولیه قبل از تعیین مدل کالیبراسیون اجرا شود، چرا که در اکثر موارد داده‌های طیفی حاصل از طیف‌سنجی حاوی برخی اطلاعات پیش‌زمینه‌ای ناخواسته و نویزهایی است که باعث می‌گردد استخراج داده‌ها و ساخت یک مدل صحیح کالیبراسیون با خطا مواجه شود [۸]. از این رو روش‌های مختلف پیش‌پردازش داده‌های طیفی برپایه روش‌های ریاضی و آماری پیشرفته با هدف حذف این گونه خطاها گسترش یافته است که از آن جمله می‌توان به برخی روش‌های هموارسازی^۱ مانند میانگین‌گیری متحرک (MA^V)، سائویتی-زکی-گولای (SG^A) و همین‌طور برخی روش‌های نرمال‌سازی مانند تصحیح پراکنش افزاینده (MSC)^۹ و توزیع نرمال استاندارد (SNV)^{۱۰} و نیز روش‌های پیش‌پردازشی که موجب افزایش قدرت تفکیک طیفی می‌شوند مانند مشتق‌گیری نوع اول و دوم^{۱۱}، اشاره کرد [۹].

در سال‌های اخیر استفاده از روش‌های پیش‌پردازش مختلف در طیف‌سنجی زیرقرمز نزدیک به منظور بهبود مدل کالیبراسیون ساخته شده و استخراج بهتر اطلاعات درباره نوع و غلظت ترکیب شیمیایی مورد نظر در صنایع مختلف غذایی [۱۰]، به خصوص در محصولات صنعت قند و شکر به طور فزاینده‌ای رو به رشد بوده است [۱۱-۱۷]. در تحقیقی که توسط روگو و همکارانش در سال ۲۰۰۲ جهت تعیین میزان ساکارز موجود در چغندر قند با استفاده از طیف‌سنجی زیرقرمز نزدیک انجام شد، نسبت به ارائه روش‌های مختلفی برای پیش‌پردازش داده‌های طیفی اقدام نمودند. نتایج

طیف‌سنجی زیرقرمز نزدیک^۱ (NIRS) روشی از انواع طیف‌سنجی است که براساس تابش فوتون‌های دارای انرژی در بازه $10^{-19} \times 2,65$ الی $10^{-20} \times 7,96$ ژول و سپس بررسی ارتعاش ملکول‌های مواد کار می‌کند. این سطح از انرژی سبب ایجاد طول موج‌هایی در محدوده ۷۸۰ تا ۲۵۰۰ نانومتر (با عدد موجی 12800 تا 400 cm^{-1}) است [۱]. طیف‌سنجی زیرقرمز نزدیک از جمله روش‌های نوظهوری است که در سال‌های اخیر به دلیل مزایا و توانمندی بسیار مورد توجه صنایع دارویی، شیمیایی و غذایی قرار گرفته است. این روش نسبت به دیگر فناوری‌های رایج تعیین کیفیت همانند روش‌های کروماتوگرافی گازی، کروماتوگرافی جرمی و کروماتوگرافی مایع، دارای برتری نسبی است. از آن جمله می‌توان به عدم نیاز نمونه آزمایشگاهی به آماده‌سازی خاص، امکان بهره‌گیری به صورت برخط^۲، هزینه آزمون کمتر، امکان اندازه‌گیری هم‌زمان چند ترکیب هدف، سهولت به کارگیری و سازگاری روش با محیط‌زیست اشاره کرد [۲].

محدوده طیفی زیرقرمز نزدیک اطلاعات مربوط به جذب پیوندهای C-H، N-H و O-H را پوشش می‌دهد، که این پیوندها جز ساختارهای اولیه و اصلی ترکیبات آلی هستند [۳]. در فرآیند طیف‌سنجی هنگامی که مولکول‌های مواد، پرتوهای الکترومغناطیسی را جذب می‌کنند، به ارتعاش در می‌آیند و در این بین به طور تقریبی تنها فرکانس‌های مشخصی از تابش زیرقرمز می‌توانند توسط مولکولی خاص، جذب و موجب مرتعش شدن آن گردند [۴]. بدین ترتیب انرژی پرتو زیرقرمز با فرکانس مشخص که با فرکانس ارتعاش طبیعی مولکول مورد نظر تطبیق دارند، جذب شده و برای افزایش دامنه حرکت ارتعاشی اتصال موجود در مولکول به کار گرفته می‌شود. سپس با استفاده از عدد موجی یا طول موج تابش جذب شده طیف NIR توسط برخی از پیوندهای شیمیایی موجود در نمونه، می‌توان نسبت به شناسایی ترکیبات شیمیایی مواد آلی و مولکول‌های زیستی اقدام نمود [۵].

