

ارزیابی روش‌های یادگیری ماشین در ریزمقیاس نمایی مکانی میانگین سالانه دمای سطح زمین و دمای هوا

آزاده عتباتی^{۱*}، حامد ادب^۲

^۱گروه علوم و مهندسی محیط زیست، دانشکده جغرافیا و علوم محیطی، دانشگاه حکیم سبزواری، سبزوار، ایران
^۲گروه سنجش از دور و سیستم اطلاعات جغرافیایی، دانشکده جغرافیا و علوم محیطی، دانشگاه حکیم سبزواری، سبزوار، ایران

تاریخ پذیرش: ۱۴۰۱/۰۳/۰۱

تاریخ دریافت: ۱۴۰۱/۰۱/۰۶

چکیده

امروزه استفاده از داده‌های شبکه‌ای پایگاه‌های اقلیمی مانند WorldClim یکی از منابع معتبر داده است که جایگزین داده‌های نقطه‌ای ایستگاه‌های هواشناسی شده است؛ اما استفاده از این پایگاه‌های اقلیمی با قدرت تفکیک مکانی پایین موجب ایجاد محدودیت برای بسیاری از مطالعات مرتبط با علوم زیست‌شناسی و بوم‌شناسی شده است. هدف از این پژوهش بررسی ارتباط دمای هوا و دمای سطح زمین و سپس بازتولید دمای سطح زمین با قدرت تفکیک مکانی بالا جهت ریزمقیاس نمایی میانگین سالانه دمای هوا با استفاده از دو محصول پرکاربرد میانگین سالانه دمای هوا از پایگاه داده WorldClim و میانگین سالانه دمای روز و شب سطح زمین MOD11A2 v061 سنجنده مادیس است. در این پژوهش، ابتدا عملکرد مدل‌های یادگیری ماشین شامل جنگل تصادفی، شبکه عصبی مصنوعی، رگرسیون شبکه الاستیک و ماشین بردار پشتیبان جهت ریزمقیاس نمایی محصول MOD11A2 v061 از ۱ کیلومتر به ۲۵۰ متر بررسی شد. برای این منظور از متغیرهای پیوسته و گسسته شامل ارتفاع از سطح دریا، عرض جغرافیایی، پوشش گیاهی، بافت خاک، جهت شیب و پوشش سطح زمین استفاده گردید. سپس میانگین سالانه دمای هوا WorldClim با استفاده از دمای سطح زمین با مدل پولی نومیال درجه ۳ از ۱ کیلومتر به ۲۵۰ متر ریزمقیاس شد. همچنین از داده‌های ۷ ایستگاه سینوپتیک جهت بررسی اعتبار محصول ریزمقیاس شده استفاده شد. نتایج نمودار تیپور نشان داد مدل جنگل تصادفی، بهترین عملکرد در ریزمقیاس نمایی میانگین سالانه دمای سطح زمین با ریشه میانگین مربعات خطا ۰/۵۴ درجه سلسیوس دارد. همچنین مدل پولی نومیال درجه ۳ میزان خطای نسبی کمتر در تولید داده ریزمقیاس دمای هوا دارد. مقدار ریشه میانگین مربعات خطا نتایج برای مدل تصحیح نشده و تصحیح شده ریزمقیاس به ترتیب ۱/۳۲ و ۱/۲۱ درجه سلسیوس به دست آمد که با توجه به آزمون t جفتی اختلاف معنی‌داری در سطح ۰/۰۵ نشان نداد. یافته‌های این پژوهش نشان می‌دهد که ریزمقیاس نمایی میانگین سالانه دمای هوا WorldClim از اعتبار لازم برخوردار است.

کلید واژگان: پایگاه اقلیمی WorldClim، داده مادیس، ریزمقیاس نمایی مکانی، یادگیری ماشین

مقدمه

مسئله‌ساز است، جایی که یک سلول از شبکه با قدرت تفکیک مکانی پایین (۱ کیلومتر) همچنان شرایط آب و هوایی یکسان را علی‌رغم چند صد متر اختلاف ارتفاع داخل آن سلول را نشان می‌دهد (Wang *et al.*, 2016).

باتوجه به مسئله ذکر شده، لزوم بهبود قدرت تفکیک مکانی پایگاه داده‌های آب و هوایی موجود برای مطالعات اکولوژیک مطرح می‌شود. که جزئیات مکانی متغیرهای اقلیمی با بکار بردن تعدادی پیش‌بینی‌کننده افزایش داده می‌شود. با توجه به قدرت تفکیک مکانی پایین پایگاه‌های داده‌های آب و هوایی شبکه‌ای مانند WorldClim که برای مطالعات قاره‌ای و جهانی مناسب است، اما در برخی موارد این مجموعه داده‌ها باید به مقیاس‌های ناحیه‌ای جهت کاربردهای تحقیقات محیط زیستی که مبتنی بر نمونه است، تبدیل شوند. از طرفی بهبود قدرت تفکیک مکانی این نوع داده‌ها برای مکان‌های خاص از مجموعه داده‌های بزرگ فرآیند ساده‌ای نیست و به‌طور معمول به مهارت‌های پیشرفته مدل‌سازی آماری و مدیریت داده نیاز دارد (Wang *et al.*, 2016).

رویکردهای زمین‌آماری مانند کریجینگ از توابع واریوگرام برای در نظر گرفتن همبستگی فضایی داده‌های تجربی برای تولید داده و سپس اعمال روش نمونه‌گیری مجدد جهت ریزمقیاس‌نمایی استفاده می‌شود، از طرفی در این روش‌ها یافتن ارتباط آماری بین متغیرها در نظر گرفته نمی‌شود (Collados-Lara *et al.*, 2021) از این‌رو، در مقام مقایسه با روش‌های یادگیری ماشین، از موفقیت کمتری برخوردار هستند. همچنین از روش‌هایی مانند کوکریجینگ با استفاده از داده‌های ثانویه برای برآورد متغیرهای آب و هوایی استفاده شده است (Collados-Lara *et al.*, 2021) اما این روش هم محدودیت در تعداد متغیر مستقل مورد استفاده دارد و امکان استفاده از داده‌های کیفی و گسسته در این نوع روش نیز وجود ندارد. بنابراین رویکرد مبتنی بر یادگیری ماشینی به‌عنوان روش تولید داده‌های ریزمقیاس در این پژوهش انتخاب گردید.

استفاده از داده‌های آب و هوایی با قدرت تفکیک مکانی بالا کاربرد زیادی در علوم زیست‌شناسی و بوم‌شناسی دارد (Chalghaf *et al.* 2018, Mwakapeje *et al.* 2019). در واقع متغیرهای زیست‌اقلیمی به‌عنوان یک نوع داده مهم در مدل‌سازی توزیع گونه‌های جانوری و گیاهی در نظر گرفته می‌شوند (Amiri *et al.*, 2020; Pineda and Lobo, 2009) که برای درک روابط بین آب‌وهوا و واکنش زیستی موجودات زنده یا الگوهای کلی سازگاری محیط زیستی با محیط‌های آب و هوایی محلی ضروری است (Wang *et al.* 2016). برای مدل‌سازی اکولوژیک، استفاده از داده‌های آب و هوایی موردنیاز است تا شرایط اقلیمی را تا آنجا که ممکن است نزدیک به محل زیستگاه موجودات زنده نشان دهد (Hamann and Wang, 2006; Rehfeldt *et al.*, 2012). در کاربردهای زیست‌شناسی و بوم‌شناسی، نمایش تنوع زمانی و مکانی دما و بارش، برای استنباط زیستگاه، فصول رشد، مهاجرت گونه‌ها یا توزیع گونه‌های در مقیاس با جزئیات زیاد اهمیت دارد. با توجه به اینکه دریافت داده‌های ایستگاه‌های هواشناسی در برخی موارد هزینه‌بر و انجام فرآیندهای آماری مانند کنترل کیفی، بازسازی و همگنی داده‌های هواشناسی، تطویل سری زمانی ایستگاه‌ها و تهیه نقشه‌های شبکه‌ای زمان‌بر و مسئله‌ساز است، استفاده از محصولات آماده اقلیمی پایگاه‌های داده‌های آب و هوایی شبکه‌ای مانند WorldClim (Hijmans *et al.* 2005) یا CHELSA (Karger *et al.*, 2017) مطرح می‌شود. بسیاری از مطالعات بوم‌شناسی با استفاده از داده‌های پایگاه اقلیمی با قدرت تفکیک مکانی ۱ کیلومتر مربع انجام شده است (Pratumchart *et al.*, 2019; Mwakapeje *et al.*, 2019). اما خطاهای موجود در مجموعه بانک داده‌های آب و هوایی مانند WorldClim که اساس مطالعات بوم‌شناسی است در مقیاس با جزئیات زیاد موجب بروز تورم خطا شود (Soria-Auza *et al.*, 2010)، چنین خطاهایی به‌ویژه در مناطق کوهستانی با تنوع توپوگرافی زیاد

غیرواقعی در ماه‌های خاص یا در مناطق خاص شود، که می‌تواند به شدت بر پارامترهای مدل رگرسیون تأثیر گذار باشد (Karger *et al.*, 2017). هنگامی که عوامل خاص، مانند دامنه‌های رو به باد یا پشت به باد در طول سال تغییر می‌کند، اما مکان ثبت داده اقلیمی تغییر نمی‌کند، ریزمقیاس نمایی آماری مبتنی بر رگرسیون، به‌عنوان مثال، ممکن است یک رابطه منفی قابل توجه بین یک ایستگاه در محل بادگیر کوه به مدت یک ماه و یک رابطه مثبت برای دیگری را تشخیص دهد، اگرچه فیزیک جوی همیشه یک رابطه مثبت را پیش‌بینی می‌کند (Karger *et al.*, 2017). با توجه به این مسئله، از روش‌های ریزمقیاس نمایی مبتنی بر یادگیری ماشین و درون‌یابی اغلب برای مناطق منفرد استفاده شده است (Daly and Taylor, 1998)، درحالی‌که مدل جهانی آن وجود ندارد. با توجه به اینکه در مناطقی داده‌های ایستگاه‌های هواشناسی از طول دوره آماری کافی برخوردار نیستند و همچنین ممکن است ایستگاه‌های هواشناسی به دلیل مسائل مرتبط با احداث در مکان‌هایی قرار داشته باشند که نمای کلی از چشم‌انداز را ارائه ندهند (Lookingbill and Urban, 2003). از اینرو با توجه به مسئله بیان شده، اهداف اصلی تحقیق (۱) افزایش قدرت تفکیک مکانی محصول میانگین سالانه دمای سطح زمین سنجنده مادیس با استفاده از روش‌های یادگیری ماشین (۲) افزایش جزئیات مکانی محصول میانگین سالانه دمای هوا پایگاه اقلیمی WorldClim با استفاده از میانگین سالانه دمای سطح زمین و مقایسه آن با ایستگاه‌های هواشناسی است. این افزایش در قدرت تفکیک مکانی میانگین سالانه دمای هوا می‌تواند در مطالعات زیست‌شناسی و اکولوژی در مقیاس ناحیه‌ای مورد استفاده قرار گیرد.

