NASA DEVELOP National Program
North Carolina – NCEI

Fall 2020

Montana Water Resources

Developing a Composite Moisture Index Utilizing NASA Earth Observations for Drought Monitoring in Montana and the Missouri River Basin

DEVELOP Technical Report

November 19th, 2020

Chloe Schneider (Project Lad)
Dean Berkowitz
Egla Ochoa-Madrid
Julie Sorfleet

Advisors:
Ronald Leeper-, NOAA National Centers for Environmental Information, North Carolina Institute for Climate Studies (Science Advisor)

1. Abstract

The Missouri River Basin provides irrigation water for a substantial part of the domestic agricultural sector within the United States. Drought events pose a significant threat to economic livelihoods of dependent individuals, industries, and ecosystems (e.g., farmers, local tribes, hydroelectric power, wildlife). In 2017 alone, the Missouri River Basin experienced a severe drought that resulted in a $2.6 billion loss to the U.S. Northern Plains. In response to such events, organizations throughout the basin, such as the Montana Climate Office, have dedicated efforts for drought monitoring and communicating relevant information to local stakeholders. In an effort to aid regional decision-making capabilities, the project team partnered with the Montana Climate Office, NOAA National Weather Service (NWS) Missouri Basin River Forecast Center, and NOAA Regional Climate Services, Central Region to create a monthly Composite Moisture Index (CMI) that relies on NASA Earth observations from the Terra Moderate Resolution Imaging Spectroradiometer (MODIS) and the Soil Moisture Active Passive (SMAP) mission. From these satellites, as well as the NOAA NWS National Operational Hydrologic Remote Sensing Center Snow Data Assimilation System (SNODAS), our team aggregated climate datasets including soil moisture, snow cover, snow depth, and snow water equivalent to compute a CMI that indicates regional moisture conditions during the winter months. The March CMI values produced over the Missouri River headwater subbasin strongly correlate ($r = 0.75$) with spring and early summer stream discharge, demonstrating the use of this metric to indicate moisture conditions for the snowmelt and growing seasons.

Key Terms
drought, composite moisture index, SMAP, MODIS, SNODAS, SWE, snow depth

2. Introduction

2.1 Background Information

The Missouri River Basin is the hydrologic region that encompasses drainage within all of Nebraska and parts of Colorado, Iowa, Kansas, Minnesota, Missouri, Montana, North Dakota, South Dakota, and Wyoming. A major global breadbasket, the Missouri River Basin comprises agricultural areas that constitute approximately 46% of wheat, 22% of grain corn, and 34% of cattle produced domestically within the United States (Mehta et al., 2012). This watershed is highly prone to drought events that impact both resource availability and economic productivity within the region. For instance, in 2017, a severe drought affected the Missouri River Basin, degrading pasture conditions to poor while resulting in a $2.6 billion loss to the U.S. Northern Plains (NOAA NCEI, 2020). To better prepare for such events, Montana, North Dakota, and South Dakota activated state drought task forces to assess drought conditions, review drought impacts, and facilitate drought relief (Jencso et al., 2019).

The recent proliferation of publicly available satellite data with high spatial, spectral, and temporal resolutions bolsters drought monitoring capacities at both global and regional scales. Previous studies have utilized these remotely sensed datasets to construct a Composite Drought Index (CDI), for various areas around the world. In this report, we will refer to our construct as a Composite Moisture Index (CMI) as we account for a range of moisture conditions that encompasses both drought and flooding conditions. Such composite indices are powerful tools for effective drought monitoring as they assimilate data from multiple indicators representing meteorological, hydrological, and agricultural drought into a single numerical value that holistically assesses moisture conditions for a region of interest (Waseem et al., 2015). For example, Bijaber et al. (2018) used parameters derived from monthly satellite data at a national scale to create a CDI for Morocco. Their weighted composite indicator incorporated remotely sensed precipitation data, land surface temperature as a proxy for soil moisture, Normalized Difference Vegetation Index anomalies from Terra Moderate Resolution Imaging Spectroradiometer (MODIS) for quantifying plant stress, and evapotranspiration anomalies from surface energy balance modeling.

Commented [A1]: Background, Study Area/ Period can all be within Background Information.

There is no need to create multiple sub-section headings within 2.1 for each bolded category. This goes for 2.2 as well. It is ok, but not necessary, to create sub-sections if each has at least a paragraph of text. **A paragraph should consist of at least 3 sentences.**

Commented [A2]: Include **at least 5** references from **peer-reviewed literature** and a **study area map.** Please cite all sources in-text using APA format.

Commented [CAL(D3]: Does this mean that these terms can be used interchangeably? For clarity in your paper, make sure to be intentional about what terms you use. Are 'composite drought indices" (previous paragraph) the same as a "Composite Drought Index (CDI)"?

Commented [EO4]: I think we should only mention one, and keep CMI to create less confusion

Commented [DB5R4]: i agree with you its definitely confusing. the problem here is that the paper we cite, Bijaber et al. 2018, refer to theirs as a Composite Drought Index (CDI). so if we want to include their study here as an example we have to refer to it by its name.

Commented [EO6]: I think we should only mention one, and keep CMI to create less confusion

Commented [DB7R6]: i agree with you its definitely confusing. the problem here is that the paper we cite, Bijaber et al. 2018, refer to theirs as a Composite Drought Index (CDI). so if we want to include their study here as an example we have to refer to it by its name.

Commented [EO8]: I feel that a line like this is necessary to address Amanda's comment. Could be re-written more posh*

The NASA DEVELOP team conducted a feasibility study to construct a monthly CMI for the Missouri River Basin (Figure 1) from January 2004 to August 2020 by incorporating soil moisture, snow depth, snow water equivalent, and snow cover measurements derived from NASA Earth observations and ancillary datasets. Specifically, the team explored the relationship between CMI outputs from December-March and annual cumulative runoff since winter soil moisture serves as antecedent conditions for potential flood and drought events in the oncoming spring and summer months.

Figure 1. Map of study area spanning the Missouri River Basin

2.2 Project Partners & Objectives

To execute this project, the team collaborated with the Montana Climate Office, the National Oceanic and Atmospheric Administration (NOAA) National Weather Service (NWS) Missouri Basin River Forecast Center, and the NOAA Regional Climate Services - Central Region. Each of these project partners works towards providing precise, quality, and timely information about the hydrological and climate conditions in the Missouri River Basin (Montana Climate Office, 2020; Central Region | National Center for Environmental Information, 2020). The Montana Climate Office hosts an Upper Missouri River Basin (UMRB) Drought Indicators Dashboard that provides information about daily, monthly, and yearly conditions of the region including temperature, precipitation, soil moisture, vegetation, and drought metrics. In light of recent severe drought events, the partners were interested in a product that quantitatively explored antecedent moisture conditions that influenced agricultural drought or anomalous runoff to better inform drought management decisions. To this end, the team built the framework for a CMI based on NASA Earth observations that provides a holistic understanding of the moisture conditions of the Missouri River Basin. The Montana Climate Office and the NOAA NWS Missouri Basin River Forecast Center were the primary decision makers in the process of creating the CMI. They provided their scientific expertise and hydrological knowledge to guide the team's integration of region-specific climate variables into the index in addition to selecting sub-watersheds for correlation analyses. In the future, the Montana Climate Office will use the CMI as a public-facing tool that provides a timely and holistic understanding of moisture conditions for their users.

Commented [DB9]: should we update this period to reflect CMI calculated earlier than when SMAP was avaialble? or maintain to indicate our CMI using all 4 datasets?

Commented [JS10R9]: Maybe put: The NASA DEVELOP team pursued a feasibility study that constructed a CMI for the Missouri River Basin (Figure 1) from December 2003 to March 2015 by incorporating snow depth, snow water equivalent, and snow cover measurements and from April 2015 to October 2020 by incorporating soil moisture, snow depth, snow water equivalent, and snow cover measurements.

Commented [JS11R9]: Or is this too wordy?

Commented [AS12]: A reminders that the 'observations' in 'Earth observations' is lower case

Commented [RB13]: For the FD, please replace this with a less grainy map! It would be best to make it yourself.

