Prognose for vandindholdet i kernemajs	Ansvarlig	HJN
	Oprettet	14-12-2020
Projekt: [4584, Projektnavn]	Side	1 af 9

Formål

Formålet med dette notat er at få beskrevet og dokumenteret datagrundlag, modeller og drift af den eksisterende majsprognose for vandindholdet i kernemajs i CropManager.
Majsprognosen er udviklet af Teknologisk Institut (tidligere AgroTech) i samarbejde med SEGES i årene 2017 til 2019. Dette notat er opstartet af Teknologisk institut men færdigskrevet af SEGES i tæt samarbejde med Teknologisk institut.

Majsprognosen

Majsprognosen er en forecast model der prædikterer en majssorts tidsmæssige udvikling af tørstofprocenten (TS%) i en given aktuel mark. Majsprognosen inddrager data fra en aktuel mark (geografisk position, sådato, sort, JB-nr.) og klimadata (historiske klimadata, 1-uges vejrudsigt og normal-klimadata for pågældende mark). Modellen er bygget på baggrund af historiske klimadata samt data fra landsforsøgene, sortsafprøvninger i DLBR-regi eller i forskningsmæssige sammenhæng (data fra Aarhus Universitet), samt data fra bedrifter indsamlet af SEGES, kvæg.
Forecast modellerne er modul-baserede i tre moduler (Præprocessering, Prædiktion og Postprocessering) så de til enhver tid kan opdateres samt forbedres.
Da majssorterne er meget forskellige i tidlighed, har sorten en afgørende betydning for tørstofindholdet, og derfor er det også nødvendigt at angive sorten. Prognosen håndterer majssorter, som har deltaget i Landsforsøgene med majssorter til helsæd de to seneste år. Hvis marken er tørkepræget, påvirket af frost, unormal eller uens udviklet eller præget af ukrudt skal man anvende prognosen med forsigtighed.

Datagrundlag og modelbeskrivelse

Teknisk specifikation:
R-version og R pakker brugt til træning og udvikling af model

```
R version 3.3.3 (2017-03-06)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows >= 8 x64 (build 9200)

locale:
[1] LC_COLLATE=Danish_Denmark.1252  LC_CTYPE=Danish_Denmark.1252    LC_MONETARY=Danish_Denmark.1252
[4] LC_NUMERIC=C                    LC_TIME=Danish_Denmark.1252

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] scam_1.2-1            randomForestSRC_2.4.2 mgcv_1.8-17           nlme_3.1-131
 [5] zoo_1.7-14            doBy_4.5-15           jsonlite_1.3          RCurl_1.95-4.8
 [9] bitops_1.0-6          dplyr_0.5.0           RevoUtilsMath_10.0.0

loaded via a namespace (and not attached):
[1] Rcpp_0.12.9       lattice_0.20-34   assertthat_0.1    MASS_7.3-45       grid_3.3.3        R6_2.2.0
 [7] DBI_0.6           magrittr_1.5      Matrix_1.2-8      splines_3.3.3     RevoUtils_10.0.3 tools_3.3.3
[13] parallel_3.3.3    tibble_1.2
```

Datasæt til træning af modellerne

- Modellerne er bygget på baggrund af følgende data:
    - Forsøgs el. sortsdata (1992 - 2016), i alt 19.389 observationer fordelt over 10 forskellige datakilder.
      Se Tabel 1 for at se hvilke datakilder TS%-data stammer fra. De stammer fra 346 forskellige lokaliteter (unikke GPS-koordinater).
      Alle datakilder er leveret af Ib Sillebak Kristensen, AU undtagen NFTS-data som er trukket fra Nordic Field Trial System og data fra kvæg, leveret fra SEGES kvæg. Tabel 2 viser et udsnit af data.

- Klimadata (1985-2016, 40 km grid).*
  Her skal det nævnes at modellen er trænet på et 40 km grid, men i praksis kører den på et 10 km grid.

- Modellerne foreligger som R-filer.
  - De aktuelle modeller er 3 steps generaliserede additive (GAM) modeller samt en randomforest model.

Tabel 1: Antal observationer indenfor hver af de 10 "serier".

NFTS	DJF	HeKo	HoHe	Kold	Kvæg	Ma_N	Sort	Tyst	UK
10229	211	15	184	6	681	102	6129	394	1438

Tabel 2: Udsnit af forsøgs/sortsdata

	id	NFTS	serie	locality	PLANNR	LBNR	Grid40Km	EASTING	NORTHING	UTMZONE	HOSTAAR	saadato
1	1	Nej	Sort	5e+05.6300000	030229292	001	7	5e+05	6300000	32	1992	1992-05-12
2	2	Nej	Sort	5e+05.6300000	030229292	001	7	5e+05	6300000	32	1992	1992-05-12
3	3	Nej	Sort	5e+05.6300000	030229292	001	7	5e+05	6300000	32	1992	1992-05-12
4	4	Nej	Sort	5e+05.6300000	030229292	001	7	5e+05	6300000	32	1992	1992-05-12
5	5	Nej	Sort	5e+05.6300000	030229292	001	7	5e+05	6300000	32	1992	1992-05-12
6	6	Nej	Sort	5e+05.6300000	030229292	001	7	5e+05	6300000	32	1992	1992-05-12

	hostdato	host	hostDayNr	saa	GrowingDays	Forfbet	JB_nr	LSNBET	TS_helsæd	TS_kolbe
1	1992-10-07	0000-10-07	49	0000-05-12	148	<NA>	1	Fronica	30.24819	NA
2	1992-10-07	0000-10-07	49	0000-05-12	148	<NA>	1	Hiro	36.44400	NA
3	1992-10-07	0000-10-07	49	0000-05-12	148	<NA>	1	Erlevo	35.24400	NA
4	1992-10-07	0000-10-07	49	0000-05-12	148	<NA>	1	Challenger	35.94400	NA
5	1992-10-07	0000-10-07	49	0000-05-12	148	<NA>	1	Aviso	36.54400	NA
6	1992-10-07	0000-10-07	49	0000-05-12	148	<NA>	1	Solidor	33.42063	NA

	VAND_kerne	saa2	host2	KlGr	JB	HOSTAAR_fac	y	tidlighed_ts
1	NA	11	6	Ukendt	JB1,JB3	1992	30.24819	3.8956492
2	NA	11	6	Ukendt	JB1,JB3	1992	36.44400	0.9686683
3	NA	11	6	Ukendt	JB1,JB3	1992	35.24400	1.2330113
4	NA	11	6	Ukendt	JB1,JB3	1992	35.94400	0.3272476
5	NA	11	6	Ukendt	JB1,JB3	1992	36.54400	-0.8688007
6	NA	11	6	Ukendt	JB1,JB3	1992	33.42063	0.8212909

Tabel 3: Udsnit af klimadata

	Grid40Km	Year	DayNo	Date	Airtemp	Mintemp	Maxtemp	Glorad	Prec
1	1	1985	1	1985-01-01	-1.5	-2.4	-0.3	0.5	5.5
2	1	1985	2	1985-01-02	-5.0	-8.9	-1.6	1.1	0.5
3	1	1985	3	1985-01-03	-6.4	-8.4	-4.7	1.6	0.1
4	1	1985	4	1985-01-04	-8.0	-8.4	-7.5	0.9	0.1
5	1	1985	5	1985-01-05	-8.7	-9.5	-7.9	1.6	0.2
6	1	1985	6	1985-01-06	-10.5	-12.3	-9.0	1.3	2.5

Datamodel til forecast

Når landmanden laver en forecast forespørgsel er der brug for følgende input.

