Method and Effect of Adjustment for Heterogeneous Variance in Holstein Type traits

K.R. Koots, B.W. Kennedy, G.C. Smith and E.B. Burnside

Centre for the Genetic Improvement of Livestock
Department of Animal and Poultry Science
University of Guelph, Guelph, Ontario

Introduction

Investigations by Sorensen and Kennedy (1985, JDS, 68: 1770) and Wade et al. (GEB, May 1992a) revealed that heterogeneous within-herd-variance does exist for type traits in Canadian Holsteins. Subsequently, Wade et al. (GEB, May 1992b) corrected four conformation traits for heterogeneous variance via the Hill Method (Hill, Anim. Prod.: 39: 473, 1984) and compared solutions and proofs to the current (unadjusted) system. The results suggested that heterogeneous variance is small and does not cause substantial misranking in the Canadian Type Evaluation.

Nevertheless, adjustment for heterogeneous within-herd variance is being considered for the January 1993 Type Evaluation Run, in light of similar changes proposed for the Production Evaluation System.

The objectives of this study are

1. To derive a better method of predicting true within-herd variances, which can be used in the Hill adjustments.

2. To compare the official (unmodified) July 1992 Holstein Type evaluations with those derived from the same data, after correcting for within-herd variance.

Background

The Hill method involves adjusting observations using the within-herd phenotypic variance. Heritabilities, therefore, are assumed constant across herds. Sorensen and Kennedy (1985) have indicated this to be the case in most type traits.

Adjustment of the data by the Hill method requires estimates of the within-herd standard deviation. Because many herds have small numbers, these within-herd standard deviations should be regressed toward some regional value. The following method has been proposed for the Production traits (Robinson, GEB May 1992).

$$\sigma^2_{p_{hyp}} = \frac{30\sigma^2_{p_{region_{yp}}} + 1/3 df_{hyp-2}\sigma^2_{p_{hyp-2}} + 2/3 df_{hyp-1}\sigma^2_{p_{hyp-1}} + df_{hyp}\sigma^2_{p_{hyp}}}{30 + 1/3 df_{hyp-2} + 2/3 df_{hyp-1} + df_{hyp}}$$

$\sigma^2_{p_{hyp}}$ = combined estimate for the current year

$\sigma^2_{p_{region_{yp}}}$ = prior for the region for this year and parity

df_{hyp-2} = degrees of freedom for this herd and parity 2 years ago

df_{hyp-1} = " " " " " " " " 1 year ago

df_{hyp} = " " " " " " " " for current year

$\sigma^2_{p_{hyp-2}}$ = estimated phenotypic variance for this herd and parity 2 years ago

$\sigma^2_{p_{hyp-1}}$ = " " " " " " " " 1 year ago

$\sigma^2_{p_{hyp}}$ = " " " " " " " " current year

That is, the regressed herd-year-parity variance for production traits is estimated as the weighted mean of variation in that herd-year-parity, variation in adjacent years for the same herd-parity, and region-year-parity variance. The only unresolved problem is to determine the appropriate weights to use for the regional prior and herd information. The U.S.D.A. production evaluation system, for example, uses a similar method but with different weightings (Wiggans and VanRaden, JDS: 4350, 1991):

$$\sigma^2_{p_{hyp}} = \frac{20\sigma^2_{p_{region_{yp}}} + 1/2 df_{hyp-2}\sigma^2_{p_{hyp-2}} + 1/2 df_{hyp-1}\sigma^2_{p_{hyp-1}} + df_{hyp}\sigma^2_{p_{hyp}}}{20 + 1/2 df_{hyp-2} + 1/2 df_{hyp-1} + df_{hyp}}$$

Appropriate weightings can be determined if one could distinguish between true inter-herd variance and sampling errors. The approach of the present study is to estimate both the herd and error variance components, accounting for region, round and classifier effects. Resulting estimates of the herd effects are BLUP (fixed effects are accounted for) and are similar to, but perhaps tidier, than the approach proposed above for production traits.

Materials and Methods

The data consisted of 1,092,387 observations for final class score from the July 1992 run of the type evaluations. These data were transformed to 1-100 (Snell) scale, pre-corrected for age and stage of lactation effects, and expressed as deviates from their mean.