اگرچه روش‌های NIRS کارآمد هستند، اما استفاده از اطلاعات طیفی به دست آمده نیازمند تجزیه و تحلیل‌های بیشتر و استفاده از روش‌های کمومتریکس^۳ یا شیمی‌سنجی است. یکی از بخش‌های عمده کمومتریکس، تجزیه و تحلیل داده چند متغیره^۴ است که به منظور انجام آزمون‌های کمی و کیفی توسط طیف‌سنجی NIR ضروری و لازم است [۶]. در میان روش‌های مختلفی که برای تجزیه و تحلیل داده‌های چند متغیره عنوان شده است، روش رگرسیونی کمینه مربعات جزئی (PLS)^۵ که به منظور تخمین و پیش‌بینی

6- Smoothing

7- Moving Average

8- Savitzky-Golay

9- Multiplicative Scatter Correction

10- Standard Normal Variate

11- First and Second Derivation

1- Near Infrared Spectroscopy

2- Online

3- Chemometrics

4- Multivariate data analysis

5- Partial least square

کارخانه قند به منظور سنجش میزان ساکارز موجود در نمونه‌ها است که بر پایه چرخش نوری قند راست گردان ساکارز استوار است [۲۰].

۲-۲-۲- طیف‌گیری از نمونه‌ها

برای طیف‌گیری از نمونه‌ها از دستگاه طیف‌سنج NIRQuest512-2.5 در حالت انتقالی^۴ که محدوده طول موجی ۸۶۹ الی ۲۵۰۰ نانومتر را پوشش می‌دهد، استفاده گردید. نمونه‌های شربت قبل از طیف‌سنجی از صافی‌های با اندازه ۰,۴۵ میکرون رد شده و به مدت ۱ ساعت در دمای اتاق نگهداری شدند تا با دمای حسگر دستگاه هم‌دم گردند. نمونه‌های شربت در سل‌هایی از جنس کوارتز با اندازه ۱,۲ سانتی‌متر در ۱,۲ سانتی‌متر و با اندازه عبوری نور ۱۰ میلی‌متر قرار گرفتند. منبع نوری مورد استفاده در دستگاه طیف‌سنج، لامپ هالوژن و آشکارساز دستگاه نیز Hamamatsu G9208-512W InGaAs بود. طیف‌های NIR توسط نرم‌افزار OceanView اخذ و به رایانه منتقل گردیدند. بدین ترتیب برای هر نمونه سه بار اندازه‌گیری انجام شد و متوسط داده طیفی در بازه ۸۶۹ الی ۲۵۰۰ نانومتر در محدوده مربوطه محاسبه شد.

۲-۲-۳- پیش‌پردازش طیف‌ها

از آنجایی که طیف‌های ناحیه NIR همواره حاوی نویز و اطلاعات پس زمینه‌ای ناخواسته هستند، به منظور بهبود عملکرد تدوین مدل کالیبراسیون از پیش‌پردازش‌های مختلف استفاده شد. در این تحقیق از روش‌های پیش‌پردازش SNV، SG، MA، MSC و مشتق اول و دوم به صورت جداگانه و همزمان استفاده گردید.

۲-۲-۴- تدوین مدل کالیبراسیون

به منظور ساخت مدل کالیبراسیون، که ارتباط بین غلظت ساکاروز اندازه‌گیری شده توسط دستگاه پلاریمتر و داده‌های طیفی به دست آمده خام و پیش‌پردازش شده از دستگاه طیف‌سنج را نشان دهد، از روش حداقل مربعات جزئی یا PLS استفاده گردید. در روش PLS با استفاده از داده‌های تخمین زنده که واریانس را در هر دو متغیر پیش‌بینی‌کننده و متغیر پاسخ توجیه می‌کند یکسری ترکیبات خطی متعامد از متغیرهای مستقل و وابسته استخراج می‌شود که به آن متغیر پنهان^۵ گفته می‌شود و در واقع مدل نهایی بر مبنای این مولفه‌ها تبیین می‌گردد [۲۱]. بدین جهت مدل PLS بر اساس اعتبارسنجی متقابل^۶ کل داده‌ها تدوین و میزان دقت مدل بر اساس سنج‌های آماری^۷ همچون ریشه میانگین مربعات خطای

تحقیق نشان داد که تقریباً تمامی روش‌های به کار گرفته شده، جهت پیش‌پردازش داده‌های طیفی بر دقت مدل نهایی تأثیر مثبت داشته است [۱۸]. در تحقیق دیگر که توسط هین و همکارانش در سال ۲۰۱۶ انجام شد، تعیین غلظت سه قند گلوکز، فروکتوز و ساکارز را در شربت‌های فرموله‌شده مختلف به کمک روش NIRS و با استفاده از روش‌های رگرسیونی PLS مورد بررسی قرار دادند. روش‌های پیش‌پردازش به کار گرفته شده در تحقیق آنها SNV و مشتق‌های اول و دوم بوده که نتایج پژوهش نشان داد میزان R^2 ، RMSEC^۱، RMSECV^۲ و SEP^۳ در همه مدل‌هایی که از پیش‌پردازش SNV و مشتق اول به صورت هم‌زمان استفاده شده بود، بالاتر است [۱۱]. در پژوهشی دیگر که توسط مقیمی و همکارانش انجام شد، تلاش گردید تا با استفاده از روش‌های طیف‌سنجی NIR و کمومتریکس میزان ماده خشک و اسیدی بودن میوه کیوی اندازه‌گیری شود. آنها در این تحقیق از روش‌های مختلف پیش‌پردازش داده‌های طیفی استفاده نموده و اهمیت بالای این روش‌ها در تبیین مدل و میزان صحت آن را نشان دادند [۱۹].

