A Manual for Use of BESTPRED: A Program for Estimation of Lactation Yield and Persistency Using Best Prediction

Release 2.0 rc 7

J. B. Cole and P. M. VanRaden

August 12, 2009
Revised April 27, 2015

Animal Genomics and Improvement Laboratory, Agricultural Research Service, United States Department of Agriculture, Room 306 Bldg 005 BARC-West, 10300 Baltimore Avenue, Beltsville, MD 20705-2350

This page intentionally left blank.

Abstract

Cole, J.B. and VanRaden, P.M. 2009. A Manual for Use of BESTPRED: A Program for Estimation of Lactation Yield and Persistency Using Best Prediction. Animal Genomics and Improvement Laboratory, Agricultural Research Service, United States Department of Agriculture.

This manual describes how to obtain, install, and run a Fortran 90 program for calculating lactation yields and persistencies using best prediction methodology. Chapters also present the theory underlying best prediction, describe the methodology used to calculate covariances among pairs of days in milk for any combination of traits, and discuss implementation details important to anyone considering modifications to the programs. The purpose is to provide both background material and step-by-step instructions for successful use of the software.

Revised April 27, 2015

CONTENTS

LIST OF TABLES

This page intentionally left blank.

LIST OF FIGURES

This page intentionally left blank.

CHAPTER

ONE

Preface

Cole, J. B., and P. M. VanRaden. 2008. A Manual for Use of BESTPRED: A Program for Estimation of Lactation Yield and Persistency Using Best Prediction.

U.S. Department of Agriculture, Agricultural Research Service, Animal Genomics and Improvement Laboratory.

This manual describes how to obtain, install, and run a Fortran 90 program for calculating lactation yields and persistencies using best prediction methodology. Chapters also present the theory underlying best prediction, describe the methodology used to calculate covariances among pairs of days in milk for any combination of traits, and discuss implementation details important to anyone considering modifications to the programs. The purpose is to provide both background material and step-by-step instructions for successful use of the software.

The authors would like to thank Brad Heins (University of Minnesota) for bug reports and porting assistance.

This page intentionally left blank.

CHAPTER

TWO

Introduction

Best prediction (BP) (VanRaden 1997; VanRaden 1998) is a computational method derived from selection index (Cameron 1997) that allows the prediction of missing yields from measured yields and condensation of test day data into lactation yields and persistencies, and was approved in 2002 for lactation record calculation by the International Committee for Animal Recording (International Committee for Animal Recording 2006). It requires only phenotypic covariances, and assumes that herd means and variances are known. Reverse prediction may be used to obtain daily yields from lactation yields and persistencies. Both single and multiple trait prediction are supported. Calculation of lactation records for milk (M), fat (F), protein (P), and somatic cell score (SCS) for cows calving in January 1997 and later using BP began in November 1999, replacing the test interval method and projection factors (Shook, Jensen, and Dickinson 1980; Shook, Johnson, and Dickinson 1980) at AIPL.

Norman et al. (1999) showed that the advantages of BP are small compared to the Test Interval Method (TIM) (Sargent, Lytton, and Wall 1968) for most 305-d lactations but larger for lactations with infrequent testing or missing component samples. Some biases in TIM were eliminated by the introduction of Shook factors (Shook, Johnson, and Dickinson 1980), but BP proved to be more useful when dealing with many alternative testing plans. It permits more precise estimation of records for SCS because test days are adjusted for stage of lactation. Yield records calculated using BP have slightly lower standard deviations (SD) because BP regresses estimates toward the herd average.

The Canadian system for estimating 305-d records, multiple trait prediction (MTP) (Schaeffer and Jamrozik 1996), is similar to BP in many respects. In a recent validation study Quist et al. (2007) compared estimates of actual daily and 305-d yields from MTP with data collected from on-farm milk meters and reported that concordance correlations increase with DIM. They also reported that MTP overestimates actual 305-d yield, which is consistent with results from BP.

The original program for computing BP, AIPLDCR, was limited to 305-d lactations, required recompilation to change parameters, used only simple linear interpolation for calculating standard curves, and could not provide BP for individual days of lactation. A new version of the program, BESTPRED, has a number of advantages over the original routines:

- Lactations of any length can be modeled.
- Lactation-to-date, 305-d, 365-d, and projected yields are provided.
- Persistencies of milk, fat, protein, and SCS are now comparable on within, rather than across, lactation groups.
- The autoregressive function used to model correlations among test day yields has been updated.
- Program options are set in a parameter file and read at runtime.
- Diagnostic plots are available for all traits.
- BP of individual daily yields, test day yields, and standard curves are now output.

- The program accepts input in, and provides output in, either pounds or kilograms for M, F, and P.

This document describes how to obtain, install, and run the BESTPRED v 2 software package for calculating yields and persistencies using best prediction. It also provides a mathematical overview of BESTPRED, references to the schenftific literature, and notes on computational details and troubleshooting.

CHAPTER

THREE

Installation

The BESTPRED program is distributed as Fortran 90 and C source code that must be compiled before it can be used. Makefiles are provided for gcc 64-bit 4.2.0/Absoft 64-bit Fortran 95 v10.1[1] on Linux-based operating systems (tested on 64-bit RedHat Enterprise Linux 4). Binary versions of the programs are not available due to the difficulty of supporting the wide array of possible compiler-operating system combinations. Table 3.1 lists operating system–compiler combinations that were tested by AIPL[2] ("AIPL tested") or reported as working by users ("User tested").

Table 3.1: Platforms on which BESTPRED has been tested

Operating System	Distribution	C Compiler	Fortran Compiler	Status
Linux	Debian 3.1	gcc	Absoft 10.1	User tested
	Kubuntu 7.10	gcc	G95	AIPL tested
	RedHat Enterprise Linux 4	gcc	Absoft 10.0	AGIL tested
	RedHat Enterprise Linux 4	gcc	Absoft 10.1	AGIL tested
	RedHat Enterprise Linux 5	gcc	Absoft 10.1	User tested
	RedHat Linux 7.1	gcc	Absoft 7.5	User tested
MacOS	OS X Yosemite (10.10.3)	gcc	gfortran	AGIL tested
Microsoft Windows	Windows XP Professional	MinGW/gcc	G95	AGIL tested

3.1 Obtaining BESTPRED

The BESTPRED program is available for download free-of-charge from the AIPL website (http://www.aipl.arsusda.gov/software/bestpred/). The programs are distributed in gzipped tarballs for *nix platforms and ZIP files for MS Windows that must be unpacked with a decompression utility such as 7-Zip (http://www.7-zip.org/). The compressed files distributed by AIPL are not password-protected but MD5 checksums are provided on the website. BESTPRED is public domain software that may be freely modified and redistributed.

3.2 Structure of BESTPRED

There are several C- and Fortran-language files in the BESTPRED distribution. Table 3.2 describes the contents of those files and provides some important usage notes, although not all subroutines and functions are described. If you

[1] Reference herein to any specific commercial products, processes, or services by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the author.

[2] Note that Windows XP Professional is the only version of MS Windows approved for use on agency computers, and that AIPL's production environment is currently Linux-based. Reports of successful deployments on Microsoft platforms are particularly welcome.

want to use BESTPRED as a standalone program for reading data from a file of Format 4 records see Section 3.3 for compilation instructions. If you wish to include one or more of the BESTPRED subroutines in one of your programs you'll need to carefully consult Table 3.2 to make sure that you include the necessary components. At a bare minimum you'll need to include the bestpred_parm.f90 and bestpred.f90 files in your project or makefile. If you are not processing Format 4 records then you do not need to include `bestpred_fmt4` and its dependencies in your program. You will also need to include the `read_parms` subroutine (from the file bestpred_parm.f90) in order to load the parameter file.

Table 3.2: BESTPRED subroutines and functions

File	Subroutine Name	Description
ageadjs.c	`ageadj1`	C-language subroutines for age adjustment of MFP. You need to include these in any program that also uses the `bestpred_fmt4` subroutine.
	`ageadj4`	To call the `ageadj1` routine from Fortran
	`adjscs`	Age adjustment of somatic cell score
	`adjscs4`	To call the `adjscs` routine from Fortran
aiplage.c	`aiplage`	C-language subroutine for obtaining age–season–previous days open adjustment factors that differ by breed, region, and time period
bestpred.f90	`bestpred`	Subroutine for calculating Data Collection Ratings (DCR), yields (projected laclen, partial, 305, 365), reliability of MFPS, 305-d persistency and reliability of persistency from test day data. If you want to use BESTPRED as a component of one of your own programs use this subroutine.
	`adjust3X`	Adjust MFPS for 3X milking
	`binsort`	Binary sort of integer list using repeated subset merges
	`covary`	Compute covariance of test day and lactation yield
	`interpolate`	Interpolate between monthly means and SD to form standard curves (see: Section 6.2)
	`invrt2`	Invert a non-symmetric matrix of order n
	`mult`	Multiply matrices
	`vary`	Compute covariance of any two test day yields (see: Section 6.3)
	`ymean`	Calculate expected daily yield from standard lactation curves
bestpred_fmt4.f90	`bestpred_fmt4`	Accepts data in AIPL's Format 4 and prepares them for processing by the `bestpred` subroutine. You do not need to use this subroutine in your own programs if your data are not in Format 4.
bestpred_log.f90	`log_check_file_exists`	Checks to see if a logfile already exists; returns 0 if a file exists, a value other than 0 if not (the value is compiler-dependent).
	`log_create_file`	Creates a logfile; returns 0 if a file exists, a value other than 0 if not (the value is compiler-dependent).
	`log_message`	Writes a message to a logfile, creating a new file if one does not currently exist.
bestpred_main.f90	`bestpred_main`	Main program for stand-alone version of BESTPRED that reads data from external files and passes them to `bestpred_fmt4`. You do not need to use this subroutine in your own programs.

continued on next page

continued from previous page

File	Subroutine Name	Description
bestpred_parm.f90	`read_parms`	Reads program parameters from the file bestpred.par. You need to use this subroutine in your own programs in order to read the parameter file.

bestpred_parm.f90 and bestpred.f90 are pure Fortran 90 and you include them in your program as you would any other Fortran 90 file. For example, in the default BESTPRED makefile bestpred_parm.f90 is first compiled to an object file:

```
bestpred_parm.o: bestpred_parm.f90
  f90 -c bestpred_parm.f90 -O3 \
    -YEXT_NAMES=LCS -s                  \
    -YEXT_SFX=""
```

and then linked into the `bestpred` executable:[3]

```
bestpred: bestpred_parm.o ...
  f90 bestpred_main.f90 -o bestpred -O3 \
    ... \
    bestpred_parm.o
```

to produce the bestpred executable.

Note that the standalone executable program is named bestpred (*nix) or bestpred.exe (MS Windows). The executable program should not be confused with the `bestpred` subroutine contained in the file bestpred.f90.

3.3 Compiling BESTPRED

If you want only to use the `bestpred` subroutine you need a Fortran 90 compiler. You must have both Fortran 90- and C-language compilers in order to build a standalone (executable) version of BESTPRED, or to use both the `bestpred_fmt4` and `bestpred` subroutines. Sample makefiles are provided with the distribution, including makefile.Absoft (for use with the gcc and Absoft Fortran v10 compilers), makefile.gnu (for use with the gcc and gfortran compilers), and makefile.g95 (for use with the gcc and G95 compilers). To compile BESTPRED, copy the appropriate makefile to a file named makefile and type make:

```
[jcole@aipl366 bp]$ cp makefile.Absoft makefile
[jcole@aipl366 bp]$ make
```

Assuming that there are no problems[4] you should see output that looks something like:

[3] This is not a complete version of the makefile directive for the bestpred executable; several statements, indicated by . . . , have been omitted for brevity.

[4] This could happen, in theory.

```
[jcole@aipl366 bp]$ make
f90 -c aipldcr.f90 -O2 \
        -YEXT_NAMES=LCS -s
        -YEXT_SFX=""   \
f90 -c fmt4dcr.f90 -O2   \
        -YEXT_NAMES=LCS -s  \
        -YEXT_SFX=""   \
        ./ageadjs.o         \
        ./aiplage.o
f90 maindcr.f90 -o maindcr -O2 \
        -YEXT_NAMES=LCS          \
        -YEXT_SFX=""                \
        -s -lm -llocal -lU77           \
        ./ageadjs.o                      \
        ./aiplage.o                      \
        ./fmt4dcr.o                      \
        ./aipldcr.o
[jcole@aipl366 bp]$
```

If you are using a compiler other than the Absoft or GNU compilers you will need to modify one of the existing makefiles to work with your compiler. In general terms, this involves adding the correct path to the compiler executable and setting compiler flags. You will need to consult your compiler documentation for additional details as a more detailed description of compiler usage is beyond the scope of this document. Makefiles for additional compilers will be gratefully accepted for distribution.

