

تأثیر روش انتخاب نمونه‌ها و تعداد داده بر دقت مدل واسنجی پیش‌بینی خواص خاک

حسنا محمدی منور^{*۱}

تاریخ دریافت: ۱۳۹۵/۰۴/۲۴

تاریخ پذیرش: ۱۳۹۵/۰۹/۲۸

چکیده

کشاورزی دقیق، نوعی فناوری است که در آن تغییرات موجود در مزرعه از جمله ویژگی‌های خاک، اندازه‌گیری و مدیریت می‌شود. در این پژوهش از روش سریع و دقیق طیف‌سنجی مرئی- مادون قرمز نزدیک (VIS-NIR) در محدوده طیف ۳۷۰ تا ۲۰۰۰ نانومتر برای تعیین خصوصیات فیزیکی و شیمیایی خاک مزرعه استفاده شد. خصوصیات خاک شامل pH، نیتروژن کل، منیزیم، پتاسیم، فسفر، آهن و کربن آلی اندازه‌گیری شدند. هدف از این مطالعه بررسی روش انتخاب تعداد نمونه‌ها در دقت مدل واسنجی است. از سه روش تصادفی، شیمیایی و روش آنالیز مؤلفه اصلی (PCA) برای انتخاب تعداد داده‌ها و مدل‌سازی استفاده شد و روش‌ها با هم مقایسه شدند. نتایج رگرسیون حداقل مربعات جزئی (PLSR) نشان داد با تعداد ۱۰۵ نمونه، بهترین روش دسته‌بندی در اکثر فاکتورهای اندازه‌گیری شده خاک، روش PCA می‌باشد. مطابق این روش نتایج پیش‌بینی مقدار نیتروژن بهترین جواب را می‌دهد ($R^2=0.77$ و $RMSE=0.21$). اگر دو سوم و کل نمونه‌ها نیز در مجموعه واسنجی استفاده شود نتایج چندان بهبود نمی‌یابد، چنانچه برای آهن آزاد خواهیم داشت: $R^2_{210}=0.72$ ، $R^2_{140}=0.77$ و $R^2_{105}=0.78$. در مورد سایر فاکتورهای خاک هم نتایج قابل استنادی به‌دست آمد. دقت پیش‌بینی منیزیم ($R^2_{140}=0.55$ ، $R^2_{105}=0.57$) و pH خاک ($R^2_{140}=0.75$ ، $R^2_{105}=0.73$) به ترتیب بیش باری و بهبود مدل را نشان داد. در حالی که در مدل‌سازی با کل داده‌ها میزان خطا برای پیش‌بینی پتاسیم و فسفر خاک به ترتیب 3.37 mg g^{-1} و 5.28 mg g^{-1} محاسبه شد. نتایج نشان داد روش غیرمخرب VIS-NIR برای پیش‌بینی فاکتورهای خاک بسیار مناسب بود و همچنین قادر به اختصاص تعداد بهینه داده‌ها در مجموعه واسنجی می‌باشد.

واژه‌های کلیدی: تعداد نمونه، خصوصیات خاک، طیف‌سنجی مرئی- مادون قرمز نزدیک، مدل واسنجی

مقدمه

فناوری غیرمخرب و سریع معمولاً همبستگی بارزی بین طیف بازتابی و خصوصیات فیزیکی و شیمیایی خاک وجود دارد (Chang and Laird, 2002). همچنین افزودن ذرات نانو رس به خاک موجب کاهش زیادی در میانگین مقدار چسبندگی و زاویه اصطکاک داخلی و به عبارت دیگر پارامترهای برشی خاک می‌شود (Sharif nasab and Abasi, 2015). از طرف دیگر پیش‌بینی کمی فاکتورهایی مانند نیتروژن و کربن خاک و در درجه دوم ظرفیت تبادل کاتیونی و میزان رس خاک در کشاورزی دقیق بسیار حائز اهمیت است (Viscarra rossel and McBratney, 2008). گذار از روش اندازه‌گیری سنتی به روش‌های غیرمخرب مانند VIS-NIR در کشاورزی دقیق نیاز به تغییر در دیدگاه مجریان دارد و به چالش فنی نیز نیازمند است. تاکنون این روش در کشاورزی دقیق برای بررسی وضعیت بذر، کود و ویژگی‌های خاک به‌کار رفته است (Mohan et al., 2005). طیف‌سنجی VIS-NIR به‌عنوان روشی تحلیلی و برتر برای آنالیز غلات، مواد غذایی انسان و دام شناخته شده است. با این وجود ماهیت پیچیده ماتریس خاک هنوز برای گروه‌های تحقیقاتی چالش برانگیز

مدیریت مزرعه بخشی از کشاورزی دقیق است که می‌تواند تأثیر محیطی و اقتصادی مثبتی بر روی کیفیت تولید گیاهان در مزرعه داشته باشد. بخش مدیریت نیاز گیاهان تنها بخشی از کشاورزی دقیق می‌باشد. شوری خاک یکی از معضلات کشاورزی دقیق است که یکی از راه‌های شناسایی آن، اندازه‌گیری در حین حرکت و تهیه نقشه هدایت الکتریکی مزرعه می‌باشد (Baradaran Moti et al., 2011). طیف‌سنجی مرئی-مادون قرمز نزدیک^۲ (VIS-NIR) نوعی فناوری است که به‌وسیله آن می‌توان به بخش وسیعی از اطلاعات خاک دست یافت و در نتیجه در مدیریت مزرعه مفید واقع شد. در این

۱- استادیار، گروه مهندسی مکانیک بیوسیستم، دانشکده کشاورزی، دانشگاه بوعلی سینا همدان

(*- نویسنده مسئول: (Email: hosna.mohamadi@basu.ac.ir

DOI: 10.22067/jam.v7i2.57501

2- Visible-Near Infrared Spectroscopy

درصدی از بازه یا چگالی طیف مربوط به فاکتورهای شیمیایی (به‌طور خلاصه گروه شیمیایی نامیده می‌شود) و انتخاب بر اساس تحلیل مؤلفه اصلی (PCA)^۴ می‌باشد و همچنین تمام نمونه‌ها نیز در مدل واسنجی به کار رفت تا دقت مدل‌سازی با تمامی داده‌ها بررسی شود. روش PCA رگرسیون بر اساس متغیرهای پنهان است که با کاهش ابعاد داده تحلیل طیف را ساده‌تر می‌سازد. در روند ساده‌سازی، تغییرات در داده مرجع یا اولیه کاهش نمی‌یابد بلکه از میزان نوفه تصادفی کاسته می‌شود. از آنجایی که افزایش تعداد نمونه‌ها معمولاً وقت‌گیر و هزینه‌بر می‌باشد، در این پژوهش بهترین روش انتخاب نمونه برای ایجاد مدل دقیق واسنجی از بین سه روش مورد مطالعه، ارزیابی شد. از طرف دیگر با افزایش گام‌به‌گام تعداد نمونه‌ها و در نهایت ساخت مدل با کل نمونه‌ها، اثر تعداد نمونه بر دقت مدل واسنجی به‌دست آمد.

