

مهارت‌سنجی ساختار بیزی مدل مارکف پنهان در برآورد بی درنگ مراحل فنولوژی گیاه ذرت

مهدی ققمقامی^۱، نوذر قهرمان^{۲*}، پرویز ایران‌نژاد^۳، حمید پزیشک^۴

تاریخ دریافت: ۱۳۹۵/۰۳/۲۳

تاریخ پذیرش: ۱۳۹۶/۰۵/۱۵

چکیده

مدل‌سازی درصد پیشرفت مراحل فنولوژی گیاهان با روش‌های کلاسیک عمدتاً با محدودیت‌هایی به ویژه، عدم کفایت یا دقت داده‌های مشاهداتی مورد نیاز مواجه است که جستجوی روش‌های جایگزین را ضروری می‌سازد. روش‌های آماری نظیر رهیافت‌های بیزی (Bayesian) و مدل مارکف پنهان دارای توانمندی‌های ساختاری مناسبی در حل مسائلی با طیف گسترده از داده‌های گسسته، پیوسته، شناخته شده و یا ناشناخته می‌باشند که قابل تلفیق با داده‌های سنجش از دور نیز هستند. هدف از تحقیق حاضر بررسی توانمندی مدل مارکف پنهان در برآورد بی‌درنگ^۵ درصد پیشرفت مراحل فنولوژی گیاه ذرت در مزرعه پردیس کشاورزی و منابع طبیعی دانشگاه تهران واقع در کرج می‌باشد. حالت‌های مختلف متشکل از مراحل فنولوژی شامل ۷ مرحله اصلی (سبز شدن تا شیری شدن) به عنوان لایه پنهان و بردار مشخصه‌های درجه-روز رشد تجمعی (AGDD) و نمایه اختلاف استاندارد شده پوشش گیاهی (NDVI) مستخرج از تصاویر ماهواره Landsat7 ETM+ به عنوان لایه قابل رویت در نظر گرفته شدند. واسنجی و صحت‌سنجی مدل براساس داده‌های دیدبانی شده در مزرعه تحقیقاتی پردیس کشاورزی و منابع طبیعی دانشگاه تهران طی دوره آماری ۲۰۱۰-۲۰۰۲ انجام شد. بر اساس نتایج، میانگین دقت مدل مارکف پنهان براساس RMSE حدود ۱۴٪ است که نشانگر امکان کاربست آن به عنوان یک ابزار کمکی در کنار مشاهدات مزرعه‌ای است. مطالعات تکمیلی جهت تدقیق مدل‌های فنولوژیک مبتنی بر روش‌های آماری-دورسنجی در سایر مناطق اقلیمی و گیاهان راهبردی پیشنهاد می‌شود.

واژه‌های کلیدی: درصد پیشرفت، فنولوژی، NDVI، AGDD

مقدمه

متولی اندازه‌گیری میدانی این کمیت مهم هستند. روش اندازه‌گیری به این صورت است که تعدادی کرت به صورت تصادفی انتخاب می‌شوند و در هر کرت وضعیت فنولوژی ۱۰ بوته بررسی می‌شود. در نهایت درصد پیشرفت هر مرحله فنولوژی با متوسط‌گیری روی کرت‌های مختلف (با احتساب زمان آغاز و پایان مرحله) محاسبه می‌شود. با وجود دقت قابل توجه، اندازه‌گیری‌های میدانی، مشکلات مختلفی نیز دارند (توکیلگلو^۷، ۲۰۰۷؛ شن و همکاران^۸، ۲۰۱۳؛ گائو و همکاران^۹، ۲۰۱۷)، ۱) عموماً این اطلاعات به صورت منظم ثبت نمی‌شوند (همواره خلأهای آماری ناشی از عدم امکان داده‌برداری به علل مختلف نظیر انجام آبیاری یا شیوه‌های مدیریت مزرعه، وجود دارد؛ ۲) روش‌های بررسی میدانی عموماً مبتنی بر فنون نمونه‌گیری و مکان ویژه هستند و تعمیم آن به کل مزرعه با گستره‌های وسیع به علت تغییرات مکانی خصوصیات

درصد پیشرفت هر گیاه^۶ در یک مرحله معین فنولوژی، نشانگر وضعیت رویش و میزان تکمیل آن مرحله از چرخه زندگی گیاه است. به عنوان مثال، هنگامی که درصد پیشرفت مرحله «شیری شدن» گیاه ذرت در تاریخ معینی ۳۳ درصد باشد، به این معنی است که ۳۳ درصد گیاهان کرت‌های مورد دیدبانی در آن تاریخ، وارد مرحله فنولوژیکی شیری شدن شده‌اند. این کمیت، نقش تعیین‌کننده‌ای در مدیریت زراعی و آگاهی از وضعیت عمومی رشد در مزارع دارد. در ایران، ایستگاه‌های تحقیقاتی هواشناسی کشاورزی

^۱ دانشجوی دکتری هواشناسی کشاورزی، دانشکده مهندسی فناوری،

دانشگاه تهران، کرج، ایران

^۲ دانشیار، گروه مهندسی آبیاری و آبادانی، دانشگاه تهران، کرج، ایران

(*نویسنده مسئول: nghahreman@ut.ac.ir)

^۳ دانشیار گروه فیزیک فضا، مؤسسه ژئوفیزیک، دانشگاه تهران

^۴ استاد، گروه علوم ریاضی و آمار، دانشگاه تهران

^۵ Real time

^۶ Crop Progress Percentage

^۷ Toukiloglou

^۸ Shen et al.

^۹ Gao et al.

ترتیب همگن یا ناهمگن^۵ می‌شود. در مورد فنولوژی، با توجه به اینکه گیاه طی فرآیند رشد فیزیولوژیکی از یک مرحله وارد مرحله جدید شده و مرحله قبلی تکرار نمی‌شود، احتمال می‌رود که مدل همگن عملکرد مناسبی نداشته باشد. به عنوان مثال، در شروع چرخه زندگی، ذرت در اولین مرحله نموی (مثلاً سبز شدن) است. بنابراین احتمال انتقال از مرحله فعلی به خودش بیشتر و به مرحله بعدی کمتر است. با گذر زمان این شرایط عکس می‌شود یعنی انتقال به مرحله بعدی از ماندن در مرحله فعلی محتمل‌تر است تا جایی که احتمال انتقال از مرحله فعلی به خود صفر می‌شود. این زمان می‌تواند به عنوان زمان انتقال مرحله فنولوژیکی در نظر گرفته شود. بر این اساس مدل مارکف پنهان استفاده شده برای پیش‌بینی فنولوژی باید یک مدل ناهمگن یا نا ایستاً^۶ با زمان باشد. برای اولین بار ویوی و همکاران^۷ (۱۹۹۲) از الگوریتم مارکف پنهان در مدل‌سازی پویای فنولوژی گیاهی استفاده کردند. روش آن‌ها مبتنی بر تحلیل سری زمانی داده‌های سنجش از دور نمایه NDVI^۸ بوده است. نتایج آن‌ها حکایت از کارکرد مناسب رهیافت در شبیه‌سازی تاریخ سه مرحله خاک بدون پوشش، رشد رویشی و بلوغ (تکمیل مرحله زایشی) داشته است. بر این اساس، اردال و همکاران^۹ (۲۰۰۵) ایده طبقه‌بندی پوشش گیاهی را با داده‌های سنجش از دور و الگوریتم مارکف پنهان پیگیری کردند. در روش آن‌ها که روی هر پیکسل اجرا می‌شود، به تعداد تیپ‌های پوشش گیاهی موجود در منطقه، مدل دو لایه مارکف پنهان شکل می‌گیرد. طبق الگوریتم پیشرو، آن پیکسل به تیپ متناظر با مدل دارای بیشترین احتمال تعلق می‌گیرد. روش مارکف پنهان در طبقه‌بندی تیپ‌های جنگلی در نروژ جنوبی دارای عملکرد بهتری (۰/۶۲/۴) نسبت به روش نظارت‌شده حداکثر درست‌نمایی (۰/۵۶/۳) است. لیت و همکاران (۲۰۱۱) با بهبود روش‌های حل مسائل موجود در الگوریتم مارکف پنهان، روشی جدید برای طبقه‌بندی پوشش گیاهی مبتنی بر فنولوژی گیاهی در برزیل ارائه دادند. در این روش انتشار هر مشاهده (نمایه رادیومتریک نظیر NDVI) از هر حالت از توزیع

فنولوژیکی، از دقت کافی برخوردار نیست؛ (۳) دقت چنین روش‌هایی به عوامل غیرمحیطی مانند مهارت داده‌بردار، نوع و کیفیت تجهیزات مورد استفاده، انتخاب مناسب مکان‌های نمونه‌گیری و نوع روش برون‌یابی بستگی دارد؛ (۴) در گستره‌های وسیع، بررسی میدانی مستلزم بکارگیری تعداد زیادی داده‌بردار است که زمان‌بر، مستلزم هزینه زیاد و متضمن خطاهای ناشی از تفاوت مهارت‌های فردی است و (۵) به علت محدودیت‌های موجود در این نوع روش‌ها، برخی مناطق فاقد اطلاعات کافی و قابل اعتماد هستند. پایش غیرمیدانی پوشش گیاهی عبارت است از توسعه و کاربست سامانه‌های برآورد بی‌درنگ جهت پیش‌آگاهی ویژگی‌های نمو نظیر درصد پیشرفت فنولوژیکی گیاه. البته برخی از این سامانه‌ها مانند مدل‌های رشد زراعی، دارای همان محدودیت‌های روش‌های میدانی مانند مکان‌ویژه^۱ بودن هستند. به علاوه، این روش‌ها درصد پیشروی را بیان نمی‌کنند بلکه صرفاً حالت فنولوژی را نشان می‌دهند. رهیافت‌های مبتنی بر استنباط بیزی^۲ در تحلیل پویای توالی داده‌ها بسیار سودمند هستند. برخلاف روش‌های کلاسیک که در آن‌ها صرفاً همبستگی بین داده‌های ورودی و خروجی و یا میزان برازیدگی یک مدل خاص معیار مدل‌سازی است، روش‌های مبتنی بر رهیافت بیزی به ارزیابی احتمالات و به خصوص احتمالات شرطی می‌پردازند. یکی از مدل‌های مبتنی بر رهیافت بیزی، مدل مارکف پنهان^۳ می‌باشد که ساختار متداول آن دارای یک لایه پنهان و یک لایه قابل رویت است. در کاربست پایش فنولوژی گیاهی، لایه پنهان عبارت است از توالی زمانی حالت‌های مختلف فنولوژی شامل فازهای سبزشدن، سه برگی‌شدن و غیره و لایه قابل‌رویت نیز شامل توالی زمانی داده‌های سنجش از دور و داده‌های هواشناسی و یا متغیر وابسته به عوامل هواشناسی نظیر درجه-روز رشد می‌باشد. با توجه به پیوستگی اطلاعات در لایه قابل رویت، به جای ماتریس احتمالات انتشار، از توزیع‌های آماری شرطی مانند توزیع گسَن استفاده می‌شود (لیت و همکاران^۴، ۲۰۱۱). ماتریس احتمال انتقال می‌تواند با زمان ثابت و یا متغیر باشد که در این صورت مدل به

^۵ Homogeneous or Non-homogeneous

^۶ Non-stationary

^۷ Viovy et al.

