

## تخمین مکانی ماده آلی خاک با داده‌های نامطمئن و کمکی خاکی و روش آنتروپی حداکثر اریب

نیکو حمزه پور<sup>۱\*</sup>، سارا ملاعلی عباسیان<sup>۲</sup>

(تاریخ دریافت: ۱۳۹۷/۰۲/۱۹ تاریخ پذیرش: ۱۳۹۷/۰۷/۲۴)

### چکیده

ماده آلی خاک یکی از مهمترین فاکتورهای کیفیت خاک می‌باشد و آگاهی از وضعیت آن در خاک، از مهمترین اقدامات در جهت مدیریت منابع اراضی است. با توجه به هزینه و زمان زیادی که در جهت پایش ماده آلی خاک نیاز است، هر روشی که بتواند با حداقل نمونه استفاده از هر نوع داده‌های خاکی، بهترین نقشه‌های ماده آلی را تولید کند، قدمی در راستای کشاورزی پایدار خواهد بود. هدف از انجام این تحقیق، تخمین مکانی ماده آلی خاک با استفاده از اندازه‌گیری‌های با دقت پایین ماده آلی، داده‌های کمکی خاکی و روش آنتروپی حداکثر اریب (BME) بود. نمونه‌برداری از ۱۲۲ نقطه و از عمق ۰-۲۰ سانتی‌متری از دشت بناب و میاندوآب صورت گرفت. سپس، ماده آلی خاک به روش والکلی و بلاک و نیز به شکل تسهیل شده آن، تعیین گردید. همچنین، برخی از پارامترهای خاکی از جمله بافت، پایداری خاکدانه‌ها و آهک نیز در نمونه‌ها اندازه‌گیری شدند. تخمین مکانی ماده آلی با داده‌های نامطمئن، مدل خطی توسعه یافته با استفاده از داده‌های کمکی خاکی و روش BME انجام شد. براساس نتایج، بالاترین R، کمترین RMSE و nRMSE با مقادیر به ترتیب ۰/۹۷، ۰/۰۷ درصد و ۰/۰۶ متعلق به تخمین مکانی ماده آلی با داده‌های نامطمئن ماده آلی با در نظر گرفتن خطا بود. همچنین استفاده از مدل خطی توسعه یافته با داده‌های کمکی خاکی و وارد کردن خطای این داده‌ها در معادلات تخمین، منجر به بهبود تخمین نسبت به زمانی گردید که از خطا استفاده نشده بود (RMSE، R و nRMSE به ترتیب از ۰/۶۵، ۰/۵۸ و ۰/۵۵ به ۰/۸۵، ۰/۳۱ و ۰/۲۹ بهبود یافت). با توجه به نتایج حاصل از این تحقیق، روش BME امکان استفاده از طیف وسیعی از اطلاعات خاکی را در تخمین فراهم می‌آورد و با ادغام خطای ناشی از استفاده از داده‌های نامطمئن در تخمین، منجر به بهبود تخمین مکانی ماده آلی گردید.

**واژه‌های کلیدی:** بافت خاک، روش والکلی و بلاک تسهیل شده، داده‌های کمکی، خطا

حمزه پور ن.، ملاعلی عباسیان. س. ۱۳۹۸. تخمین مکانی ماده آلی خاک با داده‌های نامطمئن و کمکی خاکی و روش آنتروپی حداکثر اریب. تحقیقات کاربردی خاک. جلد ۷، شماره ۴. صفحه: ۱۸-۳۴.

۱-استادیار، گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه مراغه

۲-استادیار، گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه مراغه

\*پست الکترونیک: [nhamzehpour@maragheh.ac.ir](mailto:nhamzehpour@maragheh.ac.ir)

## مقدمه

ماده آلی خاک یکی از مهمترین عوامل سلامت و باروری خاک است. وجود ماده آلی (ریشه و بقایای گیاهی) در خاک به خاکدانه‌سازی کمک می‌کند و آن هم به نوبه خود در کنار فراهم آوری برخی عناصر ضروری در رشد گیاه، نفوذ آب در خاک را نیز افزایش می‌دهد (Sullivan *et al.*, 2005; Jegajeevagan *et al.*, 2013; Forth, 1990). ماده آلی خاک یکی از اجزای کلیدی چرخه جهانی کربن است (Marchant *et al.*, 2015). وجود مقادیر کافی از ماده آلی در خاک موجب افزایش نفوذپذیری آب در خاک، جلوگیری از ایجاد رواناب، هدررفت آب و فرسایش خاک می‌شود. از سوی دیگر، در فقر پوشش گیاهی و ماده آلی خاک، ذرات ریز خاک به آسانی با باد منتقل می‌شوند (Bruun *et al.*, 2015; Saia *et al.*, 2014). مشکلی که امروزه به یکی از معضلات جدی در مناطق خشک و نیمه-خشک از جمله ایران تبدیل شده است. با توجه به میزان کم زمین‌های مناسب برای تولید غذا در جهان، داشتن دانش کافی در خصوص وضعیت ماده آلی خاک در راستای دست یافتن به استفاده پایدار از اراضی و حفاظت از محیط‌زیست اهمیت بسیاری دارد (Dono *et al.*, 2016; Novara *et al.*, 2017). از این جهت پایش تغییرات مکانی ماده آلی و تهیه نقشه‌های مربوط به وضعیت ماده آلی می‌تواند قدم مهمی در مدیریت هرچه پایدارتر اراضی کشور باشد.

برای انجام چنین کاری، نیاز به نمونه‌برداری‌های زیاد در منطقه مورد مطالعه برای ارزیابی تغییرات مکانی ماده آلی و همچنین نمونه‌برداری در زمان‌های مختلف برای پایش تغییرات زمانی آن است. یکی از مشکلات عدیده‌ای که در این زمینه وجود دارد، هزینه‌بر و زمان‌بر بودن فرآیندهای نمونه‌برداری و آنالیزهای دقیق آزمایشگاهی است. در طی دهه‌های گذشته، ترقی در زمینه آمار و افزایش قدرت محاسبه و استفاده از تصاویر ماهواره‌ای با قدرت تفکیک بالا، امکان پیشرفت در زمینه آنالیز متغیرهای وابستگی مکانی را فراهم آورده است. پژوهش‌هایی در خصوص استفاده از داده‌های سنجنش از راه دور در تخمین ماده آلی خاک صورت گرفته است (Mirzaee *et al.*, 2016; Stevens *et al.*, 2015). با وجود آن که استفاده از سنجنش از راه دور در مقایسه با روش‌های مرسوم اندازه‌گیری‌های میدانی، در نیروی کار، زمان و هزینه‌ها صرفه‌جویی می‌کند

و اطلاعات مستخرج از ابزارهای مربوط به سنجنش از راه دور، به صورت سامانمند گردآوری شده و منطقه گسترده-ای را تحت پوشش قرار می‌دهند. اما یکی از مشکلات عمده در استفاده از تصاویر ماهواره‌ای و شاخص‌های مختلف مربوط به شوری و پوشش گیاهی، دقت نسبتاً پایین تخمین‌های صورت گرفته است (Stevens *et al.*, 2012).

کریجینگ روش پذیرفته شده‌ای است که امکان تخمین متغیرهای مکانی را در نقاط نمونه‌برداری نشده، فراهم می‌کند (Li *et al.*, 2007; Giordano *et al.*, 2010; Marlet *et al.*, 2009; Acosta *et al.*, 2011). از روش‌های کریجینگ به‌طور گسترده‌تری در تخمین مکانی متغیرهای خاک استفاده شده است (Zhang *et al.*, 1997; Triantafilis *et al.*, 2001; Sullivan *et al.*, 2005; Wu *et al.*, 2009). از روش‌های کریجینگ همچنین در تخمین مکانی ماده آلی خاک نیز استفاده شده است (Wang *et al.*, 2017; Zaouche and Vaudour, 2017; Liu *et al.*, 2012; Zhang *et al.*, 2016; Mirzaee *et al.*, 2017). اما در هیچ‌کدام از روش‌های مرسوم کریجینگ، امکان محاسبه خطای ناشی از کاربرد داده‌های نامطمئن و وارد کردن خطا در هر نقطه تخمینی به‌منظور افزایش دقت تخمین وجود ندارد.

یکی از مؤثرترین راه‌های دستیابی به اطلاعات دقیق در هر نقطه، ادغام داده‌های مطمئن (hard data) و نامطمئن (soft data) با یکدیگر است. داده نامطمئن به داده‌هایی اطلاق می‌شود که چه در روش و چه در ابزار اندازه‌گیری همراه با خطا باشند (Bogaert & D-Or, 2002). کریستاکوس (Christakos, 1990) از طریق تلفیق آنتروپی حداکثر و روش‌های اریب، فرمولی بسیار عمومی برای تخمین مکانی معرفی نمود. در این روش امکان استفاده همزمان از طیف وسیعی از اطلاعات نامطمئن به صورت دامنه‌ها (intervals)، توابع احتمال تراکم‌ها (pdfs)، مدل‌ها یا نمودارها وجود دارد. از طریق توابع احتمال تراکم‌های نهایی (Posterior pdfs)، امکان تولید انواع مختلف نقشه از طریق محاسبه شاخص‌هایی که متناسب با هدف نهایی نقشه‌اند، فراهم می‌شود. از جمله این شاخص‌ها، واریانس است. با استفاده از واریانس خطای تخمین، می‌توان مرزهایی را برای تأثیر خطا روی خروجی‌های مدل تعیین کرد. به عبارت دیگر، می‌توان عدم اطمینان از مقدار واقعی

### نمونه برداری

برای انجام این پژوهش، نمونه برداری بر روی شبکه منظم با فواصل ۵۰۰ متر در منطقه‌ای به بزرگی ۶۳ کیلومتر مربع در دشت بناب صورت گرفت و حدود ۷۵ نمونه خاک جمع‌آوری شد. به منظور توسعه مدل عمومی مربوط به کالیبراسیون داده‌های نامطمئن و مطمئن ماده آلی، با توجه به تغییرات پوشش گیاهی، از ۴۷ نقطه از دشت میاندوآب نیز نمونه برداری تصادفی انجام شد تا مقادیر حداکثر و حداقل ماده آلی خاک در مدل‌سازی لحاظ شود. بنابراین در مجموع ۱۲۲ نمونه خاک تهیه شد. تمامی نمونه‌ها از عمق ۲۰-۰ سانتی‌متری تهیه شدند.