مواد و روش‌ها

انتخاب نمونه و آماده‌سازی

تعداد ۲۱۰ نمونه خاک از مزرعه شخم خورده واقع در بخش آوزمان شهرستان ملایر استان همدان برداشته شد. مزرعه مورد نظر دارای یک هکتار مساحت با طول جغرافیایی ۱۹ ۳۴ شمالی و عرض جغرافیایی ۵۱ ۴۸ شرقی بود و تقریباً از هر ۴۶ متر مربع آن یک نمونه به‌دست آمد. نمونه‌های حواشی مزرعه به‌عنوان داده پرت در نظر گرفته شد. تناوب مزرعه مورد مطالعه گندم یا جو پاییزه و زمستانه، سیب‌زمینی بهاره و تابستانه بود. نمونه‌گیری پس از آماده‌سازی زمین و قبل از کاشت سیب‌زمینی در فصل بهار انجام شد. با توجه به عمق ریشه گیاه و مخلوط نمودن کود با خاک سطحی، نمونه‌ها از عمق ۲۰ سانتی‌متری سطح زمین برداشت شد و پس از عبور از الک ۲ میلی‌متری، در هوای آزاد خشک گردید.

آنالیز شیمیایی

آنالیز شیمیایی نمونه‌های خاک در آزمایشگاه دانشکده کشاورزی دانشگاه بوعلی سینا انجام شد. میزان خاصیت اسیدی خاک در محلول 1M کلرید پتاسیم، کربن آلی به روش بهینه شده واکلی - بلک (McCleod, 1973) تعیین شد. آهن آزاد طبق روش سیترا-بی کربنات-دی تیونات^۵ (CBD) اندازه‌گیری شد (Mehra and Jackson, 1960). سایر فاکتورها شامل پتاسیم، فسفر و منیزیم در هر ۱۰۰ گرم خاک توسط طیف‌نگار انتشار پلاسما- اتمی تعیین شدند (Soltanpour et al., 1996). نیتروژن کل از روش احتراق با

است (Stenberg et al., 2010). به‌علاوه، عملیات خاک‌ورزی نه تنها بر بافت خاک بلکه بر عملکرد محصول مؤثر است. افزایش عملکرد گندم در اثر استفاده از زیرشکن مطالعه شده است (Afzali and Javaheri, 2013). از طرف دیگر با افزایش شدت خاک‌ورزی، پایداری خاکدانه‌ها کاهش یافته است (Ahmadi moghadam et al., 2014). ویژگی‌هایی از خاک که تا کنون با روش VIS-NIR پیش‌بینی شده‌اند شامل رطوبت، کربن آلی، هدایت الکتریکی (EC)، ظرفیت تبادل کاتیونی (CEC)، واکنش خاک (pH) و برخی دیگر از عناصر غذایی ماکرو و میکرو می‌باشد (Viscarrá rossel et al., 2006). جذب طیفی در محدوده NIR توسط مولکول‌هایی با پیوندهای قوی بین اتم‌های سبک اتفاق می‌افتد، به‌ویژه مولکول‌هایی که شامل پیوندهای C-H، N-H و یا O-H باشند. این خاصیت محدوده NIR را برای اندازه‌گیری کربن، نیتروژن و محتوای رطوبتی و سایر فاکتورهای وابسته به این عناصر مناسب می‌نماید. اندازه، قابل حمل بودن و آنالیز نمونه‌های مرطوب و استفاده از فیبر نوری دستگاه طیف‌سنجی VIS-NIR را مناسب استفاده در مزرعه می‌سازد. در طیف استخراج شده بیشتر تغییرات ناخواسته یا نوفه^۱ به علت پراکنش نور یا تغییرات طول موج‌ها مشاهده می‌شود (Wold et al., 1998). روش‌های مختلف پیش پردازش وجود دارند که می‌توان نوفه را حذف نمود. برخی از نوفه‌ها سبب انتقال و جابه‌جایی خط پایه طول موج به یک سمت یا بالا و پایین می‌شوند که با روش تصحیح خط پایه (BLC) اصلاح می‌شوند. برخی دیگر از روش‌های پرکاربرد در کشاورزی شامل متغیر نرمال استاندارد (SNV) تصحیح پراکنندگی افزاینده (MSC) و مشتق‌گیری^۲ از طیف می‌باشند. در روش SNV طیف با روش‌های مختلف مانند محاسبه مساحت زیر منحنی نرمال می‌شود و همه داده‌های طیف را به یک مقیاس واحد می‌برد. روش MSC بسیار شبیه روش SNV بوده و تداخل چندگانه و پراکنش طیفی را می‌کاهد. مشتق اول با فیلتر ساویتزکی - گولای^۳ شیب منحنی طیفی در هر نقطه را اندازه می‌گیرد و به‌طور مؤثر مانع از انحراف خط پایه طیف می‌شود. در فیلترها با مشتقات مرتبه بالاتر از یک چندجمله‌ای مرتبه بالاتر برای ممانعت از انحراف خط پایه استفاده می‌کند. زمانی که نمونه‌ها به‌صورت حجمی از مکانی گرفته می‌شوند انتخاب نمونه‌ای که نماینده شرایط منطقه باشد دشوار است. در این حالت مسأله مهم این است که چه تعداد از نمونه لازم است تا این تغییرات را به‌طور کامل پوشش دهد. لذا مدل واسنجی برای تخمین بهترین مدل پیش‌بینی فاکتورهای خاک، اهمیت فراوان دارد (Wetterlind and Stenberg, 2010). تأکید این مقاله در مقایسه روش انتخاب تصادفی تعداد نمونه‌های واسنجی، در نظر گرفتن

- 1- Noise
- 2- Derivatives
- 3- Savitzky-Golay

4- Principal Component Analysis
5- Citrate - bicarbonate - dithionite

جدول ۱ آمده است. قبل از شروع طیف‌سنجی، دستگاه با صفحه سفید و مشکی کالیبره شد. نمونه‌های خاک با ۲۰ اسکن طیف‌سنجی شدند و میانگین اسکن‌ها در نرم افزار x10 Unscrambler (Camo, Oslo, Norway) محاسبه شد. طبق رابطه بیر لمبرت طیف جذبی با لگاریتم‌گیری از معکوس طیف بازتابی به دست آمد (رابطه ۱). به منظور دسته‌بندی نمونه‌ها، آنالیز مؤلفه اصلی (PCA) بر روی داده‌های شیمیایی و طیفی پیاده‌سازی شد.

$$\text{Absorbance} = \log(1/\text{reflectance}) \quad (1)$$

آنالیز نیتروژن اندازه‌گیری شد (Matejovic, 1997). اندازه‌گیری تمامی فاکتورها با سه تکرار انجام شد.