در پژوهش حاضر تلاش شده است تا اثر روش‌های مختلف پیش‌پردازش داده‌های طیفی ناحیه NIR که از طیف‌سنجی محلول‌های حد واسط صنعت قند که با نام شربت خام، شربت رقیق و شربت غلیظ شناخته می‌شوند، بر روی مدل کالیبراسیون بررسی گردد. هدف این کار افزایش دقت و کاهش خطا در تخمین میزان ساکارز به عنوان مهم‌ترین ترکیب موجود در این شربت‌ها است.

۲- بخش تجربی

۲-۱- مواد

در این تحقیق تعداد ۱۰۰ نمونه از انواع شربت‌های میان‌فرآیندی کارخانه قند که از دو کارخانه همدان و قزوین نمونه‌برداری شده بودند، تهیه گردید. کلیه نمونه‌ها پس از نگهداری در دمای ۴ درجه سانتی‌گراد و عملیات صاف کردن به قدری که به صورت محلول شفاف رنگی در بیابند، مورد آنالیز شیمیایی و نیز تعیین طیف جذبی قرار گرفتند.

۲-۲- روش‌ها

۲-۲-۱- اندازه‌گیری میزان ساکارز نمونه‌ها

به منظور سنجش میزان ساکارز موجود در محلول‌های قندی از دستگاه پلاریمتر مدل Saccharomat ساخت کشور آلمان در طول موج ۵۸۹ نانومتر استفاده گردید. پلاریمتر رایج‌ترین دستگاه در یک

4- Transmission Mode

5- Latent Variable

6- Full Cross Validation

7- Statistical Factors

1- Root mean square error of calibration

2- Root mean square error of cross validation

3- Standard error of prediction

می‌شود، نویزها و نوسانات در برخی از نمونه‌ها و در بعضی از نواحی طیفی دیده می‌شود که این مسئله می‌تواند دقت و صحت مدل تدوین شده را تحت تاثیر قرار دهد. عملیات ریاضی که منجر به دستیابی مدل‌های پایدارتر و قابل اعتمادتر در آزمایش‌های طیف‌سنجی می‌شود شامل انواع روش‌های پیش پردازش طیف‌ها می‌باشد که هر کدام با هدف خاصی به کار می‌رود ولی در عین حال انتخاب یک روش مشخص پیش‌پردازش که بتواند منجر به ارائه بهترین نتایج در آزمایشات مختلف شود، وجود ندارد. از این رو انواع روش‌های پیش‌پردازش مختلف به شرح زیر روی طیف‌ها انجام گرفت:

۱- میانگین‌گیری متحرک (MA): این روش یکی از انواع روش‌های پیش‌پردازش طیفی است که در این روش از یک فیلتر میانگین‌گیر متحرک استفاده و هر نقطه از طیف با میانگین یک همسایه (نقطه‌های مجاور شامل خود نقطه) از آن جایگزین می‌شود. در این نوع پیش‌پردازش انتخاب درست عرض پنجره فیلتر^۱ به منظور افزایش نسبت سیگنال به نویز اهمیت ویژه‌ای دارد [۲۲]. در این تحقیق از عرض پنجره فیلتر ۳ برای پیش‌پردازش داده‌ها استفاده گردید. داده‌های طیفی پس از انجام میان‌گیری متحرک پیک‌های هموارتری را از خود نشان دادند (شکل A۲).

1- Segment Size

کالیبراسیون (RMSEC)، خطای پیش‌بینی (SEP)، ریشه میانگین مربعات خطای پیش‌بینی (RMSEP)، ضریب همبستگی کالیبراسیون و پیش‌بینی (r_p, r_c) مورد سنجش قرار گرفت (رابطه‌های ۱-۳). همچنین برای انجام پیش‌پردازش‌ها و تدوین مدل از نرم‌افزار The UnscramblerX نسخه ۱۰/۴ استفاده گردید.

$$SEP = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\hat{y}_i - y_i - \text{Bias})^2} \quad (1)$$

$$RMSEC = \sqrt{\frac{1}{n_c - 1} \sum_{i=1}^{n_c} (\hat{y}_i - y_i)^2} \quad (2)$$