### آنالیزهای آزمایشگاهی

#### الف- داده‌های مطمئن و نامطمئن ماده آلی

داده‌های مطمئن ماده آلی، عبارت از اندازه‌گیری ماده آلی با روش والکلی و بلاک (Walkley and Black, 1934) بود که درصد کربن آلی از فرمول زیر محاسبه شد:

$$organic, c = \frac{(ml_{blank} - ml_{sample})(M_{Fe^{2+}})(0.003)(100)}{wt. water - free soil, g} \times f \quad (1)$$

در معادله بالا،  $ml_{blank}$  میلی‌لیتر فروآمونیم سولفات مصرفی برای شاهد؛  $ml_{sample}$  میلی‌لیتر فروآمونیم سولفات مصرفی برای نمونه؛  $M_{Fe^{2+}}$  مولاریته فروآمونیم سولفات؛  $wt_{water free soil, g}$  وزن خاک خشک شده در هوای آزاد (برحسب گرم) و  $f$  ضریب تصحیح و برابر ۱/۳ می‌باشند.

روش تولید داده‌های نامطمئن که در اینجا از آن به‌عنوان روش تسهیل شده والکلی و بلاک نام‌برده شده است، همانند روش بالابود با این تفاوت که نمونه‌های خاک تنها از الک ۲ میلی‌متری عبور داده شدند و عمل ساییده شدن در هاون چینی و عبور مجدد نمونه‌ها از الک ۰/۵ میلی‌متر صورت نگرفت. اندازه‌گیری ماده آلی به روش تسهیل شده والکلی و بالک در تمام نمونه‌ها صورت گرفت، ولی تنها در ۸۰ نمونه از ۱۲۲ نمونه به‌صورت تصادفی، ماده آلی به روش والکلی و بلاک مرسوم اندازه‌گیری شد. این داده‌ها که داده‌های مطمئن در نظر گرفته شدند، جهت کالیبراسیون داده‌های نامطمئن و نیز اعتبارسنجی تخمین مکانی ماده آلی با روش‌های به‌کار گرفته شده در این پژوهش، استفاده شدند.

یک متغیر را به‌صورت دامنه‌ای از مقادیر ممکن، نشان داد (Heuvelink, 1998).

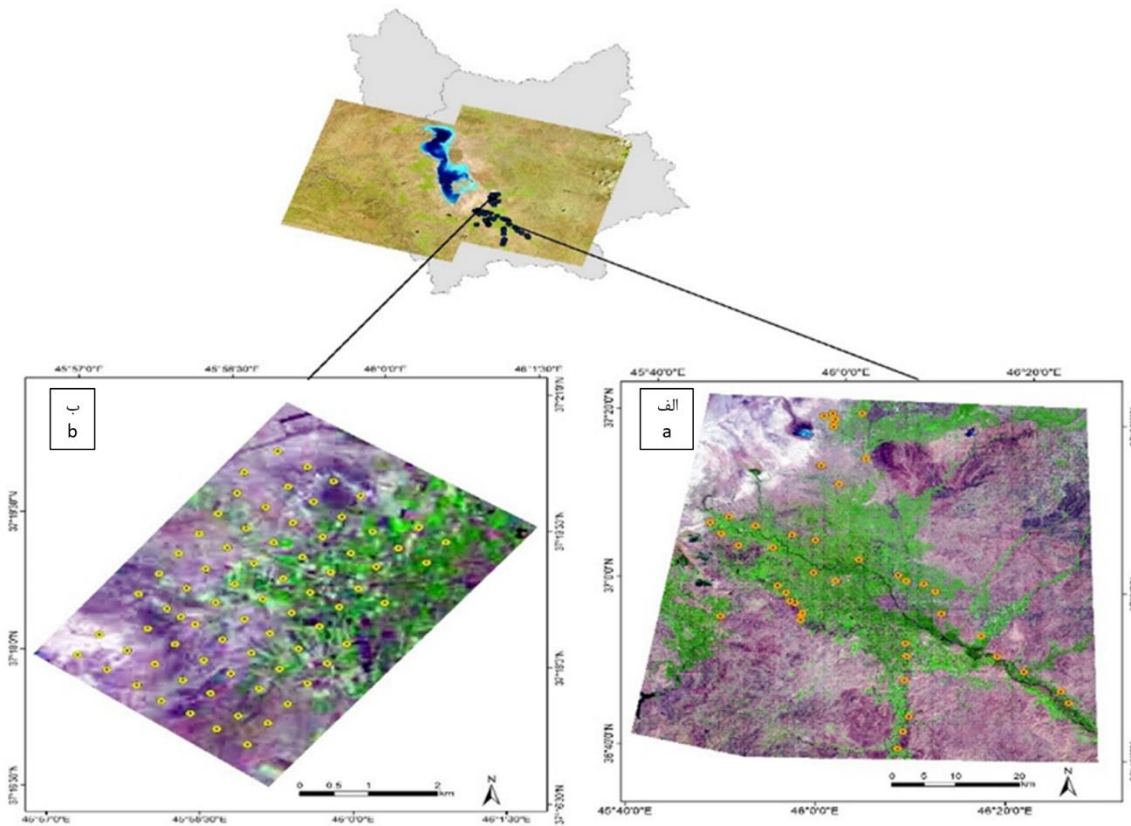
تاکنون، پژوهشگران بسیاری سعی در استفاده بهینه از داده‌های نامطمئن و وارد کردن خطای حاصل از استفاده آن‌ها در تخمین مکانی نموده‌اند (Serre & Christakos, 1999; Douaik et al., 2004; Brus et al., 2008; Hamzhepour et al., 2013; Hamzhepour et al., 2017). اما در تمامی این مطالعات، تنها یک نوع داده نامطمئن خاکی به کار گرفته شده است (به‌عنوان مثال هدایت الکتریکی ظاهری) و تاکنون مطالعه‌ای در خصوص استفاده از پارامترهای مختلف خاکی (همانند اجزای بافت خاک، آهک و غیره) و خطای حاصل از آنها، در تخمین دیگر متغیر خاکی (همانند ماده آلی خاک) صورت نگرفته است.

در پژوهش حاضر، تخمین مکانی ماده آلی خاک با مقادیر اندازه‌گیری شده آن با روش والکلی و بلاک تسهیل شده صورت گرفت. همچنین، سعی شد تا از مدل خطی توسعه یافته با استفاده از داده‌های کمکی خاکی و روش آنتروپی حداکثر اریب در تخمین ماده آلی خاک استفاده شود تا چهارچوبی در راستای استفاده از انواع داده‌های خاکی در تخمین متغیرهای پیوسته خاک با در نظر گرفتن خطا و روش آنتروپی حداکثر اریب فراهم شود.

### مواد و روش‌ها

#### منطقه مورد مطالعه

منطقه مورد مطالعه در حاشیه جنوب و جنوب شرقی دریاچه ارومیه، در دشت بناب-میاندوآب در آذربایجان شرقی واقع شده است. مجموع مساحت این منطقه در حدود ۱۵۰ کیلومتر مربع است. این منطقه در پایین دست دامنه کوه سهند و در جلگه‌ای حاصلخیز واقع شده است (شکل ۱). مختصات جغرافیایی آن بین طول جغرافیایی ۴۴° ۵۷' ۴۵" و ۴۶° ۰۱' ۰۹" شرقی و عرض جغرافیایی ۳۷° ۱۹' ۵۵" و ۳۷° ۱۹' ۵۸" شمالی در دشت بناب و بین طول جغرافیایی ۳۴° ۱۷' ۳۴" و ۴۵° ۳۲' ۰۵" شرقی و عرض جغرافیایی ۳۶° ۴۰' ۰۰" و ۳۷° ۱۵' ۲۴" شمالی در دشت میاندوآب واقع شده است. منطقه از نظر پستی و بلندی، مسطح است. ارتفاع منطقه در حدود ۱۲۷۶ متر از سطح دریاهای آزاد است.



شکل ۱- موقعیت منطقه مطالعاتی در حاشیه جنوب شرقی دریاچه ارومیه در دشت بناب و میاندوآب. الف: موقعیت نقاط نمونه‌برداری در دشت میان‌دوآب؛ ب: موقعیت نقاط نمونه‌برداری در دشت بناب

Figure 1. (a) The location of the study area in southeast Urmia Lake in Miandoab-Bonab plains; (b) the location of soil sampling points in Bonab Plain; (c) the location of soil sampling points in Miandoab Plain

در تحقیق حاضر، به منظور کالیبراسیون داده‌های نامطمئن، ابتدا با استفاده از دو سوم داده‌ها، همبستگی بین داده‌های مطمئن و نامطمئن مدل‌سازی شد. سپس با استفاده از معادله خط به دست آمده، تخمین در یک سوم دیگر داده‌ها که در مدل‌سازی استفاده نشده بودند، صورت گرفت. تفاضل بین مقادیر تخمینی و مقادیر واقعی ماده آلی، به عنوان باقی مانده‌های تخمین در نظر گرفته شدند. براساس قانون آنتروپی حداکثر اریب، چنانچه این باقی مانده‌ها توزیع نرمال داشته باشند، می‌توان تابع تراکم احتمالی (pdf) را بر روی آنها برازش کرد و از معادله pdf در هر نقطه، برای تخمین مکانی ماده آلی استفاده کرد.

#### ب- داده‌های کمکی

از آنجایی که ماده آلی خاک با برخی از پارامترهای خاک از جمله بافت خاک در اغلب مواقع رابطه معناداری نشان می‌دهد (Forth, 1991; Jegajeevagan *et al.*, 2013; Zaouche *et al.*, 2017). چنین پارامترهایی می‌توانند به عنوان داده‌های کمکی در تخمین مکانی ماده آلی خاک

داده‌های نامطمئن احتمالی به این معنا است که هر نقطه ای که در تخمین دخالت دارد، نه یک مقدار مطلق، بلکه طیفی از اعداد را شامل می‌شود که در قالب تابعی با تمام احتمالات موجود برای مقدار واقعی، با در نظر گرفتن خطا در آن نقطه، ارائه می‌شود. اولین قدم در تعریف توابع احتمالی مطرح شده، تعیین مقدار مطلق خطا در هر نقطه است. برای این منظور ابتدا تفاضل بین داده‌های مطمئن و نامطمئن به دست می‌آید. اگر باقی مانده‌های محاسبه شده، توزیعی نرمال داشته باشند و فاقد روند خاصی باشند، می‌توان واریانس باقی مانده‌ها را به عنوان خطای اندازه‌گیری در نظر گرفت و در تعریف تابع احتمالی تراکمی در هر نقطه از آن استفاده کرد. بدین منظور، از هیستوگرام باقی مانده‌ها و تابع احتمال تراکمی برای ایجاد داده‌های نامطمئن از نوع احتمالی استفاده شد و پس از برازش بهترین مدل واریوگرامی، تخمین مکانی ماده آلی خاک صورت گرفت.