3.3.1 Compiler Flags

Compiler flags are used to provide to the compiler with information about library file locations, the level of optimization desired, and similar things. While a wide array of compiler share some flags in common it is not generally the case that you can pass flags intended for one compiler to another without problems. The flags used in the default BESTPRED makefile are described in Table 3.3.1 to assist users in porting from one platform to another.

Table 3.3: Description of compiler flags in default makefile

Absoft Flag	Description	gfortran Flag
-c	Suppresses creation of an executable file – leaves compiled files in object code format.	-c
-l	Library file specification.	-l
-o *name*	Directs the compiler to produce an executable file called *name*.	-o *name*
-O2	Enables block level optimization.	-O2
-Rb	Generate code to check array boundaries.	-fbounds-check
-Rs	Generate code to check array conformance.	-fbounds-check
-s	Allocate local variables statically.	-fno-automatic
-YEXT_NAMES	Specify procedure names externally in upper, lower, or mixed case. Fortran and C compilers must be passing and receiving procedures names using the same case.	-fcase-lower (only in GCC 4.3+ ?)

continued on next page

continued from previous page

Absoft Flag	Description	gfortran Flag
-YEXT_SFX=""	Prevent the compiler from adding any characters, such as an underscore (_), to the end of procedure names (see Section **??** for more details)	-fno-underscoring

Consult your compiler documentation for additional details about compiler flags.

3.3.2 Testing the Program

You may test your build of BESTPRED using the default parameter file and example lactation data provided in the file `DCRexample.txt`. A file of results calculated by AIPL is provided in the file `DCRexample.results.dcr`; they should be compared to the results stored in `results_v2.dcr`, which will be created when you run the bestpred program. Small differences between platforms are to be expected, but large differences may indicate a problem. The input and output files are described in more detail in Section 4.8.

3.4 User Support, Feature Requests, and Bug Reports

Requests for support should be directed to John Cole (john.cole@ars.usda.gov) or Paul VanRaden (paul.vanraden@ars.usda.gov) (301-504-8334). Suggestions for improvements to the software are welcome, but no promise of implementation is made. Any user-suggested features added to BESTPRED will be made available to all users of the programs; proprietary intellectual property will not be accepted for inclusion into BESTPRED.

If you believe that you've discovered a bug please forward a written description of the behavior that's believed to be incorrect, input files that demonstrate the problem, and a thorough description of your computing environment (e.g. operating system and version, compilers and versions, etc.). You should also enable logging and attach a copy of the resulting log file (see Section 5.4).

This page intentionally left blank.

CHAPTER

FOUR

Using BESTPRED

BESTPRED may be used as a standalone program or called as a subroutine from another program. When run as a standalone program BESTPRED requires that input data be provided in AIPL's Format 4 (see Table 4.1). The `bestpred_fmt4` and `bestpred` subroutines provide more flexibility in that records may be stored in whatever format is desired as long as the input arrays and vectors are filled correctly. Interfacing Fortran 90 subroutines with other languages (e.g. C, SQL stored procedures) is beyond the scope of the manual.

4.1 Standalone Operation

BESTPRED is run in standalone mode by running the program bestpred (bestpred.exe on MS Windows) from a command line. When the program is run the parameter file is read, data file read, calculations performed, and output written. Multiple animal records can be processed from the same file. During development of BESTPRED a file of 150,000 records was routinely processed with no difficulty. When running in standalone mode data are read from an input file based on the value of the source parameter (Table 4.1). Records from sources other than 10 (AIPL Format 4) are converted into Format 4 records and passed downstream for processing.

Table 4.1: Description of input data sources

Source	Filename	Contents
10	format4.dat	See: Format 4 – Cow/Doe Lactation Data utilizing multiple test day segments and Table 4.2
11	DCRexample.txt	Testing plans coded by user
12	input.dcr	USDA master file
14	pcdart.bpi	See: INfile and OUTfile
15	format4.dat and format4.means	format4.means contains 6 space-delimited columns per record: cow ID (17 bytes), calving date (8 bytes), and herd average 305-d ME milk (5 bytes), fat and protein (4 bytes each), and SCS (3 bytes).
24	pcdart_files.txt	Reads a list of Source 14 input files from a separate file; the files are processed sequentially.

When using Source 10, only some of the fields from the complete Format 4 record are required (Table 4.2). Empty fields must be blank-filled so that column assignments correspond to the format. The required fields are 17-byte cow ID, herd code, birth date, fresh (calving) date, parity, lactation length, and the number of test day segments. For each test day segment the following fields must be included: test day DIM, number of milkings weighed, number of milkings sampled, the DHI supervision code, milk-recorded days, milk yield, fat and protein percentages, and SCS. Previous days open is optional.

Table 4.2: Fields Required for a Minimal Format 4 Record

Byte Position(s)	Num Bytes	Field Format[a]	Data Type[b]	Field Description
3 – 19	17	AA..AA	CH	17-byte animal ID (2-byte breed code, 3-byte country code, 12-byte ID number)
71–78	8	XX..XX	CH	Birth date of animal (YYYYMMDD)
107–114	8	XX..XX	CH	Herd code (2-byte state code, 2-byte county code, and 4-byte herd number)
128–135	8	XX..XX	CH	Calving date (YYYYMMDD)
136–138	3	XXX	CH	Days in milk this lactation
159–160	2	XX	CH	Lactation number for this record
246–248	3	XXX	CH	Days open in in previous lactation (OPTIONAL)
249–250	2	XX	CH	Number of segments for test days (up to 20 per Format 4 record)
251–253	3	XXX	CH	Days in milk this test day
254	1	X	CH	Supervision code
256	1	X	CH	Milking frequency (Last day of test period)
257	1	X	CH	Number of milkings at which milk was weighed (Last day of test period)
258	1	X	CH	Number of milkings at which samples were taken for comp. analysis (Last day of test period)
259–260	2	XX	CH	Number of milk-recorded days (MRD)
264–267	4	XXX.X	CH	Actual milk yield (avg. if MRD ¿ 1)
268–269	2	X.X	CH	Actual fat percentage
270–271	2	X.X	CH	Actual protein percentage
272–273	2	X.X	CH	Actual SCS

[a] 0 = Zero filler; A = Alphanumeric data possible; P = Packed decimal; X = Numeric data only (use left zero fill)
[b] CH = Character; CSL = Signed number with sign in leading (first) separate position (zero filled)

Up to 20 test day segments may be provided on a Format 4 record. Each segment is 23-bytes long: the first segment begins at column 251, the second at 274 bytes, the third at 297 bytes, etc. Records may end with the final segment in a lactation, and do not need to be padded to 710 columns.

When a complete Format 4 record is used, herd averages are calculated by subtracting the appropriate yield deviation (columns 201–216) from the standardized lactation yield (columns 188 – 200). If a lactation average is not provided, as in a minimal Format 4, a breed average value, specified in the `bestpred.f90 file`, is used.

If a valid state code is not provided in bytes 107–108 of a minimal Format 4 record then a value of 35 (corresponding to Wisconsin) will be used. This is necessary to ensure that the appropriate age adjustment factors are used.

Source 15 is identical to Source 10 except that herd average 305-d ME yields for milk, fat, protein, and SCS are read from the file `format4.means`, which should contain a record corresponding to each lactation in `format4.dat`. Both files should be sorted in the same order. The cow IDs and calving dates from the two files are checked against one another, and BESTPRED will halt if there is a mismatch.

If you want to account for days open in the previous lactation (0 to 999 d) write the value into columns 246 through 248 of your Format 4 file. The value will be passed downstream to the `bestpred_fmt4` and `bestpred` subroutines. When a value of 0 is encountered no adjustement is made.

4.2 Calling as a Subroutine

There are two ways to call BESTPRED as a subroutine: though the `bestpred_fmt4` subroutine, which prepares Format 4 data by applying age adjustments etc., and by calling the `bestpred` subroutine directly. Regardless of which routine you call you *must* either call `read_parms` in the outer routine (in order to insure that the parameter file is read) or explicitly provide values for subroutine arguments that are usually read from the parameter file. The former is preferred to the latter.

If you are calling BESTPRED from a C-language program note that C and Fortran treat multiple-dimensional arrays differently: Fortran 90 varies the leftmost index the fastest, while C varies the rightmost index the fastest (column-major versus row-major ordering). An array that is 2-by-4 in Fortran will be 4-by-2 in C, that is, the C array will be the transpose of the Fortran array.

4.2.1 Inputs

The inputs required by the `bestpred_fmt4` and `bestpred` subroutines are described in Table 4.3. Of greatest interest to users are the variables with the mode "OUT", indicating that they are values calculated and returned by the subroutine.

Table 4.3: Arguments of the `bestpred` subroutine

Parameter	Mode	Type	Dimensions[7]	Description
agefac	IN	real*8	4×1	Age adjustment factors for M, F, P, and SCS
BLUPn	OUT	integer	10×1	AIPL use only
BLUPout	OUT	character	10×1	AIPL use only
bump	OUT	real*8	4×1	Vector of reliabilities of M, F, P, and SCS adjusted for data quality (bumpiness)
cowid	IN	character	17	17-byte American ID of cow
DAILYbp	OUT	real*8	$2 \times 4 \times$ *maxlen*	Array containing cow lactation curves (BP of daily yields) for M, F, P, and SCS
DAILYherd	OUT	real*8	$4 \times$ *maxlen*	Array containing reference (herd) curves for M, F, P, and SCS
DCRc	OUT	real*8	1	Data collection rating for components (F+P) yield
DCRm	OUT	real*8	1	Data collection rating for milk yield
DCRs	OUT	real*8	1	Data collection rating for SCS
dim	IN	integer	*maxtd* $\times 1$	Vector containing days in milk at each test day
fresh	IN	character	8	8-byte fresh date (YYYYMMDD)

continued on next page

continued from previous page

Parameter	Mode	Type	Dimensions[8]	Description
herdavg	IN	real*8	2×4	Vector containing 305-d herd means for M, F, P, and SCS adjusted for age, but not 3X
length	IN	integer	1	Days from fresh date to termination date
MRD	IN	integer	*maxtd* $\times$ 1	Number of milk recorded days averaged together (LER)
parity	IN	integer	1	Lactation number (assumed mature if parity = 0)
PERSvec	OUT	real*8	4×1	Vector of persistencies for M, F, P, and SCS on a 305-d basis
READparms	IN	integer	1	Flag indicating that parameters should be those passed in from bestpred_main (0) or those stored in the file bestpred.par (1)
RELpers	OUT	real*8	4×1	Vector of reliabilities of persistency for M, F, P, and SCS on a 305-d basis
RELyld	OUT	real*8	4×1	Vector of reliabilities for M, F, P, and SCS on a 305-d basis
sample	IN	integer	*maxtd* $\times$ 1	Vector of sampling plans for each TD
super	IN	integer	*maxtd* $\times$ 1	Vector of supervision codes for each TD
tstdays	IN	integer	1	Number of test days
weigh	IN	integer	*maxtd* $\times$ 1	Vector of number of weights recorded per TD
Xmilk	IN	integer	*maxtd* $\times$ 1	Vector of milking frequency for each TD
yield	IN	real*8	$2 \times 4 \times$ *maxtd*	Contains M, F, P, and SCS data for each test day
YLDvec	OUT	real*8	2×4	Array of yields: 305-d MFPS, 365-d MFPS, maxlen-d MFPS, and laclen-d MFPS
Yvec	OUT	real*8	2×4	Expanded yields (Section 6.1) for M, F, P, and SCS

Arguments of `bestpred` not listed in Table 4.3 are user-controlled parameters discussed in Parameters (Chapter 5).

4.2.2 Output

The same output is written to the screen an/or files regardless of how BESTPRED is run. If ONscreen is 0 then no output (other than debugging messages, when *DEBUGmsgs* is 1) is written to the console; the number of graphs output is controlled by the *maxprnt* parameter.

When BESTPRED is run as a standalone program and the *source* parameter is set to 11 the file DCRexample.dat is read and cow records are created using the testing plans coded in that file. The following sub-sections use an example of a 255-d record in progress from that file to document the output printed to the screen and written to *CURVEfile* and *DATAfile*.

4.3 File Output

If an output file does not exist when BESTPRED is run the file is created and data written to it. Because output filenames include the current date successive runs of BESTPRED on the same day will append data to existing files rather than overwriting them.