^۸ Normalized Difference Vegetation Index

^۹ Aurdal et al.

^۱ Site-specific

^۲ Bayesian Inference

^۳ Hidden Markov Model (HMM)

^۴ Leite et al.

متر از سطح دریا است. متوسط درازمدت بارندگی ثبت شده در آن ۲۵۰/۱ میلی‌متر با ضریب تغییرات ۱/۲۴ درصد است. در این مزرعه گندم با ذرت به صورت تناوبی کشت می‌شوند و عموماً ماه‌های تیر و مرداد زمان‌های کاشت ذرت و ماه‌های مهر و آبان زمان‌های برداشت آن هستند. مساحت مزرعه بیش از ۲۰ هکتار بوده و ذرت رقم K704 در آن کشت می‌شود. در تحقیق حاضر از گروه داده‌هایی به شرح ذیل استفاده شد:

- توالی زمانی رخدادهای فنولوژی گیاه ذرت به همراه درصد پیشرفت هر مرحله که در قالب داده‌های دیدبانی فنولوژی توسط بخش تحقیقات اداره کل هواشناسی البرز در دوره ۹ ساله (۲۰۱۰-۲۰۰۲) ثبت شده است. این توالی شامل زمان کاشت و برداشت و درصد پیشرفت ۸ مرحله جوانه‌زنی، سبز شدن، سه برگی، برگ‌دهی، ظهور گل آذین، گل‌دهی و شیرگی شدن می‌باشد که هر دو روز یک بار ثبت شده‌اند. این اطلاعات دارای خلأهای آماری به علل عملیات زراعی یا کمبود پرسنل در زمان‌های مقرر دیدبانی هستند که برای رفع این خلأها و تکمیل مجموعه اطلاعات، اقدام به درون‌یابی با روش‌های خطی و غیرخطی شد. در شکل ۱، درصد پیشرفت مراحل مختلف فنولوژی مربوط به سال ۲۰۰۲ با استفاده از داده‌های دریافتی از اداره کل هواشناسی استان البرز ارائه شده است. تاریخ متناظر تقاطع نقطه‌چین افقی با خط منحنی درصد پیشرفت هر مرحله فنولوژی به عنوان تاریخ شروع آن مرحله در نظر گرفته می‌شود. مرحله جوانه‌زنی به دلیل کوتاه بودن و عدم دسترسی به اطلاعات کافی از مراحل کنار گذاشته شد و طول فصل رشد از ابتدای سبز شدن تا انتهای شیرگی شدن در نظر گرفته شد.

- تصاویر ماهواره‌ای چندزمانه سنجنده لندست ۷ (ETM+) پوشش مزرعه مورد مطالعه که قابل دریافت از سایت <http://reverb.echo.nasa.gov> به صورت رایگان هستند. با توجه به قرارگیری مزرعه در محل هم‌پوشانی دو تصویر با مشخصه‌های LE7164035# و LE7165035#، تعداد تصاویر موجود افزایش یافت. باندهای مورد استفاده از تصاویر باندهای ۳ و ۴ هستند که برای استخراج نمایه پوشش گیاهی استفاده شده‌اند. توالی زمانی تصاویر سنجنده منظم نبوده و از این‌رو، تعداد تصاویر در سال‌های مختلف متفاوت بوده است. کمترین تعداد تصاویر معادل ۸

گسن تبعیت می‌کند. بر این اساس احتمالات انتقال و انتشار طی نیمرخ زمانی تصاویر نا ایستا است. روش ارائه شده توسط این محققین در طبقه‌بندی پوشش گیاهی قادر به تخمین ۹۳٪ تیپ‌های گیاهی بوده است. شن و همکاران (۲۰۱۳) از رهیافت منابع چندگانه^۱ برای برآورد بی‌درنگ درصد پیشرفت مراحل فنولوژی ذرت در سطح سه ایالت آمریکا استفاده کردند. بردار سه آرایه‌ای این مشخصه‌ها به عنوان لایه قابل رویت و درصد استاندارد شده پیشرفت فنولوژیکی ذرت به عنوان لایه پنهان در نظر گرفته شد. روش آن‌ها برخلاف رهیافت پیشنهادی ویوی و همکاران (۱۹۹۲) که به صورت پیکسل به پیکسل به برآورد فنولوژی می‌پردازد، به صورت کلی و براساس میانگین طیفی و دو مشخصه دیگر اجرا شده است و حسب نتایج از انعطاف‌پذیری و دقت بیشتری برخوردار است. گائو و همکاران (۲۰۱۷) مطالعه‌ای موردی در یک مزرعه ذرت با هدف بررسی ارتباط بین معیارهای فنولوژی مستخرج از داده‌های چندزمانه سنجنده اسپکترورادئومتر تصویربردار با قدرت تفکیک متوسط یا MODIS^۲ و درصد پیشرفت انجام دادند. نتایج نشان‌گر یک ارتباط قوی بین نقاط عطف منحنی NDVI و داده‌های پیشرفت مرحله فنولوژی گزارش شده توسط سرویس ملی تامین آماره‌های کشاورزی یا NASS^۳ است. یافته‌های مطالعه تایید کرد که فنولوژی و فیزیولوژی گیاه در مقیاس مزرعه قابل پایش توسط تصاویر ماهواره‌ای (تفکیک مکانی ۳۰ متر) است. در تحقیق حاضر از ساختار بی‌زی مدل مارکف پنهان برای مدل‌سازی درصد پیشرفت فنولوژی ذرت در مقیاس مزرعه استفاده شد. دو مشخصه یکی مبتنی بر تصاویر سنجنش از دور و دیگری داده‌های هواشناسی به عنوان لایه قابل رویت بکار گرفته شده‌اند. هدف اصلی در این پژوهش، بررسی میزان مهارت مدل مارکف پنهان در برآورد بی‌درنگ فنولوژی ذرت است.

مواد و روش‌ها

منطقه مورد مطالعه، مزرعه تحقیقاتی پردیس کشاورزی و منابع طبیعی دانشگاه تهران در کرج با طول و عرض جغرافیایی به ترتیب ۵۰ درجه و ۵۶ دقیقه و ۵۸ ثانیه شرقی و ۳۵ درجه و ۴۸ دقیقه و ۱۹ ثانیه شمالی و ارتفاع ۱۳۲۰

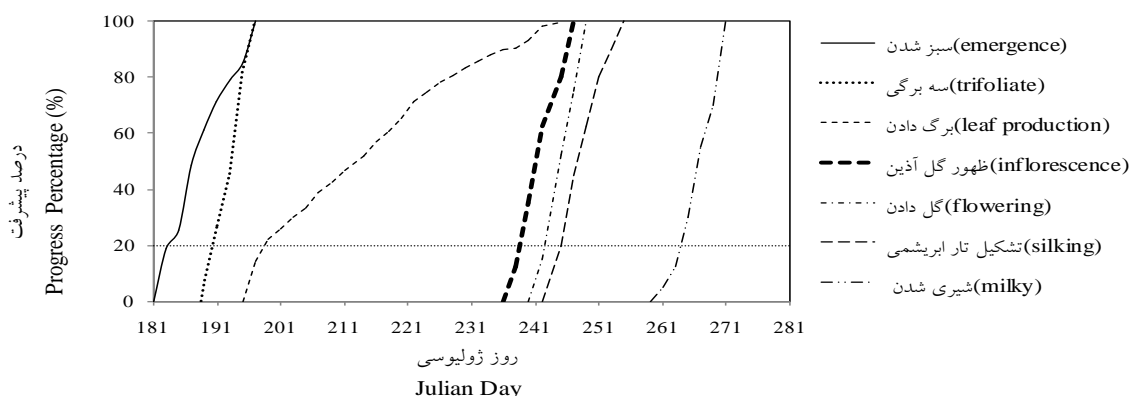
^۱ Multisource

^۲ Moderate Resolution Imaging Spectroradiometer

^۳ National Agricultural Statistics Service

- داده‌های بیشینه و کمینه روزانه دمای هوا دریافت شده از ایستگاه همدیدی کرج در مجاورت مزرعه به همراه دماهای کاردینال گیاه ذرت برای محاسبه واحدهای گرمایی.

تصویر در سال‌های ۲۰۰۲ و ۲۰۰۳ و بیشترین تعداد معادل ۲۳ تصویر در سال ۲۰۰۵ است.



شکل ۱- درصد پیشرفت مراحل مختلف فنولوژی ذرت از سبز شدن تا شیری شدن در سال ۲۰۰۲.

اول هنگامی است که دمای میانگین روزانه (T) پایین‌تر از دمای کمینه موثر و یا بالاتر از دمای بیشینه موثر است که واحد گرمایی معادل صفر می‌شود و به عبارت دیگر، آن روز نقشی در پیشرفت فنولوژیک ندارد. میزان واحد گرمایی در دو وضعیت دیگر طبق معادله ۱ به دست می‌آید.

$$TT = \begin{cases} T - T_{min} & , T_{min} \leq T \leq T_{opt} \\ \frac{(T_{max} - T)(T_{opt} - T_{min})}{(T_{max} - T_{opt})} & , T_{opt} < T \leq T_{max} \end{cases} \quad (1)$$

پس از محاسبه واحد گرمایی در هر روز، واحدهای گرمایی تجمعی یا AGDD محاسبه می‌شوند.