آنالیز طیفی

پس از حذف داده‌های پرت هشت قطعه، تعداد ۲۱۰ سیگنال، مجموعه طیفی این پژوهش را تشکیل داد. دو دستگاه طیف‌سنج مدل UV-VIS 2048- AvaSpec-ULS و FT-NIR100N برای اندازه‌گیری باندهای طیفی به کار رفت. مشخصات طیف‌سنج‌ها در

جدول ۱- مشخصات فنی طیف‌سنج‌ها

Table 1- Technical specifications of spectrometers

مشخصات فنی Technical specifications	نوع طیف‌سنج AvaSpec-ULS 2048- UV-VIS	نوع طیف‌سنج FT-NIR100N
محدوده طیفی (nm) Spectral range	200-1100	700-2500
وضوح‌پذیری (nm) Resolution	0.06-20	2×107
منبع نور Light source	هالوژن تنگستن halogen tungsten	هلیوم-نئون، هالوژن تنگستن Helium-neon, halogen tungsten
سرعت انتقال داده (اسکن در ثانیه) Data transfer rate (scan s ⁻¹)	1.1	2
جرم (گرم) Mass (g)	1135	3400
آشکارساز Detector	آرایه خطی CCD CCD Linear Array	آرایه CCD، (۴۰۰-۱۱۰۰ از جنس سیلیکون و ۱۱۰۰-۲۵۰۰ از جنس سولفید) CCD Array-(400-1100 Silicon and 1100-2500 Sulfide)

داده‌هایی برای مدل‌سازی واسنجی در نظر گرفته شد و هفت گروه مذکور در این دسته‌بندی نیز اعمال شد. روش سوم استفاده از خواص طیفی بر اساس PCA در کل محدوده اندازه‌گیری ۳۷۰ تا ۲۰۰۰ نانومتر بود. بهترین تعداد مؤلفه اصلی در نرم‌افزار Unscrambler بر اساس حداقل ۹۵ درصد داده‌ها دو عدد تعیین شد. به منظور همگن شدن روش‌ها، هفت گروه هم در این مجموعه در نظر گرفته شد. البته تقسیم‌بندی در این روش، ترتیبی و به فاصله منظم بین گروه‌ها نبود، بلکه بر اساس میزان مؤلفه دوم در PCA می‌باشد. در انتها مدل‌سازی بر اساس دو سوم (۲/۳) و کل داده‌ها نیز صورت گرفت. روش رگرسیونی PLS همبستگی مناسبی بین داده‌های طیفی و شیمیایی ایجاد نمود. ۱۸ مؤلفه (یا متغیر پنهان^۴) در مجموعه واسنجی در نظر گرفته شد. تعداد مؤلفه‌ها برای اعتبارسنجی به روش متقاطع یگانه^۵ (LOOCV) (روش متقابل با یک نمونه حذف شده) بر روی داده‌های واسنجی اعمال شد.

نحوه انتخاب تعداد نمونه‌ها در مدل واسنجی

پیش از استفاده از روش رگرسیون حداقل مربعات جزئی^۱ (PLSR) برای پیش‌بینی خواص خاک (P, K, Mg, OC, pH) جهت کاهش نسبت سیگنال به نوفه محدوده طیفی به ۳۷۰-۲۰۰۰ نانومتر کاهش یافت. سپس از میان روش‌های پیش‌پردازش موجود، روش‌های تصحیح پراکنده چندگانه^۲ (MSC) و تصحیح خط پایه^۳ (BC) توسط نرم‌افزار Unscrambler بر روی سیگنال‌های خام اعمال شد. در ابتدا نمونه‌ها به دو دسته ۱۰۵ تایی تقسیم شدند که یک گروه برای واسنجی و گروه دوم برای اعتبارسنجی به کار رفت. در روش بعدی ۱۵ نمونه به‌طور تصادفی برگزیده شد و در هر بار آزمون دقت مدل، ۱۵ نمونه تصادفی دیگر به نمونه‌های قبلی افزوده می‌شد. به این ترتیب هفت گروه (۱۵، ۳۰، ۴۰، ۱۰۵) در دسته‌بندی تصادفی قرار گرفتند. در روشی دیگر، نتایج آزمون شیمیایی نیز به‌عنوان

- 1- Partial least square regression
- 2- Multivariate scatter correction
- 3- Base line correction

4- Latent variable

5- Leave one out of cross validation

جدول ۲- خواص شیمیایی خاک

Table 2- Chemical properties of soil

ویژگی‌های خاک Soil features	داده‌های اعتبارسنجی = ۱۰۵ Validation data=105				داده‌های واسنجی = ۱۰۵ Calibration data=105				کل نمونه‌ها = ۲۱۰ All samples=210			
	SD	Mean	Max	Min	SD	Mean	Max	Min	SD	Mean	Max	Min
	کربن آلی (%) Organic carbon	0.98	2.75	3.98	0.98	0.98	2.73	3.92	0.50	0.98	2.76	4.00
نیتروژن (mg in 100g) Nitrogen	0.54	1.48	2.22	0.14	0.51	1.41	2.05	0.14	0.58	1.48	2.21	0.14
آهن آزاد (%) Free iron	0.77	0.98	2.98	0.06	0.76	0.97	2.98	0.05	0.76	0.97	3.00	0.05
پتاسیم (mg in 100g) Potassium	4.38	11.67	25.62	0.68	4.38	11.66	25.56	0.71	4.33	11.62	25.34	0.68
فسفر (mg in 100g) Phosphor	8.12	15.21	45.59	4.35	7.92	15.24	44.86	4.33	8.04	15.19	46.00	4.33
منیزیم (mg in 100g) Magnesium	5.16	33.00	89.68	2.82	5.20	29.79	89.58	3.00	5.15	33.00	89.57	2.82
pH (۱:۵ آب) pH (1:5 water)	0.58	7.38	9.12	4.62	0.55	7.39	9.09	4.65	0.57	7.38	9.09	4.62