Commented [RB14R13]: Also - figures must be referenced in text.

Commented [JS15]: Replaced image

Commented [JS16]: I replaced this figure. Does it look less grainy to y'all?

Commented [CAL(D17]: You've already introduced this acronym above so you can use MCO here.

Once you've introduced an abbreviation/acronym in your paper, you should use it throughout vs going back and forth between MRB and Missouri River Basin, for example. Think about if you even really need it – unless you are using it *a lot* it might make sense just to keep the term written out if it's not too long.

An easy trick is to use 'ctrl+f' to search for them when you are done writing.

Commented [CAL(D18]: Already introduced MRB

3. Methodology

3.1 Data Acquisition

Our team utilized NASA Earth observations to retrieve snow cover and soil moisture data while drawing upon ancillary modeled snowpack data for explanatory variables as inputs of our CMI product. All datasets were acquired at different levels of processing and aggregated to monthly means from either daily or other sub-monthly time frequency simultaneously with acquisition. Figure A1 demonstrates these datasets and their respective spatial resolutions.

The team acquired Level 3 500m gridded snow cover data from the Terra MODIS platform through a script that utilizes the Google Earth Engine (GEE) Python API. This Python script was run in a Jupyter Notebook which required the user to modify parameters specifying the month and year of interest. The snow cover image collection was filtered based on the desired time window and subsequently clipped to a shapefile of the Missouri River Basin (Hydrologic Unit Code [HUC] 2, Region 10), which was accessed from the National Resources Conservation Service (NRCS) Watershed Boundary Dataset. Lastly, the GEE reducer mean function was applied to the data to get monthly average snow cover. Similarly, surface soil moisture anomalies (SSMA) were acquired using a download link created from the GEE Python API from the NASA-USDA (United States Department of Agriculture) Soil Moisture Active Passive (SMAP) Global Soil Moisture Data product hosted in the GEE Catalog. A download URL was created for the desired raster image which was downloaded as a zipped archive and subsequently unzipped in an R script using the 'httr' package (v1.4.2; Hadley Wickham 2020). Already preprocessed as climatological anomalies, the SSMA data were accessed, clipped to the study area, and input into the GEE reducer mean function to derive monthly rasters for the available period of record. The team acquired modeled estimates of snowpack, snow water equivalent (SWE) and snow depth, from the NOAA National Weather Service's National Operational Hydrologic Remote Sensing Center (NOHRSC) Snow Data Assimilation System (SNODAS) as these data were not available in the GEE Catalog. Daily data for these sets at 1km spatial resolution for the continental United States were downloaded from the NOAA FTP server via an R script that utilized the 'httr' package for downloading zipped data (v1.4.2; Hadley Wickham 2020). Table 1 and Table 2 detail the specifications of each of the acquired datasets.

Table 1
NASA Earth observation data reference information

Platform and Sensor	Data Product	Digital Object Identifier	Period of Record	Acquisition Method
Terra MODIS	MODIS/Terra Snow Cover Daily L3 Global 500m SIN Grid, Version 6	10.5067/MODIS/ MOD10A1.006	February 24, 2000 - Present	Google Earth Engine Python API
SMAP	SMAP L3 Radiometer Global Daily 27 km EASE-Grid Soil Moisture, Version 4	10.5067/ZX7YX2 Y2LHEB	April 01, 2015 – Present	Google Earth Engine Python API

Table 2
Ancillary data reference information

Source	Data Product	Digital Object Identifier	Period of Record	Acquisition Method

Commented [CAL(D19]: Note – the 'american' spelling only has one 'l'

Commented [RB20]: Reading this, I expected Fig 2 to state the spatial resolution of each data source. Can you update the caption to include that?

Commented [JS21]: dates of availability? We don't have resolution in the table

Commented [DB22R21]: i am referring to "Figure 2" as it stands right now. i think we should keep these tables in the main body of the paper, but perhaps the maps illustrating the resolutions can be bumped to the appendix if we run into a page limit

Commented [EO23R21]: Moved this to the appendix

Commented [RB24]: not cited

Commented [EO25]: We received a comment on SMAP for the data processing. I think this detail should wait till later so we are not redundant. As in Include this in the intro paragraph of the next section

Commented [RB26]: Hey Amanda - I wasn't sure how to cite these as I couldn't find a clear instruction online.

SNODAS	Snow Data Assimilation System (SNODAS) Data Products at NSIDC, Version 1	10.7265/N5TB14TC	September 28, 2003 - Present	NOAA FTP Server

3.2 Data Processing

Given that the acquired data products differed from one another in their units of measurement, each was converted into a common, relative unit in order to create a composite index. For the three snow datasets, the team performed calculations on the monthly averaged data (outlined below in section 3.2.1) to determine the degree to which each pixel value deviated from climate normals. For soil moisture anomalies, the team calculated monthly averages across the period of record simultaneously with data acquisition. Following these processing steps, the methodology (detailed in Figure 2) enabled the integration of each dataset into a composite index despite their varying native units and disparate periods of record.

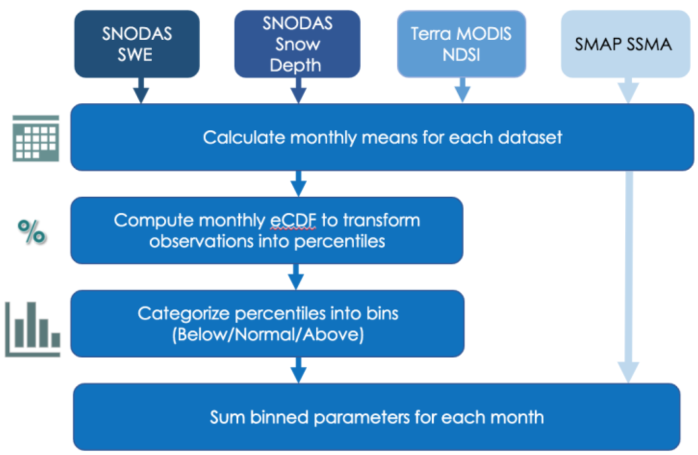


Figure 2. Data processing graphical workflow prior to calculating CMI.

3.2.1 Snow Data Percentile Generation

Drought or flood events stem from exceptional moisture states that deviate from the long-term average conditions of an area. To detect how observations deviate from historic conditions, one must establish climate normals: the average climatological conditions for a weather variable for a given time frame. Monthly deviations from climate normals were calculated utilizing R 4.0.3 (R Core Team 2020) through RStudio 1.3.1093 (R Core Team 2020). Our team used the 'stats' package (v3.6.2; R Core Team 2019) to compute an empirical cumulative distribution function (ECDF) for a given month across each snow cover historical dataset. The ECDF constructs a step function of average monthly values for each month on record for each pixel (Equation 1).

$$F_n(t) = \{x_i \leq t\}/n = \frac{1}{n}\sum_{i=1}^{n} 1_{[x_i \leq t]} \qquad (1)$$

The empirical cumulative distribution function is a step function that intakes a set of observed measurements x_i, where i represents an indexed element of that set, and n represents the total number of observations in that set (ie: the available historical data [months on record] for a given pixel). F represents the distribution of all t

Commented [EO27]: Is this subsection necessary if there is only one subsection?

Commented [EO28R27]: Or, maybe have the ECDF and then the Binning be a separate section.

Commented [CAL(D29]: Expand on this – approach to what? What is it accomplishing?

Commented [DB30]: more informative to title this section "Percentile generation" or something along those lines. the ECDF is a tool we used to create percentiles - so i think it would be better if it wasn't the title itself.

Commented [RB31]: This sentence is pretty repetitive.

Commented [RB32]: Please add a sentence or two directly after this equation describing relevant input and output variables.

Commented [RB33R32]: i.e. "Where F is equal to ____, x is equal to _____," etc.

output values. Each *t* value presents the fraction of values less than or equal to x that exists over the entire set *n*.