ب) NDVI: نمایه NDVI پاسخ طیفی تصاویر ماهواره‌ای به میزان سبزیگی گیاه است. این نمایه پس از انجام پیش‌پردازش روی تصاویر ماهواره‌ای برای هر پیکسل واقع در مزرعه محاسبه و سپس برای دستیابی به یک الگوی معرف فصل رشد میانگین مقادیر روی پیکسل‌های مختلف برآورد شد. مقادیر NDVI دارای نوبه‌های فراوانی در تصاویر چندزمانه است. علاوه بر این، محدودیت تعداد تصاویر باعث می‌شود که در همه روزها دسترسی به مقدار NDVI وجود نداشته باشد. از این رو به منظور تولید سری‌های زمانی با کیفیت بالا از نمایه‌های پوشش گیاهی، از تابع لجستیک دوگانه^۵ (وو و همکاران^۶، ۲۰۱۴) در ترکیب با روش رگرسیون وزنی^۷ (آرور و همکاران^۸، ۲۰۰۸) استفاده شده است.

۱- استخراج مشخصه‌ها

مشخصه‌های چندگانه NDVI و AGDD^۱ به عنوان ورودی مدل مارکف پنهان استفاده شده‌اند. این مشخصه‌ها برای مزرعه مورد مطالعه در طی دوره رشد ذرت استخراج شدند که دارای منحنی توزیع مختلف می‌باشند. در شکل ۲ منحنی تغییرات دو مشخصه طی فصل رشد ذرت در سال ۲۰۰۴ برای نمونه آورده شده است. منحنی NDVI یک منحنی تک مدی^۲ است و منحنی AGDD یک منحنی صعودی و یکنواخت^۳ است. مقدار NDVI از تصویر سنجش از دور و AGDD از دیدبانی‌های دما در ایستگاه همدیدی مجاور مزرعه استخراج شده است. فرآیند اجرای هر مشخصه در ادامه به تفکیک آمده است.

الف) AGDD: در مطالعات نمو گیاه، تأثیر دما با کمیت درجه-روز رشد تجمعی بیان می‌شود. بنابراین AGDD یک متغیر توصیفی از نحوه تامین گرمای موردنیاز بین تاریخ شروع یک مرحله تا تاریخی معین به صورت بخشی از چرخه زندگی گیاه است. بنابراین اساس محاسبه AGDD بر محاسبه واحدهای گرمایی در طول فصل رشد استوار است. طبق تعریف استرک و همکاران^۴ (۲۰۰۸) سه دمای کاردینال دمای کمینه موثر (T_{min})، دمای بیشینه موثر (T_{max}) و دمای بهینه (T_{opt}) در محاسبه واحدهای گرمایی نقش ایفا می‌کنند. با لحاظ این سه دما، سه وضعیت برای محاسبه واحد گرمایی وجود دارد، وضعیت

⁵ Double Logistic Function

⁶ Wu et al.

⁷ Weighted Regression Method

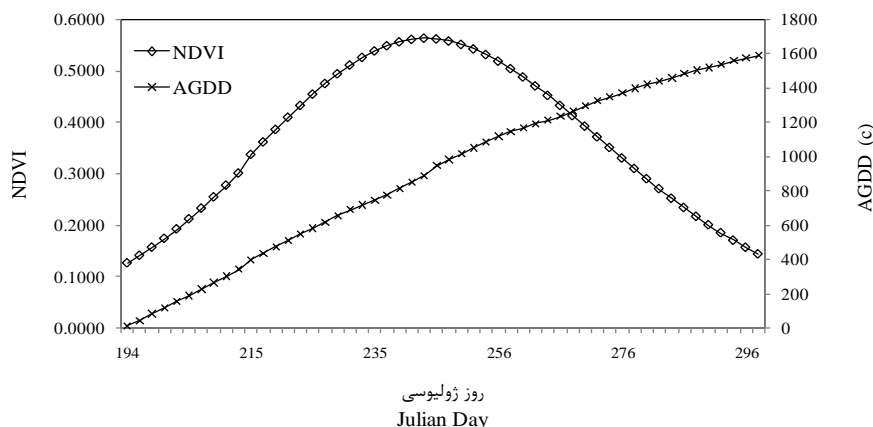
⁸ Arvor et al.

¹ Accumulated Growth Degree Day

² Unimodal

³ Monotonic

⁴ Streck et al.



شکل ۲- منحنی تغییرات NDVI و AGDD در طول فصل رشد ذرت در سال ۲۰۰۴

توالی زمانی است که با متغیر $q_t, (t=1, \dots, T)$ نمایش داده می‌شود. برای مثال درصد پیشرفت مرحله $S_i, (i=1, \dots, N)$ در زمان t به صورت $q_t = S_i$ بیان می‌شود. متناظر با توالی زمانی درصد پیشرفت، توالی مشاهدات به صورت $O = \{O_1, O_2, \dots, O_T\}$ وجود دارد. طبق تعریف مدل مارکف پنهان، درصد پیشرفت فنولوژی ذرت، براساس سه پارامتر $\lambda = (A, B, \Pi)$ تعریف می‌شود، که A ماتریس احتمال انتقال است و با زمان متغیر است و به صورت $A = a_{ij}(t) = P(q_t = S_j | q_{t-1} = S_i), (i, j = 1, \dots, N)$ نمایش داده می‌شود که بیانگر احتمال انتقال از مرحله S_i به S_j است. پارامتر B توزیع احتمال انتشار است و به صورت $B = b_j(O_t) = P(O_t | q_t = S_j)$ نمایش داده می‌شود و نشانگر احتمال تولید مشاهده O_t از مرحله S_j در زمان t است. می‌توان از یک توزیع آماری (مثلاً گسین) برای بیان این پارامتر استفاده کرد. پارامتر Π توزیع احتمال آغازین و به صورت $\Pi = \pi_i = P(q_1 = S_i)$ توصیف می‌شود و بیانگر احتمال قرارگیری مدل در ابتدای فرآیند (زمان $t=1$) در حالت S_i است. طبق الگوریتم شن و همکاران (۲۰۱۳) این پارامتر می‌تواند به احتمال پیشین^۳ مرحله S_i در زمان t تعمیم داده شود یعنی $\pi_i(t) = P(q_t = S_i)$. با دستیابی به سه پارامتر فوق و فرض مارکف مرتبه یک در توالی پنهان، طبق تعریف آمار بیزی، توزیع احتمال توام به ازای کل توالی‌های زمانی از معادله ۲ به دست می‌آید.

$$P(q_1 = S_1, \dots, q_t = S_j, O_1, \dots, O_t | \lambda) = \pi_{q_1} \cdot \prod_{k=2}^t a_{q_{k-1}, q_k}(k) \cdot \prod_{k=1}^t b_{q_k}(O_k) \quad (2)$$

توصیف این روش به طور کامل در پژوهشی توسط (قمقامی و همکاران، ۲۰۱۷) آورده شده است. با بکارگیری این الگوریتم سری زمانی هموار و بدون نوفه از این نمایه پوشش گیاهی در طی فصل رشد ذرت به دست آمد.

۲- تعریف مدل مارکف پنهان

مدل‌های مارکف پنهان تعمیم یافته مدل‌های مارکف هستند که در آن‌ها فرض می‌شود زنجیر مارکف به طور مستقیم قابل مشاهده نبوده و به عبارتی پنهان است (جکولا^۱، ۲۰۰۶). بنابراین، مدل مارکف پنهان یک فرآیند تصادفی دوگانه بوده و شامل دو سطح عدم قطعیت است: فرآیند تصادفی پنهان که به طور مستقیم قابل رویت نیست و تنها از طریق مجموعه دیگری از فرآیندهای تصادفی که توالی مشاهدات را تولید می‌کند قابل رویت است (الیوت و سیو^۲، ۲۰۱۲). در مدل به کار رفته در تحقیق حاضر، لایه قابل رویت شامل بردار دوبعدی از مشخصه‌های NDVI و AGDD است و لایه پنهان توالی زمانی درصد پیشروی فنولوژی ذرت است. درصد پیشرفت ذرت یک فرآیند مبتنی بر مدل مارکف فرض می‌شود که دارای N حالت پنهان است $S = \{S_1, S_2, \dots, S_N\}$ که شامل ۸ مرحله مختلف فنولوژی است: قبل از رشد (S_1)، سبز شدن (S_2)، سه‌برگی (S_3)، برگ‌دهی کامل (S_4)، تشکیل گل‌آذین (S_5)، گل‌دهی (S_6)، تشکیل تارهای ابریشمی (S_7) و شیری شدن (S_8). مرحله قبل از رشد حد فاصل کاشت تا سبز شدن می‌باشد و برای تسهیل طراحی مدل به مراحل فنولوژی اضافه می‌شود. این مراحل دارای یک

³ Priori probability

¹ Jaakkola

² Elliott and Siu

۳- مدل ترکیبی^۱ در مارکف پنهان

تئوری مدل ترکیبی در ساختار مارکف پنهان توسط جکولا (۲۰۰۶) بیان و توسط ویوی و همکاران (۱۹۹۲) در تحلیل پوشش گیاهی و شن و همکاران (۲۰۱۳) در برآورد درصد پیشرفت فنولوژی به کار گرفته شده است. در این مدل، فرآیند تصادفی لایه پنهان از زنجیر مارکف و پارامتر احتمال انتشار از یک توزیع آماری ترکیبی تبعیت می‌کند. مطابق نظرات ویوی و همکاران (۱۹۹۲) مارکف پنهان با توزیع ترکیبی، عدم قطعیت را در هر دو سطح فرآیند کاهش می‌دهد. همچنین سریهاری^۲ (۲۰۱۱) بیان کرد که در یک مدل مارکف پنهان با توزیع ترکیبی، مشاهدات در هر گام زمانی می‌توانند توسط مراحل متعدد پنهان (در اینجا مراحل فنولوژی) تحت تأثیر قرار گیرند و این مهم، انعطاف پذیری مدل را بالا می‌برد. براین اساس احتمال مشاهده در زمان t می‌تواند از معادله ۳ به دست آید.

$$P(O_t) = \sum_{i=1}^N \pi_i(t) \cdot b_i(O_t) \quad (3)$$

ضریب $\pi_i(t)$ (احتمالات پیشین) به عنوان وزن مولفه λ در نظر گرفته می‌شود که $\sum_{i=1}^N \pi_i(t) = 1$. چنانچه گفته شد مدل ترکیبی به دلیل ماهیت غیرخطی باعث افزایش انعطاف‌پذیری و دقت در مدل‌سازی آماره‌های اساسی پیشرفت فنولوژی می‌شود. در تحقیق حاضر به دلیل پیوستگی مشاهدات از توزیع آماری گسن برای برآورد تابع چگالی احتمال یعنی $b_i(O_t)$ و سپس از احتمالات پیشین به عنوان وزن‌هایی برای تعریف مدل ترکیبی گسن استفاده شده است.