مواد و روش‌ها

منطقه مورد مطالعه: با توجه به اهداف اصلی تحقیق، استان خراسان شمالی به دلیل تنوع توپوگرافی، پوشش زمین و اقلیمی و همچنین شرایط متفاوت پوشش‌های گیاهی و

داده میانگین دمای هوای سالانه در پایگاه‌های داده‌های آب و هوایی شبکه‌ای مانند WorldClim بر پایه متغیرهای کمکی مانند ارتفاع، دمای روز و شب سطح زمین (سنجنده مادیس)، فاصله از پهنه‌های آبی انجام شده است (Fick and Hijmans, 2017). در مطالعاتی از روش جنگل تصادفی جهت ریزمقیاس نمایی دمای سطح زمین با استفاده از متغیرهای کمکی بازتاب سطحی، متغیرهای مشتق شده از توپوگرافی و پوشش اراضی از ۱ کیلومتر تا ۲۵۰ متر استفاده شد (Hutengs and Vohland, 2016; BartkowiakCastelli and Notarnicola, 2019; Ebrahimi *et al.*, 2021) و با توجه به ارتباط دمای سطح زمین و دمای هوا، مدل برآورد دمای هوا به دست آمد (Kloog *et al.*, 2014; Vancutsem *et al.*, 2010). مطالعه‌ای Moreno و Hasenauer (۲۰۱۶) به ریزمقیاس نمایی داده‌های E-OBS با استفاده از داده‌های WorldClim پرداختند. روش‌های نمونه‌گیری مجدد مانند روش نزدیک‌ترین همسایه جهت ریزمقیاس نمایی استفاده می‌شود (HanYan and Ling, 2021) اما این روش‌ها تنها باعث کوچک‌تر شدن ابعاد پیکسل می‌شود بدون آنکه منجر به افزایش جزئیات مکانی دما شود.

ریزمقیاس نمایی آماری، به‌ویژه در مقیاس جهانی، به دلیل تغییرات زمانی در توزیع مکانی ایستگاه‌های هواشناسی مسئله‌ساز است. اگرچه اندازه‌گیری‌ها برای یک پیش‌بینی کننده معین ممکن است در یک ماه معین در دسترس باشد، اما ممکن است در ماه دیگری وجود نداشته باشد، که منجر به ناهمگونی کلی در سری زمانی ایستگاه‌ها در زمان محاسبه متغیرها اقلیمی می‌شود (Karger *et al.*, 2017). ریزمقیاس آماری هنگامی مشکل‌ساز می‌شود که پیش‌بینی کننده‌های بسیار پویا مانند میدان‌های باد نیاز به استفاده و ترکیب در مدل را داشته باشند. درحالی‌که این امر برای پیش‌بینی کننده‌های ایستا مانند ارتفاع، شیب یا جهت شیب توپوگرافی تأثیر نمی‌گذارد. ناهمگونی در توزیع زمانی و مکانی چنین عوامل پویا نیز می‌تواند منجر به همبستگی



شکل ۱- موقعیت استان مورد مطالعه، ایستگاه‌های سینوپتیک مورد استفاده و تأثیرپذیری اقلیم بزرگ‌مقیاس استان از توده‌های هوای مجاور

طی فرآیند ریزمقیاس می‌تواند به ارائه نتایج غیرواقعی توسط مدل منجر گردد. از این رو لازم است نتایج ریزمقیاس نمایی بانک داده‌ها اقلیمی با ایستگاه‌ها هواشناسی ارزیابی و در صورت نیاز اعمال تصحیحات روی داده ریزمقیاس نمایی شده انجام گردد. بدین منظور میانگین سالانه دمای هوا در دوره آماری ۲۰۱۷-۱۹۹۸ از ۷ ایستگاه هواشناسی سینوپتیک داخل استان استفاده شد (شکل ۱).

محصول دمای سطح زمین MOD11A2 v061

سنجنده مادیس: دمای سطح زمین یکی از مهم‌ترین محصولات سنجنده مادیس است که می‌توان در مطالعات مختلفی از جمله بوم‌شناختی از آن استفاده نمود. این محصول با شناسه‌های MOD11 برای ماهواره ترا و MYD11 برای ماهواره آکوا قابل استفاده است. این محصول میانگین ۸ روزه دمای روز و شب و ضریب نشر در هر پیکسل سطح زمین با قدرت تفکیک مکانی ۱ کیلومتر با پوشش ۱۲۰۰ در ۱۲۰۰ کیلومتر را ارائه می‌دهد. این دوره ترکیبی ۸ روزه به این دلیل انتخاب شد که دوره دقیق تکرار مسیر زمینی ماهواره ۱۶ روز است (MilitinoUgarte and Montesino, 2019). در این محصول سایر مجموعه داده‌ها شامل ارزیابی‌های کنترل کیفیت، زمان‌های مشاهده، زاویه‌های اوج دید، و پوشش آسمان صاف نیز ارائه می‌شود. محصول MOD11A2 v061 از بهبود و تغییراتی نسبت به نسخه‌های قبلی دمای سطح زمین برخوردار است (WanHook and Hulley, 2006).

نزدیکی به مناطق بیابانی و دریای خزر انتخاب شد (شکل ۱) زیرا این تنوع زیاد و پیچیدگی الگوها بین داده‌ها به مدل‌ها وارد می‌شود و در این شرایط توانمندی مدل‌ها بهتر قابل مقایسه و ارزیابی است.

داده‌های مورد استفاده:

پایگاه داده اقلیمی Worldclim و ایستگاه‌های

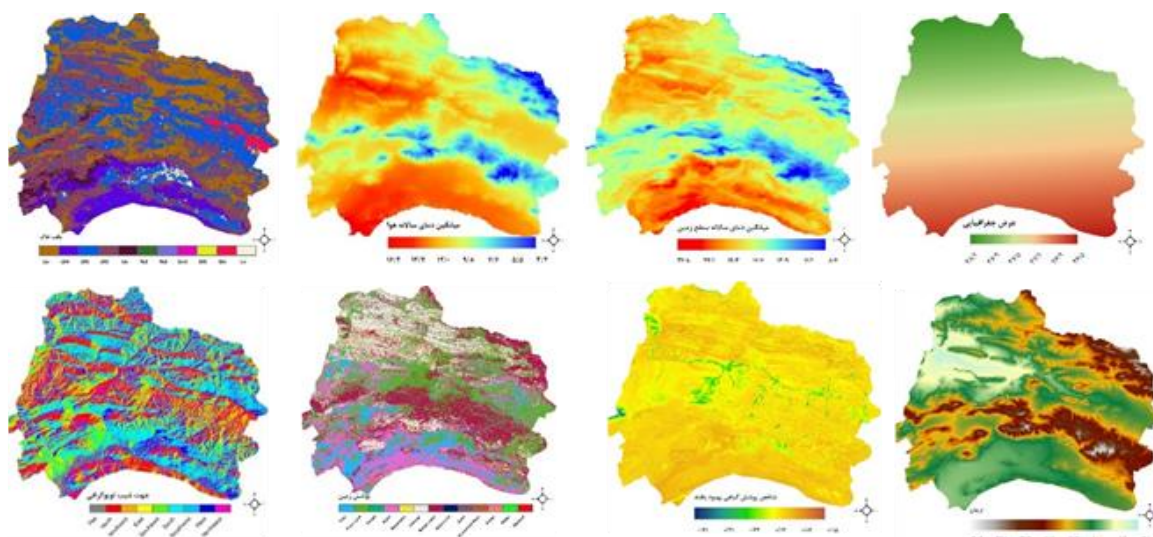
هواشناسی سینوپتیک: پایگاه داده‌های جهانی آب و هوا (Worldclim) در سه نسخه ۱/۴، ۲ و ۲/۱ مجموعه داده‌ها اقلیمی با فرمت شبکه‌ای است. اطلاعات نسخه ۲/۱ آب‌وهوا در پایگاه WorldClim برای دوره زمانی ۱۹۷۰-۲۰۰۰ با استفاده از ۹۰۰۰ تا ۶۰۰۰۰ ایستگاه هواشناسی محاسبه و در ژانویه ۲۰۲۰ منتشر شده است (Fick and Hijmans, 2017). داده‌ها در چهار قدرت تفکیک مکانی بین ۳۰ ثانیه (~ ۱ کیلومترمربع) تا ۱۰ دقیقه (~ ۳۴۰ کیلومترمربع) به صورت رایگان با پوشش جهانی در دسترس است (Hijmans et al., 2005). داده جهانی میانگین دما (wc2.1_30s_tavg) بیشترین کاربرد را در مطالعات زیست‌شناسی و بوم‌شناسی دارد، از این رو، در مطالعه حاضر این داده از نسخه ۲/۱ با قدرت تفکیک مکانی ۳۰ ثانیه از این پایگاه به آدرس <https://www.worldclim.org> دریافت شد. در فرآیند ریزمقیاس نمایی داده‌های اقلیمی از بانک داده‌ها مانند Worldclim، برآورد داده‌ها با خطای کم از اهمیت زیادی برخوردار است زیرا برآورد نادرست داده اقلیمی

جدول ۱- مشخصات متغيرهاى مورداستفاده جهت ريزمقياس نمايى دماى سطح زمين

نام متغير	منبع/روش تهيه	ابعاد پيكسل	نام متغير	منبع/روش تهيه	ابعاد پيكسل
مدل رقومى جهانى ارتفاع	نسخه ۴/۱ SRTM، كنسرسیوم داده‌هاى مكاني CGIAR	۹۰ متر	پوشش سطح زمين عرض	پوشش كشورى سطح زمين تهيه شده توسط (Ghorbanian <i>et al.</i> , 2020)	۱۰ متر
بافت خاك	نسخه ۲ محصول SoilGrids (Hengl <i>et al.</i> 2017)، تهيه شده با SAGA GIS به روش Germany KA5	۲۵۰ متر	جغرافيايى	تبديل نقاط وكتور حاوى داده عرض جغرافيايى به لايه رستري	۲۵۰ متر
جهت شيب توپوگرافى	چند جمله مرتبه دوم (Zevenbergen and Thorne, 1987)	۹۰ متر	ميانگين سالانه دماى هوا	نسخه ۲/۱ worldclim (Hijmans <i>et al.</i> , 2005)	۱ كيلومتر
شاخص پوشش گياهي بهبوديافته	استخراج ميانگين ۱۶ روزه شاخص EVI، محصول MOD13Q1 v061 ، ماديس تراه دوره زمانى ۲۰۲۱-۲۰۰۰، گوگل ارث انجين (Didan, 2021)	۲۵۰ متر	ميانگين سالانه دماى سطح زمين	استخراج ميانگين ۸ روزه دماى روز و شب سطح زمين، محصول MOD11A2 v061، ماديس تراه دوره زمانى ۲۰۲۱-۲۰۰۰، گوگل ارث انجين (WanHook and Hulley, 2021)	۱ كيلومتر