جدول ۳- ماتریس همبستگی ویژگی‌های اندازه‌گیری شده خاک

Table 3- The correlation matrix of soil features

خواص خاک Soil properties	کربن آلی Organic carbon	pH	منیزیم Magnesium	پتاسیم Potassium	فسفر Phosphor	آهن آزاد Free iron	نیتروژن Nitrogen
کربن آلی Organic carbon	1.00						
اسیدیته pH	0.66	1.00					
منیزیم Magnesium	0.58	0.17	1.00				
پتاسیم Potassium	0.71	0.03	0.48	1.00			
فسفر Phosphor	0.12	0.46	0.00	0.08	1.00		
آهن آزاد Free iron	0.48	0.16	0.22	0.02	0.06	1.00	
نیتروژن Nitrogen	0.82	0.57	0.16	0.41	0.37	0.52	1.00

$$RPD = \frac{Sd}{SEP} \quad (3)$$

$$Sd = \sqrt{\frac{\sum_{i=1}^n (\bar{y} - y_i)^2}{n - 1}} \quad (4)$$

که در آن Y_i میزان اندازه‌گیری شده متغیر، میانگین اندازه‌گیری، n تعداد نمونه‌ها، Sd انحراف معیار^۲ و SEP خطای استاندارد مجموعه پیش‌بینی می‌باشند.

تعداد مؤلفه‌ها برای اعتبارسنجی مدل پیش‌بینی بر اساس کمترین مقدار برای $RMSE$ در اعتبارسنجی روش متقابل با یک نمونه حذف شده انتخاب شد. کیفیت مدل براساس ضریب تبیین (R^2) داده پیش‌بینی در مقابل داده‌های اندازه‌گیری و مقدار $RSMEC$ و شاخص درصد انحراف نسبی^۱ (RPD) طبق روابط (۲) تا (۴) ارزیابی شد.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\bar{y} - y_i)^2}{n}} \quad (2)$$

2- Standard deviation

1- Residual predictive deviation

رگرسیون حداقل مربعات جزئی

جدول ۴ نتایج رگرسیون برخی فاکتورهای خاک را ارائه می‌دهد از آنجا که مطابق جدول ۳ همبستگی مناسبی بین کربن آلی و سایر فاکتورهای اندازه‌گیری در خاک مشاهده شد، تغییرات کربن آلی به‌عنوان شاخص بررسی شد. در دسته‌بندی تصادفی مقدار RMSE با افزایش نمونه‌های موجود از ۱۵ به ۱۰۵ و ۱۴۰ در مجموعه واسنجی از ۰/۲۰ به ۰/۱۳ کاهش یافته است اما روند آن خطی نیست. مقدار R^2 و RPD یک رفتار غیرخطی افزایشی با افزایش تعداد نمونه‌های واسنجی را نشان می‌دهند (به ترتیب از ۰/۴۹ تا ۰/۷۱ و از ۱/۱ تا ۱/۹). میزان R^2 با افزودن ۳۵ نمونه تغییر نمی‌کند، بدین معنی که دقت مدل ایجاد شده با یک دوم (۱/۲) و دو سوم (۲/۳) داده‌ها یکسان است. در مورد نیتروژن و اسیدیته، کاهش در مقدار R^2 نیز مشاهده شد که ناشی از بیش باری مدل می‌باشد. لذا می‌توان نتیجه گرفت که برای مدل‌سازی دقیق نیتروژن و اسیدیته، تعداد مناسب مقداری بین ۱۰۵ تا کمتر از ۱۴۰ نمونه می‌باشد. با افزایش تعداد نمونه‌های به‌کار رفته در مدل واسنجی در دسته آزمون شیمیایی، مقدار RMSE تقریباً به‌صورت خطی روندی نزولی از ۰/۱۷ به ۰/۱۱ را نشان داد. همزمان میزان R^2 از ۰/۴۶ به ۰/۷۲ و RPD از ۱/۳ به ۲/۰ افزایش یافتند.

نتایج و بحث

مدل رگرسیونی همه فاکتورهای خاک بر روی طیف‌های جذبی پیش‌پردازش شده به‌دست آمد. جدول ۲ مقادیر اندازه‌گیری شده نمونه‌های ۱۰۵ تا ۲۱۰ تایی خاک را در مدل واسنجی و اعتبارسنجی نشان می‌دهد. میزان کربن آلی موجود در خاک این منطقه بسیار پایین می‌باشد (4.00-0.51). مقدار متوسط و انحراف معیار (Sd) برای هر سه مجموعه تقریباً در محدوده یکسانی قرار دارد. در مورد سایر خواص خاک تغییرات زیادی مشاهده نشد. تنها میزان منیزیم و فسفر مقدار جزئی در مجموعه واسنجی و میزان پتاسیم نیز مقدار کمی را در مجموعه کل گزارش می‌دهند. همبستگی ($p < 0.05$) بین خواص خاک در جدول ۳ نشان داده شده است. بیشترین همبستگی در بین نیتروژن با کربن آلی دیده می‌شود و از آنجا که مواد آلی خاک شاخص نیتروژن می‌باشند چنین نتیجه‌ای قابل تأیید است. عنصر بعدی پتاسیم است که ۰/۷۱ با OC همبستگی دارد. منیزیم با همبستگی ۰/۵۸ در جایگاه بعدی قرار دارد.