۴- نرمال سازی درصد پیشرفت فنولوژی ذرت

درصد پیشرفت مرحله S_i فنولوژی ذرت در زمان t با p_i^t نشان داده می‌شود. این مقدار معرف درصد پیشرفت گیاه در آن مرحله است. به عبارت دیگر اگر این مقدار معادل ۲۰ درصد باشد به این معنی است که ۲۰ درصد گیاهان وارد مرحله معین S_i شده‌اند. این درصد معرف میزان پیشرفت است اما معرف نسبت تصرف^۳ هر مرحله در یک زمان معین نیست (شن و همکاران، ۲۰۱۳). برای اینکه مقادیر درصد پیشرفت به نسبت پیشرفت تبدیل

شوند یک نرمال‌سازی باید انجام شود. در تحقیق حاضر این نسبت‌های پیشرفت به عنوان ماتریس احتمالات پیشین $\pi_i(t)$ نیز در نظر گرفته شدند. قبل از محاسبه این نسبت‌ها، ابتدا باید همه درایه‌های ماتریس درصد پیشرفت را تکمیل کرد. از این رو به ازای هر مرحله معین، اگر فرآیند رشد گیاه هنوز به آن مرحله نرسیده است، مقدار درصد پیشرفت معادل صفر و اگر فرآیند رشد گیاه آن مرحله را پشت سر گذاشته باشد، مقدار درصد پیشرفت معادل ۱ یا ۱۰۰ در نظر گرفته می‌شود. با توجه به ماهیت توالی مراحل فنولوژی گیاه، فرض می‌شود که هر مرحله تنها به خود و مرحله بعد از خود انتقال می‌یابد و از این رو نسبت پیشرفت یا همان احتمالات پیشین به طور مستقیم از روی درصدهای پیشرفت قابل محاسبه خواهند بود.

$$\pi_i(t) = \begin{cases} p_i^t & , \text{if } i=N \\ p_i^t - p_{i+1}^t & , \text{if } i \neq N \end{cases} \quad (4)$$

در حالی که طبق شکل ۱ منحنی تغییرات زمانی درصد پیشرفت یک منحنی صعودی است، انتظار می‌رود که منحنی تغییرات زمانی احتمالات پیشین عموماً یک منحنی تک‌مدی باشد، به این معنی که احتمال پیشین حالت معین i با گذر زمان افزایش و سپس کاهش یابد. چنانچه در بخش ۳ گفته شد، حاصل جمع احتمالات پیشین در زمان معین t و به ازای حالات مختلف معادل یک است و بدیهی است که احتمال پیشین یا نسبت پیشرفت حالت معین i در زمان معین t هر چه بزرگتر باشد، احتمال اینکه مدل مارکف در آن زمان در حالت i قرار گیرد بیشتر خواهد بود.

۵- برآورد پارامترهای مدل مارکف پنهان

پارامترهای مدل مارکف پنهان عبارتند از توزیع احتمال اولیه، ماتریس احتمالات انتقال و توزیع احتمال انتشار که می‌توانند از روی داده‌های میدانی طبق فرآیند زیر برآورد شوند.

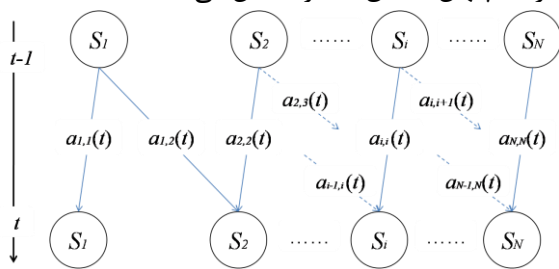
الف) ماتریس احتمال پیشین یا اولیه: توزیع احتمال اولیه یا احتمالات پیشین، حالت مدل را هنگامی که مشاهدات در نظر گرفته نشوند توصیف می‌کند (ویوی و همکاران، ۱۹۹۲). برای برآورد احتمال هر مرحله در شروع فصل رشد نیاز به تعریف مرحله «قبل از رشد» وجود دارد. این مرحله که از تاریخ کاشت تا شروع دوره جوانه‌زنی است بین ۲ تا ۱۰ روز در سال‌های مختلف متغیر است. برای هر

¹ Mixture model

² Srihari

³ Occupancy ratio

۱۹۹۲؛ سکز و کوچریک^۴، ۲۰۱۱). شکل ۴، انتقال مراحل فنولوژی ذرت را در طول چرخه حیات خود براساس مدل مارکف پنهان تحقیق حاضر نمایش می‌دهد.



شکل ۳- توصیف فرآیند انتقال مارکف مراحل فنولوژی ذرت در طول فصل رشد

اگر ماتریس احتمال انتقال به عنوان یک پارامتر محلی در نظر گرفته شود، بنابراین ماتریس A دارای درایه‌های $a_{ij}(t)$ می‌باشد که یک ماتریس سه بعدی است. از آنجا که هر حالت فنولوژی در گام زمانی یا به خود یا به حالت بعدی انتقال پیدا کند (شکل ۴)، می‌توان چرخه حیات ذرت را یک چرخه تک‌مدی فرض کرد (شن و همکاران، ۲۰۱۳). در این صورت درایه‌های ماتریس احتمال انتقال به راحتی و به طور مستقیم از روی ماتریس احتمال پیشین قابل محاسبه می‌باشد به صورت:

$$a_{ij}(t) = \begin{cases} 1 & , \text{if } i=j=N, \pi_N(t-1) > 0 \\ 1 - \sum_{k=i+1}^N \frac{(\pi_k(t) - \pi_k(t-1))}{\pi_i(t-1)} & , \text{if } i=j \neq N, \pi_i(t-1) > 0 \\ \sum_{k=i+1}^N \frac{(\pi_k(t) - \pi_k(t-1))}{\pi_i(t-1)} & , \text{if } i=j-1, \pi_i(t-1) > 0 \\ 0 & , \text{else} \end{cases} \quad (6)$$

که $\sum_{j=1}^N a_{ij}(t) = 1$ در توصیف چهار وضعیت موجود در معادله

۶ باید گفت که وضعیت اول مربوط است به احتمال انتقال به خود در انتهای زنجیر، وضعیت دوم مربوط است به احتمال انتقال به خود در طول زنجیر، وضعیت سوم مربوط است به تغییر حالت رو به جلو و وضعیت چهارم شامل بیشتر درایه‌های ماتریس احتمال انتقال می‌شود. به عنوان مثال اگر $q_t = S$ سپس همه انتقال‌ها به جز $a_{6,6}(t)$ و $a_{6,7}(t)$ صفر هستند که این دو درایه به ترتیب مربوط می‌شوند به وضعیت‌های دو و سه در معادله ۶ و جمع آن‌ها یک است. برای دستیابی به توزیع احتمالات انتقال برای کل سری زمانی رشد گیاه، دو گام عملیاتی زیر پیگیری شده است: (۱) میانگین ماتریس احتمالات

سال در دوره آموزش مدل، این ماتریس طبق معادله ۴ محاسبه می‌شود و برای هر روز از شروع فصل رشد اقدام به میانگین‌گیری روی همه سال‌ها می‌شود. طول فصل رشد (فاصله زمانی بین آغاز سبز شدن تا پایان شیری شدن) در سال‌های مختلف بین ۷۹ تا ۱۱۰ روز متغیر است، از این رو برای حذف اثر طول فصل رشد اقدام به میانگین‌گیری وزنی به صورت معادله ۵ شده است.

$$\bar{\pi}_i(t) = \frac{\sum_{k=1}^M \frac{1}{I_k^k(t)}}{\sum_{k=1}^M \frac{1}{I_k}} \quad (5)$$

که M تعداد سال‌ها در دوره آموزش مدل و I_k طول فصل رشد متناظر با سال kام می‌باشد. با توجه به وابستگی پارامتر توزیع احتمال پیشین به دو متغیر زمان و حالت، این پارامتر به صورت یک ماتریس دو بعدی بیان می‌شود که سطرهاى آن معادل گام زمانی و ستون‌های آن حالت‌های مختلف فنولوژی است.

ب) ماتریس احتمال انتقال حالت‌ها: در مدل مارکف پنهان مرسوم، ماتریس احتمال انتقال حالات یا A یک ماتریس دو بعدی و فراگیر^۱ است. به عبارت دیگر همه روزها طی فصل رشد دارای یک ماتریس مشابه هستند. ولی در مورد مدل‌سازی مراحل فنولوژی گیاهان از جمله ذرت این امر نادرست است. در این مورد باید به ماتریس احتمال انتقال اجازه داد که با زمان تغییر کند و عملکردی مشابه با یک مدل ناهمگن مارکف پنهان داشته باشد (نب و همکاران، ۲۰۰۳). بنابراین ماتریس احتمال انتقال مدل فنولوژی تحقیق حاضر یک ماتریس وابسته به زمان یا محلی^۲ است که در طی فصل رشد تغییر می‌کند. برای مثال در شروع چرخه زندگی، ذرت تمایل به ماندن در اولین حالت را دارد به عبارتی احتمال انتقال از حالت فعلی به خودش بیشتر و به حالت بعد از خود کمتر است. با این وجود با سپری شدن زمان، احتمال انتقال به خود ضعیف و به حالت بعدی بیشتر می‌شود. به طور کلی، فرآیند انتقال فنولوژیک به سازوکارهای بیوفیزیک و عوامل بیرونی مانند میزان آبیاری، ویژگی‌های خاک، عوامل هواشناسی و شیوه‌های مدیریتی وابسته است (ویوی و همکاران،

¹ Global

² Knab et al.