مشخص کرد (Lookingbill and Urban 2003). در مناطق كوهستاني، دماى هواى سطح زمين تحت تأثير شرايط جغرافيايى بسيار پيچيده مانند ارتفاع، عرض جغرافيايى، سطح خاك و پوشش گياهي است (Wang *et al.*, 2011). کاربري زمين با تغيير در تابش سطح زمين، نقش مهمى روى دماى هواى در سطح زمين ايفا مى كند (Baldocchi and Ma, 2013). کاربري زمين تأثير قابل توجهي بر آب و هواى محلي يك منطقه دارد. در سطح محلي، تغيير تعادل انرژي سطحى، به سمت سطح تيره تر و ناهموار، باعث افزايش دما مى شود (Baldocchi and Ma, 2013). از اين رو اين متغير اطلاعاتي را براي تشخيص و پيش بينى چگونگي تغييرات دماى هوا در پي تغيير کاربري يا مديريت زمين در اختيار ما قرار مى دهند. جزئيات بيشتري در مورد تأثير دماى سطحى کاربري زمين بر دماى هوا توسط Ziaul و Pal (۲۰۱۷) بحث شده است. دماى سطح نيز به دليل تغيير زاويه تابش خورشيد با عرض جغرافيايى تغيير مى كند (Wang *et al.*, 2011). در مورد دماى سطح، ارتفاع و عرض جغرافيايى با يكديگر ارتباط تنگاتنگي دارند، تشخيص اينكه ارتفاع يا عرض جغرافيايى کدام به تنهائى تغييرات دما را توضيح مى دهد، دشوار است. به همين دليل از تأثير تركيبى ارتفاع و عرض جغرافيايى بر برآورد دماى هوا استفاده مى شود (Wang *et al.*, 2011). عوامل مورد

(2021).
متغيرهاى پيش بينى كننده: به دست آوردن الگوهاى محيطي با قدرت تفكيك مكاني بالا و در مقياس چشم انداز نيمى تواند به راحتى با استفاده از روش هاى معمولى ميانيابى انجام شود. در بين متغيرهاى اقليمي، تغيير پذيرى دماى هوا كنترل مهمى بر روى بسيارى از فرايندهاى اكلوژيكي دارد (Lookingbill and Urban, 2003). متغيرهاى "پروكسى" دماى هوا مانند ارتفاع، شيب توپوگرافى و جهت شيب توپوگرافى اغلب به عنوان پيش بينى كننده به ويژه در مناطق كوهستاني مورد استفاده قرار مى گيرد، به اين دليل كه اندازه گيرى هاى مستقيم كمى در دسترس است (Lookingbill and Urban, 2003). توپوگرافى مى تواند ارتباط بين ارتفاع و دما را تا حد زيادى تغيير دهد. اثرات توپو-اقليمي اوليه ناشى از تفاوت در زاويه و جهت شيب است (Barry, 1992). اين تأثيرات تا حد زيادى توسط ارتباط جهت گيرى شيب توپوگرافى با تابش خورشيد كنترل مى شود. در نيمكره شمالى، دامنه هاى رو به شمال تابش كمترى نسبت به دامنه هاى رو به جنوب دارند (Lookingbill and Urban, 2003). از اين رو تفاوت در جهت شيب حتى مى تواند بيشتري از تأثير ارتفاع بر دما باشد (McCutchan and Fox, 1986). دامنه ها در خط الرأس ها و خط القعرها مى توان با رژيم هاى دمائي بسيار متفاوت



شکل ۲- نقشه‌های متغیرهای وابسته و مستقل مورد استفاده در پژوهش

دمای سطح زمین مجدد به پایگاه داده وارد شد تا میانگین سالانه دمای هوا برای منطقه مورد مطالعه در فرمت رستری با ابعاد ۲۵۰ مترمربع تهیه گردد.

روش‌های یادگیری ماشینی:

تنظیم روش‌های یادگیری ماشینی: اعتبارسنجی متقابل k-fold مناسب‌ترین روش برای آموزش موفق یک ماشین یادگیر است (Anguita et al., 2012). استفاده از داده‌های آزمون مشخص می‌کند که چگونه مدل یادگیر برای داده‌های خاصی که توسط مدل در فرآیند آموزش استفاده نشده است، موفق عمل می‌کند. قابلیت ماشین یادگیر با اعتبارسنجی متقابل با ۱۰ k-fold اعمال شد تا دقت مدل با ارزیابی شود. اعتبارسنجی متقاطع اطلاعات تخمینی تقریباً بی‌طرفانه را در مورد دقت مدل ارائه می‌کند، زیرا مدل با k-fold آزمایش می‌شود که در توسعه مدل دخالتی ندارد (Breiman et al., 1984). اعتبارسنجی متقاطع K-fold شامل تقسیم تصادفی داده‌ها به ۱۰ نمونه فرعی با اندازه مساوی است. به‌طور مکرر، ۸ زیرمجموعه برای آموزش مدل، ۱ زیرمجموعه به‌عنوان مجموعه داده اعتبارسنجی برای توقف آموزش، و ۱ زیرمجموعه به‌عنوان مجموعه آزمون برای ارزیابی عملکرد مدل استفاده شد. این روش ۱۰ بار برای هر به‌صورت متوالی تکرار شده است، به این معنی که هر یک از ۱۰ زیرمجموعه یک‌بار به‌عنوان یک

استفاده در این تحقیق به‌منظور ریزمقیاس‌نمایی میانگین دمای سالانه سطح زمین در جدول ۱ ارائه شده است.

آماده‌سازی داده‌ها: اندازه سلول خروجی پیش‌بینی‌کننده دمای سطح زمین به روش نمونه‌گیری مجدد نزدیک‌ترین همسایه و بر اساس بزرگ‌ترین ابعاد پیکسل در متغیرهای مستقل (شاخص پوشش گیاهی بهبودیافته) به ۲۵۰ مترمربع تنظیم شد. به‌منظور ایجاد شبکه‌ای از سطر و ستون با فاصله یکسان برای تمامی متغیرهای مستقل، از روش چسبیدن پیکسل‌های داده‌ها متغیرهای مستقل به پیکسل‌های شاخص پوشش گیاهی بهبودیافته استفاده شد تا از تراز سلولی اطمینان حاصل شود. منطقه مورد مطالعه به شبکه‌ای متشکل از ۸۳۶ ردیف در ۱۰۱۴ ستون تشکیل شد (تعداد ۵۸۰۵۴۰ سلول با توجه به مرز استان و بدون در نظر گرفتن پیکسل‌های محدوده چهارگوشه در حوزه منطقه مورد مطالعه). در مجموع با در نظر گرفتن ۶ متغیر مستقل به تعداد ۳۴۸۳۳۴۰ سلول پیش‌بینی ایجاد شد و سپس همه لایه‌ها به نقاط برداری تبدیل شد، تا مقادیر متغیرهای وابسته و مستقل در جدول با فرمت CSV ذخیره گردد. سپس فایل‌های CSV در پایگاه داده‌ها در Python 3.9 با Miniconda3 Windows 64-bit جهت مدل‌سازی استفاده شد (Zupan and Demsar, 2008). بعد از اجرای مدل‌های یادگیری، نتایج پیش‌بینی میانگین دمای سالان Z

مجموعه آزمایشی اعمال شده است (Melchiorre et al., 2011).

روش‌های یادگیری ماشین جهت برآورد میانگین سالانه دمای سطح زمین

شبکه‌های عصبی مصنوعی: پرسپترون چندلایه یک کلاس از شبکه عصبی پیش‌خور (FNN) است که برای این پژوهش انتخاب شده است. جهت پیکربندی شبکه از روش توصیه شده آزمون و خطا استفاده شد، زیرا قانون قطعی برای یافتن پیکربندی بهینه وجود ندارد (MiraboutalebiKazemi and Bahrami, 2016). تنظیمات فرآیندها به صورت ۲، ۳، ۵ نرون ورودی در ۳ لایه پنهان، پارامتر نظم α دهی برابر 0.07 جهت حل مسئله بیش‌برازش داده و حداکثر 200 تکرار معرفی شد. در این مطالعه، تابع فعال‌سازی \tanh برای همه نورون‌های لایه Z پنهان استفاده می‌شود که برابر با \tanh به عنوان تابع فعال‌سازی در لایه خروجی نیز استفاده می‌شود. BFGS یکی از پرکاربردترین الگوریتم‌های کلاسیک در یادگیری ماشین برای آموزش شبکه‌های عصبی است (Đurđević Babić, 2015). روش BFGS تحت نویز قرار نمی‌گیرد و دارای مزایای زیادی نسبت به الگوریتم پس‌انتشار (BP) است، از جمله همگرایی فوق‌خطی، توانایی تقریب و تعمیم خوب، دقت بالا و مقاومت در برابر برازش بیش‌ازحد است (XiaRusli and Kumta, 2010; Bollapragada et al., 2018, Uca et al., 2018). در این مطالعه روش تنظیم‌سازی مورد استفاده قرار گرفت که می‌تواند در فیلتر کردن نویز از داده‌ها استفاده شود و در نهایت با اضافه کردن بایاس به پارامتر وزن‌های حدی جریمه، از برازش بیش‌ازحد جلوگیری کند (Ronao and Cho, 2016).