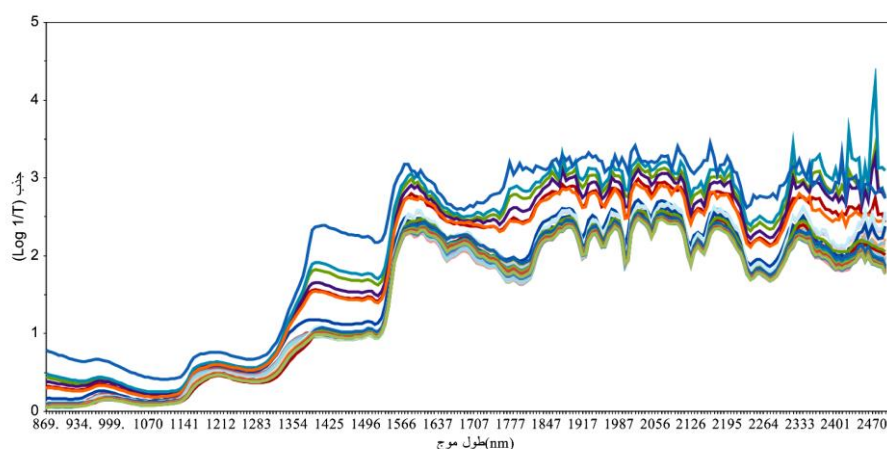
$$RMSEP = \sqrt{\frac{1}{n_p - 1} \sum_{i=1}^{n_p} (\hat{y}_i - y_i)^2} \quad (3)$$

۳- نتایج و بحث

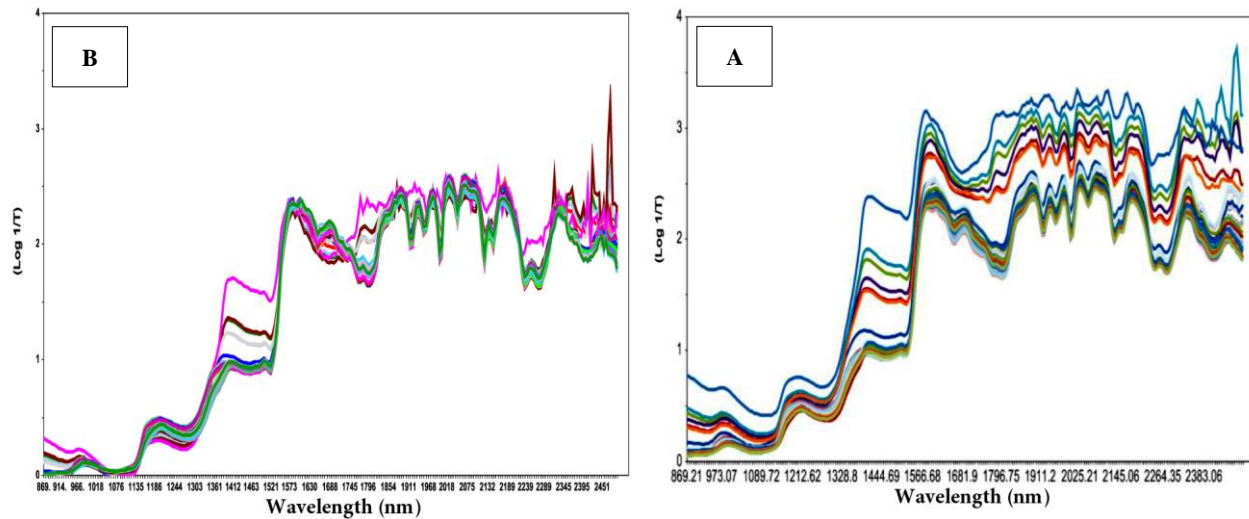
جدول ۱ نتایج مربوط به بررسی میزان ساکارز موجود در نمونه‌ها توسط دستگاه پلاریمتر را نشان می‌دهد. همان‌گونه که ملاحظه می‌شود غلظت کمینه و بیشینه ساکارز در محلول‌های قندی در دامنه وسیعی بوده و این خود می‌تواند نتایج حاصل از مدل را قابل اطمینان‌تر سازد. طیف‌های خام حاصل از نمونه‌های قندی در بازه ۸۶۹ تا ۲۵۰۰ نانومتر نیز در شکل ۱ نشان داده شده است. همان‌طور که از نمودار طیفی داده‌های خام در شکل ۱ مشاهده

جدول ۱: میزان ساکارز اندازه‌گیری شده به روش مرجع.

مورد اندازه‌گیری شده	میانگین	میانه	بیشترین	کمترین	انحراف از معیار
ساکارز پلاریمتری (/)	۲۷,۹۵	۱۵,۰۵	۵۷,۹	۱۱,۱	۱۷,۴۵



شکل ۱: طیف‌های حاصل از نمونه‌های شربت در طول موج بین ۸۶۹ تا ۲۵۰۰ نانومتر.



شکل ۲: انجام پیش‌پردازش میانگین‌گیری متحرک (A) و تصحیح پراکنش افزایشدهنده (B).