- Data input:
  - **Klimadata** (fra ClimateData tabellen se Tabel 5)
    - Klimadata fra aktuel lokalitet bestående af historisk observerede klimadata i indeværende år op til forespørgselsdatoen.
    - En uges prognose vejrdata tilgængeligt på forespørgselsdatoen.
    - Normklimadata (30 års normen) for resten af året indtil 1. november.
    - OBS. Der hentes data ud fra DMI i proceduren som gemmes ned med JobId, disse data bliver slettet igen efter prognosen så vil ikke være at genfinde i tabellen.
  - **Sortsnavn**, f.eks. "Atrium"
  - **Saadato**
  - **Forespørgselsdato**
  - **JB**-nr. i kategorierne "JB1, JB3" eller "JB2, JB4+".
  - **UTM-koordinater**. OBS: SEGES oversætter kommune til kommunens centroid-koordinat.
  - **TS_DK** i form af en JSON tabel. TS_DK dækker over tørstofprøver indhentet af SEGES i løbet af sæsonen og om muligt jævnt fordelt over Danmark for at kunne levere en mulighed for at kalibrere prædiktionen ind til indeværende år. Selvom modellen søger at indfange årsspecifikke tendenser i tørstof-udvikling via inddragelse af de historiske klimavariabler fra indeværende år, så forventes det, at inddragelsen af kalibreringsprøver i løbet af vækstsæsonen forbedrer prædiktionsevnen af modellen. JSON-tabellen skal

indeholde felterne: ”saadato", "hostdato", "UTME", "UTMN", "UTMzone", "KlGr" (kløvergræs med mulighederne ”Ja” el. Nej"), "JB" (se JB-nr), "sort" og "TS” (TS%).
    - **TS_inField** i form af en JSON tabel. TS_inField dækker over tørstofprøver indhentet af landmanden i aktuel mark for at kunne levere en mulighed for at kalibrere prædiktionen ind til den aktuelle mark. JSON-tabellen skal indeholde felterne: ”dateDay", "TS” (TS%).

- Yderlig data input:
    - **modeltabel.rda**. Modeltabellen indeholder den aktuelle model til brug for forecast, R data objekter der skal loades, klimadata variable der skal anvendes eller beregnes, navnet på den variabel, der beskriver sortens afvigelse, og endeligt navnene på de funktioner der præprocesserer, prædikterer og postprocesserer. Se Tabel 4 for et eksempel. Alle funktioner listet i modeltabellen skal være tilstede i R.filen CommonFunctions.r. Alle R-objekter, der skal loades skal findes i en undermappe ”data” i det workspace hvor der afvikles.
    - **majstabel.rda**. Majstabellen indeholder information over de enkelt sorters ”tidlighed” i form af en estimeret afvigelse fra gennemsnitstørstofprocent. Denne er estimeret på baggrund af modelprædikterede LSmeans tørstofprocenter for alle sorter over alle forsøgsår. Estimeringen af LSmeans stammer fra den årligt opdateret modelkørsel i Sortsvalg majs værktøjet, se her: https://sortsvalgmajs.dlbr.dk/da-dk. OBS: sorter der ikke er inkluderet i denne tabel kan der stadigvæk laves en prædiktion for. Disse vil så basere sig på en relativ tidlighed i TS% afvigelse fra gennemsnittet på 0, dvs. en gennemsnitssort. Se Tabel 6 for et eksempel.
      I modsætning til tidligere majsprognosemodeller, der byggede på konceptet majsdage som en skaleret afvigelse i TS% fra en referencesort, så bygger denne model på antagelsen, at den gennemsnitlige tørstofprocent på tværs af år og sorter vil være stabil og ikke ændre sig betydeligt, f.eks. en target-TS% på 32% ved høst. Hvis nu den optimale ensilage tørstof procent ændrer sig fordi der bruges andre teknikker, så vil denne antagelse ikke holde mere.
      Majstabellen bliver opdateret hvert år hos TI på baggrund af landsforsøgene via sortsvalg majs. Her er det dog kun for helsæd, men der bliver også beregnet tilsvarende for kolbe- /kernemajs.
      Modellen der her bruges for at beregne tidlighed er en mixed effects model hvor der efter korrektion for årseffekt og lokalitet beregnes en LSmeans for tørstofprocenten for hver sort. Her fratrækkes tørstofprocenten for gennemsnitssorten for at se hvor meget tidligere den givne sort er i tørstofprocent.

Tabel 4: modeltabel.rda

	__	id	afgrbet	model_name	active	update_dato	reference	rmsep
1	1	1	Majshelsaed Ibs	model 2013	0	2013-12-07 00:00:00.0000000	./dokumenter/Majsprognose_Rapport_2013_ISK.docx	2,2
2	2	2	Majshelsaed	Model v.1	1	2018-06-08 00:00:00.0000000	<NA>	<NA>
3	3	3	Kolbemajs	Model v.1	0	2018-06-08 00:00:00.0000000	<NA>	<NA>
4	4	4	Kernemajs	Model v.1	0	2018-06-08 00:00:00.0000000	<NA>	<NA>
5	5	5	Kolbemajs	Model v.2	1	2019-07-05 00:00:00.0000000	<NA>	<NA>
6	6	6	Kernemajs	Model v.2	1	2019-07-05 00:00:00.0000000	<NA>	<NA>
7	7	7	Majshelsaed	Model v.2	0	2019-09-23 00:00:00.0000000	<NA>	<NA>