مجموع، توسط انتگرال و  $p_i$  توسط مقادیر توابع احتمال تراکم، تابع  $f(z)$  جایگزین شده است. بنابراین، رابطه آنتروپی به شکل زیر در می‌آید:

$$H_z = - \int_{D_z} f(z) \log f(z) dz \quad (3)$$

که در آن  $dz$  تغییرات جزئی متغیر تصادفی  $Z$  می‌باشد. نظریه اطلاعات به طور مستقیم به نظریه آنتروپی مرتبط است. در واقع، مطابق رابطه ۴، اطلاعات و احتمال اتفاق به طور معکوس با یکدیگر مرتبط‌اند.

$$Info(A) = \frac{1}{p}(A) \quad (4)$$

که در آن،  $A$  واژه‌ای برای بیان متغیر تصادفی است. این عبارت بیان می‌کند که هر چقدر احتمال وقوع یک پدیده زیاد باشد، اطلاعات کمتری را شامل می‌شود. از آنجایی که احتمالات ممکن است بسیار کوچک باشند، استفاده از لگاریتم مناسب‌تر است (Christakos, 1990) و منجر به تعریف ریاضی زیر از اطلاعات برای یک متغیر تصادفی می‌شود.

$$Info(Y) = \log\left(\frac{1}{p(Y=y_i)}\right) = \log(P(Y=y_i)) \quad (5)$$

مفهوم معادلات ۲ الی ۵ این است که آنتروپی را می‌توان به صورت مقدار مورد انتظار اطلاعات برای تمام وقایع ممکن و انحصاری که درباره متغیرهای تصادفی  $Y$  تعریف می‌شوند، بیان کرد:

$$HY = E[Info(Y)] \quad (6)$$

این عبارت به عنوان تابع آنتروپی نیز شناخته شده است (Jaynes, 1982). با توجه به این تعریف، حداکثر کردن آنتروپی منجر به حداقل کردن اطلاعات اولیه (مورد انتظار) تحت محدودیت‌های برخی اطلاعات اولیه ناشی از دانش عمومی پایه (KG) می‌شود (Christakos, 2000).

مورد استفاده قرار گیرند. لذا، هدایت الکتریکی خاک و pH گل اشباع ( $EC_e$ )، کربنات کلسیم معادل به‌روش تیتراسیون برگشتی با اسید کلریدریک (Nelson, 1982)، کلسیم و منیزیم محلول به روش تیتراژ کردن با EDTA، سدیم محلول با استفاده از دستگاه فلیم‌فتمتر، درصد ذرات خاک به روش هیدرومتری (Gee & Bauder, 1986) و پایداری خاکدانه‌ها در حالت مرطوب<sup>۱</sup> به روش الکترون با اندازه در محدوده یک الی دو میلی‌متر تعیین شد. سپس، میزان همبستگی بین پارامترهای اندازه‌گیری شده و ماده آلی خاک بررسی شد.

به‌منظور توسعه مدل خطی با استفاده از داده‌های کمکی نیز از میان پارامترهای خاک اندازه‌گیری شده، پارامترهایی که دارای همبستگی معنادار با ماده آلی بودند، به‌عنوان پارامترهای مدل خطی انتخاب شدند. سپس مدل مناسب که بالاترین ضریب همبستگی و کمترین خطای استاندارد شده تخمین را داشت، انتخاب و پس از محاسبه خطا و برآزش مدل واریوگرامی مشابه روش مطرح شده، این داده‌ها در تخمین مکانی ماده آلی خاک با روش آنتروپی حداکثر اریب، استفاده شدند. مراحل مختلف انجام این پژوهش در شکل (۲) به‌صورت خلاصه ارائه شده است.

### تخمین مکانی ماده آلی با روش آنتروپی حداکثر اریب

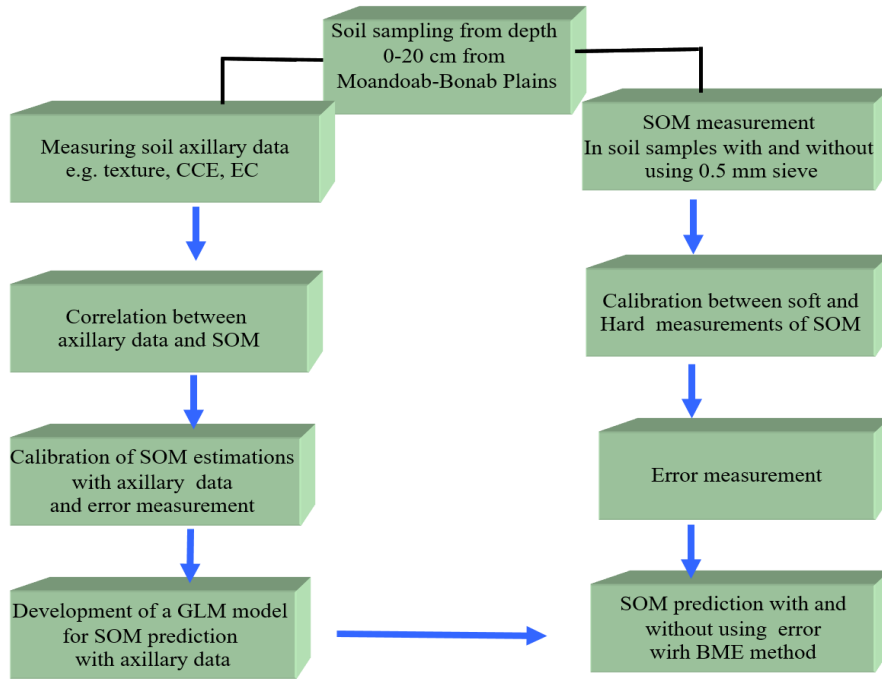
#### الف- آنتروپی و اطلاعات

آنتروپی<sup>۲</sup> اولین بار در ترمودینامیک به عنوان تغییرات در بی‌نظمی در یک سیستم شیمیایی بسته در نتیجه انجام یک واکنش، ارائه شد. یک قرن بعد، شانون مفهوم آنتروپی را در تئوری اطلاعات خود، برای بیان درجه عدم قطعیت اولیه در یک سیستم، بیان کرد (Shannon, 1948). از آن زمان، آنتروپی در طیف وسیعی از رشته‌های علمی برای اندازه‌گیری عدم قطعیت استفاده شده است. فرمول ریاضی آنتروپی برای متغیرهای تصادفی گسسته به صورت زیر است (Jaynes, 1982):

$$H_c = - \sum_{i=1}^{nc} p_i \log p_i \quad (7)$$

که در آن،  $p_i$  احتمال اینکه متغیر تصادفی  $C$  مقدار  $C_i$  را بخود بگیرد، است. برای متغیرهای تصادفی پیوسته، نماد

1. Wet Aggregate Stability  
2. Entropy



شکل ۲- فلوجارت مربوط به مراحل مختلف انجام پژوهش

Figure 2. Flowchart of the several steps of the research

که در آن  $g_\alpha(Z_{map})$  توابع معلوم  $Z_{map}$  اند.

محدودیت نرمال شدن به صورت زیر تعریف می شود:

$$g_0(z_{map}) = 1 \Rightarrow E[g_0] = 1 \quad (9)$$

با استفاده از ضریب  $\mu_\alpha$  و اعمال برخی محدودیت‌ها، معادله ۸ را می توان به صورت زیر تعریف کرد:

$$L[f_G(z_{map})] = - \int f_G(z_{map}) \log f_G(z_{map}) dz_{map} = - \sum_{\alpha=0}^{N_c} \mu_\alpha [\int g_\alpha(z_{map}) f_G(z_{map}) dz_{map} - E[g_\alpha(g_\alpha(z_{map}))]] \quad (10)$$

که در آن،  $N_c$ ، مجموع تعداد محدودیت‌ها هستند. با برابر صفر قرار دادن مشتقات جزئی معادله، و حل دستگاه معادلات نسبت به  $\mu_\alpha$ ، حل حداکثر آنتروپی برای تابع تجمعی احتمالی (pdf) اولیه به دست می‌آید (Christakos, 2000):

$$f_G(z_{map}) = \frac{1}{A} \exp\left(\sum_{\alpha=1}^{N_c} v_\alpha g_\alpha(z_{map})\right) \quad (11)$$

که در آن A تابع جزئی است که نقش محدود کننده نرمال را دارد و برابر عبارت زیر است.