When the WRITEcurve parameter is set to 1 data describing the lactation curve for each cow and trait are written to a file whose name is formed by appending the calculation method (ST or MT) and date (YYYYMMDD) to CURVEfile (e.g. cowcurve.ST.20080807). Each line in the output file contains a cow ID, lactation number, trait number, DIM, and TD yield, as well as blocks for mature equivalent and actual yields; each block includes the BP of daily yield and the

standard curve value corresponding to that day of lactation (Table 4.4). Up to 4 × *maxlen* lines will be written to the file for each cow processed. The cow's individual lactation curve is obtained by summing the daily BP and standard curve columns.

Table 4.4: Contents of the lactation curve output file

Value	Column(s)	Notes
Cow ID	1–17	
Lactation number	18–22	
Trait number	23–27	1: milk, 2: fat, 3: protein, 4: SCS
Days in milk	28–32	1–*maxlen*
Test-day yield	33–52	-999.0 if not a test-day
Best prediction of actual yield	53–72	Add the best prediction and standard curve yields to get the lactation curve
Standard curve actual yield	73–92	
Best prediction of ME yield	93–112	Add the best prediction and standard curve yields to get the lactation curve
Standard curve ME yield	113–132	

The first 10 lines of a example lactation curve data file are presented below. The data were edited to fit on the page and the resulting column boundaries do not match Table 4.4; the table is correct.

```
[jcole@aipl366 bestpred\]$ cat cowcurve.ST.20080807
HOUSA.EX.COW. 5 1  1 -999.000 8.061 40.162 8.053 40.122
HOUSA.EX.COW. 5 1  2 -999.000 8.241 46.466 8.233 46.420
HOUSA.EX.COW. 5 1  3 -999.000 8.357 50.528 8.349 50.478
HOUSA.EX.COW. 5 1  4 -999.000 8.447 53.568 8.438 53.515
HOUSA.EX.COW. 5 1  5 -999.000 8.522 56.006 8.513 55.951
HOUSA.EX.COW. 5 1  6 -999.000 8.588 58.043 8.579 57.986
HOUSA.EX.COW. 5 1  7 -999.000 8.647 59.790 8.639 59.731
HOUSA.EX.COW. 5 1  8 -999.000 8.702 61.316 8.694 61.256
HOUSA.EX.COW. 5 1  9 -999.000 8.753 62.668 8.745 62.606
HOUSA.EX.COW. 5 1 10 -999.000 8.802 63.879 8.793 63.816
```

If *CURVEsmall* is 1 then the a compact version of the file is written which contains only actual yields – the best prediction and standard curve values for ME are not written to *CURVEfile*.

Similarly, when the WRITEdata parameter is set to 1 data describing the lactation yields for each cow and trait are written to a file whose name is formed by appending the calculation method (ST or MT) and date (YYYYMMDD) to DATAfile (e.g. cowdata.ST.20080807). Each line in the output file contains a cow ID, lactation number, trait name, number of tests, blocks for mature equivalent and actual yields, DCR, method of calculation (ST or MT), persistency of yield on a 305-d basis, and reliability of persistency (Table 4.5). Each yield block includes five columns: 305-d yield, 365-d yield, *laclen*-d yield, lactation-to-date yield, and contemporary average yield. Note that the records for SCS are multiplied by 100.

Table 4.5: Contents of the cow data output file

Value	Column(s)	Notes
continued on next page		

continued from previous page

Value	Column(s)	Notes
Cow ID	1–17	
Lactation number	18–20	
Trait name	21–28	
Number of tests	29–32	
305-d ME yield	33–39	
365-d ME yield	40–46	
laclen-d ME yield	47–53	
Lactation-to-date ME yield	54–60	
Contemporary average ME yield	61–67	
305-d actual yield	68–74	
365-d actual yield	75–81	
laclen-d actual yield	82–88	
Lactation-to-date actual yield	89–95	
Contemporary average actual yield	96–102	
Data collection rating (DCR)	103–107	
Method of calculation	108–110	MT: multiple trait, ST: sigle trait
Persistency of yield	111–116	
Reliability of persistency	117–120	

An example cow data file is presented below. The data were edited to fit on the page and the resulting column boundaries do not match Table 4.5; the table is correct.

```
[jcole@aipl366 bestpred]$ cat cowdata.ST.20080807
HOUSA.EX.COW. 5 Milk 9 21753. 24744. 21753. 18824. 20000. 21731. 24720. 21731. 18805. 19980. 10
HOUSA.EX.COW. 5 Fat  9   778.   893.   778.   669.   700.   774.   889.   774.   666.   697. 10
HOUSA.EX.COW. 5 Prot 9   692.   796.   692.   595.   600.   687.   790.   687.   590.   595. 10
HOUSA.EX.COW. 5 SCS  9   221.   230.   221.   215.   308.   268.   278.   268.   260.   373. 10
```

4.4 Screen Output

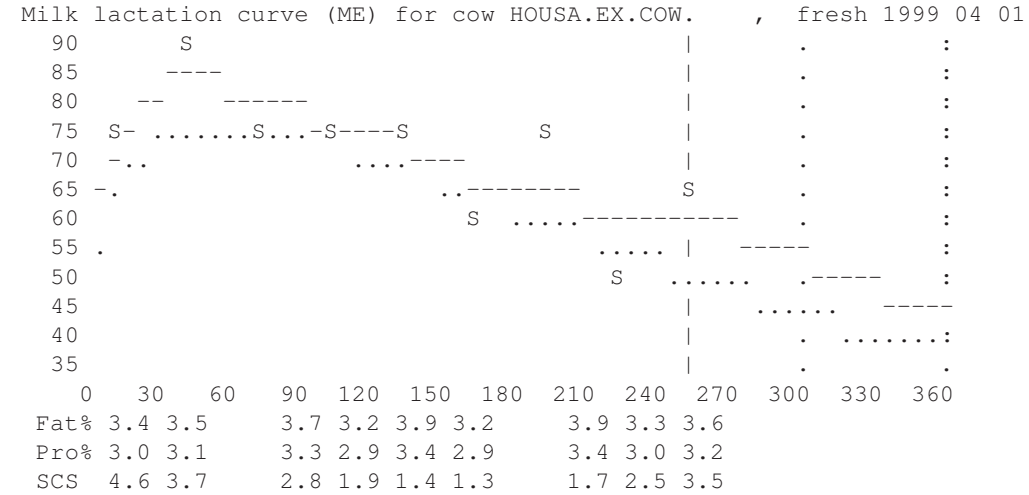
Output may be written to STDOUT, which is typically the console from which the program is run. If you are running the standalone executable `bestpred` then you'll see a banner that includes the version number of BESTPRED and contact information in the unlikely event that you need support.

```
[jcole@aipl366 bestpred]$ ./bestpred
--------------------------------------------------------------------------------
BESTPRED version 2.0b10
    Released: 08/08/2008
    Support : John B. Cole (john.cole@ars.usda.gov)
    Website : http://www.aipl.arusda.gov/software/bestpred/
--------------------------------------------------------------------------------
```

Following the banner is a header that includes the source of the data being processed, information about any requested plots, and single- versus multiple-trait methodology. It also prints breed- and parity-specific mature equivalent mean and SD of milk, fat, protein, and SCS, as well as the tipping points used for calculating persistency.

```
Reading example data from DCRexample.txt
Lactation yields are MATURE (adjusted for age, etc.) and ACTUAL (unadjusted) values
Plotted lactation curves are a mix of MATURE (adjusted for age, etc.) and ACTUAL (unadjusted) v
Multi-trait methods include data for   Milk    Fat   Prot
New 3X factors phased in over time
DCRs were not adjusted for bumpiness
=====================================================================
Breed H Lact  Mean, St.Dev.   Max.DCR   Lactation Corr.        mid-DIM
ME.Milk   2+  25658.   2946.    104.4   1.00  0.67  0.85 -0.08    161.
ME..Fat   2+    935.    119.    104.6   0.67  1.00  0.77 -0.14    159.
ME.Prot   2+    771.     97.    104.4   0.85  0.77  1.00 -0.10    166.
ME..SCS   2+    679.    295.    104.2  -0.08 -0.14 -0.10  1.00    155.
ME.Milk   1   22270.   2287.    104.0   1.00  0.67  0.85 -0.08    115.
ME..Fat   1     802.     90.    104.2   0.67  1.00  0.77 -0.14    115.
ME.Prot   1     652.     74.    104.0   0.85  0.77  1.00 -0.10    150.
ME..SCS   1     563.    267.    106.0  -0.08 -0.14 -0.10  1.00    155.
 =====================================================================
 Compute DCR, predict yield of cow ---, graph contemps ...
 Supervised =S    ampm1of2=A  2of3=B  1of3=C ver=V  LER=L dup=D
 OwnerSample=O  OSampm1of2=P  2of3=Q  1of3=R  ownerLER=U, bad=X
 Capital letters if milk was sampled, small if milk-only
```

If one or more elements of the GRAFplot array are set to a value of 1 or 2 then corresponding maxlen-d graphs will be printed to the screenm. You can request a mixture of ME and actual plots in the same run, but you can request only one plot per trait. The x- and y-axes correspond to DIM and yield, respectively. Units for M, F, and P are determined by UNITSout. The cow's lactation curve is represented by dashes (-), the standard (herd) curve by periods (.), and test days by lower- and upper-case letters (depending on the testing plan). Vertical lines of periods (.), vertical bars (—), and colons (:) represent lactation length to date, 305 DIM, and maxlen DIM, respectively. Fat and protein percent and SCS are printed immediately under the graph below their corresponding test days. Plots are of ME or actual values, depending on the values in GRAFplot, and the header will indicate if the plot contains ME or actual values. For example, the follwoing plots include ME milk and actual fat percent.



```
Fat  lactation curve (Actual) for cow HOUSA.EX.COW.    ,  fresh 1999 04 01
3.2       S                                    |       .          :
3.1                                            |       .          :
3.0       -                                    |       .          :
2.9     -- ---           S         S           |       .          :
2.8   --      ---                              |       .          :
2.7  -          S ---     -                    |       .          :
2.6 -S..........    ----- --                   |       .          :
2.5  .           ....S       -      -          |       .          :
2.4 .                ....    ------ --        S       .          :
2.3                      .....         -       |       .          :
2.2                           ....     -   -----      .          :
2.1                               ....  ---  |  ----  .          :
2.0                           S       .....  |      ----         :
1.9                                      .....     .  ----    :
1.8                                        |  .... .      ----
1.7                                     S  |      .....        :
1.6                                        |       .   ......:
1.5                                        |       .          .
    0   30   60   90  120  150  180  210  240  270  300  330  360
 Fat% 3.4 3.5     3.7 3.2 3.9 3.2     3.9 3.3 3.6
 Pro% 3.0 3.1     3.3 2.9 3.4 2.9     3.4 3.0 3.2
 SCS  5.6 4.5     3.4 2.3 1.7 1.6     2.1 3.0 4.2
```

The number of tests, 305-d yield, 365-d yield, laclen-d yield, lactation-to-date yield, contemporary average yield, DCR, method of calculation (ST or MT), persistency of yield on a 305-d basis, and reliability of persistency is also output for each trait. Yields for M, F, and P are presented in UNITSout units, pounds in this case. Somatic cell scores are multiplied by 100. Mature equivalent values are presented first, followed by actual values. Note that a cow's lactation curve and the contemporary curve may be coincident in a plot when *dailyfreq* is set to a large value; examination of the actual data shows that they are not identical lines, but overlap due to the low resolution of the plot. See Section 4.6 for details on plotting high-resolution curves.

The *plotfreq* variable specifies the interval between days plotted. The default of *6* (a point is plotted every 6 days) provides reasonable output on an 80-column display. Setting *plotfreq* to a smaller value may result in plots that wrap and are unreadable on a terminal.

```
                                                Cntmp   Adjust   Persistency   Data
             255-d  305-d  365-d  305-d  305-d   Factors  Esti- Relia-  Collectn
TRAIT Tests Yield  Yield  Yield  Yield  Mean   Age,  3X  mate  bility  Rating
ME.Milk   9 18824. 21753. 24744. 21753. 20000. 1.00 1.00 -0.06  85.%   102. MT
ME..Fat   9   669.   778.   893.   778.   700. 1.00 1.00  0.04  84.%   102. MT
ME.Prot   9   595.   692.   796.   692.   600. 1.01 1.00 -0.31  84.%   102. MT
ME..SCS   9   215.   221.   230.   221.   308. 0.83 1.00 -0.58  72.%   101. ST
   Milk   9 18805. 21731. 24720. 21731. 19980. 1.00 1.00 -0.06  85.%   102. MT
    Fat   9   666.   774.   889.   774.   697. 1.00 1.00  0.04  84.%   102. MT
   Prot   9   590.   687.   790.   687.   595. 1.01 1.00 -0.31  84.%   102. MT
    SCS   9   260.   268.   278.   268.   373. 0.83 1.00 -0.58  72.%   101. ST
```

4.5 Actual and Mature Equivalent Yields

BESTPRED outputs both actual and mature equivalent yields as of release 2.0b9. This is a notable change, as AIPLDCR and earlier versions of BESTPRED required two calls to the subroutine in order to obtain both sets of

values. This should results in a substantial gain in performance when processing large numbers of records.

