

RandomForest model of BGC. Version 10.1

A climate model of Biogeoclimatic subzone/variants was constructed from a randomForest analysis of climate surface data (ClimateBC/WNA) using the following model components.

Spatial BGC layers:

- BC_BGC ver 10.0
- Alberta Natural Subregions v2015
- USA_BGC ver 5.0

Training points:

- 2000 randomly selected training points were generated for each subzone/variant using the ARCIInfo Random Sample function.

Climate Surface:

- ClimateBC 5.41 for British Columbia
- ClimateWNA 5.41 for adjacent jurisdictions

Variables used in rF BGCv10.1 model (Descriptions in Appendix 1):

Growing Season Precipitation	Growing season temperatures	Growing season Composite	Dormant Season
MSP	Tmax_sp	AHM	PAS
PPT_sp	Tmin_sm	CMD.total	Tmin_wt
PPT06	DD5_sp	Eref	MCMT
PPT_JAS	DD_06	TD	EMT
PPT_MJ	FFP		Tmin01
PPT_sm	eFFP		

Methods to Build BGC climate randomForest model

A climate model of Biogeoclimatic Units (BEC program of British Columbia) was constructed from a randomForest analysis (Breiman and Cutler) of the modelled climate surface of Wang and others (ClimateBC/WNA). Random training points were selected from a spatial coverage of BGC(ver10) using the ARCIInfo Random Sample function. All monthly, seasonal, and annual climate variables for the 1961-1990 normal period were generated for all training points using ClimateBC for British Columbia units and ClimateWNA for adjacent jurisdictions. Some climate variables that do not influence vegetation and are likely to be mere descriptors of climate (Woodward 1987) were removed prior to the random forest. These include dormant season precipitation (other than precipitation as snow (PAS)), and chilling and heating degree days. Additionally relative humidity and solar radiation were not included in model development due to climate surface modelling issues with these variables. A list of all ClimateBC/WNA variables are listed in Appendix 1.

Several additional conceptual climate variables were created from the available climate variables. These include May/June precipitation (PPT_MJ), May/June/July/August/Sept precipitation, growing degree days, and climatic moisture deficit (PPT.grow, DD5.grow, CMD.grow), and total climate moisture deficit (CMD.total). CMD.total was estimated by adding winter recharge deficit to the total CMD and. The winter recharge deficit is (estimated zonal soil water holding capacity (500mm) - sum of the dormant season precipitation). Positive recharge deficit is added to the CMD; negative values are treated as zero recharge deficit.

An initial model was created using the randomForest package ver 4.6-12 in R (Liaw and Wiener 2015) and a reduced climate variable data set from 100 random points per BGC, ntree=101, and other parameters set to defaults. In the rF importance table output, the rank of variable Accuracy and Gini values were calculated and then summed. The top 50 ranked-summed variables were retained and an second rF model generated from these variables only. Iteratively thereafter, the variables with the lowest 5 rank-summed values were then removed and another rF model generated. This step was repeated 5 times and the model error rate graphed to determine the minimum variable set size. The importance table was examined prior to variable removal to determine if the variable had high importance (>0.2) in blocks of related BGCs and if so the variable was retained. **Redundant variables were also examined in the importance table and removed if other related variables were also consistently high. Retained variable preference was for seasonal over monthly variables and the higher over lower ranked variable. Several commonly used climate variables that had low predictive power in the rF model were included in the final model for comparative purposes. (MAT, DD5, Tave07).**

The training point data set was then iteratively increased and a new rF model produced using the final variable set and the out-of-bag (OOB) error recorded for each iteration. The optimal training points size was selected at the run where the subsequent improvement in model OOB was <1%. This final model was run with ntree= 501 and the rF Error table was examined for the best ntree size determined where there was no longer significant improvement (>1%) in any of the individual BGC error rates.

Appendix 1. ClimateWNA variables

Variables not included in the generation of the randomForest model are formatted with ~~strike through~~. Final selected variables are highlighted in **yellow**.

1) Annual variables:

Directly calculated annual variables:

MAT	mean annual temperature (°C),
MWMT	mean warmest month temperature (°C),
MCMT	mean coldest month temperature (°C),
TD	temperature difference between MWMT and MCMT, or continentality (°C),
MAP	mean annual precipitation (mm),
MSP	mean annual summer (May to Sept.) precipitation (mm),
AHM	annual heat-moisture index $(MAT+10)/(MAP/1000)$
SHM	summer heat-moisture index $((MWMT)/(MSP/1000))$

Derived annual variables:

DD<0	degree-days below 0°C, chilling degree-days
DD>5	degree-days above 5°C, growing degree-days
DD<18	degree-days below 18°C, heating degree-days
DD>18	degree-days above 18°C, cooling degree-days
NFFD	the number of frost-free days
FFP	frost-free period
bFFP	the day of the year on which FFP begins
eFFP	the day of the year on which FFP ends
PAS	precipitation as snow (mm) between August in previous year and July in current year
EMT	extreme minimum temperature over 30 years
EXT	extreme maximum temperature over 30 years
Eref	Hargreaves reference evaporation (mm)
CMD	Hargreaves climatic moisture deficit (mm)
MAR	mean annual solar radiation ($MJ\ m^{-2}\ d^{-1}$)
RH	mean annual relative humidity (%)