### ب- حداکثر کردن آنتروپی

بردار متغیرهای تصادفی  $Z_{map}=(Z_{hard}, Z_{soft}, Z_0)$  را مد نظر قرار داده و  $Z_0$  و  $Z_{soft}$ ،  $Z_{hard}$  را به ترتیب مقادیر در نقاط داده‌های مطمئن، نامطمئن و مقدار نامعلوم در نقطه تخمین در نظر بگیرید. همچنین  $f_G(Z_{map})$  را به عنوان توابع احتمال تراکم چند متغیره برای دانش عمومی  $K_G$ ، قبل از اینکه هر گونه دانش اختصاصی و ویژه  $K_S$  اضافه شود، در نظر داشته باشید. همانطور که در بخش گذشته اشاره شد، حداکثر کردن آنتروپی یک سیستم زمانی که تمام داده‌های قابل دسترس مورد استفاده قرار گرفته باشند، منجر به حداکثر کردن اطلاعات مورد انتظار آن می‌شود. بنابراین اگر آنتروپی به صورت زیر تعریف شود:

$$H(f_G(z_{map})) = E[Info_G[Z_{map}]] = - \int f_G(z_{map}) \log f_G(z_{map}) dz_{map} \quad (7)$$

ما مجبور هستیم که این عبارت را تحت محدودیت‌هایی که در اطلاعات اولیه موجود است، حداکثر کنیم. این محدودیت‌ها را می‌توان به صورت زیر تعریف کرد:

$$E[g_\alpha] = \int g_\alpha(z_{map}) f_G(z_{map}) dz_{map} \quad (8) \\ \alpha = 0, \dots, N_c$$

حال فرض کنیم داده‌های نامطمئن در معادله ۱۳ از نوع  $f_s(z_{soft})$  هستند. برای مثال یک تابع احتمال تراکم برای هر موقعیت داده در همسایگی نقطه تخمینی وجود دارد. به طور مساوی، می‌توانیم از تابع توزیع تجمعی آن یعنی  $f_s(z_{soft})$  استفاده کنیم که منجر به معادله زیر می‌شود:

$$f_G(z_0|z_{hard}, F_s(z_{soft})) = \frac{1}{F_G(z_{hard}, F_s(z_{soft}))} \frac{\partial}{\partial z_0} F_G(z_0, z_{hard}, F_s(z_{soft})) \quad (14)$$

با داشتن I به عنوان دامنه تغییرات انتگرال چند متغیره برای توابع احتمال تراکم داده‌های نامطمئن، داریم:

$$F_G(z_0, z_{hard}, F_s(z_{soft})) \quad (15)$$

$$\begin{aligned} &= \int_I \int_{D_{z_0}} f_G(z_0, z_{hard}, z_{soft}) dF_s(z_{soft}) dz_0 \\ &= \int_I \int_{D_{z_0}} f_G(z_0, z_{hard}, z_{soft}) \frac{dF_s(z_{soft})}{dz_{soft}} dz_{soft} \\ &= \int_I \int_{D_{z_0}} f_G(z_0, z_{hard}, z_{soft}) f_s dz_{soft} dz_0 \end{aligned}$$

و به طریقی مشابه:

$$F_G(z_0, z_{hard}, F_s(z_{soft})) \quad (16)$$

$$\begin{aligned} &= \int_I f_G(z_0, z_{hard}, z_{soft}) dF_s(z_{soft}) \\ &= \int_I f_G(z_0, z_{hard}, z_{soft}) \frac{dF_s(z_{soft})}{dz_{soft}} dz_{soft} \\ &= \int_{D_{z_0}} f_G(z_0, z_{hard}, z_{soft}) f_s dz_{soft} \end{aligned}$$

با جاگذاری دو معادله ۱۵ و ۱۶ در معادله ۱۴، معادله زیر حاصل می‌شود:

$$f_G(z_0|z_{hard}, F_s(z_{soft})) = \frac{\int f_G(z_0, z_{hard}, z_{soft}) f_s dz_{soft}}{\int f_G(z_{hard}, z_{soft}) f_s dz_{soft}} \quad (17)$$

معادله بالا، حل BME برای داده‌های نامطمئن از نوع احتمالی است.

$$A = \int \exp\left(\sum_{\alpha=1} \mu_{\alpha} g_{\alpha}(z_{map})\right) dz_{map} \quad (12)$$

تابع تجمعی احتمالی اولیه که از حداکثر کردن آنتروپی حاصل شده است، نشان می‌دهد که چیزی بیشتر از تمام داده‌های اولیه، مورد توجه قرار گرفته نشده‌اند. به طور کلی، توابع احتمال تراکم اولیه حاصل از معادله ۱۱، غیرگوسین است. با این حال، زمانی که تنها میانگین و تابع کوواریانس دانش عمومی  $K_G$  را تشکیل دهند؛ می‌توان نشان داد که معادله ۱۱ منجر به توزیع گوسین چند متغیره می‌شود.

ج- شرطی کردن ارباب با دانش اختصاصی  $K_s$  تا این مرحله، توابع احتمال تراکم اولیه توسط معادله ۱۱ تعیین شد. حال به دنبال بدست آوردن توابع احتمال تراکم نهایی برای متغیر  $Z$  در نقطه تخمینی  $X_0$  با در دست داشتن داده‌های مطمئن و نامطمئن و با در نظر گرفتن قانون احتمال برای احتمالات شرطی می‌توان نوشت:

$$f_k(z_0) = f_G(z_0|z_{hard}, z_{soft}) = \frac{f_G(z_0, z_{hard}, z_{soft})}{f_G(z_{hard}, z_{soft})} \quad (13)$$

سمت راست معادله فوق نشانگر تابع تجمعی احتمال نهایی شرطی بر اساس دانش عمومی و سمت چپ معادله نشانگر تابع تجمعی احتمال شرطی شده حاصل از دانش اختصاصی است (Christakos, 2002).

معادله ۱۳ دربرگیرنده اطلاعات اختصاصی، از جمله داده‌های مطمئن و انواع داده‌های نامطمئن در منطقه مورد مطالعه است. روش محاسبه این توابع احتمال تراکم، بستگی خواهد داشت به نوع داده‌های نامطمئن موجود و این چیزی است که کریستاکوس به آن قانون پردازش دانش می‌گوید (Christakos, 2000). همچنین به آن اصل شرطی کردن ارباب نیز اطلاق می‌گردد. در میان انواع داده‌های نامطمئن، در مطالعه حاضر از داده‌های نامطمئن از نوع احتمالی استفاده شده است.

د- توابع احتمال تراکم نهایی با استفاده از داده‌های نامطمئن از نوع احتمالی

## روش‌های ارزیابی

دست آمد. میانگین ماده آلی با این روش ۱/۰۶ درصد محاسبه شد. این مقادیر برای اندازه‌گیری‌های دقیق ماده آلی به روش مرسوم والکلی و بلاک به ترتیب ۰/۱۱ درصد برای مقدار حداقل، ۶/۲۸ درصد برای مقدار حداکثر و میانگین این داده‌ها ۱/۱۴ درصد به دست آمد. همان‌طور که از این نتایج قابل مشاهده است، مقادیر حداقل، حداکثر و میانگین داده‌های نامطمئن ماده آلی کمتر از مقادیر واقعی آن به دست آمد. از آنجایی که تنها تفاوت این دو روش، عبور و عدم عبور از الک نیم‌میلی‌متری است، به نظر می‌رسد که عدم استفاده از الک نیم‌میلی‌متری منجر به اندازه‌گیری‌های اغلب کمتر از مقدار واقعی شده است.

با توجه به دامنه تغییرات ماده آلی در خاک‌های مناطق خشک و نیمه‌خشک که اغلب کمتر از دو درصد است، دامنه مقادیر ماده آلی اندازه‌گیری شده در این پژوهش، طیف وسیعی از مقادیر ماده آلی قابل مشاهده در منطقه مطالعاتی را در بر گرفته است و از این نظر می‌توان نمونه‌های جمع‌آوری شده را به‌عنوان جامعه مناسبی از جامعه ماده آلی خاک در منطقه مطالعاتی در دشت بناب-میان‌دوآب دانست. بیشترین مقادیر اندازه‌گیری شده ماده آلی، در نمونه‌های جمع‌آوری شده از دشت میان‌دوآب مشاهده شد. این دشت به دلیل عبور دو رودخانه اصلی منتهی به دریاچه ارومیه به نام‌های سیمینه‌رود و زرینه‌رود، از نظر کشاورزی و حاصلخیزی وضعیت بهتری از دشت بناب دارد و از این رو، درصد ماده آلی در این دشت به‌طور میانگین بالاتر از دشت بناب است.

برای اعتبارسنجی صحت روش‌های مورد استفاده در این پژوهش، از روش اعتبارسنجی متقاطع و پارامترهای آماری ضریب تبیین ( $r$ )، میانگین مربعات خطا (MSE)، ریشه میانگین مربعات خطا (RMSE) و ریشه میانگین مربعات خطای نرمال شده (nRMSE) مدل استفاده شد. برای آموزش و اعتبارسنجی رگرسیون‌های چندگانه استفاده شده در این پژوهش، ابتدا ۸۰ درصد از نمونه‌ها به‌طور تصادفی انتخاب و برای آموزش رگرسیون‌ها و ۲۰ درصد باقی‌مانده نیز برای اعتبارسنجی روابط برآورد کننده در نظر گرفته شد. نهایتاً، تمامی عملیات آماری شامل محاسبه همبستگی‌ها، ایجاد رگرسیون چندگانه و انتخاب تصادفی داده‌ها با استفاده از نرم‌افزار SPSS انجام شد. برای انجام تمام مراحل این پژوهش از نرم‌افزار BMElib استفاده شد که این نرم‌افزار بر اساس زبان برنامه‌نویسی MATLAB (Math works, 1999) نوشته شده است.

## نتایج و بحث

تخمین ماده آلی خاک با روش آنروپی حداکثر اریب و داده‌های نامطمئن احتمالی

## الف- تجزیه و تحلیل داده‌های مطمئن و نامطمئن

خلاصه‌ای از تجزیه و تحلیل مقادیر اندازه‌گیری شده ماده آلی در جدول (۱) ارائه شده است. نتایج نشان داد در روش تسهیل شده اندازه‌گیری ماده آلی، حداقل مقدار اندازه‌گیری شده صفر و مقدار حداکثر آن ۵/۱۰ درصد به

جدول ۱- خلاصه آماری اندازه‌گیری‌های مطمئن و نامطمئن ماده آلی خاک

Table 1. Summary of statistical analysis of soil organic matter for hard and soft data

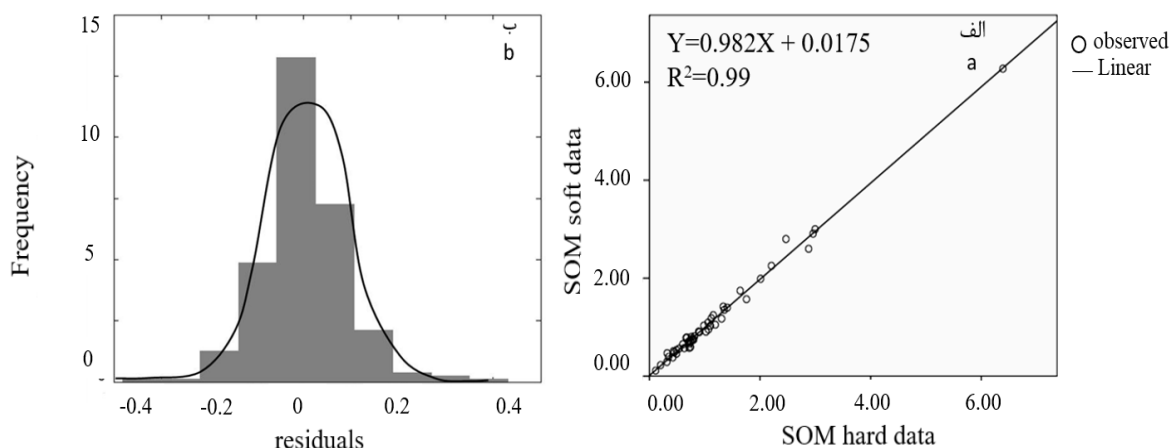
SOM	Count	Mean (%)	SD*	Minimum (%)	Maximum (%)	Skewness	Kurtosis
Hard data	80	1.14	1.02	0.11	6.28	3.10	12.91
Soft data	122	1.06	0.97	0.00	5.10	2.33	6.13

SOM: soil organic matter; SD: standard deviation

۰/۹۹۵، ۰/۹۹۱ و ۰/۹۹۱ محاسبه شد. این نتایج با خطای استاندارد شده تخمین برابر ۰/۰۹۷۷ از دقت بالایی برخوردار بود. بدین ترتیب، امکان استفاده از معادله خط به‌دست‌آمده در کالیبراسیون داده‌های نامطمئن فراهم شد.