Actual and mature equivalent yields calculated by users may differ from those calculated by AIPL even when identical parameter files are used. The edits system used by AIPL is very complex, and records processed using that system have been checked and many errors corrected. It is not possible to recreate that system in BESTPRED, which assumes that user input is correct. For example, there is a limit placed on the amount by which a test day observation can differ from expectations based on preceding and succeeding observations. Such edits must be performed before the test day data are processed by BESTPRED, regardless of how you are using the program (as a standalone program or as a subroutine call in another program).

Herd average yields also affect the best prediction calculations. When Format 4 records are processed using the standalone BESTPRED program the herd averages are back-calculated from the 305-d mature equivalent yields and the cow's deviations, and are set to 0 otherwise.

4.6 Drawing Lactation Curves

A Python-language program, bestpred_curves.py, is included in the BESTPRED archive. If you have Python (http://python.org/) and the appropriate extensions (matplotlib, numpy) installed you can use bestpred_curves.py to draw lactation curves that are written to files in Portable Network Graphics format (PNG; http://en.wikipedia.org/wiki/PNG). To run bestpred_curves.py you must provide a trait (1 = milk, 2 = fat, 3 = protein, 4 = SCS, 5 = all). and you may provide an input file name. If you do not provide a file name then all files matching the pattern "cowcurve.*" are processed:

```
[jcole@aipl366 bestpred]$ python bestpred_curves.py
[ERROR]: You must provide a trait: 1 = milk, 2 = fat, 3 = protein, 4 = SCS, 5 = all
[jcole@aipl366 bestpred]$ python bestpred_curves.py 1
Processing cowcurve.ST.20070921
Writing file curve_AYUSA000000960899_Milk_1.png
Writing file curve_AYUSA000000960899_Milk_2.png
[jcole@aipl366 bestpred]$
```

One file is written per cow-trait combination in the input file. File names are of the form "curve_¡cowid¿_¡trait¿_¡parity¿.png". An example figure is provided in Figure 4.1. The blue line in Figure 4.1 is the standard curve, the green line is the daily best prediction curve, and the green points are test day yields. The values in the top right-hand corner are the 305-d actual (green) and standard curve (blue) yields, respectively. All values in the figure are actual daily values, not mature equivalent values.

4.7 Logging

A simple logging module has been added to the software in order to assist BESTPRED users who encounter problems. This section of the manual describes the use of the logging module and the contents of the logfile; parameters specific to this module are described in Section 5.4.

There are two steps in logging an event: you first create a string that contains the message you want to write to the logfile, and you then call the LogMessage subroutine. (LogMessage will automatically prepend the correct timestamp; it's smart like that.) You will no doubt be tempted to skip creating the string containing the message, supposing that you will just pass it as a string literal in the subroutine call. If you do that each log entry will be padded out to the 80 characters with nulls, and your logfile will be a nightmare to read. Trust me. And don't assume that little feature is documented in most Fortran 90 books, either. Here's a very simple code snippet:

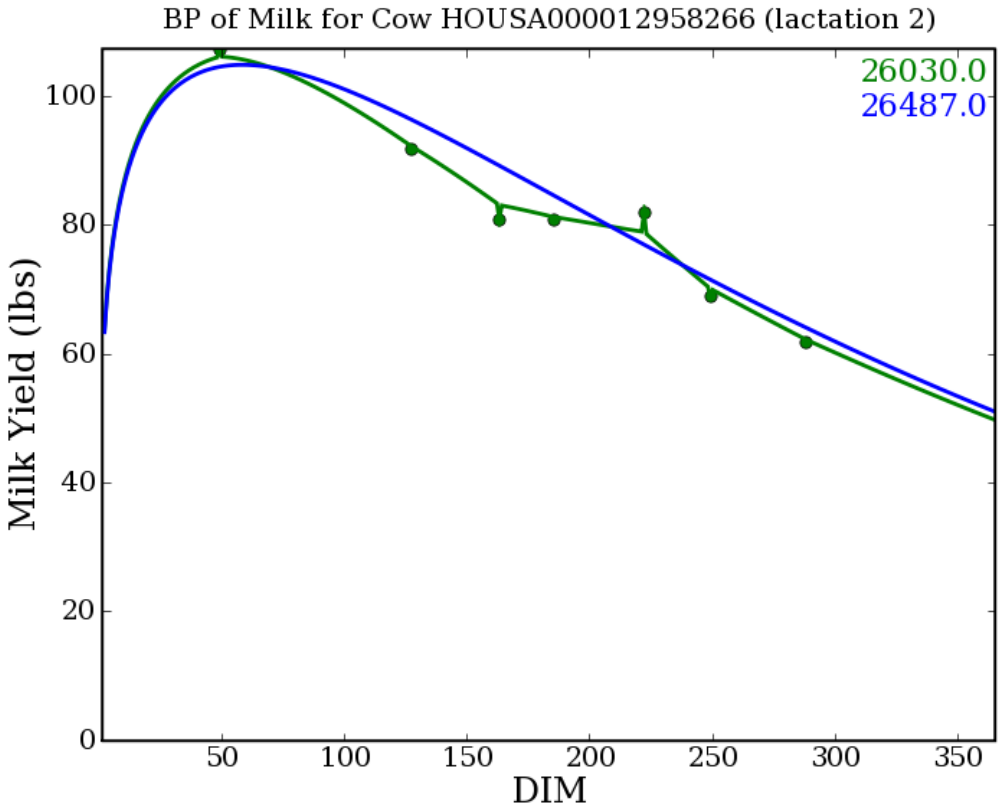


Figure 4.1: Lactation curve for milk yield (pounds) of a second parity Holstein cow

```
LogMessage = 'The quick brown fox jumped over the lazy dogs.'
call log_message(LOGfile,LogMessage)
```

LOGfile is the base name for the logfile, and is limited to 64 characters. The date and the string `.log` are appended to *LOGfile* to create the file name (e.g. `logfile.20080807.log`).

Of course, it's not for nothing that we provide the *LOGon* parameter, which toggles logging on and off, so perhaps we should respect the user's wishes and check it:

```
if ( LOGon == 1 ) then
  LogMessage = 'The quick brown fox jumped over the lazy dogs.'
  call log_message(LOGfile,LogMessage)
end if
```

You may want to add your own debugging code to BESTPRED. If you do so, you may want to check the LOGfreq parameter while inside of the bestpred() subroutine. It provides you finer-grained control of logging than on/off by specifying the frequency with which messages should be logged.

```
if ( LOGon == 1  .and. LOGfreq > 0 ) then
  LogMessage = 'This message is printed every time bestpred() is entered.'
  call log_message(LOGfile,LogMessage)
end if
```

You can get clever with the MOD() function and print only messages from specific iterations through the subroutine. For example, the following message is printed only the first time the subroutine is entered:

```
if ( LOGon == 1 .and. LOGfreq > 0 .and. mod(ncall,LOGfreq) == 0 ) then
  LogMessage = 'Allocating arrays in call '//trim(char(ncall+ICHAR('0')))
  call log_message(LOGfile,LogMessage)
end if
```

This message is logged for every *LOGfreq*-th record processed:

```
if ( LOGon == 1 .and. LOGfreq > 0 .and. mod(ncall,LOGfreq) == 0 ) then
  LogMessage = 'Allocating arrays in call '//trim(char(ncall+ICHAR('0')))
  call log_message(LOGfile,LogMessage)
end if
```

The following logfile is from an actual run of BESTPRED with *LOGon* = 1 and *LOGfreq* = 1. The contents are pretty self-explanatory.

```
[jcole@aipl3850 bestpred_onepass]$ cat example.20080807.log
# example.20080807.log created on 08/07/2008 at 10:31:41
[10:31:41]: Starting BESTPRED 2.0b9
[10:31:41]: Allocating matrices for means, SD, and covariances
[10:31:41]: Allocating arrays in call 0
[10:31:41]: Converting from lbs to kg for interal calculations
[10:31:41]: Calling interpolate with method W and breed 4
[10:31:41]: Calling interpolate with method W and breed 4
[10:31:41]: Calling interpolate with method W and breed 4
[10:31:41]: Calling interpolate with method W and breed 4
[10:31:43]: Calling interpolate with method W and breed 4
[10:31:43]: Calling interpolate with method W and breed 4
[10:31:43]: Calling interpolate with method W and breed 4
[10:31:43]: Calling interpolate with method W and breed 4
```

The first line of every logfile includes the filename and its date and time of creation. Each log entry is prepended with a timestamp. Messages longer than 80 characters will be truncated to 80 characters, not inclusive of the timestamp.

4.8 Simulating Lactation Data

BESTPRED can simulate example Holstein lactation data for testing plans described in the file `DCRexample.txt`. Each entry corresponds to separate record, and each record may have different supervision codes, milking frequency, weighing frequency, sampling frequency, number of days included in labor-efficient records, testing interval (in days), time of first sample (in days), time of last sample (in days), and lactation number. Each plan may also be given a name. Testing plans that change over time, such as every-other-month components tests, are also accommodated. The first three (3) records in `DCRexample.txt` are shown below (lines beginning with underscores are treated as comments and ignored):

```
_ super- times  times times  LER  testplan first last par-
_ vised  milked weigh sample days interval test  test ity  plan-name
_____________________________________________________________________

    1      2      2      2      1      30     15  255   5  255-day RIP

    1      3      2      2      7      60     30  300   4  LER sampled
    1      3      2      0      7      60     60  300   4  odd months

    1      2      2      2      1      30     15  285   1  Supervised
```

Herd average ME milk, fat, and protein yields are hard-coded in `bestpred_main.f90` as 20,000 pounds, 700 pounds, and 600 pounds, respectively.

Results are written to the file `results_v2.dcr`, and may be compared with the AIPL-calculated figures in `DCRexample.results.dcr` for validation that BESTPRED is working as expected. The format of `results_v2.dcr` is presented in Table 4.6, and the example values are taken from the first record in `DCRexample.results.dcr` distributred with BESTPRED.

Table 4.6: Contents of `DCRexample.results.dcr` and `results_v2.dcr`

Variable	Column(s)	Example Value	Note(s)
Animal	1–17	HOUSA.EX.COW.	Animal ID
Fresh Date	19–26	19990401	Fresh date for this lactation
DIM	27–30	255	Days in milk in current lactation
DCRm	32–36	102.	DCR for milk yield
DCRc	37–41	102.	DCR for components (mean of DCR for fat and protein)
DCRs	42–46	101.	DCR for SCS
305-d ME yield	47–53	21753.	305-d ME milk yield
	54–60	778.	305-d ME fat yield
	61–67	692.	305-d ME protein yield
	68–74	2.21	305-d ME SCS yield
365-d ME yield	75–81	24744.	365-d ME milk yield
	82–88	893.	365-d ME fat yield
	89–95	796.	365-d ME protein yield
	96–102	2.30	365-d ME SCS
laclen-d ME yield	103–109	21753.	*laclen*-d ME milk yield
	110–116	778.	*laclen*-d ME fat yield
	117–123	692.	*laclen*-d ME protein yield
	124–130	2.21	*laclen*-d ME SCS
255-d ME yield	131–137	18824.	*lactation-to-date*-d ME milk yield
	138–144	669.	*lactation-to-date*-d ME fat yield
	145–151	595.	*lactation-to-date*-d ME protein yield
	152–158	2.15	*lactation-to-date*-d ME SCS
Persistency	159–164	-0.06	Persistency of 305-d milk yield
	165–170	0.04	Persistency of 305-d fat yield
	171–176	-0.31	Persistency of 305-d protein yield
	177–182	-0.58	Persistency of 305-d SCS
REL yld	183–187	0.97	Reliability of 305-d milk yield
	188–192	0.97	Reliability of 305-d fat yield
	193–197	0.97	Reliability of 305-d protein yield
	198–202	0.97	Reliability of 305-d SCS
REL pers	203–207	0.85	Reliability of persistency of 305-d milk yield
	208-212	0.84	Reliability of persistency of 305-d fat yield
	213–217	0.84	Reliability of persistency of 305-d protein yield

continued on next page

continued from previous page

Variable	Column(s)	Example Value	Note(s)
	218-222	0.72	Reliability of persistency of 305-d SCS
Yvec (expanded yield)	223-229	21800.	Expanded 305-d ME milk yield
	230–236	780.	Expanded 305-d ME fat yield
	237–243	695.	Expanded 305-d ME protein yield
	244–250	0.00	Expanded 305-d ME SCS
herd305	251–257	20000.	Herd average 305-d ME milk yield
	258–264	700.	Herd average 305-d ME fat yield
	265–271	600.	Herd average 305-d ME protein yield
	272–278	3.08	Herd average 305-d ME SCS
Bumpiness	280–286	0.0000	A measure of “smoothness“ of test day data
	287–293	0.0000	
	294–300	0.0000	
	303–309	0.28	