اطلاعات طیفی، برطرف کند. در این روش الگوریتم SG با برازش^۲ یک چندجمله‌ای بر هر بخش متوالی طیف و جایگزین نمودن مقدارهای اصلی طیف با مقدارهای برازش شده، هموارسازی را انجام می‌دهد. در روش SG انتخاب تعداد نقاط همسایگی و درجه چند جمله‌ای نیز از اهمیت زیادی برخوردار است تا بتواند ضمن افزایش نسبت سیگنال به نویز، اطلاعات پیکها را هم حفظ کند. در این پژوهش الگوریتم ساویتزکی-گولای با در نظر گرفتن نقاط همسایگی ۵ و درجه چندجمله‌ای ۲ اجرا شد. در شکل A۳ نحوه تغییر طیف‌ها پس از اعمال این نوع هموارسازی مشخص شده است.

۴- توزیع نرمال استاندارد (SNV): روش SNV نیز یکی دیگر از روش‌های پیش‌پردازش داده است که اساس آن بر نرمال‌سازی استوار است. این روش نیز مانند روش MSC می‌تواند اثرات جمعی و افزایشدهنده نور، اندازه نمونه و تغییر فاصله نوری را برطرف کند ولی روش محاسبه آن اندکی با MSC متفاوت است (رابطه ۶) [۲۲]:

$$X_{corr} = \frac{X_{org} - a_0}{a_1} \quad (6)$$

که در آن a_0 مقادیر میانگین برای طیف نمونه‌های است که قرار است اصلاح گردند، برای نرمال‌سازی این مقدار صفر در نظر گرفته می‌شود. از این روش نیز برای پیش‌پردازش داده‌ها در محدوده طیفی استفاده گردید و نتایج آن در نمودار B۳ نشان داده شده است.

۵- مشتق اول و دوم: مشتق‌های اول و دوم از روش‌های دیگر پیش‌پردازش طیف هستند که برای افزایش قدرت تفکیک طیفی، حذف پس‌زمینه و جابه‌جایی خط پایه^۳ مورد استفاده قرار می‌گیرند. در این روش برای حذف نویزها ابتدا باید طیف را هموار نمود که برای

۲- تصحیح پراکنش افزایشدهنده (MSC): این روش رایج‌ترین و معمول‌ترین روش پیش‌پردازش داده‌های طیفی است که به منظور تصحیح اثرهای جمعی و انحراف ناشی از عوامل فیزیکی (مانند پخش غیریکنواخت در کل طیف، اندازه نامناسب نمونه و ضریب شکست نور) به کار می‌رود. در این روش حذف اثرهای پخش با خطی‌سازی طیف هر نمونه با طیف میانگین نمونه انجام می‌شود [۲۲]. تغییرات حاصل در طیف‌ها پس از انجام MSC در شکل B۲ نشان داده شده است. همان‌گونه که مشاهده می‌شود روش MSC باعث شده تا طیف‌ها به یکدیگر بیشتر نزدیک شوند و همچنین شکل کلی پیک‌ها نمایان‌تر شود. روش MSC دو مرحله مجزا را شامل می‌شود:

۱- تخمین ضرایب تصحیح^۱ داده (رابطه ۴)

$$X_{org} = b_0 + b_{ref,1} \times X_{ref} + e \quad (4)$$

۲- تصحیح طیف گرفته شده (رابطه ۵)

$$X_{corr} = \frac{X_{org} - b_0}{b_{ref,1}} = X_{ref} + \frac{e}{b_{re,1}} \quad (5)$$

که در این رابطه‌ها X_{org} یک طیف از نمونه واقعی است که توسط دستگاه طیف‌سنج گرفته شده است، X_{ref} طیف مرجع است که برای پیش‌پردازش داده استفاده شده است، e بخش مدل‌سازی نشده X_{ref} است، X_{corr} طیف تصحیح شده است، و b_0 و $b_{ref,1}$ نیز کمیت‌های عددی می‌باشند که در آن b ضریب تصحیح و $b_{ref,1}$ شیب منحنی و b_0 عرض از مبدا منحنی است [۲۲].