In order to determine the degree to which a monthly observation deviated from climate normals, our team used the ECDF to compute percentile values for each of the three snow datasets. With support from our science advisor, the team developed a custom function as an input to the calc function from the R 'raster' package (v3.3-13, Robert Hijmans 2020) so as to generate an ECDF cell-wise across a raster surface for a given month while subsequently assigning the corresponding percentile value to each pixel. Next, the team reassigned all pixel percentile values into one of five bins based on linear breaks that quantified a given monthly observation's deviation from historic climate data (Table 3). The team pursued calculating percentiles as a measure of deviation based on the availability of data. Through discussions with advisors and partners, the team found that the 16-20 years of record for the snow datasets would give confidence in calculating a representative median (50^{th} percentile) value but less confidence in identifying extreme values above or below the 90^{th} or 10^{th} percentile values. To classify all datasets into a common unit, the team generalized values above or below the median into a binning scheme to represent general values of above, below or near normal and to account for a lack of confidence in precisely identifying the 90^{th} or 10^{th} percentile values based on the years of record. This methodology was determined with input from the team's science advisor and project partners after an exploration of histograms for each data product across time and space revealed that no single distribution fit all of the data. Thus, for feasibility purposes, the team proceeded with a linear model of classification.

Commented [RB34]: missing citation

Commented [DB35]: this is what i just wrote. let me know what you think

Table 3
Binning scheme of percentile values for snow data

Bin	Bin Description	Percentile Values
-2	Far below normal	0-20%
-1	Below normal	21-40%
0	Normal	41-60%
1	Above normal	61-80%
2	Far above normal	81-100%

3.2.2 SMAP Reclassification
The SMAP surface soil moisture anomaly data product obtained from GEE contained values calculated using a 31-day moving window analysis. Thus, these data were originally calculated in a manner that increased the number of data points available across the five-year period of record, thereby strengthening the historical basis of the monthly averages calculated by the team. In order to integrate the SSMA product with snow cover data, the team devised a classification scheme that categorized monthly average anomalies into bins based on below normal, normal, and above normal values (Table 4). Our team studied histograms of SMAP over time to understand what range of values accurately captured the distribution of data. The normal bin was defined after identifying that the bulk of observations lay within the range of –0.5 to 0.5 (Figure 3). For the current SMAP data on record the minimum mean anomaly value fell just below -2 and the maximum mean anomaly value equaled 2.4. Considering this, negative and positive infinity were set as the lower and upper bounds of the "below normal" and "above normal" categories, respectively.

Table 4
Binning SMAP anomalies

Bin	Bin Description	Percentile Values
-1	Below normal	- ∞ – -0.6
0	Normal	-0.5 – 0.5

1	Above normal	0.6 – + ∞

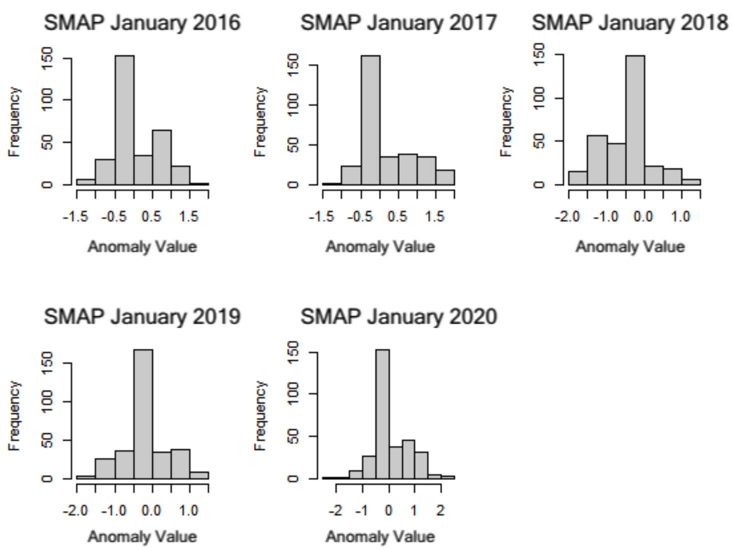


Figure 3. Histograms of January SMAP monthly mean anomaly values.

3.2.3 Spatial Aggregation & CMI Calculation

Two approaches were taken to spatially aggregate the disparate native resolutions of the four datatsets for compositing. Firstly, each binned data product was resampled to the coarsest resolution, that being SMAP at 27km, using the nearest neighbor assignment method via the 'raster' package (v3.3-13; R Core Team 2020). The second approach involved spatially aggregating input datasets at their native resolutions to the HUC-6 level in order to compute CMI values for sub-watersheds within the Missouri River Basin. To do so, the team extracted the mean pixel value for each HUC-6 polygon using the extract function in the 'raster' package (v3.3-13; R Core Team 2020). Following the spatial aggregation, the team calculated a CMI in two ways: by summing and by averaging each of the input climate variables. Therefore, the team produced two pixel-based CMI maps at 27km resolution, one computed as a sum of the indicators and another computed by averaging the indicators. Additionally, the team produced two HUC-6 level CMIs, one with CMI values computed as a sum of the indicators for each hydrological unit, and another computed by averaging the indicators for each subbasin. In our final CMI calculations, the team multiplied the binned SMAP data by 2 prior to summing or averaging. This was done to ensure that the scale of our CMI would have the same minimum and maximum range for both the period of record prior to and after the availability of SMAP data.

3.3 Correlation Analysis

To assess the explanatory power of our CMI, the team performed a correlation analysis for HUC-6 level CMI outputs with *in situ* cumulative streamflow data. We utilized the 'waterData' package (v1.0.8; Karen R. Ryberg and Aldo V. Vecchia 2017) to acquire annual cumulative streamflow from April through August across the

Commented [RB36]: .

Commented [DB37]: i rewrote this whole section to be more explicit in our methodology, however it is a bit redundant. would somebody please take a look and make some edits?

Commented [JS38R37]: I made some edits. I'm not sure if I made it more/less redundant. Everything is done in track changes so that it can be changed later

Commented [DB39]: i rewrote this whole section to be more explicit in our methodology, however it is a bit redundant. would somebody please take a look and make some edits?

Commented [DB40]: maybe we should move these sentences to data processing ? i know we do the SMAP x 2 calculation in the CMI scripts but for continuity purposes i feel like any massaging of the data belongs up there. but idk what do you think @Julie Sorfleet

Commented [RB41]: cite

study period for the following five USGS stream gaging stations of interest: station 6102050 on the Marias River, station 6329500 on the Yellowstone River, station 06478513 on the James River, station 06174500 on the Milk River, and station 6054500 on the Missouri River Headwater (Figure B1). In this way, we were able to explore the relationship between antecedent moisture conditions with stream runoff in the spring and summer months directly following wintertime. Based on expert input from project parners, the team selected stream gaging stations for our correlation analysis that drained the contents of a respective HUC-6 region within the Missouri River Basin. In the selection process, the team prioritized gaging stations that did not include dams nor other physical features which could otherwise influence the cumulative stream flow data.

We then looked at the correlation between the cumulative stream gage discharge for April through July and the December, January, February, and March CMI values for the selected regions. Each chosen gaging station was located in a river at the drainage point of a watershed contained in a single respective HUC, except for station 6329500 on the Yellowstone River which drained a total of five HUC-6 regions. The team averaged the CMI values for each respective HUC representing the aforementioned stream gaging stations in order to compute cumulative discharge correlations. Using the lm (linear model) function in the 'stats' package (v3.6.2; R Core Team 2019), our team fitted spring/summer cumulative discharge with winter CMI values using a simple linear regression. In this manner, we determined the coefficient of determination (R^2) for a winter month's CMI values and spring/summer cumulative discharge across the period of record January 2004 - March 2020 for which there was sufficient data to compute CMI values. Additionally, we computed the Pearson correlation coefficient (r) to measure the proportion of variation that was explained by our linear model for winter CMI and annual cumulative stream discharge. The resulting graphs likewise illustrated how correlations changed depending on specific CMI values for different winter months.