Results from the first three (3) records defined in `DCRexample.txt` are shown below (note that the data heve been reformatted slightly for typesetting purposes).

```
HOUSA.EX.COW. 19990401 255  102. 102. 101. 21753.   778.   692.   2.21 24744.   893.   796.
   2.30 21753.   778.   692.   2.21 18824.   669.   595.   2.15 -0.06  0.04 -0.31 -0.58 0.97
 0.97 0.97 0.97 0.85 0.84 0.84 0.72 21800.   780.   695.   0.00 20000.   700.   600.   3.08
  0.0000 0.0000 0.0000     0.28
HOUSA.EX.COW. 19990401 300  104.  97.  94. 18889.   686.   605.   2.55 21386.   790.   697.
   2.64 18889.   686.   605.   2.55 18662.   676.   597.   2.54 -0.14  0.20 -0.07 -0.20 1.00
 0.93 0.93 0.90 0.98 0.75 0.81 0.52 18894.   690.   610.   0.00 20000.   700.   600.   3.08
  0.0000 0.0000 0.0000     0.73
HOUSA.EX.COW. 19990401 285  101. 101. 103. 28599.  1019.   897.   3.33 32656.  1172.  1033.
   3.48 28599.  1019.   897.   3.33 27151.   966.   849.   3.26  0.38  0.44 -2.01 -0.24 0.97
 0.97 0.97 0.97 0.90 0.89 0.79 0.79 28885.  1029.   906.   0.00 20000.   700.   600.   3.20
  0.0000 0.0000 0.0000     0.56
```

While these results are presented in standard units (pounds) BESTPRED will return values in Metric units (kg) if so requested by the user (see Section 5.1.27).

CHAPTER

FIVE

Parameters

BESTPRED is controlled by a number of parameters that are passed to the subroutine each time it is called. When BESTPRED is called as a subroutine the parameters must be provided by the calling program. When BESTPRED is run as a standalone program the parameters are read from the file bestpred.par. Care must be taken when when making changes to this file; if the file is damaged, BESTPRED will not function correctly. Default values for each parameter are set in `a typewriter-like typeface`.

5.1 Description of Parameters

Each parameter is preceded by a description of what that parameter does, as well as a list of allowed values. As a general rule, the program DOES NOT check your inputs for correctness. Be careful. "BESTPRED" is the NAMELIST that contains all of the program parameters defined in this file; do not change it.

5.1.1 breed11

breed11 specifies the breed to be used for running the testing plans coded in DCRexample.dat. It is used only when source == 11. Breeds are as follows: 1=AY, 2=BS, 3=GU, 4=HO, 5=JE, 6=MS. The default is 4 (Holstein).

```
breed11 = 4
```

5.1.2 breedUNK

breedUNK specifies the breed to be used when an unknown breed code is found in a cow ID, which may be of particular use to anyone analyzing data from crossbred cows. Breeds are as follows: 1=AY, 2=BS, 3=GU, 4=HO, 5=JE, 6=MS. The default is 4 (Holstein).

```
breedUNK = 4
```

5.1.3 CURVEsingle

CURVEsingle controls whether or not the lactation summary and lactation curve data controlled by WRITEcurve and WRITEdata are written to separate files based on cow ID (1) or to one file per run (0).

```
CURVEsmall = 0
```

5.1.4 CURVEsmall

CURVEsmall controls which curves are written to CURVEfile, actuals only, or actuals and ME. The default is 0 (show actual and ME curves).

```
CURVEsmall = 0
```

5.1.5 dailyfreq

dailyfreq specifies the frequency with which actual daily yields are calculated. The default is 6 (actual yields calculated every 6 days). A value of 0 indicates that no daily values should be calculated. A negative value will be set to the default of 6. Note that computing time increases as dailyfreq decreases towards 1, that is, a dailyfreq of 1 will require much more processing time than a dailyfreq of 6.

```
dailyfreq = 6
```

5.1.6 DEBUGmsgs

DEBUGmsgs overrides ONscreen – debugging messages will print even if screen output has been deactivated. Two levels of debugging output are provided. If DEBUGmsgs is set to 1 then a moderate amount of debugging output is provided. Values greater than 1 will produce voluminous output - you have been warned.

```
DEBUGmsgs  = 0
```

5.1.7 DEBUGparms

When DEBUGparms is set to 1 the parameters read from bestpred.par will be written to the screen.

```
DEBUGparms  = 0
```

5.1.8 dim0

dim0 contains tipping points (see: Section 6.1) for MFPS for first and later parities. MIlk, F, P, and SCS tipping points are stored in dim(1)-dim(4) for first and dim(5)-dim(8) for later parities.

```
dim0(1) = 115
dim0(2) = 115
dim0(3) = 150
dim0(4) = 155
dim0(5) = 161
dim0(6) = 152
dim0(7) = 159
dim0(8) = 148
```

5.1.9 dim0flag

dim0flag toggles estimation of tipping points for persistency on (1) and off (0).

```
dim0flag = 0
```

5.1.10 GLOBALmtrait

Individual Source 14 records contain an mtrait switch. When set to a valid non-zero value of mtrait (1, 3, or 4) GLOBALmtrait overrides the individual switches and uses the same computational method for all records read from the same file. Only used when source = 14.

```
GLOBALmtrait = 3
```

5.1.11 GRAFplot

GRAFplot contains flags indicating which lactation curves should be plotted; 1 = plot ME values, 2 = plot actual values, and 0 = don't plot. The elements of GRAFplot are 1 = Milk, 2 = Fat, 3 = Protein, and 4 = SCS. ONscreen overrides GRAFplot – if terminal output is turned off, graphs will not be drawn even if GRAFplot is turned on.

```
GRAFplot(1) = 1
GRAFplot(2) = 0
GRAFplot(3) = 0
GRAFplot(4) = 0
```

GRAFplot also controls which traits are written to the CURVEfile in the same manner – traits with a value of 1 are output and traits with values of 0 are not.

5.1.12 INfile and OUTfile

INfile and OUTfile specify the names to be used when reading and writing Source 14 input and output files. Filenames are limited to 64 characters.

```
INfile = 'pcdart.bpi'
OUTfile = 'pcdart.bpo'
```

5.1.13 INTmethod

INTmethod specifies the interpolation method used for lactation curves for MFP. Valid values are ”L” for linear interpolation and ”W” for Woods (1967) curve. The default value is ”W”.

```
INTmethod = 'W'
```

5.1.14 INTmethodSCS

INTmethodSCS specifies the interpolation method used for lactation curves for SCS. Valid values are ”L” for linear interpolation and ”G” for Morant and Gnanasakthy’s (1989) curve C4. The default value is ”G”.

```
INTmethodSCS = 'G'
```

5.1.15 laclen

laclen specifies the length of the lactation for which calculations are being made. A value of 500 would indicate that 500-d yields should be calculated in addition to LTD, 305-d, and 365-d yields. The default is 305. A negative value will be set to 305.

```
laclen = 305
```

5.1.16 maxlen

maxlen specifies the longest possible lactation length. maxlen MUST be $>=$ laclen. A negative value will be set to 365; if maxlen $<$ laclen then maxlen will be set to laclen.

```
maxlen = 365
```

5.1.17 maxprnt

maxprnt is the number of diagnostic plots to display in the output. Set maxprnt to 1 and DEBUGmsgs to 1 to display the input variables. This affects only the display of plots; lactation summary data are controlled by the *maxshow* parameter.

```
maxprnt = 5
```

5.1.18 maxshow

maxshow is the maximum number of cow records to display. This affects only the display of lactation summary data; plots are controlled by the *maxprnt* parameter.

```
maxshow = 5
```

5.1.19 maxtd

maxtd is the maximum number of test days in a lactation.

```
maxtd = 50
```

5.1.20 mtrait

mtrait toggles between single- and multiple-trait BP: 1 does only ST to save CPU time, 3 does MT MFP, and 4 does MT MFPS. Single trait best prediction is computationally the fastest. Both methods provide similar solutions, although MT prediction is preferred for calculating fat and protein yields under alternative testing schemes when components are not sampled monthly.

```
mtrait = 1
```

5.1.21 obs

obs is the maximum number of records to read.

```
obs = 99999999
```

5.1.22 ONscreen

ONscreen is used to toggle output to the screen/terminal on (1) and off (0). ONscreen overrides maxprnt – if console output is turned off only debugging messages will print to the terminal even if maxprnt is not 0.

```
ONscreen = 1
```

5.1.23 PERSfloor and PERSceiling

PERSfloor and PERSceiling provide minimum and maximum values for persistencies calculated from Source 14 records. Values smaller than the floor and larger than the ceiling are rounded up or down, as appropriate. These options affect ONLY Source 14 records. If you set a floor or ceiling whose absolute value is greater than 9.99 then you will also need to modify the format used to write the contents of OUTfile in the bestpred.f90 program.

```
PERSfloor = -9.99
PERSceiling = 9.99
```

5.1.24 plotfreq

plotfreq specifies the frequency, in days, with which yields are plotted by the inbuilt plotting routine. A value of 10 indicates that every tenth day should be plotted. If plotfreq is 0 then no plots are drawn.

```
plotfreq = 6
```

5.1.25 READparms

READ[arms indicates whether parameter values should be read from the file `bestpred.par` (1) or taken from the subourinte call (0).

```
READparms = 6
```

5.1.26 source

Settings of interest: 10 inputs test day data in format 4 as defined in the document "Listing of Format 4 – Cow/Doe Lactation Data utilizing multiple test day segments" from the file format4.dat. 11 uses testing plans coded by user from the file DCRexample.dat. 12 inputs USDA master file records from the file input.dcr. 14 inputs TD data in DRMS format from the file INfile and writes results to the file OUTfile. 24 reads a list of filenames from 'pcdart_files.txt' and then loads TD data in DRMS format from each of those files as for Source 14 (filenames cannot be longer than 64 characters).

```
source = 11
```

5.1.27 UNITSin and UNITSout

UNITSin and UNITSout specify the units in which milk, fat and protein are provided to the program and returned to the user, respectively. Pounds ('P') and kilograms ('K') are supported; pounds is the default. If a value other than 'K' or 'P' is provided the default is used.

```
UNITSin = 'P'
UNITSout = 'P'
```

5.1.28 use3X

use3X indicates whether or not records should be adjusted for milking frequency: 0 doesn't adjust for 3X milking, 1 uses old factors, 2 new factors, and 3 uses phased-in factors over time.

```
use3X = 3
```

5.1.29 WRITEcurve and CURVEfile

If WRITEcurve is set to 1 then daily BP, YD yields, and standard curves will be written to the file CURVEfile. Note that CURVEfile is limited to 64 characters. The calculation method (ST or MT) and date (YYYYMMDD) are appended to CURVEfile to create the final output file name (e.g. cowcurve.ST.20070823).

```
WRITEcurve = 1
CURVEfile = 'cowcurve'
```

GRAFplot controls which traits are written to the CURVEfile in the same manner – traits with a value of 1 are output and traits with values of 0 are not.

5.1.30 WRITEdata and DATAfile

If WRITEdata is set to 1 then daily BP, YD yields, and standard curves will be written to the file DATAfile. Note that DATAfile is limited to 64 characters. The calculation method (ST or MT) and date (YYYYMMDD) are appended DATAfile to create the file name (e.g. cowdata.ST.20070823).

```
WRITEdata = 1
DATAfile = 'cowdata'
```

5.2 Parameter File Format

The parameter file is a Fortran NAMELIST that is read from disk when the BESTPRED program is run. Lines beginning with exclamation marks (!) denote comments, which are ignored at run time. The "&bestpred" line opens the NAMELIST and the slash (/) closes it. Option lines are of the form "¡option¿ = ¡value¿". Character and string values must be enclosed in quotation marks. Unlike C and similar languages Fortran 90 does not use end-of-line markers (e.g. semicolons). The following code sample demonstrates the format of the parameter file:

```
! Comment block
&bestpred
  ! Description of option, including valid values, etc.
  myoption = 'test'
/
```

If you are getting strange error messages when you run BESTPRED check BESTPRED.par to make sure that it has valid end-of-line characters. Some editors do not set them correctly, particularly on the last line of the file.