	R_objects
1	
2	<NA>
3	helsaed_m_step1 <- readRDS('./data/helsaed_m_step1.RDS'); helsaed_m_step2 <- readRDS('./data/helsaed_m_step2.RDS'); klima_pca_model <- readRDS('./data/klima_pca_model.RDS'); m_kalibrering <- readRDS('./data/m_kalibrering.RDS')
4	kolbe_m_step1 <- readRDS('./data/kolbe_m_step1.RDS'); kolbe_m_step2 <- readRDS('./data/kolbe_m_step2.RDS'); helsaed_m_step1 <- readRDS('./data/helsaed_m_step1.RDS'); helsaed_m_step2 <- readRDS('./data/helsaed_m_step2.RDS'); klima_pca_model <- readRDS('./data/klima_pca_model.RDS'); m_kalibrering <- readRDS('./data/m_kalibrering.RDS')
5	predict_VAND_kerne <- readRDS('./data/predict_VAND_kerne.RDS'); kolbe_m_step1 <- readRDS('./data/kolbe_m_step1.RDS'); kolbe_m_step2 <- readRDS('./data/kolbe_m_step2.RDS'); helsaed_m_step1 <- readRDS('./data/helsaed_m_step1.RDS'); helsaed_m_step2 <- readRDS('./data/helsaed_m_step2.RDS'); klima_pca_model <- readRDS('./data/klima_pca_model.RDS'); m_kalibrering <- readRDS('./data/m_kalibrering.RDS')
6	kolbe_m_step1 <- readRDS('./data/kolbe_m_step1_v2.RDS'); kolbe_m_step2 <- readRDS('./data/kolbe_m_step2_v2.RDS'); klima_pca_model <- readRDS('./data/klima_pca_model.RDS'); m_kalibrering <- readRDS('./data/m_kalibrering.RDS')
7	predict_VAND_kerne <- readRDS('./data/predict_VAND_kerne.RDS'); kolbe_m_step1 <- readRDS('./data/kolbe_m_step1_v2.RDS'); kolbe_m_step2 <- readRDS('./data/kolbe_m_step2_v2.RDS'); klima_pca_model <- readRDS('./data/klima_pca_model.RDS'); m_kalibrering <- readRDS('./data/m_kalibrering.RDS')
8	helsaed_m_step1 <- readRDS('./data/helsaed_m_step1.RDS'); helsaed_m_step2 <- readRDS('./data/helsaed_m_step2.RDS'); klima_pca_model <- readRDS('./data/klima_pca_model.RDS'); m_kalibrering <- readRDS('./data/m_kalibrering.RDS')

	climate_variables	variety_variable	preprocesFUN	predictFUN	postprocesFUN
1	Airtemp, Mintemp, Glorad, Prec	Majsdage	convert2IbsKlima	predict_TS_Ib	identity
2	Airtemp, Mintemp, Maxtemp, Glorad, Prec, MVE, T8, T0_8, T8_12, T12, coolTS	tidlighed_ts	prepareClimate_model_v1	predictTS_model_v1	postProcess_TS_model_v1
3	Airtemp, Mintemp, Maxtemp, Glorad, Prec, MVE, T8, T0_8, T8_12, T12, coolTS	tidlighed_ts	prepareClimate_model_v1	predictTS_model_v1	postProcess_TS_model_v1
4	Airtemp, Mintemp, Maxtemp, Glorad, Prec, MVE, T8, T0_8, T8_12, T12, coolTS	tidlighed_ts	prepareClimate_model_v1	predictTS_model_v1	postProcess_TS_model_v1
5	Airtemp, Mintemp, Maxtemp, Glorad, Prec, MVE, T8, T0_8, T8_12, T12, coolTS	tidlighed_ts	prepareClimate_model_v1	predictTS_model_v2	postProcess_TS_model_v1
6	Airtemp, Mintemp, Maxtemp, Glorad, Prec, MVE, T8, T0_8, T8_12, T12, coolTS	tidlighed_ts	prepareClimate_model_v1	predictTS_model_v2	postProcess_TS_model_v1
7	Airtemp, Mintemp, Maxtemp, Glorad, Prec, MVE, T8, T0_8, T8_12, T12, coolTS	tidlighed_ts	prepareClimate_model_v1	predictTS_model_v1	postProcess_TS_model_v2

Tabel 5: Klimadata

	id	jobId	createdDate	editedDate	creator	editor	airtemp	mintemp	maxtemp	glorad	prec	precNoCor	dateDay	dataSource	expectedDataSource
1	653899	83822018	2019-05-06 10:52:30	2019-05-06 10:52:30	app_dmdb	app_dmdb	14.3730	12.8	18.5	5.1170	2.0	1.8220	2018-09-20	0	0
2	654324	83822017	2019-05-06 13:46:49	2019-05-06 13:46:49	app_dmdb	app_dmdb	15.5120	11.5	20.8	13.2150	6.6	6.0000	2017-07-11	0	0
3	671719	636645875	2019-05-10 09:44:36	2019-05-10 09:44:36	app_dmdb	app_dmdb	14.0600	11.5	16.3	6.3210	9.6	8.6590	2018-09-14	0	0
4	707584	82002018	2019-05-31 13:40:23	2019-05-31 13:40:23	app_dmdb	app_dmdb	14.6101	12.9	18.5	4.4392	1.7	1.5013	2018-09-20	0	0
5	708254	82202018	2019-05-31 13:54:25	2019-05-31 13:54:25	app_dmdb	app_dmdb	20.4160	13.8	26.5	23.5340	0.0	0.0000	2018-07-11	0	0
6	741284	69502018	2019-06-19 10:25:04	2019-06-19 10:25:04	app_dmdb	app_dmdb	7.2931	5.7	8.6	1.2042	2.2	1.8000	2018-11-30	0	0
7	742881	69502017	2019-06-19 13:44:48	2019-06-19 13:44:48	app_dmdb	app_dmdb	16.3125	13.4	18.8	15.9486	2.1	1.9472	2017-07-11	0	0
8	743278	2017	2019-06-19 13:45:53	2019-06-19 13:45:53	app_dmdb	app_dmdb	-0.2710	0.0	1.1	1.1000	2.2	1.5790	2017-12-10	0	0
9	743672	55002017	2019-06-19 13:48:04	2019-06-19 13:48:04	app_dmdb	app_dmdb	15.1000	6.8	21.1	14.0326	0.9	0.7728	2017-09-05	0	0
10	745877	2019	2019-06-20 10:21:28	2019-06-20 10:21:28	app_dmdb	app_dmdb	15.9000	12.4	19.4	18.1000	2.4	2.2000	2019-07-11	2	2

Tabel 6: majstabel.

	id	lsnbet_LF1	AFGRNR	LSNNR_LF1	Majsdage	Majsdage_se	N_kardex	N_hostaar	HOSTAAR_min	HOSTAAR_max	tidlighed_ts	afgrbet
1	197	Yukon	NA	26134	NA	NA	37	10	NA	NA	-3.7085088	Kernemajs
2	198	Ambition	NA	26437	NA	NA	37	10	NA	NA	-0.1030018	Kernemajs
3	199	Pinnacle	NA	29721	NA	NA	9	3	NA	NA	-0.3858953	Kernemajs
4	200	Megusto KWS	NA	29725	NA	NA	4	2	NA	NA	0.2420831	Kernemajs
5	201	Prospect	NA	520161	NA	NA	4	2	NA	NA	-1.1579169	Kernemajs
6	202	KWS Stefano	NA	520168	NA	NA	4	2	NA	NA	1.1670831	Kernemajs
7	203	Papageno	NA	520735	NA	NA	4	2	NA	NA	0.6920831	Kernemajs
8	204	KWS Prixdor	NA	520736	NA	NA	4	2	NA	NA	-0.8079169	Kernemajs