ب- همبستگی بین اندازه‌گیری‌های ماده آلی خاک به روش مطمئن و نامطمئن  
نتایج بررسی همبستگی بین داده‌های مطمئن و نامطمئن ماده آلی خاک نشان داد که همبستگی بالایی بین آن‌ها به‌صورت خطی وجود دارد (شکل ۳، الف). به‌طوری‌که مقادیر  $R$ ،  $R^2$  و  $R^2_{aj}$  به ترتیب برابر



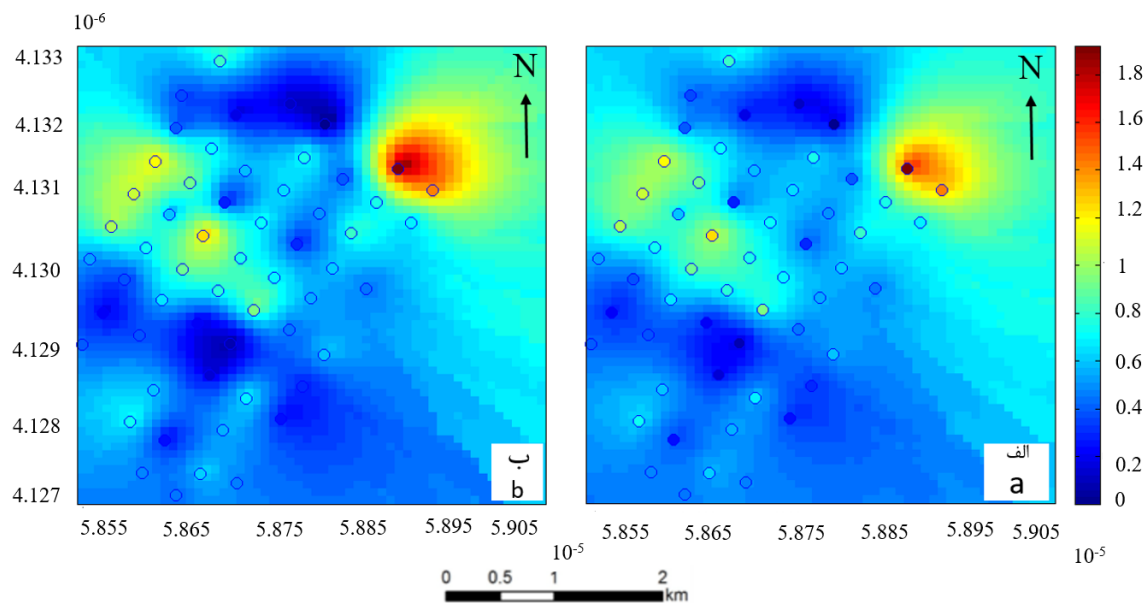


شکل ۳- الف: میزان همبستگی بین ماده آلی خاک اندازه‌گیری شده به روش مطمئن و نامطمئن. دوایر نشان‌دهنده مقادیر اندازه‌گیری شده و خط ممتد، مدل رگرسیونی برازش شده است؛ ب: هیستوگرام باقی مانده تفاضل داده‌های مطمئن و نامطمئن و تابع تراکم نهایی برازش شده آن‌ها

Figure 3. (a) The correlation between hard and soft SOM measurements. Dots correspond to the measured values while solid line is the corresponding the fitted model; (b) Histograms of the residuals between hard and soft data and its probability density function (pdf)

یکسان خطا در بین داده‌های نامطمئن است. لذا، این تابع در هر نقطه‌ای که داده‌های نامطمئن در تخمین نقاط اطراف دخالت داشت، تعریف شد. در مرحله بعدی، به‌منظور تخمین مکانی ماده آلی، پیوستگی مکانی نقاط اندازه‌گیری شده بررسی شد. بهترین مدل واریوگرامی برازش شده با گام‌های ۴۰۰ متر و تا فاصله ۳۰۰۰ متر، مدل توانی با بیشترین ضریب همبستگی (برابر ۰/۹۱۴) و کمترین خطای محاسبه انتخاب شد و پارامترهای مدل به دست آمدند. در این مدل، اثر قطعه‌ای برابر ۰/۰۴۹، دامنه تأثیر ۲۴۶۰ متر و آستانه برابر ۰/۱۰۷ به دست آمدند. سپس برای تخمین ماده آلی خاک با روش BME و داده‌های نامطمئن احتمالی، ابتدا یک شبکه با فواصل ۱۰۰ متر تعریف شد و تخمین مکانی ماده آلی با استفاده از معادله ۱۷ و با در نظر گرفتن پارامترهای مدل واریوگرامی برازش شده، صورت گرفت. این تخمین یک‌بار با دخالت دادن خطای تخمین محاسبه شده و بار دیگر بدون خطا انجام شد (شکل ۴).

ج- داده‌های نامطمئن از نوع احتمالی و تخمین مکانی نتایج مقایسه مقادیر داده‌های مطمئن و نامطمئن نشان داد که تفاوت معناداری از نظر آماری بین این دو دسته از اندازه‌گیری‌ها وجود دارد (p-value کمتر از ۰/۰۵ و بنابراین فرض صفر رد شد). هیستوگرام مربوط به تفاضل بین اندازه‌گیری‌های مطمئن و نامطمئن در شکل ۳، ب، ارایه شده است. همان‌طور که از این شکل مشهود است و نیز نتایج مربوط به ضریب معنی‌داری آزمون کلموگروف-اسمیرنوف نشان داد، باقی‌مانده‌های اندازه‌گیری شده شرط نرمال بودن (ضریب معنی‌داری کلموگروف-اسمیرنوف بزرگ‌تر از ۰/۰۵) را دارا بودند. این نشان داد که اختلاف بین اندازه‌گیری‌های مطمئن و نامطمئن ماده آلی ناشی از خطا است (خطا در ابزار اندازه‌گیری یا روش اندازه‌گیری) و روندی در بین آن‌ها مشاهده نمی‌شود. بنابراین واریانس باقی‌مانده‌ها به‌عنوان خطای اندازه‌گیری در نظر گرفته شد. در شکل ۳، ب همچنین تابع احتمال تراکمی (pdf) برازش شده بر روی باقی‌مانده‌های داده‌ها نشان داده شده است. همان‌طور که از این شکل مشهود است، این تابع از توزیع نرمال برخوردار است که نشان‌دهنده توزیع



شکل ۴- تخمین ماده آلی خاک با استفاده از داده‌های نامطمئن و روش آنتروپی حداکثر اریب در دشت بناب. الف: با استفاده از داده‌های نامطمئن؛ ب: با استفاده از داده‌های نامطمئن و خطای تخمین. شباهت بین رنگ نقاط و رنگ زمینه نشانگر دقت تخمین است

Figure 4. SOM prediction using soft data and Bayesian maximum entropy method in Bonab Plain. (a) using only soft data; (b) using soft data and error. Proximity between colours for the circles and the background map is thus an indication of the prediction quality

پژوهش‌های زیادی در خصوص استفاده از خطا در تخمین مکانی متغیرهای خاک تاکنون صورت نگرفته است. اما، نتایج اکثر این مطالعات نشان داده است که وارد کردن خطای ناشی از استفاده از داده‌های نامطمئن، منجر به بهبود تخمین مکانی شده است. در پژوهشی که توسط حمزه پور و همکاران ( Hamzhepour *et al.*, 2013) صورت گرفت، وارد کردن خطای ناشی از استفاده از داده‌های نامطمئن شوری خاک در معادلات کریجینگ، منجر به کاهش MSE از ۰/۸۲ به ۰/۶۴ شد. در خصوص داده‌های نامطمئن شوری خاک و خطا ( Douaik *et al.*, 2005, 2004; Hamzhepour & Eghbal, 2016; Hamzhepour and Bogaert, 2017) نقشه بافت خاک به‌عنوان داده نامطمئن به همراه نقشه خاک به‌عنوان داده نامطمئن ( Brus *et al.*, 2008) ، نیز نتایج نشان داده‌اند که محاسبه و وارد کردن خطا در تخمین، منجر به افزایش دقت تخمین متغیر مورد مطالعه شده است.