جدول ۴- نتایج اعتبارسنجی متقاطع یگانه در مدل رگرسیونی برخی فاکتورهای خاک

Table 4- The results of LOOCV in the regression model of some soil parameters

روش انتخاب تعداد داده Selecting method of data number	تعداد نمونه Number of sample	اسیدیته Acidity			کربن آلی Organic Carbon			نیتروژن کل Total Nitrogen		
		RPD	RMSE	R^2	RPD	RMSE	R^2	RPD	RMSE	R^2
تصادفی Random	15	1.0	0.32	0.56	1.1	0.20	0.49	1.5	0.44	0.61
	30	1.1	0.30	0.58	1.2	0.13	0.69	1.6	0.39	0.64
	45	1.2	0.29	0.61	1.5	0.14	0.55	1.6	0.41	0.66
	60	1.2	0.31	0.65	1.7	0.15	0.56	1.7	0.33	0.68
	75	1.0	0.28	0.68	1.8	0.12	0.66	1.8	0.31	0.69
	90	1.2	0.26	0.71	1.6	0.14	0.63	1.9	0.21	0.72
	105	1.4	0.26	0.73	1.9	0.13	0.71	2.0	0.21	0.75
140	1.4	0.29	0.70	1.7	0.16	0.71	1.5	0.42	0.73	
شیمیایی Chemical	15	1.1	0.33	0.55	1.3	0.17	0.48	1.4	0.41	0.63
	30	1.1	0.29	0.56	1.5	0.16	0.46	1.4	0.38	0.64
	45	1.2	0.31	0.60	1.6	0.15	0.55	1.6	0.39	0.66
	60	1.3	0.26	0.63	1.7	0.15	0.58	1.7	0.32	0.69
	75	1.2	0.28	0.69	1.9	0.13	0.64	1.9	0.32	0.70
	90	1.3	0.27	0.70	1.9	0.12	0.70	1.9	0.26	0.73
	105	1.4	0.26	0.72	2.0	0.11	0.72	2.0	0.22	0.76
140	1.3	0.30	0.69	1.6	0.16	0.72	1.7	0.33	0.74	
آنالیز مؤلفه اصلی PCA	15	1.0	0.31	0.52	1.2	0.19	0.55	1.4	0.42	0.63
	30	1.2	0.30	0.56	1.4	0.17	0.57	1.5	0.38	0.65
	42	1.1	0.29	0.59	1.5	0.15	0.61	1.6	0.33	0.66
	59	1.0	0.30	0.61	1.4	0.15	0.59	1.7	0.35	0.70
	77	1.2	0.27	0.66	1.6	0.13	0.65	1.8	0.28	0.71
	88	1.3	0.27	0.71	1.8	0.14	0.68	1.9	0.23	0.74
	105	1.4	0.26	0.73	1.9	0.13	0.71	2.1	0.21	0.77
140	1.0	0.33	0.75	1.2	0.19	0.75	1.5	0.40	0.77	

جدول ۵- نتایج ارزیابی مدل‌های واسنجی و اعتبارسنجی بر اساس PCA برخی ویژگی‌های خاک
Table 5- The calibration and validation model of some soil properties using PCA

ویژگی‌های خاک Soil features	داده‌های اعتبارسنجی=۱۰۵ Validation data=105		داده‌های واسنجی=۱۰۵ Calibration data=105		داده‌های واسنجی=۱۴۰ Calibration data=140		کل نمونه‌ها=۲۱۰ All samples=210	
	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²
منیزیم Magnesium	1.69	0.52	1.58	0.57	1.63	0.55	1.67	0.55
پتاسیم Potassium	2.69	0.52	3.41	0.48	3.38	0.48	3.37	0.48
فسفر Phosphor	6.47	0.37	6.19	0.42	6.08	0.45	5.28	0.49
آهن آزاد Free iron	0.21	0.83	0.56	0.78	0.77	0.77	0.89	0.72

نتایج و بحث

تأثیر روش انتخاب نمونه در مدل واسنجی

روش انتخاب نمونه‌ها در دقت مدل واسنجی بسیار اهمیت دارد. این انتخاب معمولاً بر اساس داده‌های شیمیایی موجود (Dunn *et al.*, 2002; Aichi *et al.*, 2009) یا به صورت تصادفی (Brown, 2005; He *et al.*, 2005) انجام می‌شود. تأثیر سه مجموعه نمونه در ساخت مدل بررسی شد و هنگامی که از بیش از ۵۷ نمونه در مدل‌سازی به کار رفت، در ریشه میانگین مربع انحراف (RMSD) کاهش مشاهده شد (Brown, 2005). این یافته‌ها نتایج این پژوهش را تأیید می‌کند، به‌ویژه در مورد دسته‌بندی تصادفی روند مشابهی مشاهده شد. در تحقیق دیگری هدف از انتخاب داده‌های بیشتر در مدل‌سازی پوشش‌دهی وسیع‌تر بیان شد تا میزان خطا کاهش یابد (Kuang and Mouazen, 2011). در این حالت استفاده از روش‌های شیمیایی بهترین نتایج را داشت که البته با نتایج پژوهش حاضر در تضاد می‌باشد. نتایج دسته‌های مختلف در این پژوهش تفاوت چندانی باهم ندارند که می‌توان علت را در تغییرات ناچیز میزان OC خاک مورد مطالعه مربوط دانست. همچنین از روش دسته‌بندی^۱ استفاده گردید (Moros *et al.*, 2009). به این صورت که ۳۹ نمونه به صورت تریبی از میان ۱۴۸ نمونه خاک انتخاب و میزان pH خاک با دقت مشابه تحقیق حاضر پیش‌بینی شد. در پژوهشی دیگر در روش طیف‌سنجی در دو محدوده NIR و MIR^۲، داده‌ها به دو دسته واسنجی و اعتبارسنجی تقسیم شد (Reeves and Smith, 2009). در روش اول از تقسیم مساوی (۵۰/۵۰) داده‌های طیفی استفاده شد و نتایج ضعیفی به دست آمد. برای بررسی این وضعیت از تقسیم مساوی داده‌های شیمیایی استفاده کردند ولی نتایج مشابهی به دست آمد که با این نتایج برای گروه مشابه همخوانی دارد. در مطالعه دیگری از دو