5.3 Default Parameter Values

Default values for all parameters are provided in Table 5.1 for ease-of-reference. If you are having trouble with BESTPRED try resetting all parameters in BESTPRED.par to their default value.

Table 5.1: Default values for BESTPRED parameters

Parameter Name	Default Value
breed11	4
breedUNK	4
CURVEfile	'cowcurve'
CURVEsingle	0
CURVEsmall	0
dailyfreq	6
DATAfile	'cowdata'
DEBUGmsgs	0
DEBUGparms	0
dim0	Please see the entry above describing dim0
dim0flag	0
GLOBALmtrait	3
GRAFplot	/1, 0, 0, 0/
INfile	'pcdart.bpi'
INTmethod	'W'
INTmethodSCS	'G'
laclen	305
mature	1
maxlen	365
maxtd	50
mtrait	3
obs	99999999
ONscreen	1
OUTfile	'pcdart.bpi'
PERSceiling	9.99
PERSfloor	-9.99
plotfreq	6
READparms	0
source	11
UNITSin	'P'
UNITSout	'P'
use3X	3
WRITEcurve	1

continued on next page

continued from previous page	
Parameter Name	Default Value
WRITEdata	1

5.4 Logging Parameters

A basic logging framework is provided in BESTPRED to assist in isolating and identifying problems. Messages are written to a plain text file in the directory in which the program is run. The logfile can grow very large very quick;y if a large number of records are being processed, so logging is turned off by default. It can be ebabled by setting 'LOGon = 1' in the parameter file. Note that you should enable logging and attach your logfile when reporting errors.

```
# logfile.20080403.log created on 04/03/2008 at 20080320 by BESTPRED
[09:18:03]: Starting BESTPRED 2.0b7
[09:31:53]: Starting BESTPRED 2.0b7
[09:35:14]: Starting BESTPRED 2.0b7
```

Logging does not have any effect on calculations pedformed by BESTPRED.

5.4.1 LOGfile

If LOGon is set to 1 then log messages (messages about the operation of BESTPRED) will be written to the file LOGfile. Note that LOGfile is limited to 64 characters. The date (YYYYMMDD) is appended to LOGfile to create the file name (e.g. logfile.20080403).

```
LOGfile = 'logfile'
```

5.4.2 LOGfreq

Write a message to the logfile when every n-th animal record is processed; 0 suppresses output of any animal record-specific log messages. If LOGfreq is set to 1 a record will be written to the logfile for each lactation processed. This can consume large amounts of storage if you're processing lots of records.

```
LOGfreq = 0
```

5.4.3 LOGon

Turn logging on (1) and off (0). Logfiles can grow very large when a large number of records is being processed, so logging is disabled by default.

```
LOGon = 0
```

5.5 Default Logging Parameter Values

Default values for the release parameters are presented in Table **??**.

Table 5.2: Default values for BESTPRED logging parameters

Parameter Name	Default Value
LOGfile	'logfile'
LOGfreq	0
LOGon	0

This page intentionally left blank.

CHAPTER

SIX

Methodology

6.1 Derivation of Best Prediction

Individual daily yield can be modeled as the expected value of a management group plus a deviation from that mean:

$$y_i = E(y_i) + t_i$$

where y_i is an individual yield on test day i, $E(y_i)$ is the expected yield for an animal in the same management group (Wiggans, Misztal, and Van Vleck 1988) on the same test day, and t_i is a deviation from the group mean on the same test day. Suppose that μ is a vector of expected values for each day of lactation for a single trait, t is a vector of 305 test day deviations for the trait, and t_m is a vector of only the measured deviations. The means and variances of t and t_m are assumed known with $V(t) = V$ and $V(t_m) = V_m$. The covariance between t and t_m, C, is also assumed known. Predictions for lactations longer than 305 d can also be made (Dematawewa, Pearson, and VanRaden 2007).

Lactation yield. A cow's true 305-d yield (y) is the sum of the expected values for each day ($1'\mu$) plus the sum of her 305 deviations from expectations ($1't$), where $1'$ is a vector of 1s of length 305. The cow's true yield, and the best prediction of that yield ($\hat{y}$), are:

$$\begin{aligned} y &= 1't \\ \hat{y} &= 1'CV_m^{-1}t_m \end{aligned}$$

Reliability of $\hat{y}$ is obtained from variances of y and $\hat{y}$, which are computed from the quadratic forms:

$$\begin{aligned} Var(y) &= 1'V1 \\ Var(\hat{y}) &= 1'CV_m^{-1}C'1 \\ Rel(\hat{y}) &= Var(\hat{y})/Var(y) \\ &= 1'CV_m^{-1}C'1/1'V1 \end{aligned}$$

Procedures to adjust V_m if data are from partial days or are computer averages from several days or are recorded by an owner instead of a supervisor were outlined by VanRaden (1997).

Persistency. Persistency may be measured by multiplying test day deviations by a linear function of days in milk (Cole and VanRaden 2006; VanRaden 1998). Let d represent a vector whose elements, d_i, represent DIM associated with y_i. A measure of persistency that is phenotypically uncorrelated with lactation may be obtained by defining coefficients $q_i = d_i - d_0$, where d_0 is a constant which acts as a balance point between yields in early and late lactation. This may be written in matrix form as $q = d - 1d_0$. The covariance between persistency and yield is: $Cov(q't, 1't) = q'V1 = (d' - 1'd_0)V1 = d'V1 - 1'V1d_0$ for all values of d_0. The tipping point, d_0, that makes yield and persistency of a trait uncorrelated is obtained by setting $Cov(q't, 1't)$ to 0 and solving for d_0: $d_0 = d'V1/1'V1$

VanRaden (1998) originally defined persistency on an across-lactation basis such that estimates were comparable across parities. The same tipping points (d_0) were used for each trait across all cows regardless of breed or parity, and were updated by Cole and VanRaden (2006). With the implementation of smooth lactation curves (6.2) new tipping

points were estimated. The new curves were found to differ enough by parity that it was not possible to select tipping points that resulted in phenotypic correlations of 0 between yield and persistency. As a result, persistency was redefined on a within-parity group basis. Persistency was originally comparable across cows of all breeds and lactations; under the new definition, cows may be compared across breeds within parity group. Tipping points for smooth and linear interpolated curves for each parity group and trait are presented in Table 6.1. Linear interpolated curves use the same tipping points for both parity groups and should produce persistencies comparable to those calculated by AIPLDCR. Persistencies are not comparable when calculated with different interpolation methods (see: INTmethod, INTmethodSCS).

Table 6.1: Persistency tipping points (d_0)

Curves	Parity	Milk	Fat	Protein	SCS
Smooth	First	115	115	150	155
	Later	161	152	159	148
Interpolated	First	161	159	156	155
	Later	161	159	156	155

Note that the default tipping points are those for the smooth curves. If you are using linear interpolated curves then you must make sure that the correct tipping points are entered into the dim0 array in the parameter file.

Matrix V accounts for changing phenotypic standard deviations and correlations across the lactation. Persistency and 305-d yield would be genetically uncorrelated if V were replaced w ith a 305-by-305 genetic variance matrix, G, which was not available or estimated in this study. Random regression models approximate G and V with polynomials, whereas best prediction approximates V with autoregressive correlations.

True persistency (p), predicted persistency ($\hat{p}$) and the expected value of persistency, $E(p)$, are given by:

$$\begin{aligned} p &= E(p) + q't \\ \hat{p} &= q'CV_m^{-1} \\ E(p) &= q'\mu \\ &= (d' - 1'd_0)\mu \\ &= d'\mu - d_0 E(y) \end{aligned}$$

Reliability of $\hat{p}$ is obtained from variances of p and $\hat{p}$, which are computed from the quadratic forms:

$$\begin{aligned} Var(p) &= Var(q't) \\ &= Var[(d' - 1'd_0)t] \\ &= d'Vd - 2d'V1d_m^2 \\ &= q'Vq \\ Var(\hat{p}) &= q'CV_m^{-1}C'q \\ Rel(\hat{p}) &= Var(\hat{p})/Var(p) \\ &= q'CV_m^{-1}C'q/q'Vq \end{aligned}$$

The accuracy of lactation records from a wide variety of test plans can be compared using a Data Collection Rating, which is calculated as the squared correlation of estimated and true yields multiplied by a factor of 104 to give monthly testing a rating of 100 and daily testing a rating of 104.

A standardized estimate of persistency, $\hat{s}$, was obtained by subtracting the population mean for persistency (μ_p) and dividing by within-herd phenotypic standard deviation:

$$\hat{s} = \frac{\hat{p} - \mu_p}{\sqrt{Var(p)}}$$

The mean and variance of $\hat{s}$ are 0 and 1, respectively. Positive values of $\hat{s}$ indicate increased persistency relative to an average cow and negative values of $\hat{s}$ indicate decreased persistency. Cows with high persistency tend to milk less

than expected at the beginning of lactation and more than expected at the end of lactation than cows with the same level of production and average persistencies.

Predicted persistency represents the component of persistency that is independent of yield. Druet et al.(2005) extracted eigenvalues from genetic covariance matrices and showed that they may be used as proxies for lactation yield and persistency in random regression test day models. While the derivation of those proxy traits is quite different that presented for $\hat{p}$, the two measures of persistency are conceptually similar in that they both represent persistency independent of yield.

Expected Values. Lactation curves may differ according to age, parity, breed, time, herd, and their interactions. In the past, adjustment factors were used to standardize lactation records. Let vector μ contain the mature-equivalent or standard lactation curve, and let vector b contain the expected 305-day values for age, parity, season, year, and herd of interest. If μ and b differ, adjustment factors can be used to standardize yield and persistency. Multiplicative factors should not be used for persistency to avoid division-by-zero and because differences in variance can be removed by creating a unitless trait. For yield, the additive adjustment is $1'(\mu - b)$. For persistency, the additive adjustment is $q'(\mu - b)$, which is equivalent to $(d' - 1'd_0)(\mu - b)$. This approach is simple, but has the disadvantage of not fully preserving curve shape. The lactation curve is scaled vertically by the yield factor and rotated by the persistency factor. For any group of interest, the assumed curve is $(d' - 1'd_0)(\mu - b)$, which is the standard curve minus the persistency factor and then divided by the yield factor.

Expansion. Predicted yields and persistencies were expanded to conform to assumptions of commonly-used statistical models. Predicted values were divided by their reliability, holding the mean constant, to produce expanded yields ($\tilde{y}$) and persistencies ($\tilde{p}$). This is analogous to the de-regression step in multiple-trait across-country evaluations. The expanded values contain the corresponding true values plus independent errors (VanRaden, Wiggans, and Ernst 1991):

$$\begin{aligned} \tilde{y} &= E(y) + [\hat{y} - E(y)]/Rel(\hat{(y)}) \\ \tilde{p} &= E(p) + [\hat{p} - E(p)]/Rel(\hat{(p)}) \end{aligned}$$

These expanded variables have greater variances than the true values, while the predicted values have lower variances:

$$\begin{aligned} Var(\tilde{y}) &= 1'CV_m^{-1}C'1/[Rel(\hat{y})]^2 \\ &= 1'V1/Rel(\hat{y}) \\ Var(\tilde{p}) &= q'CV_m^{-1}C'q/[Rel(\hat{p})]^2 \\ &= q'Vq/Rel(\hat{p}) \end{aligned}$$

Although $Cov(y, p) = 0$, best predictions or expanded estimates of y and p may covary if tests do not represent the entire lactation, such as with records in progress. The covariance of $\tilde{y}$ with $\tilde{p}$ is $Cov(\tilde{y}, \tilde{p}) = 1'CV_m^{-1}C'q/[Rel(\tilde{y})Rel(\tilde{p})]$. Expanded yield and persistency records contain the normal environmental variance present in the true record plus an additional measurement error term that is independent of the true record (VanRaden, Wiggans, and Ernst 1991). The total error variances for yield $[Var(\tilde{y} - u_y)]$ and persistency $[Var(\tilde{p} - u_p)]$ are:

$$\begin{aligned} Var(\tilde{y} - u_y) &= Var(\tilde{y}) + Var(u_y) - 2Cov(\tilde{y}, u_y) \\ Var(\tilde{p} - u_p) &= Var(\tilde{p}) + Var(u_p) - 2Cov(\tilde{p}, u_p) \end{aligned}$$

where u_y and u_p are the sum of random effects other than error contained in the models for yield and persistency, respectively. For example, u_y and u_p might contain genetic effects plus permanent environmental effects. If the measurement errors $\tilde{y} - u_y$ and $\tilde{p} - u_p$ are uncorrelated with u_y and u_p the covariance terms are variances of random effects u_y and u_p, and variances and covariances reduce to:

$$\begin{aligned} Var(\tilde{y} - u_y) &= Var(\tilde{y}) + Var(u_y) \\ Var(\tilde{p} - u_p) &= Var(\tilde{p}) + Var(u_p) \\ Cov(\tilde{y} - u_y, \tilde{p} - u_p) &= Cov(\tilde{y}, \tilde{p}) - Cov(u_y, u_p) \end{aligned}$$

Multiple Trait Prediction. VanRaden (1997) presented both single- and multiple-trait approaches for best prediction of yields but not for persistencies. The current version of BESTPRED supports single- and multiple-trait best prediction of both yield and persistency.