4. Results & Discussion

4.1 Analysis of CMI results

A comparison of the pixel-based, summed CMI to the HUC-based, summed CMI revealed similar results. For March 2016, the CMI output indicated negative values in the northern part of the Missouri River Basin, values approximating normal conditions for the lower-middle part of the basin and positive values in the southwestern most part of the basin, a phenomenon shared by both the pixel-based and HUC-based versions alike (Figure 4). This result corroborates a dryness trend throughout the region as indicated by below-median runoff values for 2016 on the annual cumulative runoff graph above Sioux City, IA, the drainage point for the upper Missouri River Basin (U.S. Army Corps of Engineers, 2019, p. 41). In comparison, 2019 was an anomalously wet year for the region, maintaining the second highest runoff value in million-acre feet since record-keeping began in 1897 at Sioux City (U.S. Army Corps of Engineers, 2019, p. 37). Our CMI values for March 2019 were predominantly positive in our pixel-based results, with a few negative values output for peripheral areas of the basin (Figure 5). Our HUC-based CMI output for March 2019 displayed similar results, in which every single HUC maintained CMI values greater than 0, thereby indicating comprehensively wetter moisture conditions across the basin (Figure 5). Thus, both the pixel-based and HUC-based CMI results for March 2019 corroborate the observed wetness trend throughout the region illustrated in the annual runoff graph (U.S. Army Corps of Engineers, 2019, p. 41). A comparison of our pixel-based with our HUC-based CMI results revealed that both methods depict similar moisture conditions for the region despite the

Commented [DB42]: began rewriting julie's block of text maintained below. work in progress

Commented [AS43]: Define this

Commented [DB44]: rewrite these sections.

Commented [RB45]: Nice placeholders!

Commented [RB46]: Nice placeholders!

Commented [RB47]: Nice placeholders!

Commented [RB48]: Nice placeholders!

Commented [RB49]: Nice placeholders!

Commented [RB50]: Nice placeholders!

Commented [RB51]: Nice placeholders!

Commented [RB52]: Nice placeholders!

Commented [RB53]: Nice placeholders!

Commented [RB54]: Nice placeholders!

Commented [RB55]: Nice placeholders!

Commented [RB56]: Nice placeholders!

fact that the respective calculations were performed at different spatial scales.

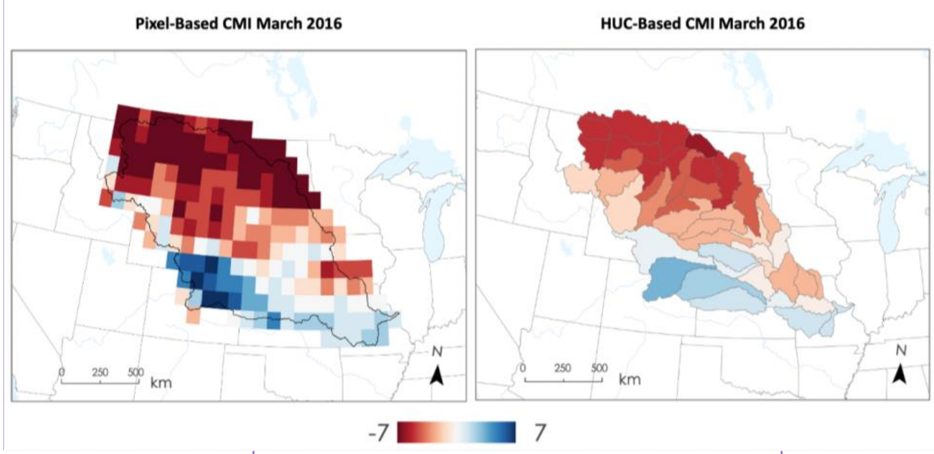


Figure 4. Left: Pixel-based CMI for March 2016 at a 27 km resolution. Right: HUC-based CMI for March 2016 with a single CMI value produced for each hydrological unit. Hydrological units are denoted with gray borders. In each map, negative CMI values indicate areas with less moisture and positive CMI values indicate areas with more moisture.

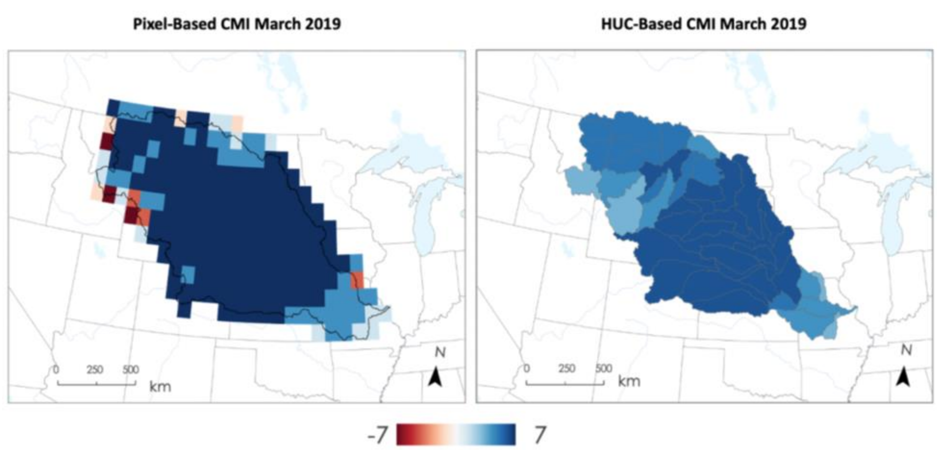


Figure 5. Left: Pixel-based CMI for March 2019 at a 27 km resolution. Right: HUC-based CMI for March 2019 with a single CMI value produced for each hydrological unit. Hydrological units are denoted with gray borders. In each map, negative CMI values indicate areas with less moisture and positive CMI values indicate areas with more moisture.

When analyzing the correlation between winter CMI and spring cumulative discharge for subbasins with the larger watershed, the team found that February and March consistently had larger coefficients of

Commented [RB57]: Nice placeholders!

Commented [DB58]: add to appendix

Commented [JS59R58]: Replace figure with one with clipped swe and snow depth?

Commented [DB60]: add to appendix

Commented [JS61R60]: Replace figure with one with clipped swe and snow depth?

Commented [DB62]: add to appendix

Commented [JS63R62]: Replace figure with one with clipped swe and snow depth?

Commented [DB64]: add to appendix

Commented [JS65R64]: Replace figure with one with clipped swe and snow depth?

Commented [DB66]: add to appendix

Commented [JS67R66]: Replace figure with one with clipped swe and snow depth?

Commented [DB68]: add to appendix

Commented [JS69R68]: Replace figure with one with clipped swe and snow depth?

Commented [DB70]: add to appendix

Commented [JS71R70]: Replace figure with one with clipped swe and snow depth?

Commented [DB72]: add to appendix

Commented [JS73R72]: Replace figure with one with clipped swe and snow depth?

determination than December and January. Figure B1 details a map for the five subbasins selected for correlation analysis in this study. The higher correlation values in March likely reflect that snow conditions closer to spring melt are the ultimate contributors to runoff. Therefore, early snow can provide predictive possibilities but the conditions right before melt reflect the most representative CMI values. This is ascertained because the coefficients of determination (R^2) for March CMI vs spring cumulative discharge were .55, .58, .60, .24, and .56 for the Yellowstone, James, Marias, Milk, and Missouri headwaters, respectively, while coefficients of determination for December CMI vs spring cumulative discharge were .40, .08, .04, .11, and .07 for the same respective regions. The average coefficient of determination was .14 for December while it was .51 for March, a 72.5% increase. The Pearson Correlation Coefficients were consistently higher in February and March than December and January for each region (Table 5). These values indicate greater reliability of our CMI calculations closer to spring snowmelt.

Table 5
Pearson's Correlation Coefficient for December, January, February, and March CMI vs spring cumulative discharge for each watershed system analyzed.