2) Seasonal variables:

Seasons:

Winter (_wt): Dec. (prev. yr) - Feb for annual, Jan, Feb, Dec for normals

Spring (_sp): March, April and May

Summer (_sm): June, July and August

Autumn (_at): September, October and November

Directly calculated seasonal variables:

Tave_wt	winter mean temperature (°C)
--------------------	---

Tave_sp spring mean temperature (°C)
Tave_sm summer mean temperature (°C)
~~Tave_at~~ ~~autumn mean temperature (°C)~~

~~Tmax_wt~~ ~~winter mean maximum temperature (°C)~~
Tmax_sp spring mean maximum temperature (°C)
Tmax_sm summer mean maximum temperature (°C)
~~Tmax_at~~ ~~autumn mean maximum temperature (°C)~~

Tmin_wt winter mean minimum temperature (°C)
Tmin_sp spring mean minimum temperature (°C)
Tmin_sm summer mean minimum temperature (°C)
Tmin_at autumn mean minimum temperature (°C)

~~PPT_wt~~ ~~winter precipitation (mm)~~
PPT_sp spring precipitation (mm)
PPT_sm summer precipitation (mm)
~~PPT_at~~ ~~autumn precipitation (mm)~~

~~RAD_wt~~ ~~winter solar radiation ($\text{MJ m}^{-2} \text{d}^{-1}$)~~
~~RAD_sp~~ ~~spring solar radiation ($\text{MJ m}^{-2} \text{d}^{-1}$)~~
~~RAD_sm~~ ~~summer solar radiation ($\text{MJ m}^{-2} \text{d}^{-1}$)~~ could be good but errors
~~RAD_at~~ ~~autumn solar radiation ($\text{MJ m}^{-2} \text{d}^{-1}$)~~

Derived seasonal variables:

~~DD_0_wt~~ ~~winter degree days below 0°C~~
~~DD_0_sp~~ ~~spring degree days below 0°C~~
~~DD_0_sm~~ ~~summer degree days below 0°C~~
~~DD_0_at~~ ~~autumn degree days below 0°C~~

~~DD5_wt~~ ~~winter degree days below 5°C~~
DD5_sp spring degree-days above 5°C
DD5_sm summer degree-days above 5°C
~~DD5_at~~ ~~autumn degree days above 5°C~~

~~DD_18_wt~~ ~~winter degree days below 18°C~~
~~DD_18_sp~~ ~~spring degree days below 18°C~~
~~DD_18_sm~~ ~~summer degree days below 18°C~~
~~DD_18_at~~ ~~autumn degree days below 18°C~~

~~DD18_wt~~ ~~winter degree days below 18°C~~

DD18_sp ——— spring degree days above 18°C
DD18_sm ——— summer degree days above 18°C
DD18_at ——— autumn degree days above 18°C

NFFD_wt ——— winter number of frost free days
NFFD_sp ——— spring number of frost free days
NFFD_sm ——— summer number of frost free days
NFFD_at ——— autumn number of frost free days

PAS_wt ——— winter precipitation as snow (mm)
PAS_sp ——— spring precipitation as snow (mm)
PAS_sm ——— summer precipitation as snow (mm)
PAS_at ——— autumn precipitation as snow (mm)

Eref_wt ——— winter Hargreaves reference evaporation (mm)
Eref_sp ——— spring Hargreaves reference evaporation (mm)
Eref_sm ——— summer Hargreaves reference evaporation (mm)
Eref_at ——— autumn Hargreaves reference evaporation (mm)

CMD_wt ——— winter Hargreaves climatic moisture deficit (mm)
CMD_sp ——— spring Hargreaves climatic moisture deficit (mm)
CMD_sm ——— summer Hargreaves climatic moisture deficit (mm)
CMD_at ——— autumn Hargreaves climatic moisture deficit (mm)

RH_wt ——— winter relative humidity (%)
RH_sp ——— spring relative humidity (%) could be OK but errors in data
RH_sm ——— spring relative humidity (%)
RH_at ——— spring relative humidity (%)

3) Monthly variables

Primary monthly variables:

Tave01 – Tave12 January - December mean temperatures (°C) (5 – 9)
TMX01 – TMX12 January - December maximum mean temperatures (°C) (5 – 9)
TMN01 – TMN12 January - December minimum mean temperatures (°C) (only TMN01)
PPT01 – PPT12 January - December precipitation (mm) (Only PPT 06-08))
RAD01 – RAD12 ——— January - December solar radiation ($\text{MJ m}^{-2} \text{d}^{-1}$)

Derived monthly variables:

DD_0_01 – DD_0_12 January - December degree days below 0°C
DD5_01 – DD5_12 January - December degree-days above 5°C (5 – 9)

DD_18_01	DD_18_12	January – December degree days below 18°C
DD18_01	DD18_12	January – December degree days above 18°C
NFFD01	NFFD12	January – December number of frost free days
PAS01	PAS12	January – December precipitation as snow (mm)
Eref01 – Eref12		January – December Hargreaves reference evaporation (mm) (5 – 9)
CMD01 – CMD12		January – December Hargreaves climatic moisture deficit (mm) (5 – 9)