Data flowet i en forecast er følgende:

1) Præprocessering af input
2) Prædiktion
3) Postprocessering af forecast

Procedurerne, der skal afvikles, er specificeret i modeltabellen i databasen i den aktive linje. Det vil sige her specificeres for hver afgrøde (majshelsæd, kolbemajs eller kernemajs) hvilke klimavariable der skal hentes, hvilke funktioner der skal køres samt hvilke R objekter der skal loades og bruges i de forskellige skridt i forecasten. Disse R objekter er RDS filer der er dannet da modellen blev trænet. I det følgende er der beskrevet de funktioner, R objekter og variable der pt. er aktive (i december 2020).
Output fra en forecast kan ses i Fig 1.

1) ***Præprocessering af input***
   I dette skridt bearbejdes klimadata fra databasen så de kan indgå i prædiktionsmodellerne.
   Der ses kun på prædiktioner af høstdatoer i høståret mellem datoerne 1. august og 1. november. Hvis der findes flere observationer på samme dag i klimatabellen i databasen da tages et gennemsnit af disse inden de bearbejdes yderligere.
   Der afvikles funktionen *prepareClimate_model_v1.* I denne funktion loades R objektet *klima_pca_model.*
   Ud fra de eksisterende klimavariable i databasetabellen: Airtemp, Glorad, Mintemp, Maxtemp, Prec og PrecNoCor beregnes der nu de følgende klimavariable: *MVE, T8, T0_8, T8_12, T12 samt coolT5*, der skal bruges i prædiktionsmodellerne for majshelsæd. Beregningerne er beskrevet nedenfor.

   - *MVE = (ymax – ymin)/2, er Majs varme enheder*
     *hvor ymax = 3.33 * (Mintemp - 10) - 0.084 * (Maxtemp - 10)^2 og*
     *ymin = 1.8 * (Mintemp - 4.4)*
     se (https://www.landbrugsinfo.dk/public/0/d/5/plante_majsvarmeenheder_produktion_majsmark) for yderligere info.
   - *T8: Hvor meget luft temperaturen afviger/er højere end 8 grader. Det vil sige hvis temperaturen er over 8 da trækkes 8 fra ellers 0.*
   - *T0_8: Den givne temperatur mellem 0 og 8, ellers 0.*
   - *T8_12: Hvor meget luft temperaturen mellem 8 o g12 afviger/er højere end 8 grader. Det vil sige hvis temperaturen er mellem 8 og 12 da trækkes 8 fra ellers 0.*
   - *T12: Hvor meget luft temperaturen afviger/er højere end 12 grader. Det vil sige hvis temperaturen er over 12 da trækkes 12 fra ellers 0.*

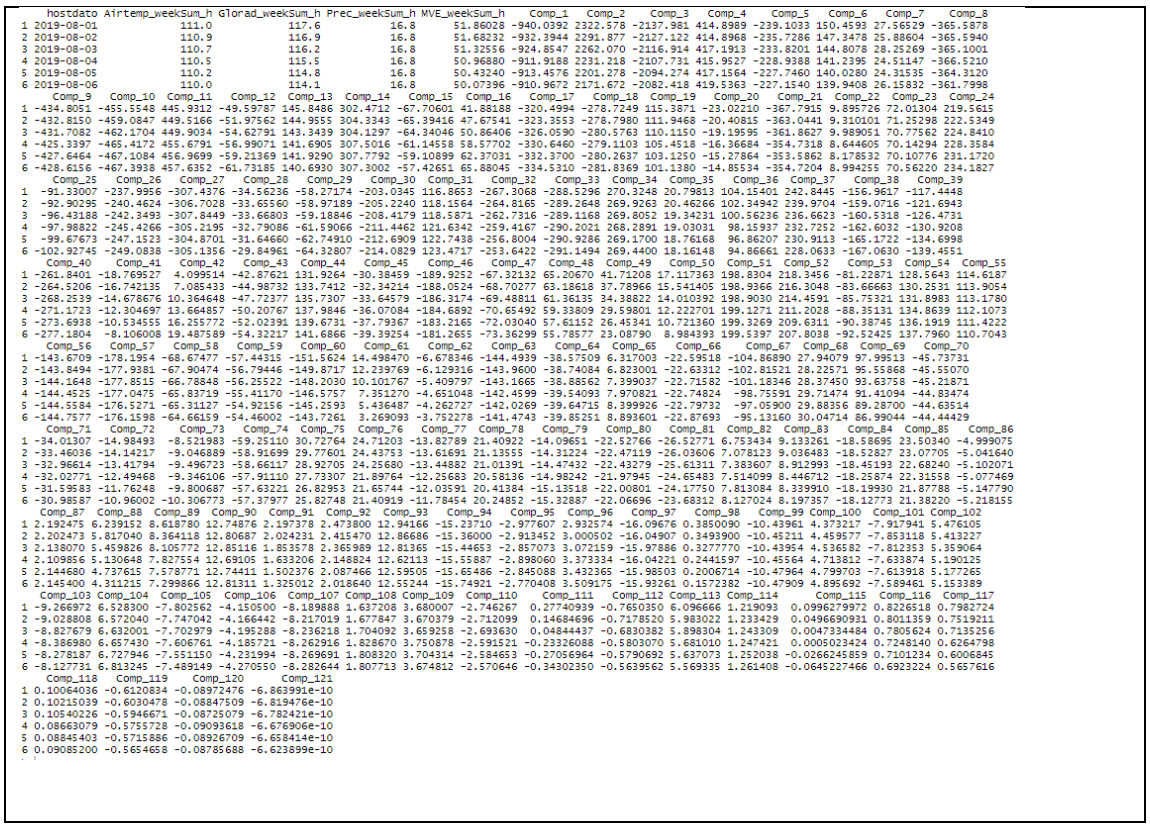
   Derudover beregnes også for hver dato en løbende uges-, måneds- og 2måneders sum, samt standardafvigelse for disse variable.
   Ud fra dette datasæt dannes der for hver mulig høstdato i vores høstperiode et nyt datasæt bestående af de ovenstående samt de løbende summer beregnet ud fra input sådatoen og de forskellige høstdage. For dette datasæt bestående af en høstdato samt 121 klimavariable bliver der prædikteret 121 principalkomponenter ud fra træningsmodellen *klima_pca_model* en principal komponentanalyse (PCA).
   klima_pca_model:

   princomp(X, covmat = cov.wt(X, wt = w/max(w)))

   Grundet den ulige fordeling af observationer på tværs af år blev hver observation vægtet med den inverse af observationsårets antal observationer. Dette forbedrer generaliserbarheden på tværs af år.