ماشین‌های بردار پشتیبان: ماشین‌های بردار پشتیبان دسته‌ای از الگوریتم‌های یادگیری نظارت‌شده هستند که از تئوری یادگیری آماری مشتق شده‌اند و اولین بار توسط Vapnik (۱۹۹۸) معرفی شد. این الگوریتم‌ها از یک تابع

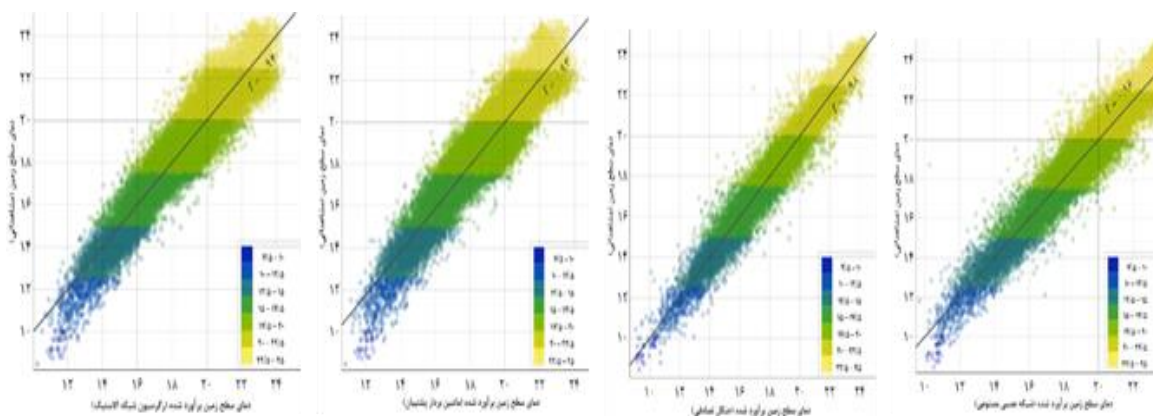
هسته‌ای برای تبدیل داده‌های ورودی استفاده می‌کند و سپس رگرسیون خطی را برای داده‌های تبدیل‌شده اعمال می‌کند (Mohandes et al., 2004). ماشین‌های بردار پشتیبان شامل دو مرحله است، (۱) انتخاب نوع هسته مناسب و تنظیم پارامتر آن (عرض هسته G) و (۲) تعیین پارامتر جریمه (Sun et al., 2011) (C). مجموعه‌ای از متغیر فرآیندها، با بهترین نتایج به دست آمده با هزینه رگرسیون $1/6$ (C)، اپسیلون $2/5$ (E) که فاصله از مقادیر واقعی را که در آن هیچ جریمه‌ای با مقادیر پیش‌بینی شده مرتبط نیست را نشان می‌دهد، تابع هسته‌ای سیگموئید (g) با $0.01, 0.01$ (c)، تلورانس 0.03 و محدودیت تکرار تا 200 انتخاب شده‌اند.

رگرسیون شبکه الاستیک: الگوریتم رگرسیون لسو و ستیغی، در رگرسیون شبکه الاستیک تلفیق می‌شود. در نتیجه قاعده‌سازی مرتبه ۱ (رگرسیون لسو) و ۲ (رگرسیون ستیغ) روی مدل به طور هم‌زمان اجرا می‌شود (Acharjee et al., 2013). قاعده‌سازی یکی از راه‌های جلوگیری از کم‌برازش و بیش‌برازش است. به این معنی که مدل رگرسیونی با توجه به مقدار پارامترهای آن جریمه می‌شود تا به یک مقدار بهینه برسد. به این صورت پیچیدگی مدل کاهش یافته، بدون آنکه از کارایی آن کاسته شود. استفاده از قاعده‌سازی زمانی که تعداد متغیرهای توصیفی زیاد باشد، تعداد متغیرها نسبت به تعداد مشاهدات زیاد باشد و هم خطی یا هم خطی چندگانه در بین متغیرهای توصیفی وجود داشته باشد، مؤثر است. بنابراین، رگرسیون شبکه الاستیک دو پارامتر تنظیم دارد: α کنترل تعادل بین $L1$ (لسو) و $L2$ (ستیغ) و λ کنترل قدرت کلی جریمه (Acharjee et al., 2013)، شدت قاعده‌سازی با افزایش مقدار λ شدت می‌گیرد. در این مطالعه با تنظیمات فرآیندها رگرسیون شبکه الاستیک $\alpha = 0.0002$ و نسبت $L1:L2 = 0.65:0.34$ اتخاذ شد. تنظیم این پارامترهای استفاده از اعتبارسنجی متقاطع k -fold و مجموعه داده‌های آزمایش به دست آمد. برای تعیین مقدار بهینه λ استفاده از آزمایش و خطا است به طوری که MSE در کمترین مقدار

از این رو برای انتخاب مقدار بهینه مدل از معیارهای ارزیابی استفاده می‌شود. خطای مدل رگرسیون چندجمله‌ای با استفاده از داده‌های ایستگاه‌های هواشناسی توسط یک تابع خطی به حداقل رسانده می‌شود و امکان مناسبی را برای تصحیح و سازگاری مدل با داده‌های واقعی فراهم می‌آورد (VernayBlanc and Pitaval, 2013).

ارزیابی و اعتبارسنجی مدل‌ها: به منظور ارزیابی کمی عملکرد مدل‌های توسعه‌یافته جهت برآورد دمای سطح زمین، معیارهای ارزیابی شامل میانگین خطای مطلق (Mean Absolute Error)، خطای جذر میانگین مربعات (Root Mean Square Error)، ضریب تغییرات مجذور میانگین مربعات خطا (Coefficient of the Variation of the Root Mean Square Error)، شاخص توافق (Index of Agreement D) و ضریب کارایی نش ساتکلیف (Nash-Sutcliffe coefficient efficiency) استفاده شده است. همچنین از نمودار تیور جهت مقایسه خروجی‌های مدل‌ها با مقادیر واقعی بر اساس میزان خطای جذر میانگین مربعات، همبستگی و همچنین انحراف معیار استفاده شد (Taylor, 2001). هرچه موقعیت مدل‌ها از نظر سه شاخص ذکرشده روی نمودار به نقطه مرجع نزدیک‌تر باشد، مدل دقت بیشتری دارد. به منظور تعیین اهمیت متغیرهای مستقل در مدل جنگل تصادفی اقدام به سنجش حساسیت مدل شد. به این منظور از نتایج مدل آموزش‌دیده و داده‌های مرجع برای محاسبه سهم هر متغیر مستقل در پیش‌بینی استفاده می‌شود. در هر مرحله یکی از متغیرهای مستقل از مدل حذف می‌گردد و سپس مدل با متغیرهای مستقل باقیمانده اجرا می‌شود. با اندازه‌گیری مقدار افزایش در خطای جذر میانگین مربعات با حذف هر متغیر، میزان حساسیت مدل در پیش‌بینی مقادیر به دست آمد. برای مقایسه نتایج حاصل از مدل برآورد میانگین سالانه دمای هوا با ایستگاه‌های هواشناسی، از آزمون t جفتی در سطح معنی‌داری ۰/۰۵ استفاده شد که در آن مقادیر واقعی ثبت شده در ایستگاه با مقادیر مدل شده آن مورد ارزیابی قرار

خود باقی بماند و یا استفاده از اعتبار یابی متقابل است. **جنگل تصادفی:** این روش به عنوان یکی از الگوریتم‌های ترکیبی یکی از روش‌های ناپارامتریک مناسب برای طبقه‌بندی و برازش داده است. این الگوریتم یادگیری ماشین با قابلیت استفاده آسان است که حتی بدون تنظیم فرآیندهای آن، می‌تواند نتایج مناسبی را ارائه کند. از جمله مزایای این روش ساده بودن فهم آن و عدم نیاز به فرض توزیع آماری برای داده‌ها، توانایی یافتن روابط غیرخطی با ابعاد بالا، استفاده از پیش‌بینی‌شده طبقه‌بندی و پیوسته، مقاومت در برابر بیش‌برازش و استحکام نسبی در برابر نویز در پیش‌بینی‌شده (Grimm et al., 2008, Liaw and Wiener, 2002). هر درخت از طریق نمونه بوت استرپ موجود در داده‌های آموزشی توسعه داده می‌شود. لازم است قبل از اجرای الگوریتم، سه فرآیند شامل تعداد درختان در جنگل، تعداد ویژگی‌هایی که در هر تقسیم‌بندی در نظر گرفته می‌شوند و کنترل رشد درختان و زیرمجموعه‌های تقسیم‌شده توسط کاربر و به صورت سعی و خطا تعیین می‌شود. در این مطالعه مقدار فرآیندهای تعداد درختان و کوچک‌ترین زیرمجموعه تقسیم‌شده به طور همزمان با استفاده از روش تکراری تغییر پیدا کرد (MiraboutalebiKazemi and Bahrami, 2016) تا در نهایت مقادیر ۱۲ و ۵ تعیین شده‌اند. محدود کردن عمق درختان برای جلوگیری از برازش بیش‌ازحد اتخاذ شده است. روش رگرسیون چندجمله‌ای جهت برآورد میانگین سالانه دمای هوا؛ روش رگرسیون چندجمله‌ای با درجه یا مرتبه‌های متفاوت انجام می‌شود که مناسب‌ترین رتبه با بررسی نمودار پراکنش و مقایسه مقدار برازش شده انتخاب می‌شود (Williamson et al., 2014)؛ اما افزایش درجه مدل ممکن است باعث تطابق بیش‌ازحد مدل گردد، زیرا تلاش می‌شود که منحنی از نقاط داده مشاهداتی (همچنین عبور از داده پرت در بین داده‌ها) بیشتری عبور کند و برای داده‌های جدید پیش‌بینی کننده خوبی نباشد. همچنین کاهش درجه مدل ممکن است باعث برازش کم‌تر از حد مدل گردد.



شکل ۳- مقادیر مشاهداتی و مدل شده میانگین سالانه دمای سطح زمین

زمین در هر ۴ روش مشابهت زیادی بایکدیگر دارند.

می‌گیرد.

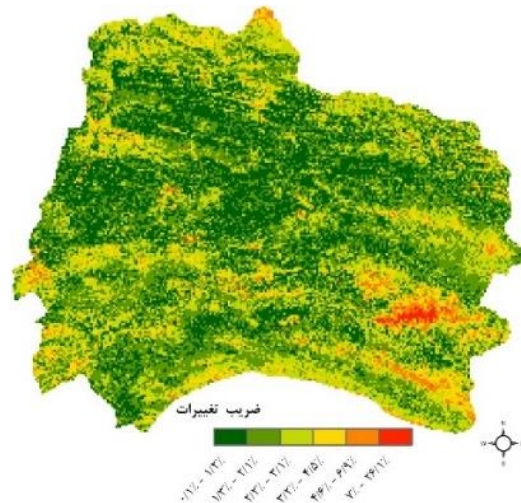
نتایج مربوط به عملکرد مدل‌های مورد استفاده توسط شاخص‌های آماری برای برآورد دمای سطح زمین در جدول ۲ ارائه گردید. با توجه به نتایج جنگل تصادفی کمترین ضریب تغییرات مجذور میانگین مربعات خطا با $2/85$ در بین مدل‌های مورد استفاده دارد. کمترین مقدار خطای جذر میانگین مربعات برای مدل جنگل تصادفی با $0/54$ درجه سلسیوس و بیشترین مقدار با $0/91$ درجه سلسیوس برای ماشین بردار پشتیبان مشاهده گردید. نتایج شاخص توافق تفاوتی بین مدل‌های مورد نشان نمی‌دهد، اما معیار نش ساتکلیف نشان‌دهنده تفاوت عمده بین مدل‌ها است. با توجه به این معیار کارایی تمام مدل‌ها بزرگ‌تر از $0/75$ است که مدل جنگل تصادفی با $0/95$ بالاترین مقدار را بین سایر مدل‌ها دارد.