6.2 Modeling Lactations Longer than 305-d

The original implementation of BP was limited to 305-d lactations, although longer lactations can be accommodated by calculating breed- and lactation-specific standard lactation curves (μ) longer than 365-d and estimating covariances among test days for DIM $>$ 365-d. Additionally, μ were constructed using the test interval method (Sargent, Lytton, and Wall 1968), which uses simple linear interpolation that assumes incorrectly that yields change at a constant rate in the interval between successive test days. Dematawewa et al. (2007) used data from long lactations to fit a number of lactation curves to test day data for M, F, and P, concluding that Wood's (1967) curves best described yield out to 999-d. Standard curves are now calculated using the "smooth" Wood's curves, although the test interval curves are still available (see: INTmethod). However, new curves for SCS yield, as well as SD of M, F, P, and SCS yield, were needed for a complete implementation of smooth curves in BESTPRED. Standard deviations of M, F, and P were modeled using Wood's (1967) curves, but there is little work in the literature describing lactation curves for SCS.

Rodriguez-Zas et al. (2000) used non-linear mixed effects models which included cow-specific lactation curves to study SCS. A number of functions were fit to SCS yield data for long lactations (Dijkstra, France, Dhanoa, Maas, Hanigan, Rook, and Beever 1997; Morant and Gnanasakthy 1989; Nelder 1966; Rook, France, and Dhanoa 1993; Wilmink 1987; Wood 1967), and a four-parameter function (C4) described by Morant and Gnanasakthy (1989) provided the best fit (Cole, 2007, unpublished data). As in Rodriguez-Zas et al.'s (2000) paper the C4 model overestimated mean SCS early and late in lactation and underestimated it in the middle. However, the other models provided much poorer fits to the data, asymptoting to positive infinity as DIM increased. The C4 function also was used to develop curves for SD of SCS yield.

Curves describing the mean and SD of all traits were developed for each of the major U.S. dairy breeds (Ayrshire, Brown Swiss, Guernsey, Jersey, Holstein, Milking Shorthorn).

6.3 Modeling Correlations Among Test Day Yields

Norman et al. (1999) used daily milk weights for Canadian cows and monthly test records of U.S. cows to estimate phenotypic correlations between test days within herd-year. They found that correlations between daily yields for a designated interval between test days generally were highest for mid-lactation and were lowest for early and late lactation. Seven possible sources of variation were examined and the resulting models for first- and later parities each included 3 sources of variation. The addition of more terms to the model generally improved fit, but gains often were very small. In addition to updating autoregressive parameters to account for changes in the shape of lactation curves over the last several years a simpler model for computing correlations matrices was developed.

Cole et al. (2007) estimated correlations among test day yields using a simplified model that included an identity matrix (I) to model daily measurement error and an autoregressive matrix (E) to account for biological change. E is defined mathematically as $E_{ij} = r^{|i-j|}$ where i and j are test day DIM and $0 < r < 1$. The r for MFP and SCS were estimated separately for first- and later-parities (Table 6.2), and values for MFP were slightly larger than previous estimates (Norman, VanRaden, Wright, and Smith 1999) due to the inclusion of the identity matrix. Parameters were not previously calculated for SCS separately from MFP.

Table 6.2: Autoregressive parameters used in modeling correlations among test days

	First	Later
MFP	0.998	0.987

	First	Later
SCS	0.992	0.995

The matrix of correlations within traits (B) was calculated as: $B = b_1 I + b_2 E$ where the b_i are regression coefficients; separate functions were used to model the yield traits and SCS. Intercept terms (b_0) were included in the calculation of B for first-parity SCS and later-parity MFP in order to guarantee the positive-definiteness of the resulting correlation matrices. In subsequent sections M and S denote the functions used to calculate correlations among MFP and SCS, respectively. Suppose that T is a 4×4 matrix of phenotypic correlations among traits partitioned as:

$$T = \left[\begin{array}{cccc:c} t_{MM} & t_{MF} & t_{MP} & t_{MS} \\ t_{FM} & t_{FF} & t_{FP} & t_{FS} \\ t_{PM} & t_{PF} & t_{PP} & t_{PS} \\ \hdashline t_{SM} & t_{SF} & t_{SP} & t_{SS} \end{array}\right] = \left[\begin{array}{c:c} T_m & T_{ms} \\ \hdashline T_{sm} & T_{ss} \end{array}\right]$$

where $t_k l$ is the phenotypic correlation of trait k with trait l. The complete correlation matrix (C) can then be obtained as:

$$C = \left[\begin{array}{c:c} M \otimes T_m & M^{\frac{1}{2}} \cdot S^{\frac{1}{2}} \otimes T_{ms} \\ \hdashline S^{\frac{1}{2}} \cdot M^{\frac{1}{2}} \otimes T_{sm} & S \otimes T_s \end{array}\right]$$

where $C_{ik,jl}$ = correlation of trait i at DIM k with trait j at DIM l, $\otimes$ is the Kronecker (direct) product operator, and $\cdot$ denotes element-wise matrix multiplication.

6.4 Computational Details

Overall Performance. On a four-way dual-core 64-bit system, and concurrent processes including DB2 and statistical packages, 150,000 records can be processed in about an hour (55 minutes) on a single core of a dual-core processor.

Memory Use. Testing at AIPL indicates that BESTPRED uses about 128 MB of RAM when running. Memory consumption should not grow as a function of the number of records processed unless there is an undetected memory leak. If such behavior does occur please file a bug report.

Batch Processing. Batch processing efficiency can be improved by blocking records by breed. When there is a change in breed within a file the correlations among test days, as well as means and SD of yield, must be recalculated. Sorting the input file by breed will avoid unnecessary calculations resulting from back-and-forth changes in breed codes.

Editing Records. The AIPL in-house processing system includes a very complex series of edits that are performed before records are passed to BESTPRED. These edits catch, for example, overlapping LER segments in input records. Those edits are not replicated in the best prediction programs, although BESTPRED will identify some errors and print warnings to STDOUT (if output is not suppressed). It is essential that users check their logfiles to catch problems arising from problems with input records. BESTPRED may process a record and still return an incorrect answer if there are problems with the input data.

This page intentionally left blank.

CHAPTER

SEVEN

Troubleshooting

7.1 Absoft 64-bit Fortran 95 compiler

7.1.1 Link Failed

A change was made in the default suffix applied to prodecure names between versions 10.0 and 10.1. In the more recent version an underscore (_) is automatically appended to procedure names at compile time causing problems for the linker.

```
f90 bestpred_main.f90 -o bestpred -O3   \
        -YEXT_NAMES=LCS                         \
        -s -lm -llocal -lU77                    \
        ageadjs.o                                       \
        aiplage.o                                       \
        bestpred_fmt4.o                                 \
        bestpred.o                                      \
        bestpred_parm.o
/tmp/ZHupkF/bestpred_main.o: In function `MAIN__':
/work5/john/jcole/bestpred/bestpred_main.f90:133: undefined reference to `read_parms_'
/work5/john/jcole/bestpred/bestpred_main.f90:380: undefined reference to `bestpred_fmt4_'
/work5/john/jcole/bestpred/bestpred_main.f90:365: undefined reference to `bestpred_fmt4_'
collect2: ld returned 1 exit status
link failed.
make: *** [bestpred] Error 1
[jcole@aipl366 bestpred]$
```

The problem is easily solved by adding the compiler option `-YEXT_SFX=""` to the makefile for each Fortran program in the makefile.

7.2 Missing File

7.2.1 adjust.scs

A text file is distributed with BESTPRED that contains adjustment factors for SCS. If that file is missing or cannot be located the standalone program will not run.

```
--------------------------------------------------------------------------------
BESTPRED version 2.0b6
    Released: 10/31/2007
    Support : John B. Cole (john.cole@ars.usda.gov)
    Website : http://www.aipl.arusda.gov/software/bestpred/
--------------------------------------------------------------------------------
adjscs.c failed to open adjust.scs
```

This problem can be solved in two ways. Either copy the file `adjust.scs` from the installation tarball into the directiry in which BESTPRED is being run or make sure that the directory in which `adjust.scs` is located is on the user's path (the latter is preferred).

7.3 Output

7.3.1 Suppress Screen Output

How do I suppress all output to the console?

Setting the *maxprnt*, *maxshow*, and *ONscreen* parameters to 0 will suppress all output from BESTPRED that is normally written to the screen.

7.3.2 Suppress Screen Output

My lactation curve output files are huge – on the order of GB – and I'd like to make them smaller.

There are two ways to decrease the size of your *CURVEfile*s: reducing the columns of data written out by setting *CURVEsmall* to 1, which will result in only actual yield data being written out; and reducing the traits written by setting elements of *GRAFplot* to 0. Using these two options together can reduce *CURVEfile* sizes by more than 75%.

7.4 Performance

7.4.1 Very Slow Processing

BESTPRED is running **very** slow. What can I do to improve its performance?

Make sure that your records are sorted by breed, parity, and calving date.

CHAPTER

EIGHT

Changes

Table 8.1: Changes to BESTPRED

Date	Program	Version	Description
04/24/2015	Documentation	2.0rc7	Updated the documentation on minimal Format 4 records (source 10 processing).
04/24/2015	Makefiles	2.0rc7	Tweaked compiler options.
04/24/2015	all .f90 files	2.0rc7	There are about 1,000,000 compiler warnings that need to be fixed. Or I can change the makefile to turn off a bunch of them.
04/24/2015	bestpred_fmt4.dat	2.0rc7	If a minimal Format 4 does NOT include a state code in bytes 107–108 of the record then it's set to 0. This causes the aiplage routine to return some...odd...values for the age adjustment factors. Going forward, if the state is 0 then it will be set to 35 (Wisconsin) and a message printed to STDOUT. This fixes a bug reported by Miel Hostens and his colleagues.
04/21/2015	bestpred.f90	2.0rc7	In the vary() function code 4 had a value of 0., which results in very low DCR for robotic herds. Changed to 1. to match supervised herds. This resolves a bug reported by AgSource.
04/21/2015	bestpred.f90	2.0rc7	Made some minor formatting changes to the lactation summaries.
11/19/2014	bestpred_main.f90	2.0rc7	Added CURVEsingle to the parm file. When 0, all cow curve and lactation summary results are written to the same file. When 1, a separate file is written for each cow.
11/19/2014	bestpred_parm.f90	2.0rc7	Added CURVEsingle to the parm file. When 0, all cow curve and lactation summary results are written to the same file. When 1, a separate file is written for each cow.
11/19/2014	bestpred_fmt4.f90	2.0rc7	Added CURVEsingle to the parm file. When 0, all cow curve and lactation summary results are written to the same file. When 1, a separate file is written for each cow.
11/19/2014	bestpred.f90	2.0rc7	Added CURVEsingle to the parm file. When 0, all cow curve and lactation summary results are written to the same file. When 1, a separate file is written for each cow.