³ Time-dependent or Local

⁴ Sacks and Kucharik

$$\ln L(\Theta|O_t) = \sum_{t=1}^T \ln \sum_{j=1}^N \pi_j(t) \cdot N_j(O_t | \mu_j, \Sigma_j) \quad (8)$$

بردار احتمالات پیشین نیز چنانچه گفته شد از روی مشاهدات زمینی به دست می‌آید. برای رسیدن به همگرایی در برآورد پارامترهای معادله ۷ می‌توان از الگوریتم‌های تکرار^۲ مانند بیشینه‌سازی امید ریاضی یا EM^۳ (دمستر و همکاران^۴، ۱۹۷۷) استفاده کرد. این الگوریتم دو گام را برای دریافت یک جواب با بیشترین مقدار در معادله ۸ طی می‌کند: (۱) E-step و (۲) M-step. الگوریتم با مقادیر اولیه برای پارامترهای μ_i و Σ_i اجرا می‌شود و سپس دو گام فوق را تا رسیدن به همگرایی در بیشینه محلی تابع درست‌نمایی ادامه می‌دهد. مبنای محاسبه پارامترها در تکرارهای بعدی، برآورد کمیت تکرارپذیر $\beta_i^q(t)$ است که q نشان‌دهنده شماره تکرار است. این کمیت در حقیقت به صورت یک ضریب عمل نموده و در هر تکرار جدید این ضریب تعدیل می‌شود تا اینکه بهترین برآورد را از پارامترهای تابع داشته باشد.

۶- برآورد درصد پیشرفت فنولوژی

هدف تحقیق حاضر برآورد درصد پیشرفت مراحل فنولوژی ذرت است. این مساله با محاسبه احتمالات پسین^۵ پسین^۵ روی حالات پنهان در هر زمان معین t به ازای پارامتر λ و همه مشاهدات تا زمان فعلی می‌تواند پیگیری شود (یعنی $P(q_t = S_j | O_1, \dots, O_t)$). این برآورد به صورت بی‌درنگ می‌باشد و در هر گام زمانی (در تحقیق حاضر، روز است) به ازای مشاهدات تا آن گام زمانی و پارامترهای مدل با بکارگیری الگوریتم پالایش^۶ انجام می‌شود. در اینجا لازم است تأکید شود که سه وضعیت برای فرآیند برآورد درصد پیشرفت فنولوژی به ازای مشاهدات O_1, \dots, O_h وجود دارد: (۱) اگر $t < h$ از الگوریتم هموارسازی^۷ هموارسازی^۷ استفاده می‌شود، (۲) اگر $t = h$ از الگوریتم پالایش استفاده می‌شود که عبارت است از برآورد بی‌درنگ و (۳) اگر $t > h$ از الگوریتم‌های پیش‌بینی^۸ استفاده می‌شود. در بکارگیری الگوریتم پالایش، احتمالات

پیشین برای سال‌های آموزش (طبق معادله ۱) برای متوسط فصل رشد محاسبه می‌شود و (۲) فرآیند محاسبه توزیع احتمالات انتقال بر روی خروجی گام یک خواهد بود.

(ج) توزیع احتمال انتشار: توزیع احتمال انتشار، احتمال وقوع مشاهدات را از یک حالت معین توصیف می‌کند. در تحقیق حاضر بردار وضعیت مشاهدات شامل دو مشخصه یا متغیر AGDD و NDVI است. مقادیر مشاهدات به طور مستمر در اثر تناوب فنولوژیکی تغییر می‌کنند (شن و همکاران، ۲۰۱۳). در یک واحد زمانی از شروع فصل رشد، هر مرحله دارای احتمال پیشین معینی است و بنابراین چنانچه در بخش ۳ نیز گفته شد، مدل‌سازی احتمالات انتشار مبتنی بر یک مدل ترکیبی است که هر مرحله در احتمال انتشار مشاهده معینی سهم مشخصی دارد. این سهم توسط وزن‌هایی تعیین شده که نسبت فضایی اشغال هر مرحله را تعیین می‌کنند و مطابقت با همان احتمالات پیشین هر مرحله یعنی $\pi_i(t)$ دارند. (شن و همکاران، ۲۰۱۳) از توزیع چندمتغیره گسن به عنوان یک تابع چگالی احتمال برای بیان احتمالات انتشار استفاده کردند. در تحقیق حاضر نیز از همین روش استفاده شده است. بنابراین در معادله ۳ $P(O_t)$ برهم‌نهی خطی^۱ توزیع گسن و $b_i(O_t)$ با بردار میانگین μ_i و ماتریس کوواریانس Σ_i پارامتری می‌شود.

$$b_i(O_t) = N_i(O_t | \mu_i, \Sigma_i) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_i|}} \cdot \exp\left(\frac{-(O_t - \mu_i)' \Sigma_i^{-1} (O_t - \mu_i)}{2}\right) \quad (7)$$

که در آن، d تعداد بعد فضای مشاهدات است که در اینجا به دلیل انتخاب مشخصه‌های AGDD و NDVI معادل ۲ است. معادله ۷ نشان می‌دهد که پارامترهای این تابع، فراگیر و مستقل از زمان هستند. این پارامترها هم می‌توانند از روی مشاهدات زمینی محاسبه شوند و هم می‌توانند ناشناخته در نظر گرفته شوند که باید با روش‌هایی برآورد می‌شوند. در وضعیت اول یعنی محاسبه پارامترها، پارامترهای هر حالت از روی بردار مشاهدات به ازای آن حالت ($0 < \pi_i(t) < 1$) به دست می‌آیند. در شرایط برآورد پارامترها، به ازای توالی معین O_1, \dots, O_T مقادیر پارامترها با استفاده از روش حداکثر درست‌نمایی برآورد می‌شوند. تابع لگاریتمی درست‌نمایی براساس بردار پارامتر $\Theta = \{\mu, \Sigma\}$ از قرار زیر است.

² Iterative algorithm
³ Expectation Maximization
⁴ Dempster et al.
⁵ Posteriori probabilities
⁶ Filtering
⁷ Smoothing
⁸ Prediction

¹ Linear Superposition

ماتریس دو بعدی احتمالات پیشین طی ۸ سال محاسبه شد. در شکل ۵، تغییرات زمانی میانگین احتمالات پیشین برای مراحل مختلف برای سال صحت‌سنجی ۲۰۰۲ با استفاده از دوره واسنجی ۲۰۰۳ تا ۲۰۱۰ آورده شده است. چنانچه مشخص است؛ مرحله «قبل از سبز شدن» دارای افت در احتمال پیشین با گذر زمان بوده و سایر مراحل دارای افت و خیزهای متعدد می‌باشند. این افت و خیزها به صورتی است که در هر واحد زمانی مجموع بردار احتمالات معادل ۱ خواهد بود. محاسبه ماتریس سه بعدی احتمالات انتقال، گام بعدی فرآیند تحقیق بوده است. بیشتر درایه‌های این ماتریس صفر بوده و هر گام زمانی یک ماتریس مختص به خود دارد. مقادیر بزرگ‌تر از صفر ماتریس به طور عمده حول قطر اصلی آن بوده و با گذر زمان از بالای قطر به سمت پایین قطر جابجا می‌شوند. همچنین برای آزمون کارکرد الگوریتم EM در برآورد پارامترهای تابع توزیع ترکیبی گسن $(\Theta = \{\mu, \Sigma\})$ از آزمون‌های آماری t و χ -square استفاده شد. برای این منظور، پارامتر Θ یک بار از روی داده‌های تاریخی محاسبه و بار دیگر با الگوریتم EM برآورد می‌شوند. با مقایسه این دو نمونه، می‌توان به کیفیت عملکرد الگوریتم برآورد دست یافت. در جدول ۱ آماره‌های آزمون‌های آماری به تفکیک مراحل مختلف فنولوژی آورده شده است. برای هر مرحله ۹ بردار میانگین μ و ۹ ماتریس کوواریانس Σ وجود دارد که آزمون آماری روی این مجموعه داده انجام شد. چنانچه مشخص است براساس هر دو آزمون، تا مرحله ظهور گل آذین و همچنین مرحله شیری شدن، برآورد پارامترهای مدل با الگوریتم EM اختلاف معنی داری در هیچکدام از سطوح عدم اطمینان ۰/۰۵ و ۰/۰۱ با مقادیر محاسبه شده ندارد. در دو مرحله گل‌دهی و تشکیل تارهای ابریشمی معنی‌داری در سطح ۰/۰۵ در برخی موارد مشاهده شده است. با این وجود می‌توان اذعان کرد که در سطح عدم اطمینان ۰/۰۱، الگوریتم EM توانسته است تمامی پارامترها را بدون تفاوت معنی‌دار با پارامترهای محاسبه شده از سری اطلاعات تاریخی برآورد نماید. با بکارگیری پارامترهای برآورد شده از الگوریتم EM تابع توزیع احتمال انتشار که از تابع گسن تبعیت می‌کند برآورد شد. در شکل ۶، این توابع برای هر هشت مرحله و همچنین برای تابع توزیع ترکیبی آورده شده است.

پسین طبق قانون احتمالات بیزی از معادله ۹ به دست می‌آیند.

$$P(q_t = S_j | O_1, \dots, O_t) = \frac{P(q_t = S_j, O_1, \dots, O_t)}{\sum_{i=1}^N P(q_t = S_i, O_1, \dots, O_t)} \quad (9)$$

صورت کسر عبارت است از احتمال توأم حالت معین j و توالی مشاهدات تا زمان t که به وسیله الگوریتم پیشرو^۱ به طور مستقیم محاسبه می‌شود.

$$P(q_t = S_j, O_1, \dots, O_t) = \begin{cases} \pi_j \cdot b_j(O_t) & , \text{if } t=1 \\ \left(\sum_{i=1}^N k_i(t-1) \cdot a_{ij}(t) \right) \cdot b_j(O_t) & , \text{if } t>1 \end{cases} \quad (10)$$