در دسته آزمون شیمیایی با افزایش داده به ۱۴۰ در هیچ‌کدام از فاکتورهای اسیدیته، کربن آلی و نیتروژن کل افزایشی مشاهده نشد و در دو مورد (اسیدیته و نیتروژن کل) اثرات بیش باری نیز دیده شد. در نهایت در دسته‌بندی بر اساس PCA در میزان RMSE کاهش تقریباً خطی (۰/۱۹-۰/۱۳) مشاهده می‌شود. حداقل میزان خطا در روش PCA در تعداد ۷۷ نمونه به کار رفته برای ساخت مدل واسنجی (۰/۱۳ درصد) گزارش شده است. با افزایش تعداد ۱۴۰ نمونه برای ساخت مدل واسنجی مقدار R² و RMSE افزایش نشان داد. اسیدیته نیز مانند کربن آلی، با افزایش نمونه به ۱۴۰ عدد، افزایش (۰/۷۵) R² را نشان داد در حالی که در نیتروژن بدون افزایشی در R² مقدار RMSE افزایش یافت. تغییرات RPD و میزان R² روندی غیرخطی افزایش یافته دارند. سایر فاکتورهای خاک نیز در هر دسته‌بندی روندی افزایش یا کاهش و خطی یا غیرخطی را نشان می‌دهند. نتایج مدل رگرسیونی سایر ترکیبات خاک وقتی دو سوم (۲/۳) و تمام داده‌ها برای ساخت مدل واسنجی بر اساس PCA به کار رفته در جدول ۵ آورده شده است. در مقایسه مدل‌های واسنجی با تعداد ۱۰۵ و ۱۴۰ نمونه، با افزایش تعداد نمونه به ۲/۳ برابر، در فسفر کاهش خطا (RMSE₁₀₅=6.19, RMSE₁₄₀=6.08) و افزایش (R²₁₀₅=0.42, R²₁₄₀=0.45) مشاهده شد. در حالی که با افزایش نمونه‌ها به ۱۴۰ عدد، آهن آزاد و منیزیم با کاهش R² مواجه شدند. در مدل‌سازی با کل نمونه‌ها، مدل واسنجی پتاسیم با کمترین مقدار R² (۰/۴۸) و فسفر با بیشترین میزان خطا (RMSE=5.28) ضعیف‌ترین مدل‌ها می‌باشند و بعد از آن منیزیم با R²=0.55 و RMSE=1.67 در رتبه سوم قرار دارد. سایر خواص خاک دارای ضریب تبیین بالاتر از ۰/۵۰ (R²>0/5) می‌باشند. لذا دقت قابل قبولی در مدل پیش‌بینی آن‌ها وجود دارد.

1- Clustering

2- Mid-Infrared

پیشین است (Ladoni et al., 2010; Vasques et al., 2008). از طرف دیگر تعداد بیشتری از پژوهش‌ها میزان R^2 بالاتری برای OC گزارش نموده‌اند (Chang and Laird, 2002; Shepherd and Walsh, 2002). بررسی OC خاک در محدوده ۴۰۰-۹۵۰ نانومتر نتیجه ضعیفی نشان می‌دهد (Aïchi et al., 2009). اما در مورد سایر فاکتورهای مهم برای گیاه (منیزیم، پتاسیم، نیتروژن و فسفر) تحقیقات نشان می‌دهد پیش‌بینی مدلی برای آن‌ها بسیار مشکل می‌باشد که علت آن را می‌توان در وابستگی این عناصر به سایر ترکیبات خاک جستجو نمود. دلیل مشابهی برای مدل ضعیف سدیم (Na) ارائه شد (Dunn et al., 2002). درحالی‌که در همین تحقیق دقت پیش‌بینی منیزیم قابل قبول است اما مطالعات پیش‌بینی میزان فسفر مورد نیاز گیاه را با مدل ضعیفی تأیید نمودند (Shepherd and Walsh, 2002). علت آن را می‌توان به تشخیص مشکل فسفر غیرآلی موجود در خاک‌های معدنی مرتبط دانست (Chodak, 2008). همچنین افزایش تعداد داده‌های مدل واسنجی به دو سوم یا کل نمونه‌ها، دقت مدل را برای تمامی فاکتورها بهبود نداد. در حالی‌که در مطالعات پیشین با در نظر گرفتن کل داده‌ها در مجموعه واسنجی مدل دقیق‌تری ارائه شد (Wetterlind et al., 2010). از طرف دیگر وقتی ۲۵ نمونه خاک از یک یا دو مزرعه متفاوت برای مدل‌سازی در نظر گرفتند، نتایج ضعیفی به دست آمد. در این پژوهش علت نتایج قابل قبول را می‌توان به بافت خاک مربوط دانست. در میان فاکتورهای خاک اسیدیته و وابستگی چندان با سایر ویژگی‌های خاک مورد بررسی نداشت. در جدول ۳ این نتایج در مورد pH بر اساس نتایج مطالعات قبلی قابل تأیید می‌باشد (Moros, 2009). اما در این پژوهش میزان خطای بیشتری گزارش شد که علت این امر این است که pH محدوده مشخصی در طیف NIR ندارد و معمولاً از روی سایر خواص خاک قابل تشخیص نسبی می‌باشد (Kuang and Mouazen, 2011).

نتیجه‌گیری

نتایج این مطالعه تأییدی بر کاربرد مفید فناوری طیف‌سنجی VIS-NIR به عنوان روشی سریع و دقیق برای پیش‌بینی فاکتورهای مختلف خاک می‌باشد. استفاده از روش‌های مختلف برای انتخاب مجموعه واسنجی نتایج مشابهی در پی داشت اما در این میان تکنیک PCA بهترین جواب را ارائه داد. گرچه پیش‌بینی برخی از فاکتورها مانند منیزیم، پتاسیم، نیتروژن و فسفر با مدل ضعیف گزارش شد. از طرف دیگر همواره افزایش داده‌ها سبب افزایش قابل قبول دقت مدل‌سازی نمی‌شود و می‌توان با تعداد مؤلفه پیشنهادی آنالیز مؤلفه اصلی، به جواب قابل قبولی دست یافت.

روش تقسیم بر پایه داده‌های طیفی و شیمیایی استفاده نمودند (Guerrero et al., 2010) و نتایج آن‌ها نشان داد مدل دقیق‌تری بر پایه داده‌های طیفی به دست می‌آید.