مقایسه بصری نتایج برآورد دمای سطح زمین روی نمودار عملکرد تیلور نشان می‌دهد که مدل جنگل تصادفی بالاترین کارایی را دارد، زیرا نزدیک‌ترین فاصله تا انحراف استاندارد از داده‌های مرجع را دارد و کمترین خطای جذر میانگین مربعات با بالاترین ضریب همبستگی را نشان می‌دهد (شکل ۵). بنابراین مطابق کلیه معیارهای ارزیابی می‌توان مشخص نمود که روش جنگل تصادفی در مقایسه با سایر روش‌ها به‌عنوان روش مناسب و مطلوب در برآورد دمای سطح زمین مطرح می‌گردد.

نتایج سنجش حساسیت مدل جنگل تصادفی با حذف جداگانه متغیرهای مستقل و اجرای مجدد مدل نشان داد که

نتایج

ارزیابی روش‌ها در برآورد میانگین سالانه دمای سطح زمین: نمودارهای پراکنش بین مقادیر مشاهداتی و مقادیر مدل شده دمای سطح زمین نشان می‌دهد که همبستگی بالایی ($R > 0/9$) برای تمامی چهار روش مورد استفاده وجود دارد (شکل ۳). بر اساس نتایج، مقادیر مدل شده دمای سطح زمین در روش رگرسیون شبکه الاستیک و ماشین بردار پشتیبان از پراکنش بیشتر به‌ویژه در دمای حداکثر (رنگ‌های زرد در نمودار) برخوردار هست؛ اما پراکنش مقادیر مدل شده در شبکه عصبی مصنوعی و به‌ویژه در جنگل تصادفی به‌مراتب کمتر (مقادیر حداقل و حداکثر دما) از رگرسیون شبکه الاستیک و ماشین بردار پشتیبان است. مقایسه نتایج ضریب تغییرات (به‌عنوان معیار نسبی از میزان تغییرات) بین چهار مدل استفاده‌شده نشان می‌دهد که تغییرات در مقادیر برآورد شده دمای سطح زمین بین چهار مدل مورد استفاده در اکثر مناطق استان خراسان شمالی کمتر از ۲ درصد (نواحی سبزرنگ) است (شکل ۴). این امر بیانگر آن است که چهار مدل مورد استفاده در برآورد دمای سطح زمین نتایج یکسانی را داشته‌اند؛ اما بیشترین اختلافات بین مدل‌ها در حدود ۲۶ درصد در جنوب شرقی استان مشاهده شد (نواحی قرمز رنگ). این به آن معنا است که در قسمت زیادی از منطقه مورد مطالعه (نواحی سبزرنگ)، برآورد دمای سطح



شکل ۴- ضریب تغییرات دمای سطح زمین مدل شده بین چهار مدل استفاده شده

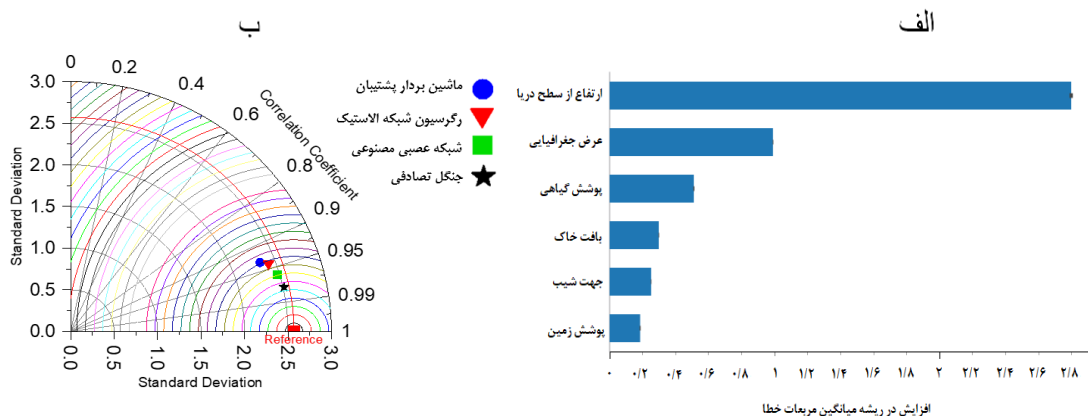
جدول ۲- متریک‌های نیکویی برازش جهت بررسی تفاوت بین مقادیر مشاهده شده و = مدل شده میانگین سالانه دمای سطح زمین

ماشین بردار پشتیبان	رگرسیون شبکه الاستیک	شبکه عصبی مصنوعی	جنگل تصادفی	متریک نیکویی برازش
۰/۷۳	۰/۶۷	۰/۵۳	۰/۳۸	میانگین خطای مطلق
۰/۹۱	۰/۸۶	۰/۷	۰/۵۴	خطای جذر میانگین مربعات
۰/۹۶	۰/۹۶	۰/۹۸	۰/۹۸	شاخص توافق
۴/۸	۴/۵۱	۳/۶۹	۲/۸۵	ضریب تغییرات مجذور میانگین مربعات خطا
۰/۸۷	۰/۸۸	۰/۹۲	۰/۹۵	معیار نش ساتکلیف

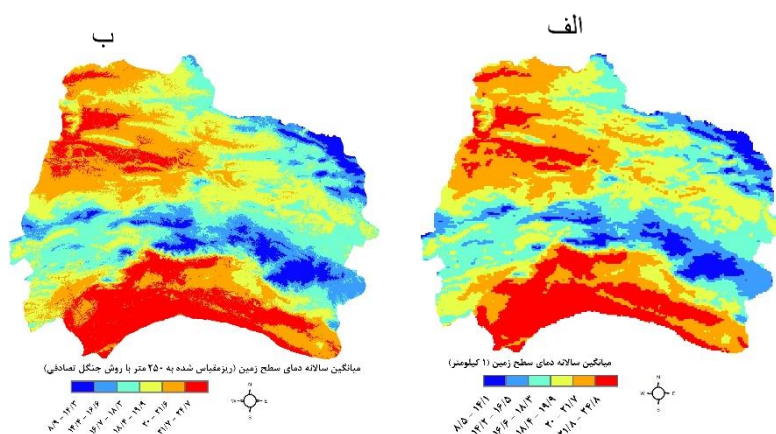
اختلاف زیادی بین طبقات میانگین سالانه دمای سطح زمین داده مشاهداتی و ریزمقیاس شده وجود ندارد. بر اساس نتایج حساسیت سنجی متغیرهای مستقل، ارتفاع بیشترین تأثیر را بر دمای سطح زمین دارد که این تأثیر بر روی داده ریزمقیاس شده دمای سطح زمین مشخص و به وضوح قابل تشخیص است. تأثیر سایر متغیرهای مستقل روی داده ریزمقیاس شده نیز وجود دارد که به وضوح قابل تشخیص نیست، مثلاً تأثیر پوشش گیاهی در قسمت‌هایی از منطقه (جنوب شرقی) قابل مشاهده است. علت این موضوع می‌تواند در درصد تأثیر متغیرهای مستقل در یک پیکسل ریزمقیاس شده باشد که در آن، اثر ارتفاع بیش از سایر متغیرها است.

ارزیابی روش رگرسیون چندجمله‌ای در برآورد میانگین سالانه دمای هوا و اعمال تصحیحات: در این پژوهش برای انتخاب بهترین مرتبه رگرسیون چندجمله‌ای

متغیر ارتفاع بیشترین تأثیر را بر دمای سطح زمین دارد به طوری که مقدار خطای جذر میانگین مربعات با حذف این متغیر از مدل به مقدار ۲/۷ درجه سلسیوس افزایش قابل توجهی پیدا می‌کند (شکل ۵). در مدل متغیرهای عرض جغرافیایی با ۱ درجه سلسیوس و پوشش گیاهی با ۰/۵ درجه سلسیوس اثر تعیین کننده‌ای بر دمای سطح زمین دارند میزان زیرا با حذف این متغیرها مقدار خطای جذر میانگین مربعات افزایش می‌یابد. متغیرهای بافت خاک، جهت شیب و پوشش زمین از اثرات مشابه روی مدل برخوردار هستند اما مقدار افزایش خطای جذر میانگین مربعات قابل توجه نیست. داده مشاهداتی و ریزمقیاس نمایی میانگین سالانه دمای سطح زمین در شکل ۶ نشان داده شده است. همان گونه که ملاحظه می‌شود میانگین سالانه دمای سطح زمین قبل و بعد از ریزمقیاس نمایی روند یکسانی را نشان می‌دهد و



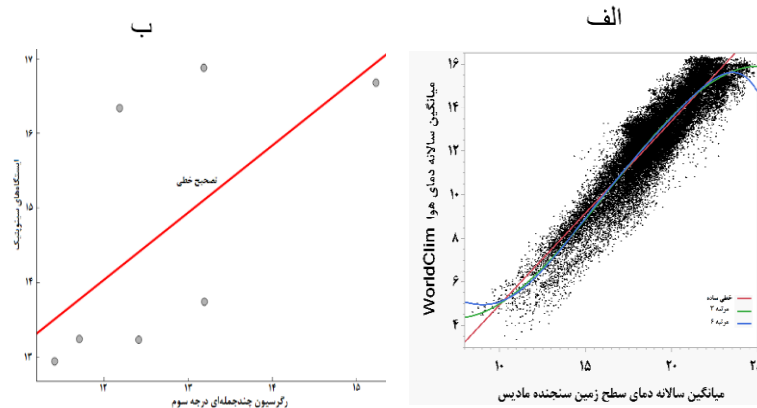
شکل ۵- الف) اثرات حذف متغیرهای مستقل بر مدل جنگل تصادفی جهت برآورد میانگین سالانه دمای سطح زمین، ب) موقعیت چهار مدل مورد استفاده و میزان نزدیکی آن نسبت به مقدار مرجع روی نمودار تیلور