   Disse 121 principalkomponenter samt hostdato, Airtemp_weekSum_h, Glorad_weekSum_h, Prec_weekSum_h og MVE_weekSum_h danner tilsammen det nye klimainput til prædiktionen. Se Tabel 7.

   Tabel 7



```
    hostdato Airtemp_weekSum_h Glorad_weekSum_h Prec_weekSum_h MVE_weekSum_h    Comp_1   Comp_2   Comp_3   Comp_4   Comp_5   Comp_6   Comp_7   Comp_8
1 2019-08-01             111.0             117.6           16.8          51.86028 -940.0392 2322.578 -2137.981 414.8989 -239.1033 150.4593 27.56529 -365.5878
2 2019-08-02             110.9             116.9           16.8          51.68232 -932.3944 2291.877 -2127.122 414.8968 -235.7286 147.3478 25.88604 -365.5940
3 2019-08-03             110.7             116.2           16.8          51.32556 -924.8547 2262.070 -2116.914 417.1913 -233.8201 144.8078 28.25269 -365.1001
4 2019-08-04             110.5             115.5           16.8          50.96880 -911.9188 2231.218 -2107.731 415.9527 -228.9388 141.2395 24.51147 -366.5210
5 2019-08-05             110.2             114.8           16.8          50.43240 -913.4576 2201.278 -2094.274 417.1564 -227.7460 140.0280 24.31535 -364.3120
6 2019-08-06             110.0             114.1           16.8          50.07396 -910.9672 2171.672 -2082.418 419.5363 -227.1540 139.9408 26.15832 -361.7998
    Comp_9  Comp_10  Comp_11  Comp_12  Comp_13  Comp_14  Comp_15  Comp_16  Comp_17  Comp_18  Comp_19  Comp_20  Comp_21  Comp_22  Comp_23  Comp_24
1 -434.8051 -455.5548 445.9312 -49.59787 145.8486 302.4712 -67.70601 41.88188 -320.4994 -278.7249 115.3871 -23.02210 -367.7915 9.895726 72.01304 219.5615
2 -432.8150 -459.0847 449.5166 -51.97562 144.9555 304.3343 -65.39416 47.67541 -323.3553 -278.7980 111.9468 -20.40815 -363.0441 9.310101 71.25298 222.5349
3 -431.7082 -462.1704 449.9034 -54.62791 143.3439 304.1297 -64.34046 50.86406 -326.0590 -280.5763 110.1150 -19.19595 -361.8627 9.989051 70.77562 224.8410
4 -425.3397 -465.4172 455.6791 -56.99071 141.6905 307.5016 -61.14558 58.57702 -330.6460 -279.1103 105.4518 -16.36684 -354.7318 8.644605 70.14294 228.3584
5 -427.6464 -467.1084 456.9699 -59.21369 141.9290 307.7792 -59.10899 62.37031 -332.3700 -280.2637 103.1250 -15.27864 -353.5862 8.178532 70.10776 231.1720
6 -428.6156 -467.3938 457.6352 -61.73185 140.6930 307.3002 -57.42651 65.88045 -334.5310 -281.8369 101.1380 -14.85534 -354.7204 8.994255 70.56220 234.1827
    Comp_25  Comp_26  Comp_27  Comp_28  Comp_29  Comp_30  Comp_31  Comp_32  Comp_33  Comp_34  Comp_35  Comp_36  Comp_37  Comp_38  Comp_39
1  -91.33007 -237.9956 -307.4376 -34.56236 -58.27174 -203.0345 116.8653 -267.3068 -288.5296 270.3248 20.79813 104.15401 242.8445 -156.9617 -117.4448
2  -92.90295 -240.4624 -306.7028 -33.65560 -58.97189 -205.2240 118.1564 -264.8165 -289.2648 269.9263 20.46266 102.34942 239.9704 -159.0716 -121.6943
3  -96.43188 -242.3493 -307.8449 -33.66803 -59.18846 -208.4179 118.5871 -262.7316 -289.1168 269.8052 19.34231 100.56236 236.6623 -160.5318 -126.4731
4  -97.98822 -245.4266 -305.2195 -32.79086 -61.59066 -211.4462 121.6342 -259.4167 -290.2021 268.2891 19.03031  98.15937 232.7252 -162.6032 -130.9208
5  -99.67673 -247.1523 -304.8701 -31.64660 -62.74910 -212.6909 122.7438 -256.8004 -290.9286 269.1700 18.76168  96.86207 230.9113 -165.1722 -134.6998
6 -102.92745 -249.0838 -305.1356 -29.84961 -64.32807 -214.0829 123.4717 -253.6422 -291.1494 269.4400 18.16148  94.86661 228.0633 -167.0630 -139.4551
    Comp_40  Comp_41  Comp_42  Comp_43  Comp_44  Comp_45  Comp_46  Comp_47  Comp_48  Comp_49  Comp_50  Comp_51  Comp_52  Comp_53  Comp_54  Comp_55
1 -261.8401 -18.769527  4.099514 -42.87621 131.9264 -30.38459 -189.9252 -67.32132 65.20670 41.71208 17.117363 198.8304 218.3456 -81.22871 128.5643 114.6187
2 -264.5206 -16.742135  7.085493 -44.98732 133.7412 -32.34214 -188.0524 -68.70277 63.18618 37.78966 15.541405 198.9366 216.3048 -83.66663 130.2531 113.9054
3 -268.2539 -14.678676 10.364648 -47.72377 135.7307 -33.64579 -186.3174 -69.48811 61.36135 34.38822 14.010392 198.9030 214.4591 -85.75321 131.8983 113.1780
4 -271.1723 -12.304697 13.664857 -50.20767 137.9846 -36.07084 -184.6892 -70.65492 59.33809 29.59801 12.222701 199.1271 211.2028 -88.35131 134.8639 112.1073
5 -273.6938 -10.534555 16.255772 -52.02391 139.6731 -37.79367 -183.2165 -72.03040 57.61152 26.45341 10.721360 199.3269 209.6311 -90.38745 136.1919 111.4222
6 -277.1804  -8.106008 19.487589 -54.32217 141.6866 -39.39254 -181.2655 -73.36299 55.78577 23.08790  8.984393 199.5397 207.8038 -92.52425 137.7960 110.7043
    Comp_56  Comp_57  Comp_58  Comp_59  Comp_60  Comp_61  Comp_62  Comp_63  Comp_64  Comp_65  Comp_66  Comp_67  Comp_68  Comp_69  Comp_70
1 -143.6709 -178.1954 -68.67477 -57.44315 -151.5624 14.49470 -6.678346 -144.4939 -38.57509 6.317003 -22.59518 -104.86890 27.94079 97.99513 -45.73731