Pearson's Correlation Coefficient (r)	**Yellowstone**	**Marias**	**James**	**Missouri**	**Milk**	**Average**
December	.64	.21	.28	.27	.33	.35
January	.57	.12	.62	.24	.47	.40
February	.67	.56	.78	.50	.52	.61
March	.74	.78	.76	.75	.50	.71

The team determined the lower 10th and upper 90th percentiles of spring cumulative discharge for each of the five subbasins to assess whether CMI values accurately capture extreme moisture states as indicated by anomalously high and low runoff (Figure 6). The team referred to CMI values that accurately described extreme discharge events as 'hits,' which were below zero for the lower 10th percentile of cumulative discharge data or above zero for the upper 90th percentile of cumulative discharge data. CMI values that did not accurately describe extreme discharge events were considered 'misses,' represented by values in the lower 10th and upper 90th percentiles of cumulative discharge that did not satisfy either of the previously stated conditions. For the month of March across our five regions of analysis, the results indicated 17 'hits,' 1 in miss,' and 2 inconclusive (CMI=0). Other "hits" and "misses" for December, January, and February are visually denoted (Figures 6, B2 - B5).

Commented [DB74]: add to appendix

Commented [JS75R74]: Replace figure with one with clipped swe and snow depth?

Commented [DB76]: add to appendix

Commented [JS77R76]: Replace figure with one with clipped swe and snow depth?

Commented [DB78]: it looks like we have at least 3 misses based on the figures below. can somebody please confirm @Julie Sorfleet @Chloe Schneider

Commented [CS79R78]: Those are for all of the March's

Commented [DB80R78]: ok im going to rewrite this sentence to make that more clear

Commented [DB81]: this result needs to be contextualized more, otherwise we are just cherry picking data. if we want to discuss why March was successful, I think we also have to be explicit in why December and January weren't.

Commented [DB82]: this result needs to be contextualized more, otherwise we are just cherry picking data. if we want to discuss why March was successful, I think we also have to be explicit in why December and January weren't.

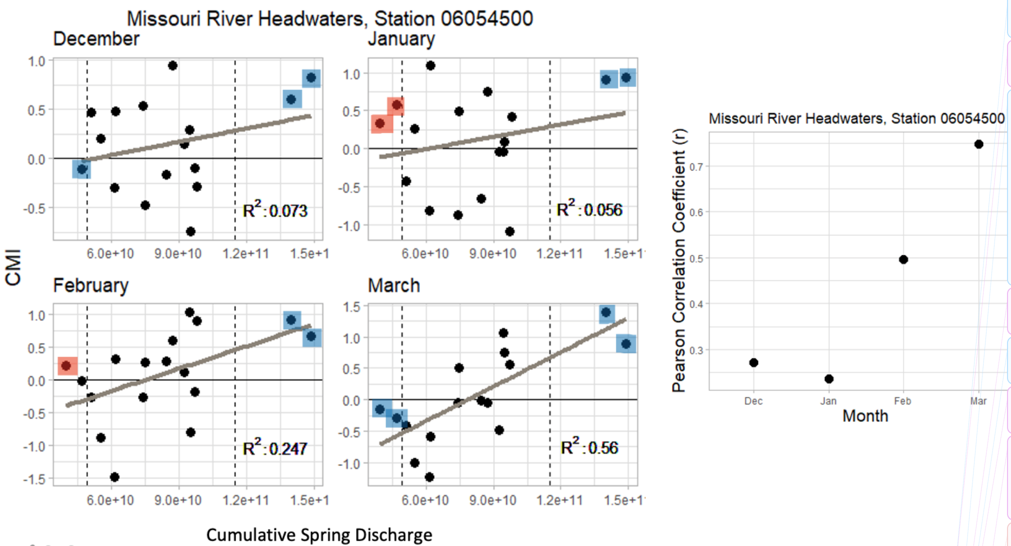


Figure 6. Left: Cumulative spring/summer discharge (April-August) taken at gage station 06054500 at the outlet of the Missouri River Headwaters plotted against the CMI values for the Missouri River Headwater HUC-6 region for December (2004-2019), January, February, and March 2004-2020. Vertical dashed lines represent the 10th and 90th percentiles of cumulative spring discharge. Colored boxes indicate whether the CMI performs as expected based on anomalous observed runoff, where blue points represent 'hits', and red points indicate 'misses'. Right: Pearson's Correlation Coefficient (r) for each month's CMI vs cumulative discharge plot.

4.2 Future Work

To expedite the communication process to end users, the CMI workflow should be further developed to enable automation. Currently, the framework our team developed requires researchers to manually acquire, process, and analyze datasets through a pipeline of individual scripts, each of which involves changing input temporal and indicator variables. An automated CMI workflow would be capable of programmatically ingesting data, computing the intermediary analysis, and outputting the final CMI map products, thereby decreasing processing times while increasing overall use value. While the team's present data management scheme and analysis methods were sufficient for completing this feasibility study, restructuring scripts and data storage would be necessary to achieve a level of computational efficiency necessary for future automation.

Additionally, the team suggests adding functionality that would allow users to select how the parameters are calculated (i.e., percentiles vs. standardized anomalies). It would also be worth investigating how a change in computing these relative measures impact the results of the correlation analysis. Similarly, a sensitivity analysis of how the input parameters is binned into categories and weighted in the CMI summation would help demonstrate which methods produce the strongest correspondence with spring and summer runoff. Conducting a principal components analysis of how additional climate variables and drought indicators contribute to runoff would help develop a dynamic weighting scheme tailored to specific subbasins and adjusted for time of year (e.g., early fall, late fall, early winter, etc.).
Analysis comparing winter CMI values to spring cumulative discharge could be improved by identifying when exactly snowmelt is happening in the different watersheds. Cumulative spring discharge was generalized through spring and early summer months, which may not have captured the realistic snowmelt timeframe,

Commented [DB83]: add to appendix

Commented [JS84R83]: Replace figure with one with clipped swe and snow depth?

Commented [DB85]: do we need to restate what hits and misses mean here or can i just say "hits" where CMI outputs accurately capture wet or dry conditions indicated by anomalous discharge values"

"red points indicate misses, where CMI outputs mischaracterize basin moisture conditions indicated by discharge values"

its a wall of text that i dont like to be so redundant

Commented [JS86R85]: I'd leave out the explanation of what a hit/miss is since we essentially defined it earlier

Commented [DB87]: add to appendix

Commented [JS88R87]: Replace figure with one with clipped swe and snow depth?

Commented [JS89]: Change to "An automated CMI would be capable of programmatically ingesting, data computing the intermediary analysis, and outputting the final CMI map products."?

Commented [E[90]: @Dean Berkowitz thoughts?

Commented [EO91]: LOooong sentence

Commented [JS92R91]: Change to "An automated CMI would be capable of programmatically ingesting, data computing the intermediary analysis, and outputting the final CMI map products."?

Commented [EO93R91]: @Dean Berkowitz thoughts?

Commented [DB94]: should we include this here? i feel l ...

Commented [AS95R94]: They can access it ...

Commented [EO96]: Yes, I think Dean said he would ...

Commented [EO97]: Yes, I think Dean said he would ...

Commented [EO98]: Yes, I think Dean said he would ...

Commented [EO99]: Yes, I think Dean said he would ...

Commented [EO100]: Yes, I think Dean said he would ...

Commented [EO101]: Yes, I think Dean said he would ...

Commented [EO102]: Yes, I think Dean said he would ...

Commented [EO103]: Yes, I think Dean said he would ...

Commented [EO104]: Yes, I think Dean said he would ...

Commented [EO105]: Yes, I think Dean said he would ...

Commented [EO106]: Yes, I think Dean said he would ...

Commented [EO107]: Yes, I think Dean said he would ...

Commented [EO108]: Yes, I think Dean said he would ...

Commented [EO109]: Yes, I think Dean said he would ...

Commented [EO110]: Yes, I think Dean said he would ...

Commented [EO111]: Yes, I think Dean said he would ...

and was likely largely influenced by liquid precipitation noise. To accurately capture changes in cumulative discharge that is a result of winter snowmelt, future teams could conduct a time-lag analysis between CMI and runoff, or calculate changes in snow cover or SWE to identify periods when snow is melting.