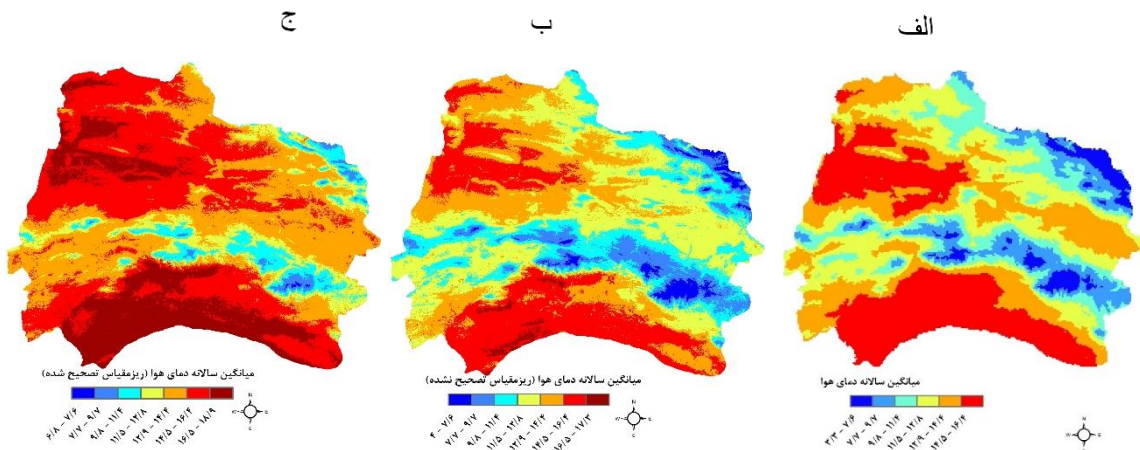
شکل ۶- الف) میانگین سالانه دمای سطح زمین داده اصلی مادیس (۱ کیلومتر)، ب) داده ریزمقیاس نمایی (۲۵۰ متر)

مناسب‌ترین روش جهت برآورد میانگین سالانه دمای هوا بر اساس میانگین سالانه دمای سطح زمین در منطقه مورد مطالعه، روش رگرسیون با تابع چندجمله‌ای درجه سوم است. به‌منظور اعمال تصحیحات بر روی نتایج رگرسیون با تابع چندجمله‌ای درجه سوم از روش تصحیح خطی با استفاده از ایستگاه‌های هواشناسی استفاده شد (شکل ۷ ب). در شکل ۸، نتایج ریزمقیاس میانگین سالانه دمای هوا با ابعاد سلولی ۲۵۰ متر و افزایش جزئیات مکانی نسبت به مقادیر WorldClim با ابعاد سلولی ۱ کیلومتر مشاهده می‌گردد. به‌عنوان مثال طبقه دمای هوا بین ۱۲/۹ تا ۱۴/۴ درجه سلسیوس (مناطق نارنجی شکل ۸) در جنوب استان از تغییرات مکانی زیادی در داده ریزمقیاس شده نسبت به داده اصلی برخوردار است. میزان اختلاف میانگین سالانه دمای هوا بین مقادیر ایستگاه‌های سینوپتیک، مدل تصحیح‌شده و

از خطای جذر میانگین مربعات و ضریب تعیین استفاده شد. نتایج نشان می‌دهد که مرتبه ۳ از میان ۶ درجه، از مناسب‌ترین حالت برازش برخوردار است (شکل ۷ الف) به‌طوری‌که مقدار خطای جذر میانگین مربعات برای درجه ۳ معادل ۰/۸۸ درجه سلسیوس با ضریب تعیین ۰/۸۶ و برای درجه ۶ معادل ۰/۸۷ درجه سلسیوس با ضریب تعیین ۰/۸۶۱ به‌دست آمد. از آنجا که خط برازش درجه ۶ یک رابطه با تابع چندجمله‌ای مرتبه ۶ است، تلاش می‌شود منحنی از نقاط بیشتر عبور کند که موجب مسئله بیش برازش می‌گردد. همچنین مرتبه ۶ در مقادیر حداکثر موجب ایجاد رابطه معکوس بین میانگین سالانه دمای هوا و دمای سطح زمین شده است (شکل ۷ الف). خاطر نشان می‌شود که انتخاب بهترین مرتبه بر اساس روش اعتبار سنجی متقاطع K-fold به‌دست آمده است. از این رو مشخص گردید که



شکل ۷- الف) ارتباط دمای هوا با دمای سطح زمین با مرتبه ۱، ۳، ۶، ب) تصحیح مدل چندجمله‌ای درجه ۳ با ایستگاه‌های هواشناسی



شکل ۸- الف) میانگین سالانه دمای هوا WorldClim (۱ کیلومتر)، ب) داده ریزمقیاس نمایی تصحیح نشده (۲۵۰ متر) ج) داده ریزمقیاس تصحیح شده (۲۵۰ متر)

و نتایج حاصل از مدل تصحیح شده و تصحیح نشده ریزمقیاس اختلاف معنی داری در سطح ۰/۰۵ وجود ندارد (۱) p-value اما بین مقادیر پایگاه Worldclim و مقادیر ایستگاه‌های سینوپتیک با p-value (۰/۰۴) اختلاف معنی دار است. با استناد به نتایج به دست آمده مدل ریزمقیاس نمایی از ۱ کیلومتر به ۲۵۰ متر در هر دو حالت تصحیح شده و تصحیح نشده نتایج قابل قبولی ارائه کرده است.

بحث و نتیجه گیری

متغیرهای مستقل متعددی در روش‌های ریزمقیاس نمایی مکانی دمای سطح زمین با رویکرد آماری و یادگیری ماشین مورد استفاده قرار می‌گیرد. یک راه حل رایج و معمول استفاده از رگرسیون خطی ساده با در نظر گرفتن گرادیان ارتفاع یا

تصحیح نشده ریزمقیاس و پایگاه Worldclim نشان داد که مدل تصحیح شده ریزمقیاس انطباق نسبتاً خوبی با مقادیر مشاهداتی (ایستگاه‌های سینوپتیک) دارد. اختلاف بین مقادیر داده پایگاه Worldclim، مدل تصحیح نشده و مدل تصحیح شده ریزمقیاس با ایستگاه‌های سینوپتیک بر اساس خطای جذر میانگین مربعات به ترتیب ۱/۵۸ درجه سلسیوس، ۱/۳۲ درجه سلسیوس و ۱/۲۱ درجه سلسیوس به دست آمد. همچنین معیار شاخص توافق به ترتیب ۰/۷، ۰/۷۲ و ۰/۷۳ به دست آمد. در این تحقیق همچنین میزان اختلاف معنی داری بین مقادیر پایگاه Worldclim، مدل تصحیح نشده و تصحیح شده ریزمقیاس با ایستگاه‌های سینوپتیک با استفاده از آزمون t جفتی مورد بررسی قرار گرفت. نتایج نشان داد که بین مقادیر ایستگاه‌های سینوپتیک

شاخص‌های پوشش گیاهی است (Maeda, 2014). اما نتایج تجزیه و تحلیل این پژوهش نشان می‌دهد که علاوه بر متغیر ارتفاع، استفاده از سایر متغیرهای کمکی مانند عرض جغرافیایی در مدل نیز موجب بهبود در برآورد دمای سطح زمین می‌شود. این اختلاف یا عدم استفاده محققین در استفاده از داده‌های متنوع در ریزمقیاس نمایی دمای سطح زمین به دلیل مقیاس و مکان مورد مطالعه است. به عنوان مثال Maeda (۲۰۱۴)، به ریزمقیاس نمایی دمای سطح زمین در منطقه‌ای نزدیک خط استوا با ابعاد ۹۰ در ۹۰ کیلومتر پرداخته است. در چنین مقیاسی که تنوع عوارض ممکن است کم باشد و از طرفی دیگر برخی متغیرها مانند عرض جغرافیایی به دلیل شرایط جغرافیایی بی اثر هستند، استفاده از یک متغیر ارتفاع برای ریزمقیاس نمایی استفاده می‌شود. اما همان‌طور که در این پژوهش اشاره شد، ریزمقیاس نمایی تحت شرایط متنوع توپوگرافی، پوشش زمین و اقلیمی انجام شد تا اثرات متغیرهای کمکی قابل بررسی باشد. از این‌رو نتایج این پژوهش نشان داد که توزیع فضایی دمای سطح زمین در منطقه مورد مطالعه از طریق تأثیر متقابل سه عامل اصلی مرتبط قابل توضیح است: (۱) ارتفاع، و جهت شیب توپوگرافی که تأثیر قوی بر دریافت تابش ورودی خورشید را نشان می‌دهد (۲) پوشش سطح زمین و پوشش گیاهی، که تنوع پوشش، خواص ساختار تاج پوشش، و تعرق (۳) رطوبت خاک که ارتباط نزدیکی با بافت خاک دارد که این نتایج نیز توسط He و همکاران (۲۰۱۹) در دو منطقه کوهستانی واقع در شمال شرقی و جنوب غربی چین با شرایط متنوع عوارض گزارش شده است (He et al., 2019). نتایج سایر محققین نیز نشان داد که توزیع مکانی-زمانی دمای سطح زمین به‌ویژه در مناطق کوهستانی تحت تأثیر متغیرهای توپوگرافی و همچنین شرایط پوشش گیاهی و خاک قرار دارد (He et al., 2019). بر اساس نتایج این پژوهش، متغیر ارتفاع بیشترین تأثیر را بر دمای سطح زمین دارد، همچنین جهت شیب توپوگرافی با تأثیر کمتر نیز مشاهده می‌گردد که با یافته‌های نتایج تحقیق حاضر همسو

است.

نتایج پژوهش حاضر نشان داد که عرض جغرافیایی به‌عنوان دومین متغیر تأثیرگذار بر میانگین سالانه دمای سطح زمین است. نتایج Zhao و Duan (۲۰۲۰) در جنوب غربی اروپا نشان داد که عرض جغرافیایی به‌عنوان مهم‌ترین متغیر مؤثر بر دمای سطح زمین در بین متغیرهای توپوگرافی، پوشش گیاهی، تابش خورشیدی و آلبیدو سطح است. عرض جغرافیایی همچنین یک عامل مهم بر توزیع فضایی مقادیر دمای سطح زمین است. یک گرادیان دمای عرضی مشخص وجود دارد به‌طوری‌که میانگین دمای سالانه به‌طور متوسط ۰/۷ درجه سانتی‌گراد برای هر درجه از عرض جغرافیایی در نیمکره شمالی و ۰/۵ درجه سانتی‌گراد در نیمکره جنوبی به‌جز مناطق گرمسیری کاهش می‌یابد (La Sorte et al., 2014). این مقدار برای منطقه مورد مطالعه معادل ۰/۶۵ درجه سانتی‌گراد برای هر درجه از عرض جغرافیایی به‌دست آمد. در این مطالعه، پوشش گیاهی به‌عنوان سومین متغیر مؤثر بر دمای سطح زمین شناخته شد. واضح است که تغییرات دمای سطح زمین تا حد زیادی با تغییرات عوامل توپوگرافی مرتبط است. با این حال، این رابطه منحصر به فرد نیست و پوشش گیاهی نیز نقش مهمی در تغییرات دمای سطح زمین ایفا می‌کند. همان‌طور که SandholtRasmussen و Andersen (۲۰۰۲) نشان دادند، پوشش گیاهی نه تنها دمای سطح را از طریق تعادل انرژی سطح کنترل می‌کند، بلکه شرایط پوشش گیاهی نیز بر دمای تابشی خاک در اطراف آن مؤثر است. با افزایش مقدار پوشش گیاهی، دمای سطح در یک محدوده دمای نسبتاً کوچک کاهش می‌یابد و ممکن است به‌نوبه خود از شدت اثر توپوگرافی بر دمای سطح زمین کاهش یابد. نتایج مطالعه حاضر نشان داد که بافت خاک به‌عنوان متغیر تأثیرگذار بر دمای سطح زمین است. نتایج سایر محققین نشان داد که بافت خاک یک پارامتر کلیدی در مدل‌سازی سطح زمین و فرآیند اتمسفر است و ارتباط نزدیکی با دمای سطح زمین دارد (DuanSchaake and Koren, 2001).