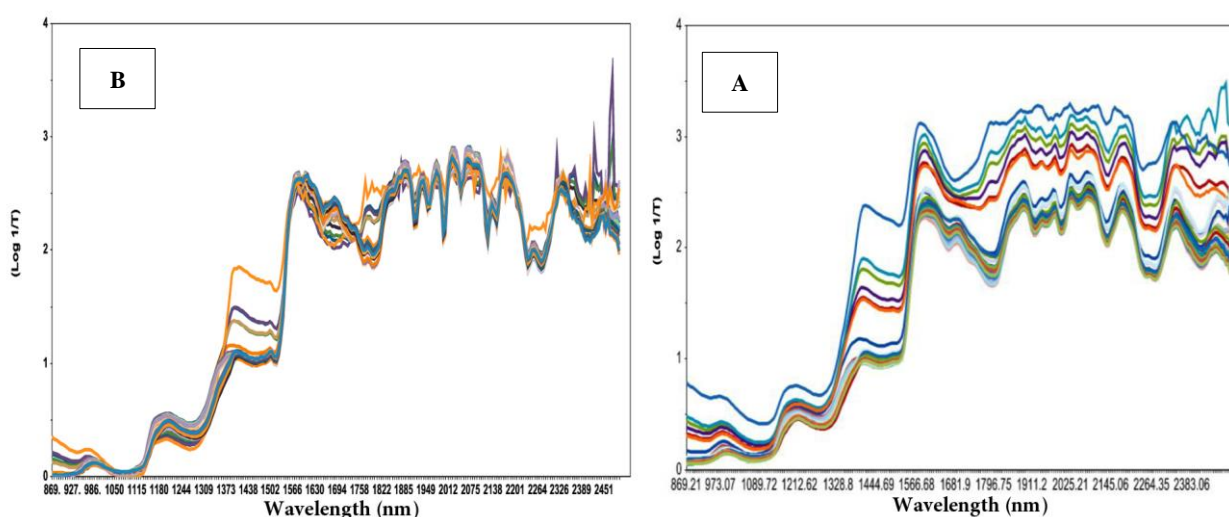
۳- هموارسازی ساویتزکی-گولای (SG): این روش می‌تواند به طور مفید و موثری نوسانات و نویزهای تیز طیف را بدون از بین بردن

که در آن X_i' و X_i'' به ترتیب مشتق اول و دوم را در نقطه (طول موج) i نشان می‌دهد. شکل ۴ مشتق‌های اول و دوم را که روی طیف‌ها اعمال شده است نشان می‌دهد. برای این کار تعداد نقاط همسایگی روی ۲ و درجه چندجمله‌ای روی ۳ تنظیم شدند.

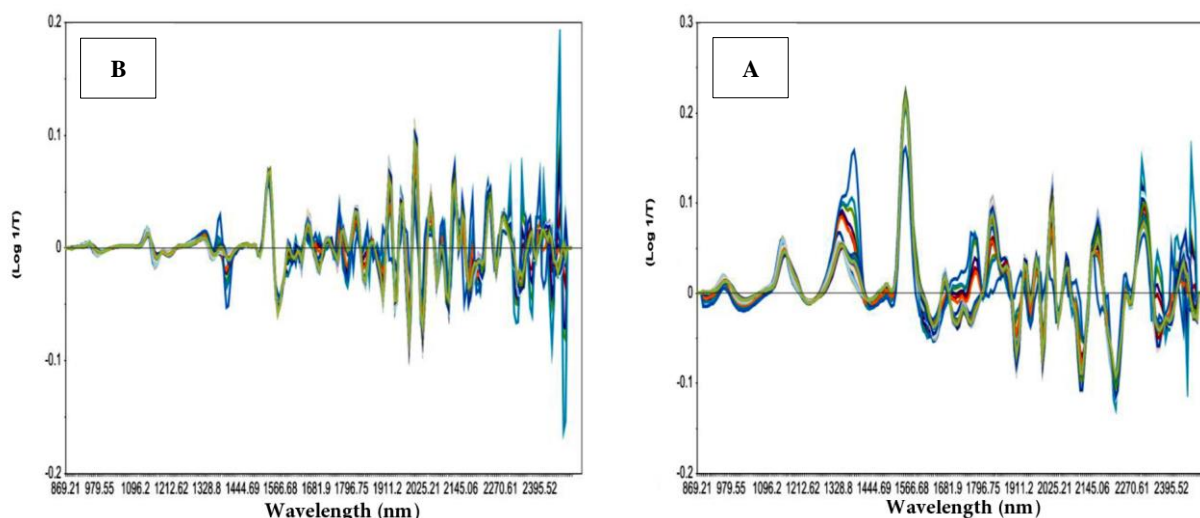
- 1- correction coefficients
- 2- Fitting
- 3- Baseline

این منظور هموارسازی SG از معمول‌ترین روش‌هایی است که استفاده می‌شود. با هر بار مشتق‌گیری از طیف و افزایش درجه آن پیک‌های تیز نسبت به پیک‌های پهن بیشتر می‌شوند. به همین دلیل است که مشتق دوم نسبت به مشتق اول به منظور افزایش قدرت تفکیک طیف کارآمدتر است. مشتق اول فقط خط پایه را حذف می‌کند، اما مشتق دوم هم خط پایه و هم انحراف خطی را برطرف می‌سازد. رابطه‌های مربوط به مشتق اول و دوم در پیش‌پردازش طیف‌ها بر طبق رابطه ۷ می‌باشند.

$$\begin{aligned} x_i' &= x_i - x_{i-1} \\ x_i'' &= x_i - x_{i-1} = x_{i-1} - 2x_i + x_{i+1} \end{aligned} \quad (7)$$



شکل ۳: پیش‌پردازش سوابت‌زکی گولای (A) و SNV (B).