به‌منظور اعتبارسنجی دقت تخمین، از اندازه‌گیری‌های دقیق ماده آلی که از آن‌ها در مدل‌سازی و محاسبه خطا استفاده نشده بود، بهره گرفته شد. از نتایج ارائه شده در جدول (۲) می‌توان استنباط نمود استفاده از خطای تخمین و روش آنتروپی حداکثر اریب، منجر به بهبود تخمین مکانی ماده آلی شده است. برخلاف همبستگی بالای بین اندازه‌گیری ماده آلی به روش مطمئن و نامطمئن در این پژوهش، و نیز برخلاف مقدار عددی کم خطای محاسبه‌شده، وارد کردن خطا در تخمین، منجر به بهبود تخمین مکانی ماده آلی خاک شد. به‌طوری‌که مقدار nRMSE از ۰/۲۲ به ۰/۰۶ کاهش یافت. بنابراین، به نظر می‌رسد با وجود آنکه اندازه‌گیری ماده آلی به روش تسهیل شده والکلی و بلاک (داده‌های نامطمئن) در این پژوهش، تفاوت زیادی با مقادیر اندازه‌گیری شده به روش مرسوم والکلی و بلاک نداشت، اما نتایج تخمین مکانی ماده آلی با داده‌های نامطمئن، نشان داد که خطای ناشی از استفاده از این روش می‌تواند دقت تخمین را کاهش دهد و وارد کردن خطای اندازه‌گیری در معادلات تخمین مکانی، منجر به افزایش دقت تخمین شد (جدول ۲).

جدول ۲- مقادیر کمی روش‌های مقایسه اعتبارسنجی دقت BME با استفاده از داده‌های نامطمئن و داده‌های نامطمئن به همراه خطا

Table 2. Quantitative values of the validation methods of BME accuracy with soft data and soft data with error

Comparison criteria	Soft data	soft data with error
R	0.89	0.97
MSE (%)	0.055	0.005
RMSE (%)	0.23	0.07
nRMSE	0.22	0.06

کلسیم معادل، شن و سیلت (معادله ۱۸) منجر به بهترین مدل خطی تخمین ماده آلی با ضریب تبیین ۰/۵۹ شد:

$$OM = 0.361 + 0.04 \text{ CCE} - 0.013 \text{ silt} - 0.005 \text{ sand} \quad (18)$$

که در این معادله، CCE، درصد کربنات کلسیم معادل، silt، درصد سیلت و sand درصد شن در نمونه‌های خاک مورد مطالعه می‌باشند. برازش مدل واریوگرامی بر روی داده‌های تخمینی ماده آلی از مدل خطی نشان داد که این داده‌ها از پیوستگی مکانی قوی برخوردار هستند. بهترین مدل واریوگرامی برازش شده با گام‌های ۴۰۰ متر و تا فاصله ۳۰۰۰ متر، مدل کروی با بیشترین ضریب همبستگی (برابر ۰/۹۱۷) و کمترین خطای محاسبه انتخاب شد و پارامترهای مدل به دست آمدند. در این مدل، اثر قطعه‌ای برابر ۰/۰۰۰۱، دامنه تأثیر ۱۲۵۴ متر و آستانه برابر ۰/۰۱۵۲ به دست آمدند.

ب- داده‌های نامطمئن به دست آمده از مدل رگرسیون خطی تعمیمی

در شکل ۵، الف، همبستگی بین مقادیر تخمینی ماده آلی که از مدل رگرسیون خطی تعمیمی به دست آمده‌اند و مقادیر واقعی ماده آلی، ارائه شده است. مقادیر واقعی ماده آلی که در این بخش استفاده شده است، در توسعه معادله ۱۸، به کار گرفته نشد. نتایج این بخش از مطالعه نشان داد که با وجود همبستگی ضعیف‌تر این دسته از داده‌های نامطمئن ماده آلی با مقادیر واقعی آن در مقایسه با نتایج ارائه شده در بخش پیشین، با این حال این یافته‌ها با ضریب تبیین ۰/۵۹ و معناداری در سطح یک درصد، از دقت کافی برای استفاده از آن‌ها در کالیبراسیون مقادیر تخمینی ماده آلی برخوردار بودند.

هیستوگرام مربوط به تفاضل بین مقادیر تخمینی ماده آلی با مدل رگرسیون خطی تعمیمی و مقادیر واقعی ماده آلی در شکل ۵، ب، ارایه شده است. همان‌طور که از این شکل و نیز نتایج مربوط به ضریب معنی‌داری آزمون

تخمین مکانی ماده آلی خاک با استفاده از رگرسیون خطی تعمیمی

الف- انتخاب پارامترهای مدل و برازش بهترین مدل خطی

پارامترهای خاکی اندازه‌گیری شده در این پژوهش و خلاصه‌ای از تجزیه تحلیل‌های آماری آن‌ها در جدول (۳) ارائه شده است. در جدول (۴) نیز همبستگی بین ماده آلی خاک و پارامترهای اندازه‌گیری شده ارائه شده است. همان‌طور که از این جدول قابل استنباط است، ماده آلی در سطح یک درصد با آهک، پایداری خاکدانه-ها در حالت مرطوب، سیلت و شن و در سطح پنج درصد، با رس همبستگی معناداری نشان داد. با توجه به اینکه پایداری خاکدانه‌ها در حالت مرطوب، خود متغیر وابسته به ماده آلی است و اندازه‌گیری آن بسیار دشوارتر از اندازه‌گیری خود ماده آلی می‌باشد، لذا این پارامتر کنار گذاشته شد و از دیگر پارامترها با همبستگی معنادار برای توسعه مدل خطی به منظور تخمین مکانی ماده آلی استفاده شد. نتایج مطالعه شیلاسی و همکاران (Schillaci et al., 2017) نیز نشان داد که در کنار پارامترهای مربوط به توپوگرافی، اجزای بافت خاک نیز همبستگی معناداری با ماده آلی خاک دارند. استیونز و همکاران (Stevens et al., 2015) نیز گزارش کرده‌اند که در حدود ۶۵٪ از تغییرات مکانی ماده آلی که با روش REML محاسبه شده بود، با استفاده از بافت خاک قابل توضیح بود. این نتایج توسط افراد دیگر نیز تأیید شده است (Kempen et al., 2010; Zhang et al., 2012; Meersmans et al., 2012).

نتایج مدل‌سازی مرحله به مرحله ماده آلی خاک با رگرسیون خطی تعمیمی با پارامترهای خاکی نشان داد که حذف درصد رس از معادله منجر به افزایش ضریب تبیین گردید. لذا این پارامتر نیز از مدل کنار گذاشته شد و در نهایت استفاده از پارامترهای درصد کربنات

کلموگروف-اسمیرنوف نشان داد، باقی مانده های اندازه-گیری شده شرط نرمال بودن (ضریب معنی داری کلموگروف-اسمیرنوف بزرگ تر از ۰/۰۵) را دارا بودند و روندی در بین آن ها مشاهده نمی شود. بنابراین، واریانس باقی مانده ها به عنوان خطای اندازه گیری در نظر گرفته شد.

جدول ۳- خلاصه ای از تجزیه و تحلیل های آماری برخی پارامترهای خاکی اندازه گیری شده به عنوان داده های کمکی در تخمین مکانی ماده آلی خاک

Table 3. Summary of statistical analysis of measured soil properties as axillary data in soil organic matter spatial prediction

Soil property	Mean	SD	Min	Max	Skewness	Kurtosis
EC (dS m <sup>-1</sup> )	10.95	20.48	0.33	107.5	2.94	9.01
PH	8.16	0.44	7.10	8.97	-0.5	-0.24
Na <sub>eq</sub> (meq.l <sup>-1</sup> )	29.58	67.60	0.92	0.44	26.34	4.72
(Ca+Mg) <sub>eq</sub> (meq.l <sup>-1</sup> )	24.58	16.19	4.00	108.00	11.28	2.88
SAR بدون واحد	5.99	8.82	0.23	48.25	16.02	3.67
CCE (%)	18.57	3.91	12.64	30.58	0.72	0.95
WAS (%)	16.41	20.26	0.00	87.11	3.24	1.83
Sand (%)	61.55	18.19	14.00	90.00	2.41	1.39
Silt (%)	26.70	13.03	6.00	56.00	-0.65	0.39
Clay (%)	11.75	7.12	2.00	36.00	-0.34	-0.66

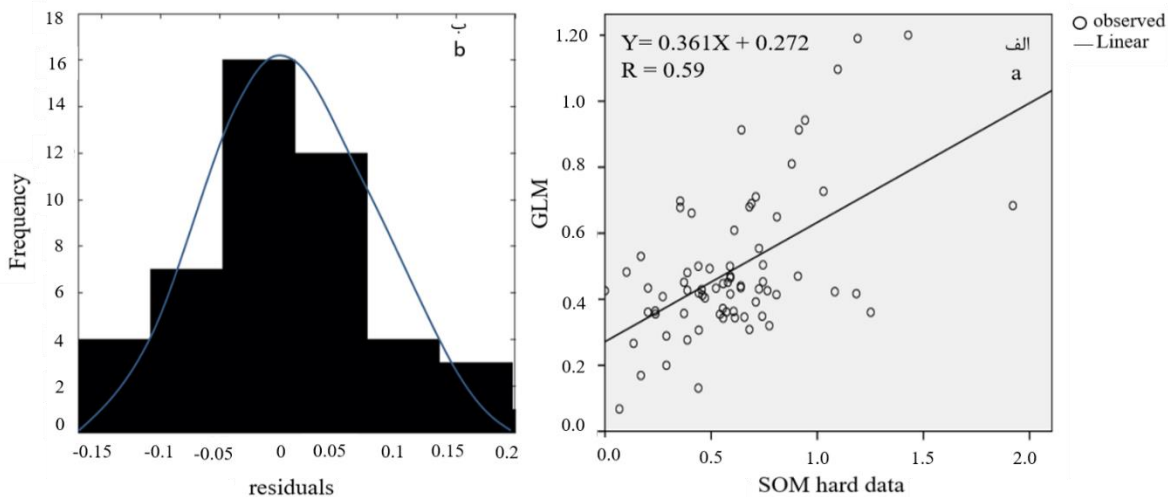
EC: electrical conductivity; SAR: sodium adsorption ratio; CCE: calcium carbonate equivalent; WAS: wet aggregate stability

جدول ۴- همبستگی بین ماده آلی خاک و برخی از پارامترهای خاکی اندازه گیری شده

Table 4. The correlation between soil organic matter and some of the studied soil properties

		EC	pH	Na	(Ca+Mg)	SAR	CCE	WAS	Clay	Silt	Sand
SOM	Correlation Coefficient	0.027	-0.058	0.04	0.18	-0.03	**0.51	**0.33	*0.27	**0.44	**0.42

SOM: soil organic matter



شکل ۵- الف: همبستگی بین ماده آلی خاک اندازه گیری شده و مقادیر تخمینی آن از رگرسیون خطی تعمیمی (GLM). دوایر نشان دهنده مقادیر اندازه گیری شده و خط، مدل برازش شده است؛ ب: هیستوگرام تفاضل داده های مطمئن و نامطمئن و تابع تراکم نهایی برازش شده بر روی آن ها

Figure 5. (a) The correlation between SOM measurements and its estimated values from GLM. Dots correspond to the measured values while solid line is the corresponding the fitted model; (b) Histograms of the residuals between hard and soft data and fitted probability density function (pdf)

کمکی و مدل خطی، منجر به بهبود تخمین مکانی ماده آلی شد. به طوری که MSE، RMSE و nRMSE از ۰/۳۴، ۰/۵۸ و ۰/۵۵ به ترتیب به ۰/۰۹۵، ۰/۳۱ و ۰/۲۹ کاهش یافتند.