تأثیر تعداد نمونه بر دقت مدل پیش‌بینی

معمولاً در استفاده از داده‌های طیفی در روش VIS-NIR تعداد نمونه‌های واسنجی بیشتر از داده‌های اعتبارسنجی و یا با هم برابر می‌باشد (یک چهارم نمونه‌ها در اعتبارسنجی (Dunn et al., 2002)، یک سوم نمونه‌ها در اعتبارسنجی (Brown et al., 2005; Kuang and Mouazen, 2011) and دو نمونه‌ها در اعتبارسنجی (Reeves & Smith, 2009)). در این روش نمونه‌های کمتری مورد بررسی قرار گرفت زیرا یکی از برتری‌های این روش نیاز به نمونه‌های کمتر برای مدل‌سازی با دقت قابل قبول می‌باشد. نتایج نشان می‌دهد که حداقل ۱۴۰ نمونه واسنجی برای پیش‌بینی کربن آلی و pH و ۱۰۵ نمونه برای مدل‌سازی نیتروژن مورد نیاز است. با توجه به شرایط متفاوت مزارع، مقایسه نتایج این بخش با یافته‌های پیشین بسیار پیچیده است. نتایج مشابهی در دو مزرعه مطالعه شده برای مدل‌سازی فاکتورهای خاک با حداقل ۲۵ نمونه از کل ۱۳۰ نمونه به دست آمد (Wetterlind et al., 2010). تفاوت این نتایج با پژوهش حاضر تعداد نمونه مورد بررسی در مساحت مزرعه می‌باشد. برخی پژوهشگران به این نتیجه رسیدند که تغییرات اندک در RMSE به علت تفاوت در شرایط مزرعه می‌باشد (Wetterlind and Stenberg, 2012; Kuang and Mouazen, 2010). این یافته‌ها با نتایج به دست آمده برای OC در پژوهش حاضر مطابقت دارد. در نهایت کاهش RMSE به عنوان معیاری برای انتخاب تعداد نمونه و آنالیز تنش می‌باشد (Brown et al., 2005). طبق نتایج به دست آمده تقریباً ۷۷ تا ۱۴۰ نمونه واسنجی برای بهترین مدل OC و pH و ۱۰۵ نمونه برای سایر فاکتورها مورد نیاز است. تفاوتی که در این اعداد با یافته‌های محققین دیده می‌شود به علت تفاوت در عناصر غذایی خاک می‌باشد. می‌توان نتیجه‌گیری نمود که روش انتخاب و تعداد نمونه‌های واسنجی تأثیر مهمی بر اعتبار مدل پیش‌بینی خواص خاک در روش طیف سنجی VIS-NIR خواهد گذاشت.

نتایج مدل فاکتورهای خاک

در حالت کلی محتوای OC با مقدار کمی خطا پیش‌بینی شد. این نتایج مشابه گزارشی می‌باشد که از ۸۵ نمونه خاک برای واسنجی استفاده نمود (Brown et al., 2005). مدلی با حداقل RMSE در این پژوهش دارای ۷۷ نمونه خاک بود که بسیار نزدیک به یافته Brown می‌باشد. همچنین RMSE به دست آمده در این پژوهش ۰/۱۲ می‌باشد که بسیار کمتر از موارد گزارش شده در تحقیقات

References

1. Afzali, M. J., and E. Javaheri. 2013. The effect of tillage on soil penetration resistance, technical indices and yield of wheat. *Journal of Agricultural Machinery* 3 (1): 24-31. (In Farsi).
2. Ahmadi Moghadam, P., L. Eftekhari, A. Mardani, and H. Khodaverdilou. 2014. Determination the amount of plant debris, physical and mechanical properties of soil in different tillage systems. *Journal of Agricultural Machinery* 6 (1): 102-113. (In Farsi).
3. Aichi, H., Y. Fouad, C. Walter, R. A. Viscarra Rossel, Z. Lili-Chabaane, and M. Sanaa. 2009. Regional predictions of soil organic carbon spectral reflectance measurements. *Biosystems Engineering* 104 (3): 442-446.
4. Baradaran Motie, J., M. H. Aghkhani. M. H. Abaspour Fard, and A. Lakzian. 2011. Construction and assessment the on the go mapping system of soil electrical conductivity. *Journal of Agricultural Machinery Engineering* 1 (1): 25-33. (In Farsi).
5. Brown, D. J., R. S. Bricklemeyer, and P. R. Miller. 2005. Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C prediction in Montana. *Geoderma* 129 (3-4): 251-267.
6. Chang, C. W., and D. A. Laird. 2002. Near-infrared reflectance spectroscopy analysis of soil C and N. *Soil Science* 167 (2): 110-116.
7. Chodak, M. 2008. Application of near infrared spectroscopy for analysis of soils, litter and plant materials. *Polish Journal of Environmental Studies* 17 (5): 631-642.
8. Dunn, B. W., H. G. Beecher, G. D. Batten, and S. Ciavarella. 2002. The potential of near-infrared reflectance spectroscopy for soil analysis- a case study from the Riverine Plain of south-eastern Australia. *Australian Journal of Experimental Agriculture* 42 (5): 607-614.
9. Guerrero, C., R. Zornoza, I. Gómez, and J. Mataix-Beneyto. 2010. Spiking of NIR regional models using samples from target sites: effect of model size on prediction accuracy. *Geoderma* 158 (1-2): 66-77.
10. He, Y., H. Y. Song, A. G. Pereira, and A. H. Gómez. 2005. Measurements and analysis of soil nitrogen and organic matter content using near-infrared spectroscopy techniques. *Journal of Zhejiang University B*. 6B (11): 1081-1086.
11. Kuang, B., and A. M. Mouazen. 2011. Calibration of visible and near infrared spectroscopy for soil analysis at the field scale on three European farms. *European Journal of Soil Science* 62: 629-636.
12. Kuang, B. and A. M. Mouazen. 2012. Influence of the number of samples on prediction error of visible and near infrared spectroscopy of selected soil properties at the farm scale. *European Journal of Soil Science* 36: 421-429.
13. Ladoni, M., H. A. Bahrami, S. K. Alavipanah, and A. A. Norouzi. 2010. Estimating soil organic carbon from soil reflectance; a review. *Precision Agriculture* 11 (1): 82-99.
14. Matejovic, I. 1997. Determination of carbon and nitrogen in samples of various soils by the dry combustion. *Commun. Soil Sci. Plan.*, 28 (17-18): 1499-1511.
15. McCleod, S. 1973. Studies on wet oxidation procedures for the determination of organic carbon in soil. In 'Notes on soil techniques'. pp. 73-79. (CSIRO Division of Soils: Melbourne).
16. Mehra, O. P., and M. L. Jackson. 1960. Iron oxide removal from soils and clays by a dithionite-citrate system buffered with sodium bicarbonate. *Clay. Clay. Miner* 7: 317-327.
17. Mohan, L. A., C. Karunakaran, D. S. Jayas, and N. D. G. White. 2005. Classification of bulk cereals using visible and NIR reflectance characteristics. *Biosystems Engineering* 7 (7): 7-14.
18. Moros, J., M. J. Martínez-Sánchez, C. Pérez-Sirvent, S. Garrigues, and M. de la Guardia. 2009. Testing of the Region of Murcia soils by near infrared diffuse reflectance spectroscopy and chemometrics. *Talanta* 78: 388-398.
19. Reeves, J. B., and D. B. Smith. 2009. The potential of mid- and near-infrared diffuse reflectance spectroscopy for determining major- and trace-element concentrations in soils from a geochemical survey of North America. *Applied Geochemistry* 24 (8): 1472-1481.
20. Sharif nasab, H., N. Abasi. 2015. Assessment the impact of nano-clay particles on some physical and mechanical properties of soil. *Journal of Agricultural Machinery* 6 (1): 250-258. (In Farsi).
21. Shepherd, K. D., and M. G. Walsh. 2002. Development of reflectance spectral libraries for