The February and March CMI values most accurately demonstrated an ability to capture spring discharge events, while December and January values lacked strong correlations with spring discharge. Given that three of the four input variables to the CMI involve snow conditions, our team's current analysis framework is primarily useful for the few months preceding spring snowmelt. To expand the temporal use of our CMI index, we suggest that our partners or future DEVELOP teams incorporate additional seasonal drivers of drought and flooding to increase flexibility of the index to best account for variance in seasonal conditions.

5. Conclusions

The Montana Climate Office was interested in a product that could express a holistic understanding of moisture conditions of the Missouri River Basin. The conditions antecedent to spring were especially of interest as snowmelt runoff largely influences moisture states along with drought or flood events in the early growing season. With this in mind, we developed a Composite Moisture Index that incorporated three snow variables: snow cover, snow depth, and snow water equivalent, along with soil moisture to describe moisture storage states during winter months. In analyzing the explanatory power of our CMI with *in situ* stream gage data, we found that our CMI values for February and March were most highly correlated with spring cumulative discharge within a sample group of subbasins in the watershed. In this way, the CMI was able to provide information on winter conditions that contributed to spring moisture states, addressing the primary need expressed by the Montana Climate Office, and developing the foundation for a holistic CMI. The CMI will aid in analyzing and indicating seasonal moisture conditions that can inform drought and flood planning efforts. The CMI methodology that our team developed with the guidance of our partners will be adopted by the Montana Climate Office to serve as a base product for further development and integration into the Montana Climate Office's live platform, providing an interface for members of the public to interact.

Commented [CAL(D112]: Use of spelled out terms vs. acronyms?

6. Acknowledgments

This project was made possible through the support and mentorship of select individuals and organizations. We extend a big thanks to the following:

- Zachary Hoylman, Research Hydrologist at the Montana Climate Office
- Kelsey Jencso, Montana State Climatologist at the Montana Climate Office
- Doug Kluck, NOAA, Regional Climate Services Director, Central Region
- Ronald Leeper, Researcher at NOAA, National Centers for Environmental Information, North Carolina Institute for Climate Studies
- Kevin Low, Hydrologist at NOAA, National Weather Service, Missouri Basin River Forecast Center
- Andrew Hoell, NOAA Physical Sciences Lab, Meteorologist
- Britt Parker, National Integrated Drought Information System, Regional Drought Information Coordinator
- Andrew Shannon, NASA DEVELOP Fellow/Center Lead at NOAA National Centers for Environmental Information

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Aeronautics and Space Administration.

This material is based upon work supported by NASA through contract NNL16AA05C.

7. Glossary

API – Application Programming Interface

CMI – Composite Moisture Index

ECDF – Empirical Cumulative Distribution Function

Earth Observations – Satellites and instruments that collect remotely-sensed information about the Earth's physical, chemical, and biological systems over space and time

FTP – File Transfer Protocol

GEE – Google Earth Engine: a cloud-based geospatial processing platform

MODIS – Moderate resolution Imaging Spectroradiometer

NOAA – National Atmospheric and Oceanic Administration

SMAP – Soil Moisture Active Passive

SNODAS – Snow Data Assimilation System

SWE – Snow Water Equivalent

8. References

Bijaber, N., El Hadani, D., Saidi, M., Svoboda, M., Wardlow, B., Hain, C., Poulsen, C., Yessef, M., & Rochdi, A. (2018). developing a remotely sensed drought monitoring indicator for Morocco. *Geosciences, 8*(2), 55. https://doi.org/10.3390/geosciences8020055

Ecdf function | R Documentation. (n.d.). Retrieved October 13, 2020, from https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/ecdf

Hall, D. K., Salomonson, V. V., and Riggs, G.A. (2016). MODIS/Terra Snow Cover Daily L3 Global 500m SIN Grid. Version 6. [Data set]. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/MODIS/MOD10A1.006

Mehta, V. M., Knutson, C. L., Rosenberg, N. J., Olsen, J. R., Wall, N. A., & Bernadt, T. K. (2012). Decadal climate information needs of stakeholders for decision support in water and agriculture production secotrs: A case study in the Missouri River Basin. *Weather, Climate, and Society, 5*(1), 27-42. https://doi.org/10.1175/WCAS-D-11-00063.1

NOAA National Centers for Environmental Information (NCEI). (n.d.). *Billion-dollar weather and climate disasters: Overview*. https://www.ncdc.noaa.gov/billions/

NOAA National Centers for Environmental Information (NCEI). (n.d.). *Central Region.* Retrieved October 14, 2020, from https://www.ncdc.noaa.gov/rcsd/central

Montana Climate Office. (n.d.). *UMRB Drought Indicators Dashboard*. Retrieved October 13, 2020, from https://drought.climate.umt.edu/

Montana Climate Office. (2018). *Montana drought & climate: Monitoring and forecasting for agriculture producers,* (Spring 2018). Retrieved from: https://climate.umt.edu/mtdrought/Montana-Drought-Climate-Spring2018.v3.pdf

Jencso, K., B. Parker, M. Downey, T. Hadwen, A. Howell, J. Rattling Leaf, L. Edwards, and A. Akyuz, D. Kluck, D. Peck, M. Rath, M. Syner, N. Umphlett, H. Wilmer, V. Barnes, D. Clabo, B. Fuchs, M. He, S. Johnson, J. Kimball, D. Longknife, D. Martin, N. Nickerson, J. Sage and T. Fransen. (2019). *Flash* *flood drought: Lessons learned from the 2017 drought across the U.S. Northern Plains and Canadian Prairies.* NOAA National Integrated Drought Information System.

National Operational Hydrologic Remote Sensing Center. 2004. Snow Data Assimilation System (SNODAS) Data Products at NSIDC, Version 1. [Indicate subset used]. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center. https://doi.org/10.7265/N5TB14TC. [September 2020].

O'Neill, P. E., S. Chan, E. G. Njoku, T. Jackson, and R. Bindlish (2016). SMAP L3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture, Version 4 [Data set]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi:10.5067/ZX7YX2Y2LHEB

Waseem, M., Ajmal, M., & Kim, T. W. (2015). Development of a new composite drought index for multivariate drought assessment. *Journal of Hydrology, 527,* 30-37. https://doi.org/10/1016/j.jhydrol.2015.04.044

U.S. Army Corps of Engineers, Northwestern Division, Missouri River Basin Water Management Division. (2019). *Missouri river mainstem reservoir system: Summary of actual 2019 regulation.* https://usace.contentdm.oclc.org/utils/getfile/collection/p266001coll1/id/9928