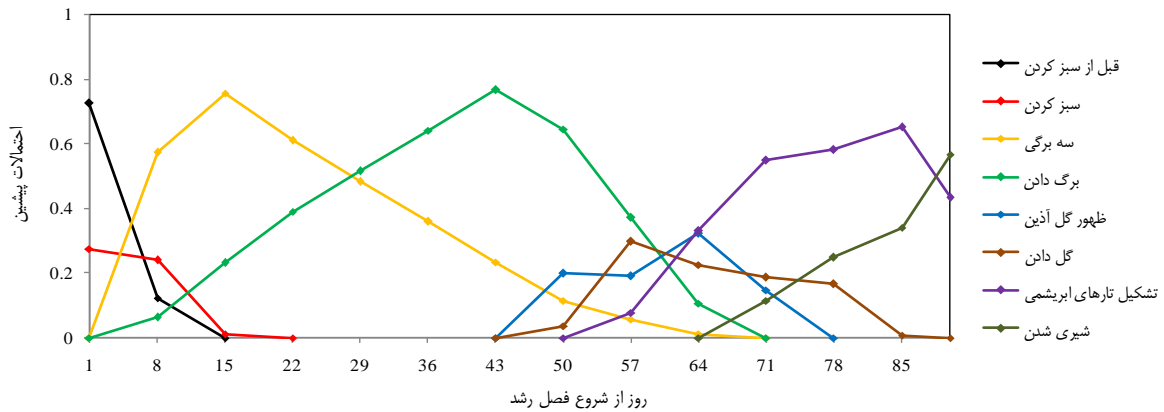
احتمالات پسین همچنین می‌توانند به عنوان نسبت تصرف یک حالت معین نیز در نظر گرفته شوند که با یک تبدیل نرمال‌سازی معکوس می‌توانند به درصد پیشرفت مرحله معین فنولوژی در زمان مشخص یعنی $\delta_i(t)$ تبدیل شوند.

$$\delta_i(t) = \sum_{k=1}^N P(q_t = S_k | O_1, \dots, O_t) \quad (11)$$

نتایج و بحث

با شکل‌دهی چارچوب مدل مارکف پنهان طبق ساختار فنولوژی گیاه ذرت با مشخصه‌های چندگانه، همه هفت مرحله اصلی فنولوژی ذرت و درصد پیشرفت هر مرحله در یک وضعیت بی‌درنگ برآورد شدند. مدل، ۹ بار اجرا شد و در هر بار ۸ سال به مرحله واسنجی و ۱ سال به مرحله صحت‌سنجی اختصاص یافت. در اولین گام، ماتریس احتمالات پیشین برای هر سال محاسبه شد. به عنوان مثال در روز ۵۹ از شروع فصل رشد در سال ۲۰۰۲، درصد پیشرفت مراحل فنولوژی از سبز شدن تا شیری شدن به ترتیب ۰/۱۰۰، ۰/۱۰۰، ۰/۹۹/۵۴، ۰/۴۴/۹۹، ۰، ۰ و ۰ بوده است. طبق معادله ۴ احتمال پیشین در این روز و سال به ترتیب برای مراحل فنولوژی معادل ۰، ۰، ۰/۴۵، ۰/۵۴۶، ۰/۱۰۰ و ۰ به دست آمده است. برای توصیف این احتمال می‌توان گفت که احتمال اولیه اینکه در روز ۵۹ سال ۲۰۰۲ گیاهان در مرحله سه برگی، برگ‌دهی و ظهور گل‌آذین باشند به ترتیب ۰/۴۶، ۰/۵۴/۵۵ و ۰/۴۴/۹۹ است. باقی مراحل در این روز دارای احتمال اولیه صفر بوده به این معنی که شانس برای وقوع در این روز ندارند. در مرحله واسنجی، ابتدا میانگین

¹ Forward algorithm

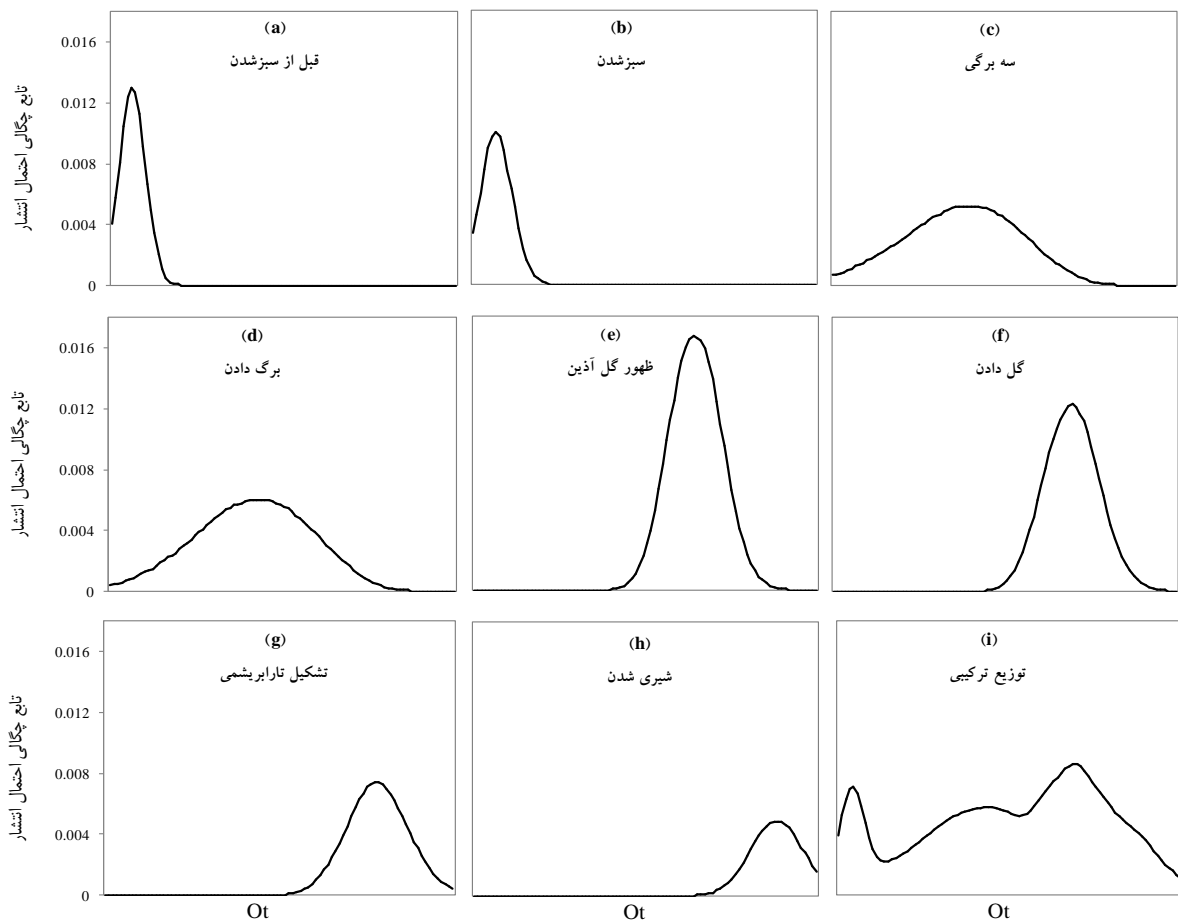


شکل ۴- تغییرات زمانی میانگین مقادیر ماتریس احتمالات پیشین برای سال صحت‌سنجی ۲۰۰۲

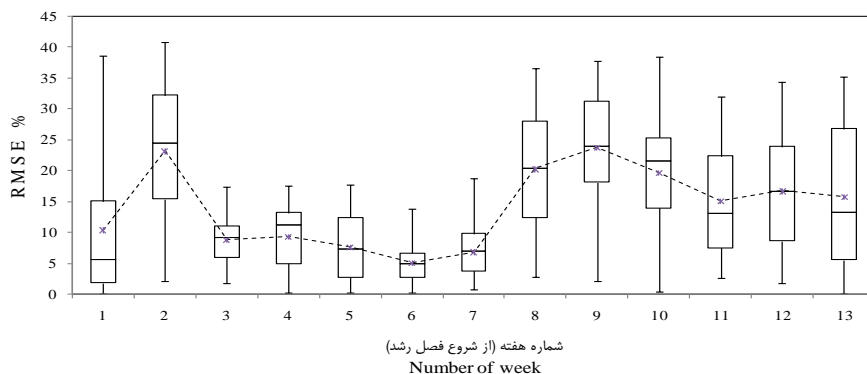
جدول ۱- آماره آزمون های t و χ^2 مقایسه دو روش محاسبه و برآورد پارامترهای مدل ترکیبی

قبل از سبز شدن	سبز شدن	سه برگی	برگ دادن	ظهور گل آذین	گل دهی	تشکیل تارهای ابریشمی	شیری شدن	NDVI	p-value
۰/۸۴۱	۰/۷	۰/۳۰۲	۰/۴۶۶	۰/۰۵۹	*۰/۰۴	۰/۰۶۱	۰/۰۸		
۲/۵۶۱	۰/۴۵	۰/۱۰۶	۰/۱	۰/۲۳۶	۰/۱۴	*۰/۰۱۲	۰/۰۶۳	AGDD	(t-test)
۱/۲۴۱	۰/۹۶۱	۲/۹۶۸	۳/۵۰۸	۳/۲۴۱	*۵/۰۰۳	*۴/۰۶۳	۳/۷۵۹		χ^2

* معنی داری در سطح اطمینان ۹۵٪ و ** معنی داری در سطح اطمینان ۹۹٪



شکل ۵- توابع توزیع احتمال انتشار در مراحل فنولوژی (a-h) به همراه توزیع ترکیبی (i)



شکل ۶- نمودارهای جعبه‌ای و ویسکر^۱ میانگین هفتگی مقادیر RMSE (در سطح اطمینان ۹۵٪)

یک کاهش در روند خطا تا انتهای فصل رشد مشاهده گردید. بیشترین میانگین خطا حدود ۲۵٪ و مربوط به هفته نهم بوده که مقارن است با مراحل از ظهور گل آذین تا تشکیل تارهای ابریشمی (بسته به سال) و کمترین میانگین خطا حدود ۷٪ و مربوط به هفته ششم بوده که مقارن است با اواخر مرحله برگ دهی. با کمک شکل ۵ می‌توان دریافت که هم پوشانی مراحل تأثیر شایانی بر اندازه خطا دارد. چنانچه در شکل ۵ دیده می‌شود، از هفته سوم تا هفته ششم کمترین هم‌پوشانی مراحل فنولوژی وجود دارد و فقط دو مرحله «سه‌برگی» و «برگ‌دهی» در بردار احتمالات پیشین دارای مقداری بزرگ‌تر از صفر هستند. از هفته هفتم به بعد میزان هم‌پوشانی مراحل افزایش می‌یابد. با مقایسه نتایج شکل‌های ۵ و ۶ می‌توان دریافت با افزایش هم‌پوشانی مراحل، خطای مدل افزایش می‌یابد. شکل ۶ ترکیب همه سال‌های مورد مطالعه است. در حالی که دامنه خطا در سال‌های مختلف متفاوت است. به‌عنوان یک ارزیابی کلی، میانگین RMSE برای سال‌های مختلف همواره کوچک‌تر از ۲۰٪ و بین ۸/۸٪ و ۱۹/۴٪ متغیر بوده است. تحقیق حاضر در مقیاس مزرعه انجام شده است، با این وجود می‌توان نتایج تحقیق حاضر را با مطالعاتی که روی درصد پیشرفت فنولوژی در مقیاس منطقه‌ای تمرکز داشته‌اند، مقایسه نمود. برای مثال می‌توان به تحقیق یو و همکاران^۲ (۲۰۱۲) اشاره کرد که یک روش مبتنی بر کرنل را برای برآورد درصد پیشرفت ذرت بکاربردند. مقدار RMSE روش آن‌ها هیچ‌گاه از ۲۴٪ کمتر نشد. شن و همکاران (۲۰۱۳) نیز روش مورد استفاده در تحقیق را (با اندکی تغییرات) بکار بردند و برای سه ایالت آمریکا میانگین RMSE معادل ۱۳/۲۷٪