- characterization of soil properties. *Soil Science Society of America* 66 (3): 988-998.
22. Soltanpour, P. N., G. W. Johnson, S. M. Workman, and J. B. Jones. 1996. Inductively Coupled Plasma Emission Spectrometry and Inductively Coupled Plasma-Mass Spectrometry. In: *Methods of Soil Analysis. Part 3 Chemical Methods*. Soil Science Society of America Inc. pp. 91-139.
 23. Stenberg, B. 2010. Effects of soil sample pretreatments and standardized rewetting as interacted with sand classes on Vis-NIR predictions of clay and soil organic carbon. *Geoderma* 158 (1-2): 15-22.
 24. Stenberg, B., R. A. Viscarra Rossel, A. M. Mouazen, and J. Wetterlind. 2010. Visible and Near Infrared Spectroscopy in Soil Science. *Advances in Agronomy* 107: 163-215.
 25. Vasques, G. M., S. Grunwald, and J. O. Sickman. 2008. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near infrared spectra. *Geoderma* 146 (1-2): 14-25.
 26. Viscarra Rossel, R. A., D. J. J. Walvoort, A. B. McBratney, L. J. Janik, and J. O. Skjemstad. 2006. Visible, near-infrared, mid-infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131 (1-2): 59-75.
 27. Viscarra Rossel, R. A., and A. B. McBratney. 2008. Diffuse reflectance spectroscopy as a tool for digital soil mapping. PP 165-172 in: Hartemink, A.E., A.B. McBratney, and L. Mendonça-Santos, (Eds.), *Digital Soil Mapping with Limited Data*. Developments in Soil Science series. Elsevier Science, Amsterdam.
 28. Wetterlind, J., and B. Stenberg. 2010. Near-infrared spectroscopy for within-field soil characterization: small local calibrations compared with national libraries spiked with local samples. *European Journal of Soil Science* 61 (6): 823-843.
 29. Wetterlind, J., B. Stenberg, and M. Söderström. 2010. Increased sample point density in farm soil mapping by local calibration of visible and near infrared prediction models. *Geoderma* 156 (3-4): 152-160.
 30. Wold, S., H. Antti, F. Lindgren, and J. Ohman. 1998. Orthogonal signal correction of near-infrared spectra. *Chemometrics and Intelligent Laboratory Systems* 44: 175-185.

The Effect of Sample Size and Data Numbering on Precision of Calibration Model to predict Soil Properties

H. Mohamadi Monavar^{1*}

Received: 14-07-2016

Accepted: 18-12-2016

Introduction: Precision agriculture (PA) is a technology that measures and manages within-field variability, such as physical and chemical properties of soil. The nondestructive and rapid VIS-NIR technology detected a significant correlation between reflectance spectra and the physical and chemical properties of soil. On the other hand, quantitatively predict of soil factors such as nitrogen, carbon, cation exchange capacity and the amount of clay in precision farming is very important. The emphasis of this paper is comparing different techniques of choosing calibration samples such as randomly selected method, chemical data and also based on PCA. Since increasing the number of samples is usually time-consuming and costly, then in this study, the best sampling way -in available methods- was predicted for calibration models. In addition, the effect of sample size on the accuracy of the calibration and validation models was analyzed.

Materials and Methods: Two hundred and ten soil samples were collected from cultivated farm located in Avarzaman in Hamedan province, Iran. The crop rotation was mostly potato and wheat. Samples were collected from a depth of 20 cm above ground and passed through a 2 mm sieve and air dried at room temperature. Chemical analysis was performed in the soil science laboratory, faculty of agriculture engineering, Bu-ali Sina University, Hamadan, Iran. Two Spectrometer (AvaSpec-ULS 2048- UV-VIS) and (FT-NIR100N) were used to measure the spectral bands which cover the UV-Vis and NIR region (220-2200 nm). Each soil sample was uniformly tiled in a petri dish and was scanned 20 times. Then the pre-processing methods of multivariate scatter correction (MSC) and base line correction (BC) were applied on the raw signals using Unscrambler software. The samples were divided into two groups: one group for calibration 105 and the second group was used for validation. Each time, 15 samples were selected randomly and tested the accuracy of models, then other 15 samples were added randomly to the previous set and it was done continuously. Finally, seven groups (15, 30... 105) were placed in each category.

Results and Discussion: All regression models on the whole pre-processed soil spectra were obtained in absorption mode. By increasing the number of samples in the calibration set of random group, RMSE was decreased from 0.2 to 0.13 nonlinearly. RMSE in the chemical test was also decreased almost linearly from 0.17 to 0.11. At the same time, R^2 and RPD were increased from 0.46 to 0.72 and from 1.3 to 2.0 respectively. Finally, in categories based on PCA, the RMSE fell down almost linearly (0.19-0.13). Potassium prediction model with the least amount of R^2 (0.48) and phosphorus with highest number of errors (RMSE = 5.28) were the weakest models in whole data. Other properties of the soil had a higher coefficient of determination ($R^2 > 0.5$). Therefore, prediction models had acceptable accuracy. At least 77, 105 and 105 samples are required for precise calibration model of OC, nitrogen and pH respectively. Due to the different conditions of farms, comparing these results with previous findings is very complex. Furthermore, model accuracy did not improve by increasing data of calibration models to the total number of samples. While in previous studies, more precise model was calibrated by considering the entire data sets. Among all factors of soil, acidity has little dependence on the other soil properties. The pH modeling is also confirmed by Moros (2009) however, the more error was reported here. There is no certain pH range in the NIR spectra, and usually it is distinguishable from the other properties of the soil (Kuang and Mouazen, 2011).

Conclusions: Spectroscopic methods exhibited good potential for detecting soil properties. MSC and BC can effectively remove irrelevant information to improve prediction accuracy. Using different methods to select numbers of data for the calibration models presented similar results, but in the meantime PCA technique provided the best answer. Supplementary, the ever-increasing number of data does not always improve modeling accuracy. It is better to choose numbers of data according to principal components (PCs) of PCA to obtain acceptable answer. It must be noted that every crops requires a specific soil and nutrients, so it is necessary to develop models for other soil properties.

Keywords: Calibration model, Number of sample, Soil properties, Visible and NIR spectroscopy

1- Assistant Professor, Mechanics of Biosystem Engineering, Department of Agriculture, Bu-ali Sina University
(*- Corresponding Author Email: hosna.mohamadi@basu.ac.ir)