شکل ۴: پیش‌پردازش مشتق اول (A) و دوم (B) طیف‌ها.

جدول ۲: نتایج حاصل از کالیبراسیون و تخمین مدل با انجام پیش‌پردازش‌های مختلف

نوع پیش‌پردازش	RMSEC	SEC	r_c	RMSEP	SEP	r_p
بدون پیش‌پردازش	۱,۶۲	۱,۶۳	۰,۹۹	۳,۷۷	۳,۷۹	۰,۹۵
MA	۲,۴۵	۲,۴۸	۰,۹۸	۳,۶۳	۳,۶۶	۰,۹۶
MSC	۲,۷۷	۲,۸	۰,۹۷	۳,۲۳	۳,۲۶	۰,۹۶
SG	۲,۳۴	۲,۳۶	۰,۹۸	۳,۲۴	۳,۲۶	۰,۹۶
SNV	۲,۷۶	۲,۷۹	۰,۹۷	۳,۱۹	۳,۲۲	۰,۹۷
مشتق اول	۳,۰۶	۳,۰۹	۰,۹۷	۳,۳۹	۳,۴۲	۰,۹۶
مشتق دوم	۲,۷۳	۲,۷۶	۰,۹۸	۳,۲۲	۳,۲۵	۰,۹۷

پایین‌ترین میزان RMSEP (به ترتیب ۳,۱۹ و ۳,۲۲) بهترین روش‌های پیش‌پردازش بودند.

۴- نتیجه‌گیری

در این تحقیق از روش طیف‌سنجی زیرقرمز نزدیک به منظور تشخیص میزان ساکارز موجود در شربت‌های قندی کارخانه و با تمرکز بر استفاده از روش‌های مختلف پیش‌پردازش استفاده گردید. این مسأله بسیار مهم است که در هر آزمون طیف‌سنجی روش یا روش‌های مناسب پیش‌پردازش داده‌ها گزینش شود، چرا که هر نمونه‌ای با توجه به شرایط کلی آزمایش، می‌تواند نتایج متفاوتی ارائه دهد و این مسئله صرفاً با آزمون و خطا روی انواع روش‌های پیش‌پردازش و مقایسه دقت مدل‌های حاصل به دست می‌آید. نتایج این تحقیق نشان داد که روش‌های مختلف پیش‌پردازش داده‌های طیفی می‌توانند با حذف اثرات ناخواسته بر طیف‌ها، دقت و صحت مدل به دست آمده را با افزایش میزان ضریب همبستگی و کاهش مقادیر RMSEP، افزایش دهند. در این پژوهش مشخص گردید که روش‌های پیش‌پردازش SNV و مشتق‌گیری دوم می‌توانند اثرات بیشتری در بهبود وضعیت مدل تدوین شده داشته باشند.

در گام بعدی به منظور بررسی اثرات پیش‌پردازش و ارائه مدل‌های کالیبراسیون بهتر و با دقت و قابل اعتماد تر با هدف پیشگویی میزان قند موجود در شربت‌ها بر پایه طیف‌های NIR، پس از اعتبارسنجی، مدل‌های رگرسیون چندمتغیره تدوین شده با این روش‌های پیش‌پردازش انجام شد. برای این منظور از مدل رگرسیونی PLS بر پایه روش‌های مختلف پیش‌پردازش استفاده گردید. جدول ۲ نتایج حاصل از کالیبراسیون و پیش‌بینی مدل‌های PLS تدوین شده بر پایه ترکیب روش‌های مختلف پیش‌پردازش اعمال شده روی طیف را برای تخمین غلظت ساکارز موجود در نمونه‌ها نشان می‌دهد.

مدل مناسب در این راستا مدلی است که دارای بیشترین ضریب همبستگی و پایین‌ترین میزان RMSEP بین نتایج واقعی و داده‌های پیش‌بینی شده باشد [۲۳]. همانطور که در جدول ۲ نشان داده شده است، مدل‌ها بر پایه روش‌های متفاوت پیش‌پردازش تدوین شده‌اند و به عنوان مثال مدلی که هیچ‌گونه پیش‌پردازش اولیه‌ای روی آن انجام نشده است، کمترین میزان ضریب همبستگی پیشگویی (r_p) و بالاترین میزان RMSEP در آن مشاهده شده است. با این حال در تیمارهایی که پیش‌پردازش روی داده‌ها اعمال شود میزان ضریب همبستگی افزایش و خطای پیشگویی کاهش پیدا کرده است. روش SNV و مشتق دوم با بالاترین میزان ضریب همبستگی (۰,۹۷) و

۵- مراجع

1. C. Pasquini, Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. *J. Braz. Chem. Soc.* 14(2003), 198-219.
2. M. Golic, K. Walsh, P. Lawson, Short-wavelength near-infrared spectra of sucrose, glucose, and fructose with respect to sugar concentration and temperature. *Appl. Spectrosc.* 57(2003), 139-45.
3. M. Manley, Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. *Chem. Soc. Rev.* 14(2014), 14-43.
4. L. U. Porep, D. R. Kammerer, R. Carle, On-line application of near infrared (NIR) spectroscopy in food production.