2017, *et al.*). همچنین این الگوریتم در مقایسه با سایر روش‌های یادگیری ماشین که در این پژوهش استفاده شده از مزایای پارامترهای کمتر برخوردار است که موجب می‌شود کاربر به‌دوراز پیچیدگی‌ها به تنظیم ساده مدل بپردازد (Tran *et al.*, 2017).

نتایج مطالعه حاضر نشان داد که داده میانگین سالانه دمای هوای ریزمقیاس شده که بر اساس متغیر میانگین سالانه دمای سطح زمین انجام شد، دارای اطلاعات بیشتری نسبت به داده‌های Wordclim با تفکیک مکانی پایین است زیرا تنوع و پیچیدگی توپوگرافی و عوارض مختلف سطح زمین با قدرت تفکیک مکانی بالا در نظر گرفته شده است. اگرچه نتایج سایر محققین نشان داده‌اند که استفاده از نرخ‌های استاندارد افت آهنگ زمانی که ارتباط واضحی بین توزیع دما و ارتفاع وجود داشته باشد، قابل استفاده است (Liston and Elder, 2006; Kunkel, 1989)، اما در مناطقی با توپوگرافی و شرایط پیچیده سطح زمین از دقت کافی برخوردار نیست (Lookingbill and Urban, 2003). نتایج Collados-Lara و همکاران (۲۰۲۱) نیز نشان دادند که برآورد دمای هوا با تفکیک مکانی بالا با استفاده از داده‌های ارتفاع و دمای سطح زمین امکان‌پذیر است. روش پیشنهادی در این مطالعه برای شرایطی که نیاز توأمان به استفاده از داده‌های دما پایگاه Wordclim و همچنین قدرت تفکیک مکانی بالا است، قابل استفاده می‌باشد و می‌توان از این روش جهت تولید داده ریزمقیاس دما در ابعاد متفاوت نیز استفاده کرد و نیازی به اعمال تصحیحات با استفاده از داده‌های ایستگاه‌های هواشناسی نمی‌باشد. با توجه به نتایج به‌دست آمده از داده ریزمقیاس شده می‌توان برای مطالعات زیست‌شناسی و اکولوژی استفاده نمود. از آنجا که در این مطالعه از داده‌های پایگاه Wordclim استفاده شده است، همچنین به دلیل تفاوت در داده‌های پایه و روش‌های استفاده شده در سایر پایگاه‌ها داده مانند CHESA، بررسی قابلیت ریزمقیاس نمایی میانگین سالانه دمای هوای سایر پایگاه‌ها نیز پیشنهاد می‌شود. لازم به ذکر است که دقت تولید داده

(Wang *et al.*, 2012). زیرا بافت خاک با میزان نگهداری آب در خاک مرتبط است از این رو تغییرات در رطوبت خاک نزدیک به سطح ارتباط مؤثری بر دمای سطح خاک دارد (Wang *et al.*, 2012). نتایج تجزیه و تحلیل این پژوهش نشان می‌دهد که پوشش سطح زمین که بیانگر تنوع فیزیکی عوارض سطح زمین و تعیین‌کننده ظرفیت ذخیره حرارت سطوح است، کمترین تأثیر را در برآورد میانگین سالانه دمای سطح زمین دارد. اما این تأثیر اندک به دلیل تغییر در ابعاد سلول داده سطح زمین از ۱۰ متر به ۲۵۰ متر مربع (جدول ۱) باشد که موجب ریز اطلاعات فضایی در طبقات پوشش سطح زمین شده است. این موضوع نیز توسط Tran و همکاران (۲۰۱۷) نیز مشاهده شده است که قدرت تفکیک کمتر منجر به یک نتیجه بسیار کلی و حذف تعدادی از عوارض می‌شود که موجب کاهش دقت برآورد دمای سطح زمین می‌شود. علت بعدی نیز روش مورداستفاده در طبقه‌بندی تصویر ماهواره‌ای جهت استخراج طبقات پوشش سطح زمین است که بر دقت برآورد دمای سطح زمین مؤثر در نظر گرفته شده است (Tran *et al.*, 2017). بنابراین حساسیت‌سنجی روش‌های مختلف طبقه‌بندی بر برآورد دمای سطح زمین مطرح می‌گردد که خارج از اهداف این پژوهش بوده است.

مقایسه روش‌های مختلف یادگیری ماشین در این پژوهش نشان داد که الگوریتم جنگل تصادفی از توانمندی بهتری در ایجاد مدل ریزمقیاس جهت برآورد میانگین سالانه دمای سطح زمین با استفاده از متغیرهای پیش‌بینی‌کننده را دارد. نتایج سایر محققین نیز نشان داد که الگوریتم جنگل تصادفی سطح بالاتری از دقت را نسبت به سایر روش‌ها در برآورد دمای سطح زمین ارائه کرده است (Tran *et al.*, 2017; BartkowiakCastelli and Notarnicola; Hutengs and Vohland, 2016, 2019) زیرا این الگوریتم یک فرا تخمین‌گر است که درخت‌های تصمیم‌گیری را به نمونه‌های مختلف مجموعه داده برازش غیر خطی می‌دهد و از میانگین‌گیری آن برای بهبود دقت پیش‌بینی و کنترل بیش‌ازحد برازش استفاده می‌کند (Tran

داده‌های پایگاه داده و انجام تصحیحات متناسب با اهداف پژوهش در نظر گرفته شود.

ریزمقیاس تا حد زیادی به کیفیت مجموعه داده وابسته است و از آنجا که این پایگاه‌های داده از اعتبار یکسانی برای مطالعات مختلف برخوردار نیستند، استفاده صحیح از

References

- Acharjee, A., Finkers, R., Visser, R.G., Maliepaard, C.J.M., 2013. Comparison of regularized regression methods for omics data. *Metabolomics* 3(3), 1-9.
- Amiri, M., Tarkesh, M., Jafari, R., Jetschke, G., 2020. Bioclimatic variables from precipitation and temperature records vs. remote sensing-based bioclimatic variables: Which side can perform better in species distribution modeling?. *Ecological Informatics* 57, 1-45.
- Anguita, D., Ghelardoni, L., Ghio, A., Oneto, L., Ridella, S., 2012. The 'K' in K-fold cross validation. In: 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN) Bruges, Belgium, pp. 441-446.
- Baldocchi, D., Ma, S., 2013. How will land use affect air temperature in the surface boundary layer? Lessons learned from a comparative study on the energy balance of an oak savanna and annual grassland in California, USA. *Tellus B: Chemical and Physical Meteorology* 65(1), 2-9.
- Barry, R.G., 1992. Mountain weather and climate. Third Edition. Routledge, London, pp. 1-49.
- Bartkowiak, P., Castelli, M., Notarnicola, C., 2019. Downscaling Land Surface Temperature from MODIS Dataset with Random Forest Approach over Alpine Vegetated Areas. *Remote Sensing* 11(11), 1-19.
- Bollapragada, R., Nocedal, J., Mudigere D., Shi, H.-J., Tang, P.T.P., 2018. International Conference on Machine Learning (ICML). A progressive batching L-BFGS method for machine learning, Stockholm, Sweden. pp. 1-10.
- Breiman, L., J. Friedman, H., Olshen, R.A., Stone, C.G. 1984. Classification and regression trees. Wadsworth International Group, Belmont, California, USA. 16(3), 199-215.
- Chalghaf, B., Chemkhi, J., Mayala, B., Harrabi, M., Benie, G.B., Michael, E., Ben Salah, A., 2018. Ecological niche modeling predicting the potential distribution of Leishmania vectors in the Mediterranean basin: impact of climate change. *Parasites Vectors* 11(461), 1-9.
- Collados-Lara, A.-J., Fassnacht, S. R., Pardo-Igúzquiza, E., Pulido-Velazquez, D., 2021. Assessment of high resolution air temperature fields at rocky mountain national park by combining scarce point measurements with elevation and remote sensing data. *Remote Sensing* 13(1), 1-26.
- Daly, C., Taylor, G. H., 1998. The prism approach to mapping precipitation and temperature. 10th Conference on Applied Climatology and Workshop on Extreme Value Analysis in Climatology 1-4.
- Didan, K., 2021. MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V061. N. E. L. P. DAAC (Ed.). <https://www.earthdata.nasa.gov/Accessed> 20th February 2021
- Duan, Q., Schaake, J., Koren, V., 2001. A priori estimation of land surface model parameters. in: land surface hydrology, meteorology, and climate: observations and modeling. American Geophysical Union 1-215.
- Durđević Babić, I., 2015. Predicting student satisfaction with courses based on log data from a virtual learning environment—a neural network and classification tree model. *Croatian Operational Research Review* 6(1), 105-120.
- Ebrahimi, H., Aghighi, H., Azadbakht, M., Amani, M., Mahdavi, S., Matkan, A.A., 2021. Downscaling MODIS Land Surface Temperature Product Using an Adaptive Random Forest Regression Method and Google Earth Engine for a 19-Years Spatiotemporal Trend Analysis Over Iran. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14(1), 2103 - 2112.
- Fick, S. E., Hijmans, R. J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37(12), 4302-4315.