continued on next page

continued from previous page

Date	Program	Version	Description
11/18/2014	bestpred_main.f90	2.0rc7	Corrected a typo relating to persistency floors and ceilings.
11/18/2014	bestpred_main.f90	2.0rc7	Added new input format, source 24, that reads a file named pcdart_-files.txt. Each line in that file is the name of a file that contains source 14 records. This lets a user process animals from different herds in one run of bestpred to avoid repeatedly paying startup costs.
10/02/2014	bestpred_main.f90	2.0rc7	Corrected the format string so that the columns in Format 14 (.bpo) files are no longer run together. Thanks to Albert de Vries at UF for reporting the problem.
10/02/2014	bestpred_main.f90	2.0rc7	Added an INTERFACE block for the interpolate() subroutine so that the oprional arguments (region and season) are now correctly handled. Thanks to John Beaulieu at ATA for reporting the problem.
08/13/2009	bestpred_main.f90	2.0rc6	Removed maxshow from the parameter statement so that the correct number of lactation summaries is now shown when using the standalone version.
08/13/2009	bestpred.f90	2.0rc6	Summary information, such as the table of lactation average yields, is now printed only the first time bestpred() is called when processing many records from format4.dat.
08/13/2009	bestpred.f90	2.0rc6	Cow identification is now printed with the table of ME and actual yields.
08/11/2009	ageadjs.c	2.0rc6	Updated to the latest version, which uses additive rather than multiplicative factors for SCS.
08/11/2009	bestpred_fmt4.f90	2.0rc6	Changed edits on TD segments w/status 2 to let SCS through rather than blanking it.
08/11/2009	bestpred_fmt4.f90	2.0rc6	SCS herd averages now divided by 100 just before the call to bestpred().
08/11/2009	bestpred.f90	2.0rc6	Herd average SCS are now on the phenotypic scale, e.g., 3.20, rather than multiplied by 100. This is necessary for compatibility with the AIPL edits system.
03/16/2009	bestpred_main.f90	2.0rc5	Fixed a format statement that caused G95 to throw warnings on the WIndows build.
03/16/2009	bestpred.f90	2.0rc5	Corrected use of non-portable inequality operators that were breaking Windows builds.
03/13/2009	Documentation	2.0rc4	Updated documentation to reflect changes to parameters and file formats. Added new material to the Troubleshooting chapter.
03/13/2009	bestpred.f90	2.0rc4	Added the *CURVEsmall* parameter, which controls what variables are written to the *CURVEfile*.
03/13/2009	bestpred.f90	2.0rc4	*GRAFplot* now toggles traits written to *CURVEfile* on and off.
03/13/2009	bestpred.f90	2.0rc4	Output curve files now contain only one TD yield column.
03/13/2009	bestpred_parm.f90	2.0rc4	Moved all of the version data out of the parameter file and into the bestpred subroutine.
03/13/2009	bestpred.f90	2.0rc4	Fixed a bug in which missing values in the DAILYbp were mistakenly set to 0.0, which resulted in very strange plots.
10/17/2008	bestpred_main.f90	2.0rc3	Merged-in Brad Heins's code for reading previous days open from columns 246 through 248 of Format 4 (source 10 and 15) records.

continued on next page

continued from previous page

Date	Program	Version	Description
10/16/2008	bestpred_main.f90	2.0rc3	Added code to support source 15, which is the same as source 10 but 305-d herd averages are also read from the file format4.means. The means are stored in the first row of the herd305 array (herd305(1,:).
08/19/2008	All makefiles	2.0rc2	Rearranged the makefiles such that all dependencies are built and linked in the correct order. I don't know the Absoft linker ever worked with the default file!
08/19/2008	bestpred.f90	2.0rc2	Added a few lines of code to better handle records with no test days. Thanks to Jim Boyer for noting the bugs and providing working code.
08/19/2008	bestpred_fmt4.f90	2.0rc2	Removed logging and parameter file dependencies.
08/19/2008	bestpred.f90	2.0rc2	Removed logging dependency.
08/19/2008	makefile.g95	2.0rc2	Fixed makefile for use with the G95 Fortran compiler.
08/18/2008	bestpred.f90	2.0rc2	Removed mature parameter from the ymean() subroutine header.
08/18/2008	makefile.gnu	2.0rc2	Replaced references to maindcr with references to bestpred; removed the -static directive from calls to gfortran becausethey wreck [sic] havoc with OSX.
08/11/2008	Best Prediction Manual.pdf	2.0rc1	Added material on using simulated data to test a new installation.
08/08/2008	Best Prediction Manual.pdf	2.0rc1	Added material to some tables; corrected typos; updated default parameter values.
08/08/2008	bestpred_main.f90	2.0rc1	Added debugging code to display the data returned by the bestpred() subroutine; it is activated by setting maxprnt to -1.
08/08/2008	bestpred_main.f90	2.0rc1	Added maxshow to the subroutine header. This fixes graph display problems in which maxshow and maxprnt were not working properly.
08/08/2008	bestpred_fmt4.f90	2.0rc1	Added maxshow to the call to bestpred().
08/08/2008	bestpred_main.f90	2.0rc1	Added maxshow to the call to bestpred_fmt4().
08/07/2008	bestpred_main.f90	2.0rc1	Made modifications to the code that prints incoming data for debugging. The data were being displayed in kilograms regardless of the input units. The printout now uses same units as the incoming data.
08/07/2008	Best Prediction Manual.pdf	2.0b9	Updated to reflect new file formats, removal of the mature parm, uses of GRAFplot, and the logging module.
08/07/2008	bestpred_parm.f90	2.0b9	Added code to check GRAFplot for permissible values; if invalid values are found the default of 2 (actual values) is used.
08/07/2008	bestpred.f90	2.0b9	GRAFplot is now checked for permissible values and used as an index in plot drawing to toggle between ME and actual plots.
08/06/2008	bestpred.f90	2.0b9	Fixed an array with incorrect dimensions that was corrupting herd averages.
08/06/2008	bestpred_fmt4.f90	2.0b9	Changed how 305d herd averages are calculated.
08/06/2008	bestpred_main.f90	2.0b9	Fixed another bug w/lactation length handling.
08/05/2008	bestpred.f90	2.0b9	Removed 'mature' parameter from the subroutine. Mature and ME yields are now returned from a single call to the subroutine.
08/05/2008	bestpred_fmt4.f90	2.0b9	Removed 'mature' parameter.
08/05/2008	bestpred_parm.f90	2.0b9	Removed 'mature' parameter.

continued on next page

continued from previous page

Date	Program	Version	Description
08/05/2008	bestpred_main.f90	2.0b9	Removed 'mature' parameter.
08/05/2008	bestpred_main.f90	2.0b9	Fixed a bug in which the lengths of sample lactations read from DCRexample.txt were not being correctly calculated.
04/03/2008	Best Prediction Manual.pdf	2.0b9	Updated manual.
04/03/2008	bestpred_log.f90	2.0b8	Fixed a glitch in the header of newly-created log files.
03/19/2006	bestpred*.f90	2.0b8	Modified parameter lists to support logging tools.
03/13/2006	bestpred_main.f90	2.0b8	Removed unused variable declarations to get a clean compile.
03/12/2006	makefile	2.0b8	Added entries for compiling the logging module.
03/12/2006	bestpred_log.f90	2.0b8	Added logging module.
03/12/2006	bestpred_main.f90	2.0b8	Made changes to input record processing so that the last record read from source 10 is no longer skipped. Also verified that I did not break processing of records w/0 TD.
12/19/2007	bestpred.f90	2.0b8	Updated all lactation curves (mean and SD for MFPS) to use parameters based on 1 versus 2-and-later parities using up to 500-d data for MFP and 999-d data for SCS instead of 1 versus 3+ parities.
12/14/2007	bestpred.f90	2.0b8	Updated the regression coeffs. in the vary() function with new values from a corrected program.
12/10/2007	bestpred.f90	2.0b7	Updated mean and SD lactation curves for all breeds and traits.
12/05/2007	bestpred_fmt4.f90	2.0b7	Modified the format4 string to take up to 50 TD rather than 20 as in the Format 4 specification.
12/05/2007	bestpred_main.f90	2.0b7	Modified the format4 string to take up to 50 TD rather than 20 as in the Format 4 specification. Commented out JBC code for writing lactation curves b/c (1) the code didn't work correctly, and (2) curves are now written to files by a function in bestpred.f90.
12/04/2007	bestpred.f90	2.0b7	Updated lactation curve parms for later parity M, F, and P in the interpolate subroutine from the 305 d Dematawewa et al. (2007) parms to the 999 d parms.
12/04/2007	standard_curves.py	2.0b7	Updated lactation curve parms for later parity HO M, F, and P from the 305 d Dematawewa et al. (2007) parms to the 999 d parms.

BIBLIOGRAPHY

Cameron, N. D. (1997). *Selection Indices and Prediction of Genetic Merit in Animal Breeding*. CAB International.

Cole, J. B., P. VanRaden, and C. M. B. Dematawewa (2007). Estimation of yields for long lactations using best prediction. *Journal of Dairy Science 90(Suppl. 1)*, 421.

Cole, J. B. and P. M. VanRaden (2006). Genetic evaluation and best prediction of lactation persistency. *Journal of Dairy Science 89*, 2722–2728.

Dematawewa, C. M. B., R. E. Pearson, and P. M. VanRaden (2007). Modeling extended lactations of holsteins. *Journal of Dairy Science 90*, 3924–3936.

Dijkstra, J., J. France, M. S. Dhanoa, J. A. Maas, M. D. Hanigan, A. J. Rook, and D. E. Beever (1997). A model to describe growth patterns of the mammary gland during pregnancy and lactation. *Journal of Dairy Science 80*, 2340–2354.

Druet, T., F. JaffrÃ©zic, and V. Ducrocq (2005). Estimation of genetic parameters for test day records of dairy traits for the first three lactations. *Genetics, Selection, Evolution 37*, 257–271.

International Committee for Animal Recording (2006, June 9). *International Agreement of Recording Practices*. Rome, Italy: International Committee for Animal Recording.

Morant, S. V. and A. Gnanasakthy (1989). A new approach to the mathematical formulation of lactation curves. *Animal Production 49*, 151–162.

Nelder, J. A. (1966). Inverse polynomials, a useful group of multifactor response functions. *Biometrics 22*, 128–141.

Norman, H. D., P. M. VanRaden, J. R. Wright, and J. S. Clay (1999). Comparison of test interval and best prediction methods for estimation of lactation yield from monthly, a.m.-p.m., and tri-monthly testing. *Journal of Dairy Science 82*, 438–444.

Norman, H. D., P. M. VanRaden, J. R. Wright, and L. A. Smith (1999). Mathematical representations of correlations among yield traits and somatic cell score on test day. *Journal of Dairy Science 82*, 2205–2211.

Quist, M. A., S. J. LeBlanc, K. J. Hand, D. Lazenby, F. Miglior, and D. F. Kelton (2007). Agreement of predicted 305-day milk yields relative to actual 305-day milk weight yields. *Journal of Dairy Science 90*, 4684–4692.

Rodriguez-Zas, S. L., D. Gianola, and G. E. Shook (2000). An approximate bayesian analysis of somatic cell score curves in holsteins. *Acta Agriculturae Scandinavia A 50*, 291–299.

Rook, A. J., J. France, and M. S. Dhanoa (1993). On the mathematical description of lactation curves. *Journal of Agricultural Science 121*, 97–102.

Sargent, F. D., V. H. Lytton, and O. G. Wall, Jr. (1968). Test interval method of calculating dairy herd improvement association records. *Journal of Dairy Science 51*, 170–179.

Schaeffer, L. R. and J. Jamrozik (1996). Multiple-trait prediction of lactation yields for dairy cows. *Journal of Dairy Science 79*, 2044–2055.

Shook, G. E., E. L. Jensen, and F. N. Dickinson (1980). Factors for estimating sample-day yield in am-pm plans. *DHI Letters 56(4)*, 25–30.

Shook, G. E., L. P. Johnson, and F. N. Dickinson (1980). Factors for improving accuracy of estimates of test-interval yield. *DHI Letters 56(4)*, 9–24.

VanRaden, P. M. (1997). Lactation yields and accuracies computed from test day yields and (co)variances by best prediction. *Journal of Dairy Science 80*, 3015–3022.

VanRaden, P. M. (1998). Best prediction of lactation yield and persistency. *Proc. 6th World Congr. Genet. Appl. Livest. Prod., Armidale, Australia XXIII*, 347–350.

VanRaden, P. M., G. R. Wiggans, and C. A. Ernst (1991). Expansion of projected lactation yield to stabilize genetic variance. *Journal of Dairy Science 74*, 4344–4349.

Wiggans, G. R., I. Misztal, and D. Van Vleck (1988). Implementation of an animal model for genetic evaluation of dairy cattle in the united states. *Journal of Dairy Science 71(Suppl. 2)*, 54–69.

Wilmink, J. B. M. (1987). Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livestock Production Science 16*, 335–348.

Wood, P. D. P. (1967). Algebraic model of the lactation curve in cattle. *Nature 216*, 164–165.

INDEX