2 -143.8494 -177.9381 -67.90474 -56.79446 -149.8717 12.239769 -6.129316 -143.9600 -38.74084 6.823001 -22.63312 -102.81521 28.22571 95.55868 -45.55070
3 -144.1648 -177.8515 -66.78848 -56.25522 -148.2030 10.101767 -5.409797 -143.1665 -38.88562 7.399037 -22.71582 -101.18346 28.37450 93.63758 -45.21871
4 -144.4525 -177.0475 -65.83719 -55.41170 -146.5757  7.351270 -4.651048 -142.4599 -39.54093 7.970821 -22.74824  -98.75591 29.71474 91.41094 -44.83474
5 -144.5584 -176.5271 -65.31127 -54.92156 -145.2593  5.436487 -4.262727 -142.0269 -39.64715 8.399926 -22.79732  -97.05900 29.88356 89.28700 -44.63514
6 -144.7577 -176.1598 -64.66159 -54.46002 -143.7261  3.269093 -3.752278 -141.4743 -39.85251 8.893601 -22.87693  -95.13160 30.04714 86.99044 -44.44429
    Comp_71  Comp_72  Comp_73  Comp_74  Comp_75  Comp_76  Comp_77  Comp_78  Comp_79  Comp_80  Comp_81  Comp_82  Comp_83  Comp_84  Comp_85  Comp_86
1 -34.01307 -14.98493  -8.521983 -59.25110 30.72764 24.71203 -13.82789 21.40922 -14.09651 -22.52766 -26.52771 6.753434 9.133261 -18.58695 23.50340 -4.999075
2 -33.46036 -14.14217  -9.046889 -58.91699 29.77601 24.43753 -13.61691 21.13555 -14.31224 -22.47119 -26.03606 7.078123 9.036483 -18.52827 23.07705 -5.041640
3 -32.96614 -13.41794  -9.496723 -58.66117 28.92705 24.25680 -13.44882 21.01391 -14.47432 -22.43279 -25.61311 7.383607 8.912993 -18.45193 22.68240 -5.102071
4 -32.02771 -12.49468  -9.346106 -57.91110 27.73307 21.89764 -12.25683 20.58136 -14.98242 -21.97945 -24.65483 7.514099 8.446712 -18.25874 22.31558 -5.077469
5 -31.59583 -11.76248  -9.800687 -57.63221 26.82953 21.65744 -12.03591 20.41384 -15.13518 -22.00801 -24.17750 7.813084 8.339910 -18.19930 21.87788 -5.147790
6 -30.98587 -10.96002 -10.306773 -57.37977 25.82748 21.40919 -11.78454 20.24852 -15.32887 -22.06696 -23.68312 8.127024 8.197357 -18.12773 21.38220 -5.218155
    Comp_87  Comp_88  Comp_89  Comp_90  Comp_91  Comp_92  Comp_93  Comp_94  Comp_95  Comp_96  Comp_97  Comp_98  Comp_99 Comp_100 Comp_101 Comp_102
1 2.192475 6.239152 8.618780 12.74876 2.197378 2.473800 12.94166 -15.23710 -2.977607 2.932574 -16.09676 0.3850090 -10.43961 4.373217 -7.917941 5.476105
2 2.202473 5.817040 8.364118 12.80687 2.024231 2.415470 12.86686 -15.36000 -2.913452 3.000502 -16.04907 0.3493900 -10.45211 4.459577 -7.853118 5.413227
3 2.138070 5.459826 8.105772 12.85116 1.853578 2.365989 12.81365 -15.44653 -2.857073 3.072159 -15.97886 0.3277770 -10.43954 4.536582 -7.812353 5.359064
4 2.109856 5.130648 7.827554 12.69105 1.633206 2.148824 12.62113 -15.55887 -2.898060 3.373334 -16.04221 0.2441597 -10.45564 4.713812 -7.633874 5.190125
5 2.144680 4.737615 7.578771 12.74411 1.502376 2.087466 12.59505 -15.65486 -2.845088 3.432365 -15.98503 0.2006714 -10.47964 4.799703 -7.613918 5.177265
6 2.145400 4.311215 7.299866 12.81311 1.325012 2.018640 12.55244 -15.74921 -2.770408 3.509175 -15.93261 0.1572382 -10.47909 4.895692 -7.589461 5.153389
   Comp_103 Comp_104 Comp_105 Comp_106 Comp_107 Comp_108 Comp_109 Comp_110    Comp_111  Comp_112 Comp_113 Comp_114    Comp_115 Comp_116 Comp_117
1 -9.266972 6.528300 -7.802562 -4.150500 -8.189888 1.637208 3.680007 -2.746267  0.27740939 -0.7650350 6.096666 1.219093  0.0996279972 0.8226518 0.7982724
2 -9.028808 6.572040 -7.747042 -4.166442 -8.217019 1.677847 3.670379 -2.712099  0.14684696 -0.7178520 5.983022 1.233429  0.0496690931 0.8011359 0.7519211
3 -8.827679 6.632001 -7.702979 -4.195288 -8.236218 1.704092 3.659258 -2.693630  0.04844437 -0.6830382 5.898304 1.243309  0.0047334484 0.7805624 0.7135256
4 -8.386980 6.657430 -7.606761 -4.185721 -8.262916 1.828670 3.750878 -2.591521 -0.23326088 -0.5803070 5.681010 1.247421  0.0005023424 0.7248140 0.6264798
5 -8.278187 6.727946 -7.551150 -4.231994 -8.269691 1.808320 3.704314 -2.584653 -0.27056964 -0.5790692 5.637073 1.252038 -0.0266245859 0.7101234 0.6006845
6 -8.127731 6.813245 -7.489149 -4.270550 -8.282644 1.807713 3.674812 -2.570646 -0.34302350 -0.5639562 5.569335 1.261408 -0.0645227466 0.6923224 0.5657616
    Comp_118  Comp_119  Comp_120     Comp_121
1 0.10064036 -0.6120834 -0.08972476 -6.863991e-10
2 0.10215039 -0.6030478 -0.08847509 -6.819476e-10
3 0.10540226 -0.5946671 -0.08725079 -6.782421e-10
4 0.08663079 -0.5755728 -0.09093618 -6.676906e-10
5 0.08845403 -0.5715886 -0.08926709 -6.658414e-10
6 0.09085200 -0.5654658 -0.08785688 -6.623899e-10
```

2) Prædiktion

I dette skridt afvikles funktionen *predictTS_model_v1*. I denne funktion loades R objekterne *kolbe_m_step1_v2*, *kolbe_m_step2_v2, m_kalibrering og predict_VAND_kerne.*
Her bruges data input beskrevet længere oppe, den nye præprocesserede klima datatabel, samt data fra majstabellen for den forespurgte sort.