در سالیان اخیر، محققین زیادی سعی در استخراج اطلاعات مختلف خاکی از نقشه‌های خاک و یا تصاویر ماهواره‌ای در برای تخمین مکانی ماده آلی داشته‌اند (Sindayihebura *et al.*, 2017; chillaci *et al.*, 2017; Stevens *et al.*, 2015; 2012). با توجه به اینکه هر نوع اطلاعات استخراجی به این روش به همراه خطا است، تلاش‌های زیادی در راستای کالیبره کردن این داده‌ها به منظور افزایش دقت استفاده از آن‌ها صورت گرفته است. به‌عنوان مثال، میرزایی و همکاران (Mirzaee *et al.*, 2016) از داده‌های به‌دست‌آمده از تصاویر ماهواره‌ای برای تخمین مکانی ماده آلی خاک استفاده کردند و از روش‌های مختلفی در راستای افزایش دقت داده‌های به‌دست‌آمده استفاده کردند. شیلاچی و همکاران (Schillaci *et al.*, 2017) نیز از داده‌های کمکی مستخرج از نقشه‌های توپوگرافی، کاربری اراضی و بافت خاک به همراه تصاویر ماهواره‌ای برای تخمین مکانی-زمانی ماده آلی استفاده کردند. در این مطالعات با وجود استفاده از روش‌های آماری متعدد برای کالیبراسیون داده‌ها و کاهش عدم قطعیت استفاده از داده‌های نامطمئن، خطای ناشی از داده‌های نامطمئن در تخمین ماده آلی خاک لحاظ نشده است. لذا پژوهش حاضر قدمی رو به جلو در جهت بهبود استفاده از داده‌های ارزان و در دسترس برای تخمین سایر متغیرهایی است که اطلاعات کافی از آن‌ها در دست نیست.

در شکل ۵، ب همچنین pdf برازش شده بر روی باقی-مانده‌ها نشان داده شده است. این pdf از توزیع گوسی برخوردار است و توزیع یکسان خطا را در بین داده‌های نامطمئن نشان می‌دهد. این تابع در هر نقطه‌ای که داده‌ای نامطمئن در تخمین نقاط اطراف دخالت داشت، تعریف شد.

### ج- تخمین ماده آلی خاک با روش آنتروپی حداکثر اریب و داده‌های کمکی خاکی

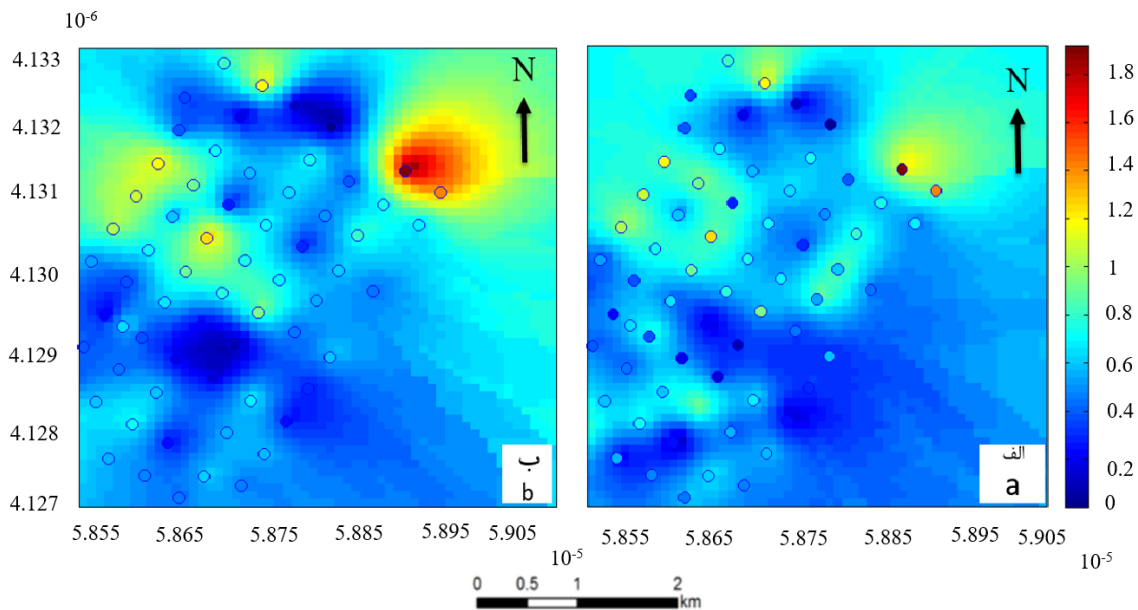
برای تخمین ماده آلی خاک با روش BME و داده‌های نامطمئن احتمالی توسعه یافته با استفاده از مدل رگرسیون خطی تعمیمی، ابتدا یک شبکه با فواصل ۱۰۰ متر تعریف شد. سپس تخمین با استفاده از معادله ۱۷ صورت گرفت. در شکل ۶، نقشه تخمین مکانی ماده آلی با استفاده از داده‌های نامطمئن توسعه یافته با پارامترهای خاکی به همراه نقاط اعتبارسنجی که در آن‌ها اندازه‌گیری‌های ماده آلی به صورت داده‌های مطمئن موجود بودند، ارائه شده است. تفاوت در طیف رنگی نقاط بارنگ پستی، به معنای خطای تخمین است. در شکل ۶، الف، تخمین مکانی ماده آلی با استفاده از مدل توسعه یافته با استفاده از داده‌های کمکی و بدون در نظر گرفتن خطا ارائه شده است. همان‌طور که از این شکل قابل برداشت است، اختلافات مشخصی بین نقاط اعتبارسنجی و نقشه تخمینی وجود دارد. نتایج مربوط به اعتبارسنجی این نقشه در جدول ۵ ارائه شده است. مقایسه این نتایج با نتایج به‌دست آمده از اعتبارسنجی نقشه تخمین مکانی ماده آلی با داده‌های مشابه اما با در نظر گرفتن خطا (شکل ۶، ب) نشان داد که استفاده از خطای محاسبه شده حاصل از استفاده از داده‌های

جدول ۵- مقادیر کمی روش‌های مقایسه اعتبارسنجی دقت BME با استفاده از داده‌های نامطمئن به دست آمده از مدل رگرسیون

#### خطی تعمیمی با و بدون خطا

Table 5. Quantitative values of the validation methods of BME accuracy with soft data resulted from GLM with and without error.

Comparison criteria	Soft data	soft data with error
R	0.65	0.85
MSE (%)	0.34	0.095
RMSE (%)	0.58	0.31
nRMSE	0.55	0.29



شکل ۶- تخمین ماده آلی خاک با استفاده از داده‌های نامطمئن تخمینی با داده‌های کمکی خاکی و مدل رگرسیون خطی تعمیمی و روش آنتروپی حداکثر اریب در دشت بناب. الف: با استفاده از داده‌های نامطمئن؛ ب: با استفاده از داده‌های نامطمئن و خطای تخمین. شباهت بین رنگ نقاط و رنگ زمینه نشانگر دقت تخمین است

Figure 6. SOM prediction using estimated soft data with axillary soil data and generalized linear model and BME method in Bonab Plain. (a) using only soft data; (b) using soft data and error. Proximity between colours for the circles and the background map is thus an indication of the prediction quality

داده‌های ارزان قیمت خاکی و روش آنتروپی حداکثر اریب می‌توانند در کنار کاهش هزینه‌های مطالعات، از دقت کافی در تخمین متغیرهای پیچیده نیز برخوردار باشد.

#### سپاسگزاری

این پژوهش با حمایت مالی مدیریت پژوهشی دانشگاه مراغه با شماره مصوب ۹۵/د/۵۰۷۸ انجام یافته است و بدین وسیله تشکر و قدردانی می‌شود.

#### نتیجه‌گیری کلی

به‌عنوان روش آنتروپی حداکثر اریب که امکان وارد کردن خطا در محاسبات تخمین مکانی را فراهم کرده، روشی کارآمد در تخمین مکانی متغیری مثل ماده آلی خاک در نقاطی است که داده‌های دقیق مربوط به اندازه‌گیری آن فراهم نیست. در این پژوهش سعی شد تا چهار چوبی برای استفاده از روش آنتروپی حداکثر اریب و داده‌های خاکی به‌منظور تخمین متغیر خاکی دیگری همانند ماده آلی ارائه شود. با توجه به اینکه شرط توزیع نرمال داده‌ها نیز در این روش ضروری نیست، استفاده از

#### Reference

- Acosta JA., Faz A., Jansen B., Kalbitz K., and Martinez-Martinez S. 2011. Assessment of salinity status in intensively cultivated soils under semiarid climate, Murcia, SE Spain. *Journal of Arid Environments*, 75: 1056-1066.
- Bogaert P., and D'Or D. 2002. Estimating soil properties from thematic soil maps: The Bayesian Maximum Entropy. *Soil Science Society of American Journal*, 66: 1492-1500.
- Bruun TB., Elberling B., Neergaard A., and Magid AJ. 2015. Organic carbon dynamics in different soil types after conversion of forest to agriculture. *Land Degradation and Development*, 26: 272-283.
- Brus DJ., Bogaert P., and Heuvelink GBM. 2008. Bayesian maximum entropy prediction of soil categories using a traditional soil map as soft information. *European Journal of Soil Science*, 59(2): 166-177.