9. Appendices

Appendix A

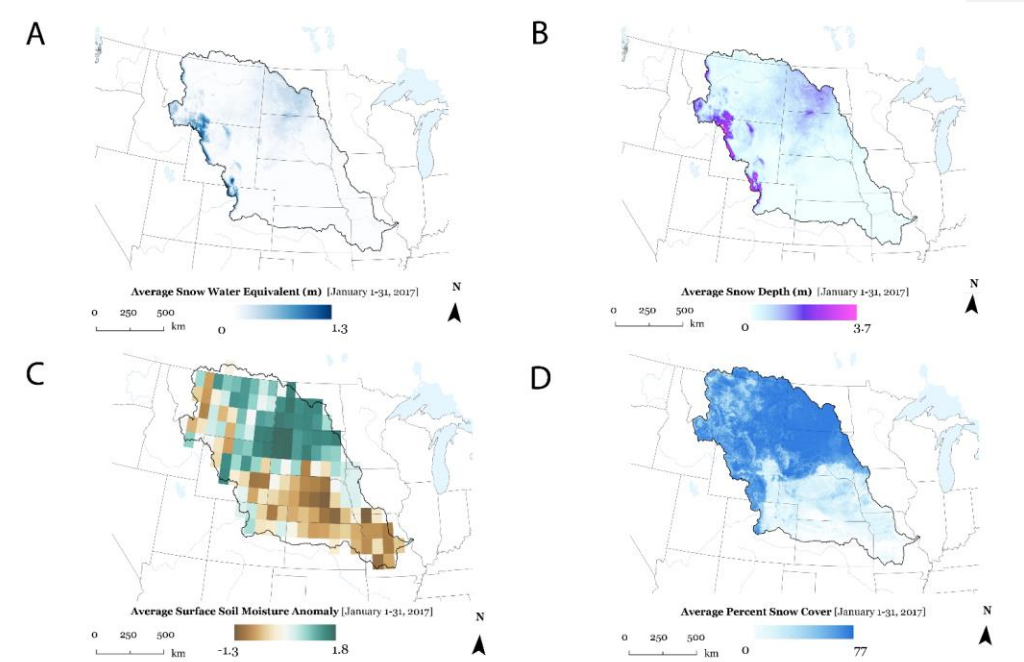


Figure A1. Maps of January 2017 monthly average values of (A) SNODAS Snow Water Equivalent, (B) SNODAS Snow Depth, (C) SMAP soil moisture anomalies, (D) MODIS NDSI snow cover. The black polygonal line indicates the extent of the Missouri River Basin (HUC2 Region 10). These maps demonstrate all four data sets acquired at their respective native resolutions over the Missouri River Basin. Terra Moderate Resolution Imaging Spectroradiometer (MODIS) provided snow cover data at a 500m resolution. The team acquired soil moisture anomalies from Soil Moisture Active Passive (SMAP) at approximately 27km resolution. Snowpack snow water equivalent (SWE) and snow depth, from the NOAA National Weather Service's National Operational Hydrologic Remote Sensing Center (NOHRSC) Snow Data Assimilation System (SNODAS) were acquired at a 1km resolution.

Commented [DB113]: add to appendix

Commented [JS114R113]: Replace figure with one with clipped swe and snow depth?

Appendix B

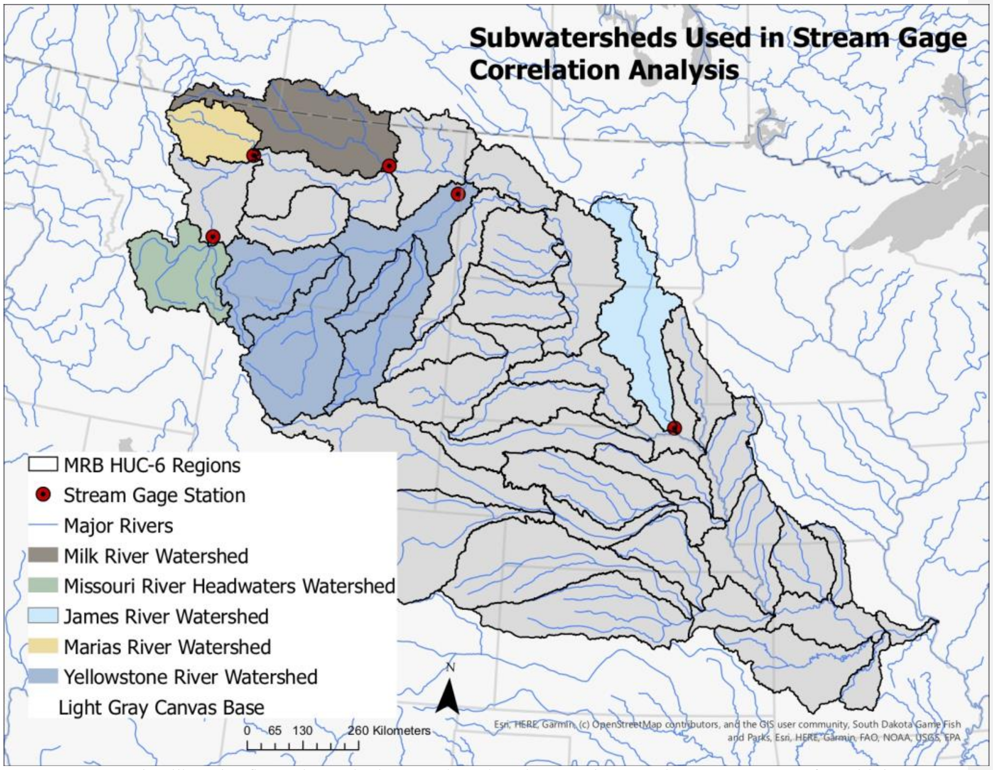


Figure B1. Map detailing the five Missouri River Basin subregions and respective stream gages for which analysis was conducted.

Commented [DB115]: add to appendix

Commented [JS116R115]: Replace figure with one with clipped swe and snow depth?

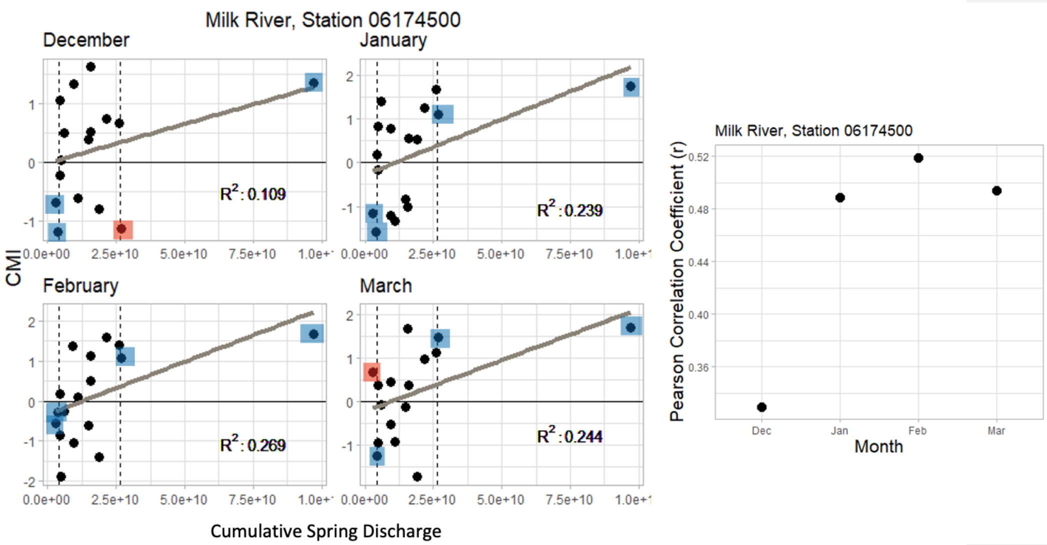


Figure B2. Left: Cumulative spring discharge (April-August) taken at gage station 06174500 at the outlet of the Milk River HUC-6 Watershed plotted against the CMI values for the Milk River HUC-6 region for December (2004-2019), January, February, and March 2004-2020. Blue highlighted points represent points in the lower 10th percentile of cumulative discharge data that correlate with CMI values below 0 or points in the upper 90th percentile of cumulative discharge data that correlate with CMI values above zero. Red values represent points in the lower 10th and upper 90th percentiles of cumulative discharge that do not satisfy either of the previously stated conditions. Right: Pearson's Correlation Coefficient(r) for each month's CMI vs cumulative discharge plot.

Commented [DB117]: add to appendix

Commented [JS118R117]: Replace figure with one with clipped swe and snow depth?

Commented [DB119]: add to appendix

Commented [JS120R119]: Replace figure with one with clipped swe and snow depth?