Hvis der i løbet af sæsonen er foretaget tidlige tørstofprøver for majs da vil de eller bør de blive indrapporteret af administrator (SEGES) og de vil da være at finde i tabellen *TS_DK*. Ved hjælp af disse prøver vil der blive beregnet en årsvariation som så bliver brugt som kalibrering for årsniveauet. Tanken er her at prøver taget tidligt i løbet af sæsonen på forskellige lokationer kan bruges til at niveauforskyde/parallelforskyde den prædikterede tørstofprocent.
Man kender tidligheden ud fra majsforsøgene, men da man ved der er en stor variation fra år til år, så kan man, ud over ved inddragelse af klimadata fra det aktuelle år, ud fra disse tidlige prøver se hvordan tørstofprocenten reagerer med det pågældende års klimadata. Man kan altså tage højde for en årsvariation der indikerer et højere/eller lavere niveau og dermed forbedre prædiktionsevnen. Der findes ikke prøver for 2019 og 2020, men det vil blive set ind i, om der kan oprettes nogle procedurer således at prøver kan indhentes til databasen fremover for de sorter hvor vi også kender tidligheden, da dette kun vil forbedre prædiktionsmodellen.
Ud fra alle udtagne prøver i det givne prædiktionsår kobles tidlighedsdata tidlighed_ts fra majstabellen på samt forskel i dage til en reference sådato 1. maj og en forskel i dage til en reference høstdato 1. oktober. Ud fra træningsmodellen *m_kalibrering* prædikteres tørstofprocenten for prøverne for et gennemsnitsår på den lokalitet det drejer sig om og der beregnes en gennemsnitlig årsafvigelse for den prædikterede og observerede tørstofprocent. Modellen er defineret som nedenfor:
m_kalibrering:

TS ~ s(week, EASTING, NORTHING, k = 35) + KlGr + JB +
s(locality, bs = "re") + s(HOSTAAR_fac, bs = "re") +
saa2 + tidlighed_ts

Modellen modellerer TS% i helsæd som en glat funktion (spline funktion) af det tre-dimensionelle rum udspændt af geografisk beliggenhed (*EASTING, NORTHING*) og tid (week),

indikatorvariablene for JB-nr. kategori, kløvergræs som forfrugt, en lineær effekt af såtidspunkt (saa2) og sorternes tidlighed (tidlighed_ts) beregnet som TS%-afvigelsen fra gennemsnits TS% på tværs af alle sorter. Derudover er der modelleret en tilfældig normalfordelt effekt af lokalitet og høstår. Grundet den ulige fordeling af observationer på tværs af år blev hver observation vægtet med den inverse af observationsårets antal observationer. Dette forbedrer generaliserbarheden på tværs af år. Den gennemsnitlige årsafvigelse på baggrund af TS_DK beregnes som den gennemsnitlige residual, dvs. forskel mellem prædikteret (*m_kalibrering*) og observeret (TS_DK), for kalibreringsprøverne.

Det nye datasæt består nu af de præprocesserede klimadata, en gennemsnitlig årsafvigelse, en tidlighed_ts for sorten, input koordinater, en forskel i dage til reference sådatoen (1. maj) og en forskel i dage til reference høstdatoen (1. oktober). Der prædikteres nu tørstofprocent i en 2 skridts model hvor tørstofprocenten prædikteres ud fra de fittede værdier i en GAM model *kolbe_m_step1_v2* (1) samt residualerne fra en random forest model *kolbe_m_step2_v2* (2).
(1) kolbe_m_step1_v2:

y ~ s(EASTING, NORTHING, k = 7, fx = T) + TS_year_deviation +
s(host2, k = k, fx = T) + tidlighed_ts + saa2 + JB

Denne GAM (Generalized additive model) model består af en glat funktion af koordinaterne easting og northing med en fixed udglatningsgrad, en lineær effekt af årsvariationen der indgår som en parallelforskydning, en glat funktion af host2 som er en numerisk værdi af høstdatoen, en lineær effekt af tidligheden, såtidspunkt samt jordbundstype.
Motivationen bag step 1 modellen er, at levere en model der kan levere prædiktioner for alle geografiske placeringer ved interpolering langs den estimerede glatte flerdimensionelle modelflade.

(2) kolbe_m_step2_v2:

res ~ tidlighed_ts + JB + fit_step1 + EASTING + NORTHING + saa2 + host2 +
TS_year_deviation + Comp_1 + Comp_2 + Comp_3 + Comp_4 + Comp_5 +
Comp_6 + Comp_7 + Comp_8 + Comp_9

Denne model er en random forest model. En random forest er en machine learning prædiktionsmodel, der prædikterer responsvariablen som den gennemsnitlige prædiktion på tværs af en stor samling (skov) af decision tree modellers prædiktioner, hvor hver decision tree model er trænet på baggrund af et subsample af observationer og et subsample af forklarende variabler for derigennem at opnå dekorrelerede modeller i skoven. En decision tree model er en model der består af en hierarkisk samling af if-else statements (tvegrenende træ). En random forest er dermed en ikke-parametrisk model, som kan approksimere en hvilken som helst sammenhæng mellem input og outputvariabler. Modellens formel skal forstås på den måde, at variablerne listet på højre side af ~ er brugt som forklarende variabler i random forest modellens decisions tree modeller.
Motivationen bag step 2 modellen er, at indfange lokale effekter både i tid (klimaeffekter som PCA-scores) og rum (*EASTING, NORTHING*) ikke indfanget af den glatte step 1 model, dvs. at levere nuancer, der hvor step 1 modellen er upræcis.
Den samlede TS% prædiktionen er summen af step1 og step 2 prædiktionen.

OBS: i Juli 2019 ser det ud til der blev taget en version 2 i brug af de to ovenstående. Der er for step1 delen ændret til at medtage JB nummer i træningsmodellen, derudover er der ikke en glat funktion klimavariablen precipitation (nedbør) mere, og ej heller en glat funktion af pred_helsæd (det vil sige at helsædsprøverne blev brugt til at styrke prædiktionen for kolbe og kernemajs i forhold til lokation og klima da der ikke har været så stort et datagrundlag). Der er i stedet inddraget en glat funktion af lokation, det vil sige EASTING og NORTHING. For step2 er der i version 2 inkluderet JB samt EASTING og NORTHING variablene i randomforest modellen.