هر چه کشیدگی منحنی کمتر باشد به معنی کوتاه‌تر بودن مرحله فنولوژی است. پایین‌ترین منحنی سمت راست مربوط به توزیع ترکیبی است که برای اجرای الگوریتم برآورد درصد پیشرفت فنولوژی در نظر گرفته شده است. از ریشه میانگین مربعات خطا یا RMSE برای سنجش مهارت مدل بیزی در برآورد درصدهای پیشرفت فنولوژی استفاده شده است (هیج و همکاران^۳، ۲۰۰۴). این سنجه بیانگر اختلاف بین مقادیر برآورد شده و مشاهده شده است. مقادیر کوچک‌تر نشان‌دهنده واریانس باقیمانده کمتر هستند. ارزیابی RMSE فصل رشد ذرت را از «سبز شدن» تا «شیری شدن» پوشش می‌دهد. مرحله «قبل از سبز شدن»، در ارزیابی خطا در نظر گرفته نمی‌شود چراکه برای تسهیل در اجرای مدل به عنوان لایه مرزی تعریف شده است. شکل ۷، دامنه تغییرات RMSE را نشان می‌دهد که مربوط به همه هفت مرحله فنولوژی است. نتایج به صورت هفتگی بوده و شامل هر ۹ اجرای مدل می‌باشد. نتایج گزارش‌شده در نمودار جعبه‌ای معرف سطح احتمال ۹۵ درصدی اطمینان و نتایج گزارش شده در نمودار خطی معرف میانگین گام زمانی (هفته) می‌باشد. با تحلیل خطاها در فرآیند تعیین درصد پیشرفت از شکل ۷ دریافته شد که در هفته اول که همزمان با «سبز شدن» است میانگین خطا حدود ۱۰٪ بوده که با یک افزایش در هفته دوم همراه است. سپس از هفته سوم میزان خطا کاهش یافته و این روند تا هفته هفتم پس از شروع فصل رشد ادامه دارد. متوسط خطاها در این بازه زمانی زیر ۱۰٪ می‌باشد. از هفته هشتم روند خطا یک روند افزایشی بوده که تا هفته نهم ادامه داشته و سپس

¹ Box and Whisker Plots

² Heij et al.

³ Yu et al.

کردن زیرروالی به مدل که نسبت به طول فصل عکس العمل نشان دهد می‌تواند به کارآمدی آن کمک کند.

ب) کیفیت داده‌های فنولوژی دیدبانی شده: داده‌های دیدبانی شده اصولاً وابستگی زیادی به ارزیابی شخصی دیدبان دارند. از این رو خطای اریبی به طور اجتناب‌ناپذیری در برداشت این اطلاعات وجود خواهد داشت. علاوه بر این، وجود خلأهای آماری ناشی از جنبه‌های اقلیمی و مدیریتی باعث بکارگیری روش‌های بازسازی می‌شود که این روش‌ها می‌توانند بخشی از منابع عدم قطعیت را با خود به همراه داشته باشند.

ج) کیفیت داده‌های سنجش از دور: نوبه‌های ناشی از پوشش ابرناکی، داده‌های مفقود، پیکسل‌های بی‌کیفیت و یا برخی خطاهای سامانمند^۱ به راحتی می‌توانند کیفیت داده‌های اولیه سنجش از دور را کاهش دهند؛ هر چند با بکارگیری روش‌های هموارسازی تا حدی می‌توان بر این مشکل فائق آمد.

د) روش محاسبه AGDD: هر چند در این تحقیق تلاش شد که آستانه‌های بالا، بهینه و پایین دمایی برای محاسبه AGDD در نظر گرفته شوند، با این وجود بکارگیری روش خطی برای این امر ممکن است در هر مجموعه داده‌ای صحیح نباشد. به عبارت دیگر ترکیب روش‌های خطی و غیر خطی برای محاسبه واحدهای گرمایی می‌تواند در کاهش منابع عدم قطعیت موثر باشد.

ه) بی‌نظمی در الگوهای دما و بارش: تغییر الگوی دمایی به طور مستقیم بر تأمین نیاز گرمایی گیاه و به تبع آن بر طول فصل رشد تأثیر می‌گذارد. حتی با وجود کشت فاریاب ذرت، تغییر الگوهای بارش و وقوع خشکسالی‌های شدیدتر می‌تواند ناهمگنی در رشد ذرت طی سال‌های مختلف ایجاد نماید. دلیل آن این است که، رخداد حدی در یک سال خاص اتفاق می‌افتد و روی ثبات نتایج تأثیر می‌گذارد. این امر به دلیل تأثیر این بی‌نظمی بر برآورد پارامترهای مدل مانند ماتریس احتمالات انتقال در مرحله آموزش است.

و) تابع چگالی احتمال انتشار: این تابع، انتشار لایه قابل رویت را مدیریت می‌کند و میزان برآزش آن بر داده‌ها می‌تواند بر کیفیت نتایج تأثیر داشته باشد. در تحقیق حاضر فرض شده است که برای هر مرحله فنولوژی، بردار قابل رویت از توزیع نرمال تبعیت می‌کند. توزیع نرمال

۱۶/۱۴٪ و ۱۲/۹٪ را به دست آوردند که با میانگین RMSE به دست آمده در تحقیق حاضر (۱۴٪) کاملاً قابل قیاس است. مزیت مطالعه شن و همکاران (۲۰۱۳) دسترسی به داده‌های همگن و مدیریت شده (دریافت شده از NASS) بود که تأثیر بسزایی در کاهش خطا دارد. داده‌های مربوط به فنولوژی در تحقیق حاضر بسیار ناهمگن از نظر زمان کاشت، جنبه‌های مدیریتی و کیفیت داده برداری بودند که نتایج را بسیار تحت تأثیر قرار داد. باتوجه به نتایج می‌توان گفت که مدل استفاده شده در تحقیق حاضر قادر به برآورد بی‌درنگ درصد پیشرفت فنولوژی ذرت است. مراحل فنولوژی به طور معمول دارای هم‌پوشانی هستند. یعنی زمانی که برخی بوته‌ها وارد مرحله برگ‌دهی شده‌اند، هنوز برخی از بوته‌ها ممکن است در مرحله سه برگگی باقی مانده باشند. در نظر گرفتن این جنبه فنولوژی در روش‌های مدل‌سازی غیربیزی ناممکن است. با این وجود عوامل زیر تأثیر در دریافت نتایج داشته‌اند که بهبود آن‌ها می‌تواند منجر به نتایج بهتر شود.

الف) طول فصل رشد: طول فصل رشد که حد فاصل سبز شدن تا انتهای شیری شدن است، نقش بسزایی در برآورد درصد پیشرفت فنولوژی دارد. در داده‌های فنولوژی دیدبانی شده، طول فصل رشد از ۷۹ تا ۱۱۰ روز متغیر بوده است و این امر، نوعی ناهمگنی در داده‌های سال‌های مختلف ایجاد می‌کند. در تحقیق حاضر برای تسهیل در اجرای مدل، به میانگین طول فصل رشد (۹۰ روز) بسنده شد. به این معنی که همه سال‌ها به طور پیش‌فرض دارای ۹۰ روز فصل رشد ذرت هستند. هر چند با بکارگیری میانگین‌گیری وزنی از ماتریس احتمالات پیشین سعی شد تا حدی اثر طول فصل رشد جبران شود با این وجود کم برآوردی‌ها و بیش‌برآوردی‌ها در درصد پیشرفت منجر به تشدید خطای مدل می‌شود. برای مثال در سال ۲۰۰۹ که طول فصل رشد ذرت ۷۹ روز است، روز ۷۹ به عنوان انتهای شیری شدن (درصد پیشرفت شیری شدن معادل ۱۰۰٪ است) در نظر گرفته شده است. با این وجود درصد پیشرفت شیری شدن در این روز توسط مدل معادل ۳۰٪ برآورد شده است. حالت عکس این امر زمانی است که طول فصل رشد دیدبانی شده بزرگتر از میانگین باشد. در اینجا عدم انعطاف‌پذیری مدل در برخورد با مقادیر مختلف طول فصل رشد روشن می‌شود. برای حل این معضل، اضافه

^۱ Systematic

پارامترها با زمان تغییر کنند؛ ۲) به دلیل پیوستگی لایه قابل رویت (مشخصه‌های چندگانه)، از یک تابع چگالی احتمال انتشار به جای ماتریس احتمالات انتشار استفاده شده است. ۳) از آنجا که مراحل مختلف فنولوژی می‌توانند به صورت مشترک روی مشاهدات لایه قابل رویت تاثیر بگذارند از یک مدل ترکیبی احتمال انتشار استفاده شده است که سهم هر تابع به صورت ضریبی از آن تعیین می‌شود. در واقع یک مدل تعمیم‌یافته مارکف پنهان می‌تواند قادر به برآورد درصد پیشرفت فنولوژی باشد. مقدار متوسط RMSE به دست آمده از ۹ بار اجرای مدل حدود ۱۴٪ در مقیاس روزانه است. وجود منابع متعدد خطا ریشه در داده‌های مورد استفاده، الگوریتم‌های متعدد مورد استفاده در مدل و عدم قطعیت در وقوع رخداد‌های هواشناسی دارد؛ با این وجود، مدل مارکف پنهان قادر است که درصد پیشرفت مراحل مختلف فنولوژی ذرت را به صورت بی‌درنگ در دامنه قابل قبول خطا در مقایسه با مطالعات دیگر برآورد نماید.