- Trends Food Sci. Technol.* 46(2015), 211-230.
5. H. Wang, J. Peng, C. Xie, Y. Bao, Y. He, Fruit quality evaluation using spectroscopy technology: a review. *Sensors*. 15(2015), 11889-11927.
 6. Y. Ozaki, W. F. McClure, A. A. Christy, Near-infrared spectroscopy in food science and technology: John Wiley & Sons; (2006).
 7. L. Xu, Y. P. Zhou, L. J. Tang, H. L. Wu, J. H. Jiang, G. L. Shen, R. Q. Yu. Ensemble preprocessing of near-infrared (NIR) spectra for multivariate calibration. *Anal Chim Acta*. 616(2008), 138-143.
 8. B. Jamshidi, E. Mohajerani, J. Jamshidi, S. Minaei, A. Sharifi. Non-destructive detection of pesticide residues in cucumber using visible/near-infrared spectroscopy. *Food Addit. Contam. Part A*. 32(2015), 857-863.
 9. B. Jamshidi, S. Minaei, E. Mohajerani, H. Ghassemian, Prediction of soluble solids in oranges using visible/near-infrared spectroscopy: Effect of peel. *Int. J. Food Prop.* 17(2014), 1460-1468.
۱۰. ز. حقیقی، ا. ح. کریمی، غ. وطن خواه، شناسایی بست دیوارنگاره‌های تاریخی خانه سوکیاس اصفهان با استفاده از طیف بینی زیرقرمز تبدیل فوریه و کمومتریکس. *مجله علمی پژوهشی علوم و فناوری رنگ*. (۱۳۹۴) ۹، ۸۲-۷۵.
11. R. Henn, A. Schwab, C. W. Huck, Evaluation of benchtop versus portable near-infrared spectroscopic method combined with multivariate approaches for the fast and simultaneous quantitative analysis of main sugars in syrup formulations. *Food Control*. 68(2016), 97-104.
 12. S. Heppner, K. Thielecke, K. Buchholz, D. Wullbrandt, Potential applications of NIR spectrometry in the sugar industry. *Zuckerindustrie*. 125(2000), 325-330.
 13. R. Lu, Predicting firmness and sugar content of sweet cherries using near-infrared diffuse reflectance spectroscopy. *Trans. ASAE*. 44(2001), 1265.
 14. R. Mehrotra, H. W. Siesler, Application of mid infrared/near infrared spectroscopy in sugar industry. *Appl. Spectrosc. Rev.* 38(2003), 307-354.
 15. S. Minaei, H. Bagherpour, M. Abdollahian Noghbi, M. Khorasani Fardvani, F. A. Forughimanesh Comparative study concerning linear and nonlinear models to determine sugar content in sugar beet by near infrared spectroscopy (NIR). *J. Food Biosci. Technol.* 6(2016), 13-22.
 16. Y. Roggo, L. Duponchel, B. Noe, J. Huvenne, Sucrose content determination of sugar beets by near infrared reflectance spectroscopy. Comparison of calibration methods and calibration transfer. *J. Near Infrared Spectrosc.* 10 (2002), 137-150.
 17. Y. Roggo, L. Duponchel, J. P. Huvenne. Quality evaluation of sugar beet (*Beta vulgaris*) by near-infrared spectroscopy. *J. Agric. Food Chem.* 52(2004), 1055-1061.
 18. Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, N. Jent. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *J. Pharm. Biomed. Anal.* 44(2007), 683-700.
 19. A. Moghimi, M. H. Aghkhani, A. Sazgarnia, M. Sarmad. Vis/NIR spectroscopy and chemometrics for the prediction of soluble solids content and acidity (pH) of kiwifruit. *Biosyst. Eng.* 106(2010), 295-302.
 20. Wojtczak M. ICUMSA-International commission for uniform methods of sugar analysis. *Gazeta Cukrownicza*. 6(2003), 191-192.
 21. A. Rinnan, F. Van den Berg, S. B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends Anal. Chem.* 28(2009), 1201-1222
۲۲. ب. جمشیدی، س. مینایی، ع. مهاجرانی، ح. قاسمیان. بررسی اثر پیش‌پردازش‌های مختلف طیفی بر ارزیابی غیر مخرب کیفیت پرتقال با اسپکتروفوتومتری فروسرخ نزدیک (NIRS). *مجله تحقیقات مهندسی کشاورزی*. (۲۰۱۴) ۱۵.
23. B. M. Nicolai, K. Beullens, E. Bobelyn, A. Peirs, W. Saeys, K. I. Theron, J. Lammertyn. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology. Technol.* 46(2007), 99-118.