- Christakos G. 1990. A Bayesian / maximum –entropy view to the spatial estimation problem. *Mathematical Geological*, 22(7): 763-777.
- Christakos G. 2000. Modern spatiotemporal geostatistics. Oxford University Press, New York. 312pp.
- Christakos G. 2002. On the assimilation of uncertain physical knowledge bases: Bayesian and non-Bayesian techniques. *Advances in Water Resources*, 25(8-12): 1257-1274.
- Douaik A., Van Meirvenne M., and Toth T. 2004. Spatio-temporal kriging of soil salinity rescaled from bulk soil electrical conductivity. In: Sanchez Vila X., Carrera J. and Gomez-Hernandez J. (Ed.), *GeoEnv IV: Geostatistics for Environmental Applications*. Netherlands: Kluwer Academic Publishers.
- Dono G., Cortignani R., Dell D., Deligios P., Doro L., Lacetera N., Mula L., Pasqui M., Quaresima S., Vitali A., and Roggero PP. 2016. Winners and losers from climate change in agriculture: insights from case study in Mediterranean basin. *Agricultural systems*, 147: 65-75.
- Forth H. 1990. Fundamentals of soil science. 8<sup>th</sup> Ed. New York: Wiley. ISBN: 0-471-52279-1.
- Giordano R., Liersch S., Vurro M., and Hirsch D. 2010. Integrating local and technical knowledge to support soil salinity monitoring in the Amudarya river basin. *Journal of Environmental Management*, 91: 1718-1729.
- Gee GW., and Bauder JW. 1986. Particle size analysis. In: Klute A. (Ed.), *Methods of Soil Analysis. Part I. Physical and Mineralogical Methods*, 2nd ed., Agronomy Monograph. No: 9. ASA and SSSA. Madison, WI.
- Hamzehpour N., and Bogaert P. 2017. Improved spatiotemporal monitoring of soil salinity using filtered kriging with measurement errors: an application to the West Urmia Lake, Iran. *Geoderma*, 275: 22-33.
- Hamzehpour N., and Eghbal MK. 2016. Spatiotemporal prediction of soil salinity boundary using kriging with measurement errors. *Journal of Ecology, Environment and Conservation*, 22(3): 57-67.
- Hamzehpour N., Eghbal MK., Bogaert P., Toomanian N., and Oskoui RS. 2013. Spatial prediction of soil salinity using kriging with measurement errors and probabilistic soft data. *Arid Land Research and Management*, 27 (2): 128-139.
- Heuvelink GBM. 1998. *Error Propagation in Environmental Modeling with GIS*, London: Taylor and Francis.
- Jaynes ET. 1982. On the rationale of Maximum-Entropy methods. *Proceeding of the IEEE*, 70: 939-952.
- Jegajeevagan K., Sleutel S., Ameloot N., Kader MA., and De Neve S. 2013. Organic matter fractions and N mineralization in vegetable-cropped sandy soils. *Soil Use and Management*, 29(3): 333-343.
- Kempen B., Heuvelink GBM., Brus DJ., and Stoorvogel JJ. 2010. Pedometric mapping of soil organic matter using a soil map with quantified uncertainty. *European Journal of Soil Science*, 61: 333-347.
- Li Y., Shi Z., and Li F. 2007. Delineation of site-specific management zones based on temporal and spatial variability of soil electrical conductivity. *Pedosphere*, 17 (2): 156–164.
- Liu Sh., An N., Yang J., Dong Sh., Wang C., and Yin Y. 2017. Prediction of soil organic matter variability associated with different land use types in mountainous landscape in southwestern Yunnan province, China. *Catena*, 133: 137-144.
- Marchant P., Villanneau J., Arrouays D., Sabyn P. A., and Rawlins BG. 2015. Quantifying and mapping topsoil inorganic carbon concentrations and stocks: approaches tested in France. *Soil Use and Management*, 31: 29-38.
- Marlet S., Bouksila F., and Bahri A. 2009. Water and salt balance at irrigation scheme scale: A comprehensive approach for salinity assessment in a Saharan oasis. *Agricultural Water Management*, 96: 1311-1322.
- Meersmans J., Martin MP., Ridder FD., Lacarce E., Wetterlind J., Baets S. D., Bas CL, Louis BP., Orton TG., Bispo A., and Arrouays D. 2012. A novel soil organic C model using climate, soil type and management data at national scale in France. *Agronomy and Sustainable Developments*, 32: 873-888.

- Mirzaee S., Ghorbani-Dashtaki S., Mohammadi J., Asadi H., and Asadzadeh F. 2016. Spatial variability of soil organic matter using remote sensing data. *Catena*, 145: 118-127.
- Nelson RE. 1982. Carbonate and Gypsum. P. 181-196. In: Page A.L. (Ed.), *Methods of Soil Analysis. Part 2. Chemical and Microbiological Properties*, 2<sup>nd</sup> Edition, Agronomy Monograph. No: 9. ASA and SSSA. Madison, WI.
- Novara A., Gristina L., Sala G., Galati A., Crescimanno M., Cerda A., Badalamenti E., La Mantia T. 2017. Agricultural land abandonment in Mediterranean environment provides ecosystem services via soil carbon sequestration. *Science of Total Environment*, 574: 420-429.
- Saia., S, Benitez E., Garcia-Garrido JM., Settanni L., Amato G., Giambalvo D. 2014. The effect of arbuscular mycorrhizal fungi on total plant nitrogen uptake and nitrogen recovery from soil organic material. *Journal of Agricultural Science*, 152: 370-378.
- Schillaci C., Acutis M., Lombardo L., Lipani A., Fantappie M., Marker M., and Saia S. 2017. Spatio-temporal topsoil organic carbon mapping of a semi-arid Mediterranean region: the role of land use, soil texture, topographic indices and the influence of the remote sensing data to modeling. *Science of the Total Environment*, 601-602: 821-832.
- Serre ML., and Christakos G. 1999. Modern geostatistics: computational BME in the light of uncertain physical knowledge-the Equus Beds study. *Stochastic Environmental Research and Risk Assessment*, 13: 1-26.
- Shanon, CE. 1948. A mathematical theory of communication. *Bell System Technical Journal*, 27: 379-423.
- Stevens F., Bogaert P., and Wesemael BV. 2015. Detecting and quantifying field-related spatial variation of soil organic carbon using mixed-effect models and airborne imagery. *Geoderma*, 259-260: 93-103.
- Stevens A., Miralles I., van Wesemael B. 2012. Soil organic carbon predictions by air-borne imaging spectroscopy: comparing cross-validation and validation. *Soil Science Society of American Journal*, 76: 2174-2183.
- Sullivan DG., Shaw JN., and Rickman D. 2005. IKONOS imagery to estimate surface soil property variability in two Alabama physiographies. *Soil Science Society of American Journal*, 170: 954-968.
- Triantafilis J., Odeh IOA., and McBratney AB. 2001. Five geostatistical models to predict soil salinity from electromagnetic induction data across irrigated cotton. *Soil Science Society of American Journal*, 65: 869-878.
- Walkley A., and Blackm IA. 1934. An examination of the Degtjareff method for determining organic carbon in soils: Effect of variations in digestion conditions and of inorganic soil constituents. *Soil Science*, 63: 251-263.
- Wang T., Kang F., Cheng X., Han H., Bai Y., and Ma J. 2017. Spatial variability of organic carbon and total nitrogen in the soils of a subalpine forested catchment at Mt. Taiyue, China. *Catena*, 155: 41-52.
- Wu C., Wu J., Luo Y., Zhang L., and DeGloria SD. 2009. Spatial prediction of soil organic matter content using cokriging with remotely sensed data. *Soil Science Society of American Journal*, 73: 1202-1208.
- Zaouche M., Bel L., and Vaudour E. 2017. Geostatistical mapping of topsoil organic carbon and uncertainty assessment in Western Paris croplands (France). *Geoderma Regional*, 10: 126-137.
- Zhang S., Huang Y., Shen C., Ye H., and Du Y. 2012. Spatial prediction of soil organic matter using terrain indices and categorical variables as auxiliary information. *Geoderma*, 171-172: 35-43.
- Zhang R., Shouse P., and Yates S. 1997. Use of pseudo-crossvariograms and cokriging to improve estimates of soil solute concentrations. *Soil Science Society of American Journal*, 61: 1342-1347.



## Spatial Prediction of Soil Organic Matter with Soft and Axillary Data Using Bayesian Maximum Entropy Method

Nikou Hamzhepour<sup>1\*</sup>, Sara Mola Ali Abasiyan<sup>2</sup>

Received: May 2018 Accepted: October 2018)

### Abstract

Soil organic matter (SOM) is one of the important soil quality factors and knowledge of its condition in soil, is one of the most important steps in the management of land resources and controlling soil losses. As SOM monitoring is an expensive and time-consuming task, any method which can produce high quality maps of SOM with available axillary soil data and less samples, would be a step forward in reaching the goals of sustainable agriculture. The aim of this research is to predict SOM using soft data, auxiliary data and Bayesian maximum entropy method (BME). Soil samples were gathered from the Bonab-Miandoab plain, and almost 122 samples were collected from 0-20 cm depth of surface soil. SOM and some other soil properties including soil texture, aggregate stability, and calcium carbonate equivalent were measured. Later spatial prediction of SOM was done using SOM soft data, auxiliary data and generalized linear model (GLM) using BME method. Results showed that the highest R, lowest RMSE and nRMSE with values of 0.97, 0.07 and 0.12 respectively, belonged to spatial prediction of SOM with soft data and error. Results also revealed that the developed GLM model with calculated error, resulted in better R, RMSE and nRMSE in comparison to predictions with GLM model without error (R, RMSE and nRMSE improved from 0.65, 0.58 and 0.55 to 0.85, 0.31 and 0.29 respectively). As a conclusion, BME method has provided the possibility of merging error resulted from the use of soft data, in spatial prediction equations and through that, has helped to improve spatial prediction of SOM.

**Keywords:** Auxiliary Data, Error, Modified Walkley and Black Method, Soil Texture

Hamzhepour N., Mola Ali Abasiyan S. 2019. Spatial prediction of soil organic matter with soft and axillary data using bayesian maximum entropy method. *Applied Soil Research*.7(4): 18-34.

1. Assistant professor, Soil Science Department, Faculty of agriculture, University of Maragheh

2. Assistant professor, Soil Science Department, Faculty of agriculture, University of Maragheh

\*. Corresponding Author Email: [Nhamzhepour@maragheh.ac.ir](mailto:Nhamzhepour@maragheh.ac.ir)