سپاسگزاری

داده‌های فنولوژی مورد استفاده در این پژوهش توسط اداره کل هواشناسی استان البرز تأمین شده است. مؤلفین به این وسیله مراتب تشکر و قدردانی خود را اعلام می‌دارند.

منابع

- Arvor, D., Jonathan, M., Meirelles, M. S. P., Dubreuil, V., Lecerf, R. 2008. Comparison of multitemporal MODIS-EVI smoothing algorithms and its contribution to crop monitoring. in Geoscience and Remote Sensing Symposium. IGARSS 2008. IEEE International, 2: 958-961, 7-11 July, doi: 10.1109/IGARSS.2008.4779155.
- Aurdal, L., Huseby, R. B., Eikvil, L., Solberg, R., Vikhamar, D., Solberg, A. 2005. Use of Hidden Markov Models and phenology for multitemporal satellite image classification: applications to mountain vegetation classification, published in: Analysis of Multi-Temporal Remote Sensing Images, 2005 International on the Date 16-18 May, 220-224.
- Dempster, A. P., Laird, N. M., Rubin, D. B. 1977. Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Stat. Soc. B., 39: 1-38.
- Derrode, S., Carincotte, C., Bourenane, S. 2004. Unsupervised image segmentation based on high-order hidden MARKOV chains. In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing

قادر نیست چولگی داده‌ها را لحاظ نماید. این در حالی است که ممکن است از همه یا برخی حالت‌های فنولوژی لایه قابل رویت با توزیعی غیر از نرمال منتشر شوند. به نظر می‌رسد استفاده از توابع چگالی انعطاف‌پذیرتر بتواند در کیفیت نتایج تأثیر داشته باشد.

ز) مرتبه مارکفی: راه حل بالقوه دیگر برای حصول نتایج دقیق‌تر استفاده از مدل‌های مارکف پنهان با مرتبه بالاتر^۱ است. در زمینه‌های مختلفی مانند تشخیص کلمات (مری و همکاران^۲، ۱۹۹۷)، مدل‌سازی وابستگی‌های مکانی بین مناطق کروموزومی (سیفرت و همکاران^۳، ۲۰۱۲) و طبقه‌بندی نظارت نشده تصاویر (درود و همکاران^۴، ۲۰۰۴) از مرتبه‌های بالاتر مدل مارکف پنهان استفاده شده است. وارد کردن وابستگی‌های زمانی بیشتر در مدل مارکف پنهان می‌تواند منجر به مدل‌سازی بهتر پیچیدگی‌های فیزیولوژیک چرخه حیات گیاه شود. در نتیجه بخشی از عواملی که در یک مدل ساده در نظر گرفته نمی‌شوند، در یک مدل پیچیده‌تر مورد توجه قرار می‌گیرند.

نتیجه‌گیری

درصد پیشرفت فنولوژی به عنوان یک ویژگی پوشش گیاهی، نشانگر آهنگ رشد و نمو گیاه بوده و نقش کلیدی در تعیین پارامترهای رشد دارد. بنابراین پایش بی‌درنگ آن می‌تواند بر آگاهی محققین از نحوه رشد گیاه در طول فصل رشد بیفزاید. علاوه بر این، پایش میدانی این ویژگی مستلزم صرف وقت، نیرو و هزینه زیاد است. داده‌های چندزمانه سنجش از دور و داده‌های هواشناسی هر کدام به طور مجزا در کاربردهای مختلفی از جمله پایش ویژگی‌های پوشش گیاهی استفاده شده‌اند. در تحقیق حاضر از این داده‌ها به عنوان مشخصه‌های چندگانه برای گرفتن سنجش پیشرفت ذرت استفاده شده است و جهت برآورد بی‌درنگ، مدل مارکف پنهان پیشنهاد گردید. این مدل با مدل مرسوم مارکف پنهان دارای تفاوت‌هایی به شرح ذیل می‌باشد:

۱) به دلیل ویژگی ذاتی رشد گیاه، یک مدل ناهمگن است. مفهوم ناهمگن زمانی به کار می‌رود که پارامترهای مدل وابسته به زمان یا محلی باشند و به عبارت دیگر این

¹ High-order HMMs

² Mari et al.

³ Seifert et al.

⁴ Derrode et al.

- Sacks, W. J., Kucharik, C. J. 2011. Crop management and phenology trends in the U.S. corn belt: Impacts on yields, evapotranspiration and energy balance. *Agr. Forest Meteorol.*, 151: 882–894.
- Seifert, M., Cortijo, S., Colomé-Tatché, M., Johannes, F., Roudier, F., Colot, V. 2012. MeDIP-HMM: Genome-wide identification of distinct DNA methylation states from high-density tiling arrays. *Bioinformatics*. doi:10.1093/bioinformatics/bts562.
- Shen, Y., Di, L., Wu, L., Yu, G., Tang, H., Yu, G., Shao, Y. 2013. Real-time estimation of corn progress stages using hidden markov models with multisource features. *Remote Sens.*, 5: 1734-1753; doi:10.3390/rs5041734.
- Srihari, S. N. 2011. Machine Learning and Probabilistic Graphical Models Course: Hidden Markov Models. Available online: <http://www.ced.ar.buffalo.edu/srihari/CSE574/index.html> (accessed on 18 November 2012).
- Streck, N. A., Lago, I., Gabriel, L. F., Samboranza, F. K. 2008. Simulating maize phenology as a function of air temperature with a linear and a nonlinear model. *Pesquisa Agropecuária Brasileira*, 43: 449–455.
- Toukiloglou, P. 2007. Comparison of AVHRR, MODIS and VEGETATION for land cover mapping and drought monitoring at 1 km spatial resolution. Ph. D. Thesis, Cranfield University, Bedford, UK.
- Viovy, N., Arino, O., Belward, A. S. 1992. The best index slope extraction (BISE): a method for reducing noise in NDVI time series. *Int. J. remote sens.*, 13(8): 1585-1590.
- Wu, C., Gonsamo A., Gough C. M., Chen J. M., Xu S. 2014. Modeling growing season phenology in North American forests using seasonal mean vegetation indices from MODIS. *Remote Sens. Environ.*, 147: 79–88.
- Yu, G., Di, L., Yang, Z., Shen, Y., Zhang, B., Chen, Z. 2012. Corn growth stage estimation using time series vegetation index. In Proceedings of 2012 First International Conference on Agro-Geoinformatics (Agro-Geoinformatics), Shanghai, China, 2–4 August 2012: 1–6.
- (ICASSP), Marseille, France, 17–21 May 2004: 769–772.
- Elliott, R. J., Siu, T. K. 2012. An HMM approach for optimal investment of an insurer. *Int. J. Robust Nonlinear Contr.*, 22: 778–807.
- Gao, F., Anderson, M. C., Zhang, X., Yang, Z., Alfieri, J. G., Kustas, W. P., Mueller, R., Johnson, D. M., Prueger, J. H. 2017. Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. *Remote Sens. Environ.*, 188: 9–25.
- Ghamghami, M., Ghahreman, N., Irannejad, P., Ghorbani, K. 2017. Application of multi-temporal satellite images to improve prediction models of maize phenology. *J. water and soil res.*, In press.
- Heij, C., de Boer, P., Franses, P. H., Kloek, T., van Dijk, H. K. 2004. *Econometric methods with applications in business and economics*; Oxford University Press Inc.: New York, NY, USA, 2004.
- Jaakkola, T. S. 2006. Machine Learning, Lecture Notes 19: Hidden Markov Models (HMMs); Available online: <http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-867-machine-learning-fall-2006/lecture-notes/lec19.pdf> (accessed on 18 November 2012).
- Knab, B., Schliep, A., Steckemetz, B., Wicher, B. 2003. Model-based clustering with Hidden Markov Models and its application to financial time-Series data. In *Between Data Science and Applied Data Analysis*; Schader, M., Gaul, W., Vichi, M., Eds.; Springer: New York, NY, USA: 561–569.
- Leite, P., Feitosa, R., Formaggio, A., Costa, G., Pakzad, K., Sanches, I. 2011. Hidden Markov models for crop recognition in remote sensing image sequences. *Pattern Recognition Lett.*, 32: 19–26.
- Mari, J. F., Haton, J. P., Kriouile, A. 1997. Automatic word recognition based on second-order hidden Markov models. *IEEE Trans. Speech Audio Proc.*, 5: 22–25.

Assessment of Bayesian structure of hidden Markov model for real time prediction of maize phenology

M. Ghamghami¹, N. Ghahreman^{2*}, P. Irannejad³, H. Pezeshk⁴

Received: 12/06/2016

Accepted: 06/08/2017

Abstract

The Crop Progress Percentage (CPP) in a given phenology stage reflects growth status in life cycle. Generally, routine field measurements of this parameter are lacking, hence various alternative approaches have been proposed for its estimation. The statistical methods such as Bayesian approaches and hidden Markov models (HMMs) have appropriate structural skills for solving problems with variety of continuous or discrete data and can be combined with remotely sensed data also. The aim of this study is evaluation of hidden Markov models' skill in real time prediction of maize progress percentage in research field of university of Tehran located in Karaj. The HMMs follow the Bayesian structure in which, there are usually two layers; hidden and observable. Different phenological stages including Emergence to Milky were considered as the hidden layer and both Normalized Difference Vegetation Index (NDVI) and Accumulated Growth Degree-Day (AGDD) features, extracted from air temperature and LANDSAT7 ETM+ images, as a vector variable of observable layer. Calibration and evaluation of the model was performed using a 9 years (2002-2010) data set of the field phenology observations and meteorological data. According to the results, in general, for all phenological stages, the HMM was able to estimate the CPPs with average RMSE of 14%, which confirms the applicability of this approach as a suitable tool. Further studies in other climatic regions of the country are recommended for more scrutiny of phenological prediction models using remotes sensing and statistical approaches.

Key words: Progress percentage, Phenology, AGDD, NDVI



¹ Ph. D. student of Agrometeorology, University of Tehran, Karaj, Iran

² Associate Professor, Department of Irrigation and Reclamation, University of Tehran, Karaj, Iran

(*Corresponding Author's Email Address: nghahreman@ut.ac.ir)

³ Associate Professor, Department of Space Physics, Geophysics Institute, University of Tehran

⁴ Professor, Department of Statistics, College of Mathematics, Statistics and Computer Sciences, University of Tehran