For Kernemajs er det ikke tørstofprocenten men derimod vandindholdet der skal afrapporteres. Derfor bruges de prædikterede tørstofprocenter for kolbemajs i funktionen *predict_VAND_kerne* som omregner til vandindholdet.
predict_VAND_kerne:

*Vandindhold = 138.776 + -13.46026 * sqrt(Tørstofprocent for kolbemajs)*

3) ***Postprocessering af forecast***

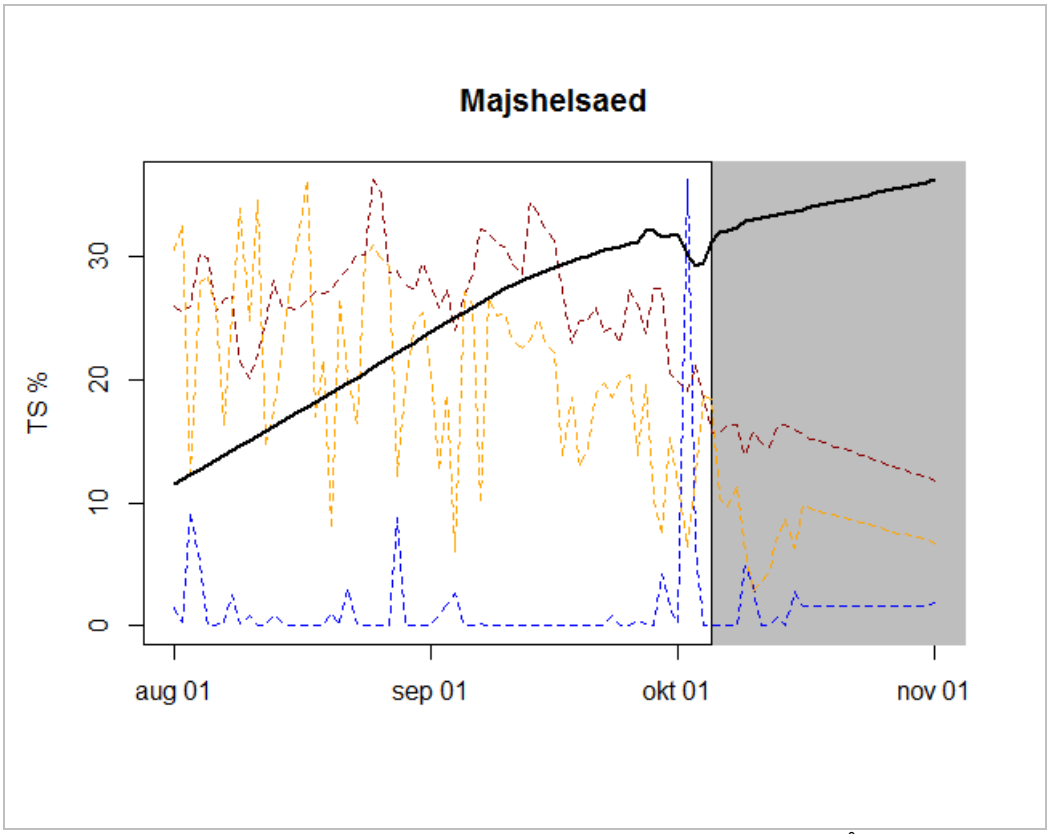
Funktionen *postProcess_TS_model_v1* afvikles. Tanken her er at der ønskes en udglatning af modellens prædiktioner da der er flere svingninger fra dag til dag end der var ønsket fra opgavestiller. Der blev vurderet at dette ville være forvirrende for landmanden at tolke på, og derfor blev der ønsket en udglatning af vandindholdsprædiktionerne.
Der er både en logit vægtning af observationer samt en SCAM model (Shape constrained additive model, som er en GAM model men hvor der er sat restriktioner på de splines der fittes) hvor der her for kernemajs er krævet en monotont faldende form og for de andre en monotont stigende form.
Det prædikterede vandindhold (yhat) bruges i en scam model

yhat ~ s(host)

for at smoothe prædiktionerne. Høstdatoer tættere på midten af høstsæsonsperioden er vægtet højere end først eller sidst på sæsonen.
Ud over dette er disse brugt i en udglatningsfunktion. Se *postProcess_TS_model_v1* funktionen for nærmere uddybelse.

Dertil kommer at findes der majs tørstofprøver indrapporteret af landmanden selv for aktuelle lokalitet, da beregnes der en gennemsnitligt afvigelse mellem det udglattede estimerede vandindhold og det observerede vandindhold. Det prædikterede vandindhold parallelforskydes derefter med denne korrektion. Det vil sige niveauet bliver korrigeret for lokationen såfremt at der er set et lavere eller højere niveau end forventet og dette styrker dermed prædiktionen.
Desuden afgrænses vandindholdet til at være mellem 0 og 100 hvis de prædikterede værdier skulle ligge udenfor dette interval.

Tilsidst samles outputtet så der for hver høstdato i intervallet består af sort, prædikteret vandindhold, sådato, JB-nummer, UTM koordinater samt klimadata.



Figur 1: Eksempel af forecast profilen for en majshelsædssort. (Sort = Atrium, sådato = 2016-05-01, forespørgselsdato = 2016-10-05, JB = ”JB1, JB3”, UTME = 606571, UTMN = 6131654, UTM zone = 32.) Grafen vil være tilsvarende for kolbe- og kernemajs.

Afvikling, drift og årlig opdatering
Den mere tekniske del som SEGES, digital står for vedrørende afvikling af R scripts i crop manager vil ikke blive beskrevet i dette dokument.
Hvert år den 1. december da genkøres sortsvalg majs hos TI. Her dannes der input til Sortsvalg majs programmet. Her beregnes også input til majstabellen for majshelsæd. I samme forbindelse afvikles der lignende for kolbe- og kernemajs. Resultater sendes til SEGES der opdaterer majstabellen i databasen til brug for majshøstprognosen.
Der har tidligere været indberettet kalibreringsprøver som SEGES har fået taget. Disse prøver er ikke blevet taget i 2019 og 2020. Det ville styrke prædiktionsmodellen hvis disse prøver blev taget og der burde i den forbindelse sikres en automatiseret proces, sådan at kalibreringsprøver bliver læst ind i tabellen TS_DK.

Udviklingsmuligheder
Der er trænet på 40 km grids hvor man i den kørende prædiktion har 10 km grids. Dette kan betyde at prædiktionerne rammer skævt.
Hvis der prædikteres for en kolbe/kerne sort hvor der ikke findes helsædsprøver, da parrer man op mod en gennemsnitlig sort, her kunne der overvejes om man skulle vælge en kategori af sorter der minder mere om sorten.
Der er trænet på modeller hvor der findes kalibreringsprøver. Disse har været manglende i både 2019 og 2020. Der fjernes noget styrke fra modellen.
Der er ikke afrapporteret usikkerheder på prædiktionsestimaterne. Dette kunne man overveje, da der naturligvis er noget usikkerhed forbundet med estimaterne. Der skal dog gøres nogle overvejelser hvordan man præcis beregner disse usikkerheder for eksempelvis en blanding af en GAM model og en randomforestmodel.
Version to modellen taget i brug i 2019 burde måske ses på igen i forhold til at den ikke længere inddrager viden omkring lokalitet og klima fra helsædsprøverne.