- Ghorbanian, A., Kakooei, M., Amani, M., Mahdavi, S., Mohammadzadeh, A., Hasanlou, M., 2020. Improved land cover map of Iran using Sentinel imagery within Google Earth Engine and a novel automatic workflow for land cover classification using migrated training samples. *ISPRS Journal of Photogrammetry and Remote Sensing* 167, 276-288.
- Grimm, R., Behrens, T., Märker, M., Elsenbeer, H., 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island-Digital soil mapping using Random Forests analysis. *Geoderma* 146 (1-2), 102-113.
- Hamann, A., Wang, T., 2006. Potential effects of climate change on ecosystem and tree species distribution in British Columbia. *Ecology* 87, 2773-2786.
- Han, F., Yan, J., Ling, H., 2021. Variance of vegetation coverage and its sensitivity to climatic factors in the Irtys River basin. *Peer J* 9, 1-24.
- He, J., Zhao, W., Li, A., Wen, F., Yu, D., 2019. The impact of the terrain effect on land surface temperature variation based on Landsat-8 observations in mountainous areas. *International Journal of Remote Sensing* 40(5-6), 1808-1827.
- Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: Global gridded soil information based on machine learning. *Plos One* 12(2), 1-40.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25, 1965-1978.
- Hutengs, C., Vohland, M., 2016. Downscaling land surface temperatures at regional scales with random forest regression. *Remote Sensing of Environment* 178, 127-141.
- Karger, D. N., Conrad, O., Böhrner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H. P., Kessler, M., 2017. Climatologies at high resolution for the earth's land surface areas. *Scientific Data* 4, 170122.
- Kloog, I., Nordio, F., Coull, B.A., Schwartz, J., 2014. Predicting spatiotemporal mean air temperature using MODIS satellite surface temperature measurements across the Northeastern USA. *Remote Sensing of Environment* 150, 132-139.
- Kunkel, K.E., 1989. Simple procedures for extrapolation of humidity variables in the mountainous western united states. *Journal of Climate* 2, 656-669.
- La Sorte, F.A., Butchart, S.H., Jetz, W., Böhning-Gaese, K., 2014. Range-wide latitudinal and elevational temperature gradients for the world's terrestrial birds: implications under global climate change. *PloS one* 9(5), 1-13.
- Liaw, A., Wiener, M.J.R., 2002. Classification and regression by random Forest. *R News* 2-3, 18-22.
- Liston, G.E., Elder, K., 2006. A meteorological distribution system for high-resolution terrestrial modeling (MicroMet). *Journal of Hydrometeorology* 7, 217-234.
- Lookingbill, T.R., Urban, D.L., 2003. Spatial estimation of air temperature differences for landscape-scale studies in montane environments. *Agricultural and Forest Meteorology* 114 (3-4), 141-151.
- Maeda, E.E., 2014. Downscaling MODIS LST in the East African mountains using elevation gradient and land-cover information. *International Journal of Remote Sensing* 35(9), 3094-3108.
- McCutchan, M.H., Fox, D.G., 1986. Effect of elevation and aspect on wind, temperature and humidity. *Journal of Applied Meteorology and Climatology* 25(12), 1996-2013.
- Melchiorre, C., Castellanos Abella, E.A., van Westen, C.J., Matteucci, M., 2011. Evaluation of prediction capability, robustness, and sensitivity in non-linear landslide susceptibility models, Guantánamo, Cuba. *Computers Geosciences* 37 (4), 410-425.
- Militino, A. F., Ugarte, M.D., Montesino, M., 2019. Filling missing data and smoothing altered data in satellite imagery with a spatial functional procedure. *Stochastic Environmental Research and Risk Assessment* 33, 1737-1750.
- Miraboutalebi, S.M., Kazemi, P., Bahrami, P., 2016. Fatty Acid Methyl Ester (FAME) composition used for estimation of biodiesel cetane number employing random forest and artificial neural

- networks: A new approach. *Fuel* 166, 143-151.
- Mohandes, M.A., Halawani, T.O., Rehman, S., Hussain, A. A., 2004. Support vector machines for wind speed prediction. *Renewable Energy* 29(6), 939-947.
- Moreno, A., Hasenauer, H., 2016. Spatial downscaling of European climate data. *International Journal of Climatology* 36(3), 1444-1458.
- Mwakapeje, E.R., Ndimuligo, S.A., Mosomtai, G., Ayebare, S., Nyakarahuka, L., Nonga, H.E., Mdegela, R.H., Skjerve, E., 2019. Ecological niche modeling as a tool for prediction of the potential geographic distribution of *Bacillus anthracis* spores in Tanzania. *International Journal of Infectious Diseases* 79 142-151.
- Pal, S., Ziaul, S., 2017. Detection of land use and land cover change and land surface temperature in English Bazar urban centre. *The Egyptian Journal of Remote Sensing and Space Science*, 20 (1), 125-145.
- Pineda, E. Lobo, J. M., 2009. Assessing the accuracy of species distribution models to predict amphibian species richness patterns. *Journal of Animal Ecology* 78(1), 182-190.
- Pratumchart, K., Suwannatrai, K., Sereewong, C., Thinkhamrop, K., Chaiyos, J., Boonmars, T., Suwannatrai, A. T., 2019. Ecological niche model based on maximum entropy for mapping distribution of *bithynia siamensis goniomphalos*, first intermediate host snail of *opisthorchis viverrini* in thailand. *Acta Tropica* 193, 183-191.
- Rehfeldt, G. E., Crookston, N.L., Sáenz-Romero, C., Campbell, E.M., 2012. North American vegetation model for land-use planning in a changing climate: a solution to large classification problems. *Ecological applications* : a publication of the Ecological Society of America 22(1), 119-141.
- Ronao, C. A., Cho, S.-B., 2016. Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Systems with Applications* 59, 235-244.
- Sandholt, I., Rasmussen, K., Andersen, J., 2002. A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. *Remote Sensing of Environment* 79 (2-3), 213-224.
- Soria-Auza, R. W., Kessler, M., Bach, K., Barajas-Barbosa, P. M., Lehnert, M., Herzog, S. K., Böhner, J., 2010. Impact of the quality of climate models for modelling species occurrences in countries with poor climatic documentation: a case study from Bolivia. *Ecological Modelling* 221(8), 1221-1229.
- Sun, Z., Guo, H., Li, X., Lu, L., Du, X., 2011. Estimating urban impervious surfaces from Landsat-5 TM imagery using multilayer perceptron neural network and support vector machine. *J. of Applied Remote Sensing* 5(1), 1-18.
- Taylor, K. E., 2001. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres* 106 (D7), 7183-7192.
- Tran, D.X., Pla, F., Latorre-Carmona, P., Myint, S.W., Caetano, M., Kieu, H.V., 2017. Characterizing the relationship between land use land cover change and land surface temperature. *ISPRS Journal of Photogrammetry and Remote Sensing* 124, 119-132.
- Uca, T., Ekhwan, J., Othman, M., Amal Rosmini, A., Ansari Saleh, A., 2018. Daily suspended sediment discharge prediction using multiple linear regression and artificial neural network. *Journal of Physics: Conference Series* 954, 1-20.
- Vancutsem, C., Ceccato, P., Dinku, T., Connor, S.J., 2010. Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. *Remote Sensing of Environment* 114(2), 449-465.
- Vapnik, V., 1998. *Statistical learning theory.*, 1998. Wiley, New York. pp. 25-40.
- Vernay, C., Blanc, P., Pitaval, S., 2013. Characterizing measurements campaigns for an innovative calibration approach of the global horizontal irradiation estimated by HelioClim-3. *Renewable Energy* 57, 339-347.
- Wan, Z., Hook, S., Hulley, G., 2021. MODIS/Terra land surface temperature/emissivity 8-day 13 global 1km sin grid v061. N. E. L. P. DAAC (Ed.). <https://www.earthdata.nasa.gov/Accessed> 20th February 2021
- Wang, D.-C., Zhang, G.-L., Pan, X.-Z., Zhao, Y.-G., Zhao, M.-S., Wang, G.-F., 2012. Mapping soil texture of a plain area using fuzzy-c-means clustering method based on

- land surface diurnal temperature difference. *Pedosphere* 22(3), 394-403.
- Wang, K., Sun, J., Cheng, G., Jiang, H., 2011. Effect of altitude and latitude on surface air temperature across the Qinghai-Tibet Plateau. *Journal of Mountain Science* 8, 808-816.
- Wang, T., Hamann, A., Spittlehouse, D., Carroll, C., 2016. Locally downscaled and spatially customizable climate data for historical and future periods for North America. *PloS one* 11 (6), 1-17.
- Williamson, S.N., Hik, D.S., Gamon, J.A., Kavanaugh, J.L., Flowers, G.E., 2014. Estimating temperature fields from modis land surface temperature and air temperature observations in a sub-arctic alpine environment. *Remote Sensing* 6 (2), 946-963.
- Xia, J., Kumta, A.S., 2010. Feedforward Neural Network trained by BFGS algorithm for modeling plasma etching of silicon carbide. *IEEE Transactions on Plasma Science* 38(2), 142-148.
- Zevenbergen, L.W., Thorne, C.R., 1987. Quantitative analysis of land surface topography. *Earth surface processes and landforms* 12(1), 47-56.
- Zhao, W. Duan, S.-B., 2020. Reconstruction of daytime land surface temperatures under cloud-covered conditions using integrated MODIS/Terra land products and MSG geostationary satellite data. *Remote Sensing of Environment* 247, 1-15.
- Zupan, B., Demsar, J., 2008. Open-Source tools for data mining. *Clinics in Laboratory Medicine* 28(1), 37-54.

Evaluation of machine learning methods in spatial downscaling of average annual land surface temperature and air temperature

Azadeh Atabati*¹, Hamed Adab²

¹Department of Environmental Sciences and Engineering, Faculty of Geography and Environmental Sciences, Hakim Sabzevari University, Sabzevar, Iran

²Department of Remote Sensing and Geographic Information System, Faculty of Geography and Environmental Sciences, Hakim Sabzevari University, Sabzevar, Iran

*Corresponding author: a.atabati@hsu.ac.ir

Abstract

Today, the use of raster data from climate databases such as WorldClim is one of the reliable data sources that have been used instead of the point data of weather stations. However, the use of these climate databases with low spatial resolution has generated limitations for many studies related to biological and ecological studies. This study aims to investigate the relationship between air temperature and land surface temperature and then reproduce the land surface temperature with the high spatial resolution for downscaling of annual average air temperature using two widely used products, namely, annual average air temperature from the WorldClim database and annual average day and night temperature from MOD11A2 v061 MODIS sensor. In this study, firstly, the performance of machine learning models including random forest, artificial neural network, elastic network regression, and support vector machine for downscaling of MOD11A2 v061 product from 1 km to 250 meters was assessed. For this purpose, continuous and discrete predictor variables including height above sea level, latitude, vegetation cover, soil texture, slope direction, and land cover were used. Then, WorldClim's annual average air temperature was downscaled from 1 km to 250 meters using the land surface temperature with a 3rd-degree polynomial regression model. Also, the weather data of seven synoptic stations were used to check the validity of the downscaled product. The results of the Taylor diagram represented that the random forest model has the best performance for the downscaled land surface temperatures with a root mean square error of 0.54 degrees Celsius. Also, the 3rd-degree polynomial regression model has a lower relative error rate in producing downscaled air temperature. The value of the root mean square error of the results for the uncorrected and corrected downscaled air temperature was 1.32 and 1.21 degrees Celsius, respectively, which did not show a significant difference at the 0.05 level according to the paired t-test. The findings of this research show that the downscaling of the mean annual air temperature of WorldClim has the required validity.

Keywords: WorldClim database, MODIS data, Spatial downscaling, Machine learning