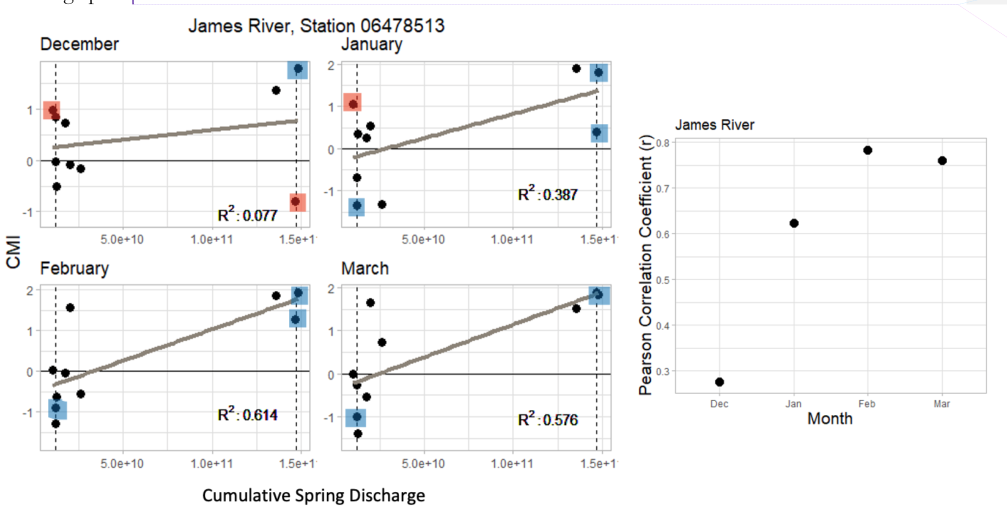


Figure B3. Left: Cumulative spring discharge (April-August) taken at gage station 06478513 at the outlet of the James River HUC-6 Watershed plotted against the CMI values for the James River HUC-6 region for

December (2004-2019), January, February, and March 2004-2020. Colored boxes indicate whether the CMI performs as expected based on anomalous observed runoff, where blue points represent ‘hits’, and red points indicate ‘misses’. Right: Pearson’s Correlation Coefficient(r) for each month’s CMI vs cumulative discharge plot.

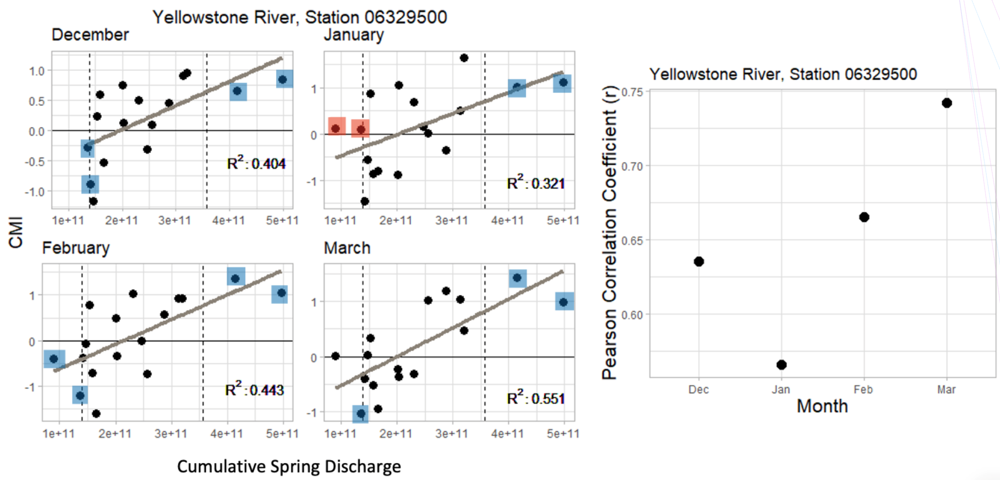
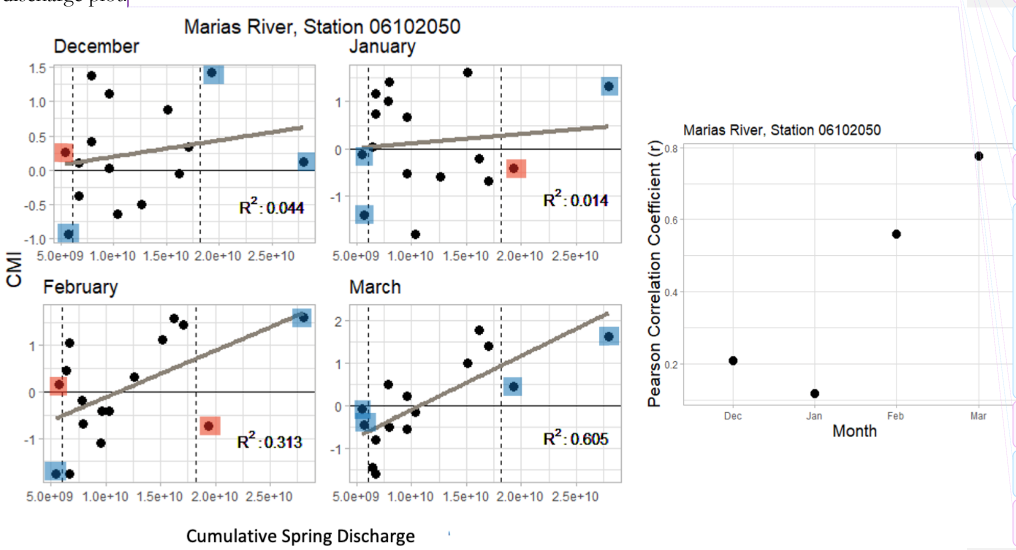


Figure B4. Left: Cumulative spring discharge (April-August) taken at gage station 06329500 at the outlet of the Yellowstone River Watershed plotted against the average CMI values for the five contributing HUC-6 regions for December (2004-2019), January, February, and March 2004-2020. Colored boxes indicate whether the CMI performs as expected based on anomalous observed runoff, where blue points represent ‘hits’, and red points indicate ‘misses’. Right: Pearson’s Correlation Coefficient(r) for each month’s CMI vs cumulative discharge plot.



Commented [DB121]: add to appendix

Commented [JS122R121]: Replace figure with one with clipped swe and snow depth?

Commented [DB123]: add to appendix

Commented [JS124R123]: Replace figure with one with clipped swe and snow depth?

Commented [DB125]: do we need to restate what hits and misses mean here or can i just say "hits" where CMI outputs accurately capture wet or dry conditions indicated by anomalous discharge values"

"red points indicate misses, where CMI outputs mischaracterize basin moisture conditions indicated by discharge values"

its a wall of text that i dont like to be so redundant

Commented [JS126R125]: I'd leave out the explanation of what a hit/miss is since we essentially defined it earlier

Commented [DB127]: add to appendix

Commented [JS128R127]: Replace figure with one with clipped swe and snow depth?

Commented [DB129]: add to appendix

Commented [JS130R129]: Replace figure with one with clipped swe and snow depth?

Commented [DB131]: add to appendix

Commented [JS132R131]: Replace figure with one with clipped swe and snow depth?

Commented [DB133]: add to appendix

Commented [JS134R133]: Replace figure with one with clipped swe and snow depth?

Commented [DB135]: do we need to restate what hits and misses mean here or can i just say "hits" where CMI outputs accurately capture wet or dry conditions indicated by anomalous discharge values"

"red points indicate misses, where CMI outputs mischaracterize basin moisture conditions indicated by discharge values"

its a wall of text that i dont like to be so redundant

Commented [JS136R135]: I'd leave out the explanation of what a hit/miss is since we essentially defined it earlier

Commented [DB137]: add to appendix

Commented [JS138R137]: Replace figure with one with clipped swe and snow depth?

Figure B5. Left: Cumulative spring discharge (April-August) taken at gage station 06102050 at the outlet of the James River HUC-6 Watershed plotted against the CMI values for the James River HUC-6 region for December (2004-2019), January, February, and March 2004-2020. Colored boxes indicate whether the CMI performs as expected based on anomalous observed runoff, where blue points represent 'hits', and red points indicate 'misses'. Right: Pearson's Correlation Coefficient(r) for each month's CMI vs cumulative discharge plot.

Commented [DB139]: add to appendix

Commented [JS140R139]: Replace figure with one with clipped swe and snow depth?

Commented [DB141]: add to appendix

Commented [JS142R141]: Replace figure with one with clipped swe and snow depth?

Commented [DB143]: add to appendix

Commented [JS144R143]: Replace figure with one with clipped swe and snow depth?

Commented [DB145]: do we need to restate what hits and misses mean here or can i just say "hits" where CMI outputs accurately capture wet or dry conditions indicated by anomalous discharge values"

"red points indicate misses, where CMI outputs mischaracterize basin moisture conditions indicated by discharge values"

its a wall of text that i dont like to be so redundant

Commented [JS146R145]: I'd leave out the explanation of what a hit/miss is since we essentially defined it earlier

Commented [DB147]: add to appendix

Commented [JS148R147]: Replace figure with one with clipped swe and snow depth?