A second data set was constructed of standard deviations of final class score for each Herd-Round-Classifier (HRC) subclass. After removing 7 594 HRCs with single observations, 103 563 HRCs remained. A linear model was fitted to this data set, treating herds as random effects. The observations (HRC standard deviations) are repeated over rounds, therefore a estimate of the herd variance component can be obtained. The following linear model was fitted using Maximum Likelihood (Table 1):

$$\sigma_{p_{hrc_{ijklm}}} = \mu + Round_i + Classifier_j + Region_k + Herd_{kl} + b(N_{ijklm}) + e_{ijklm}$$

where,

$\sigma_{p_{hrc_{ijklm}}}$	- individual HRC standard deviation
μ	- overall mean standard deviation
$Round_i$	- fixed effect of the *i*th round (*i*=1,2...21)
$Classifier_j$	- fixed effect of the *j*th classifier (*j*=1.2...24)
$Region_k$	- fixed effect of the *k*th region (*k*=1,2...5)
$Herd_{kl}$	- random effect of the *k*th herd nested within the *l*th region
b	- linear effect of number of cows in each HRC
N_{ijklm}	- number of animals in each HRC
e_{ijklm}	- random residual effect

A simulated data set with the same data structure as the original data set from the July 1992 run of the type evaluation run was created with homogeneous within-herd variances. When compared to the field data set, this gives an indication of how much of a problem exists from inter-herd heterogeneous variance. That is, it would indicate what proportion of within-herd variance is random and what proportion is really due to herd.

In addition to an estimate of the herd variance component, Maximum Likelihood provided the best estimate of these herd standard deviations (BLUP). These BLUP estimates are already corrected for all fixed effects (region, round and classifier) and are regressed by the relative amount of information (number of rounds of herd data) and by the appropriate variance ratio. These BLUP estimates of individual herd standard deviations have a mean of zero, and can be used to construct estimates of the expected herd standard deviation, as follows:

$$\hat{\sigma}_{p_{hrc_{ijklm}}} = \hat{\mu} + \widehat{Round}_i + \widehat{Classifier}_j + \widehat{Region}_k + \widehat{Herd}_{kl} \quad (1)$$

Because only the herd effect was of interest in the present study, the following herd standard deviation was estimated:

$$\sigma_{Phre_{ijklm}} = \beta + \tilde{Herd}_{kl} \tag{2}$$

The original data set from the July 1992 type evaluation run was adjusted using the Hill formula, employing the herd standard deviation as estimated in (2) above. Albeit, (1) would probably be the appropriate estimate to used if such an adjustment is implemented into the type evaluation system.

Results

The results (Table 2) indicate that heterogeneous within-herd variance is not a problem for type traits. The repeatability of herd variance was found to be around 3 percent.

However, if adjustment for inter-herd variance is warranted (to be consistent with what is done for production traits), BLUP estimates of the individual herd standard deviations may be used in the Hill formula. The data (1 092 387 records) from the July 1992 run of the Type evaluation system was adjusted using the Hill method, employing the BLUP estimates of the herd standard deviations and results are presented in Tables 3 and 4.

The correlations between the two systems are high, as expected. The distribution of changes in proofs (Table 4) indicates that proofs changed much less due to the adjustment than they generally do between semi-annual runs. The absence of large changes in proofs is further evidence that an adjustment for heterogeneous herd variance is not necessary for type traits.

One assumption of this approach is that within-herd variance does not change significantly over rounds. This assumption can be alleviated by using more sophisticated methods such as Nearest Neighbour Adjusted BLUP (NNABLUP) (Stroup and Mulitze, 1991, Amer. Stat., 45: 194).

Conclusions

The results above support earlier research that heterogeneous within herd variance is not a problem for type traits in Canadian Holsteins. However, if adjustment is deemed necessary in the future, BLUP estimates of the herd standard deviations are one way to retrieve within herd variance. It should be noted, however, that this method is computationally demanding and significantly increases computing time.

The relative weight given to the regional prior versus the herd standard deviation appears to be equal. The average effect due to region was .238 (Appendix 1); and the standard deviation of the herd estimates was .242.

Recommendations

If, and when, an adjustment for heterogeneous within-herd variance is to be implemented for production traits, a similar procedure should be implemented for conformation traits. We propose the procedure described above be used to adjust all 31 conformation traits in the Holstein breed, using appropriate herd variances (ie different) for each trait. It may not be possible to implement such a procedure in the small breeds due to small number of herds and difficulty in getting convergence.

Table 1. Analysis of variance for HRC standard deviation.

Source	df	MS	F
Round	20	1248.3	94.95**
Classifier	23	973.7	74.06**
Region	4	104.2	7.93**
Covariate	1	10162.7	773.0**
Error	103 514	13.1	

** P<.01

Table 2. Estimated components of variance.

	Field data	Simulated data
σ^2_{herd}	0.4308	0.0000
σ^2_{error}	13.1410	14.3976
$\sigma^2_{error} / \sigma^2_{herd}$	35.507	∞

Table 3. Correlations between proofs obtained under the current system and those obtained after adjusting data for heterogeneous herd variance.

	N	Final Class Score
Bull solutions	23 774	.999
Bull proofs	23 774	.997
Cow indexes	1 095 849	.997

Table 4. Frequency distribution of changes in official sire proofs (>55% repeatability) and cow indexes (>20% repeatability) after adjusting for heterogeneous variance.

Change	Frequency	
	Sire Proofs	Cow Indexes
-4	0	0
-3	0	1
-2	0	1
-1	121	96351
0	3305	993153
1	149	6340
2	0	3
3	0	0
4	0	0

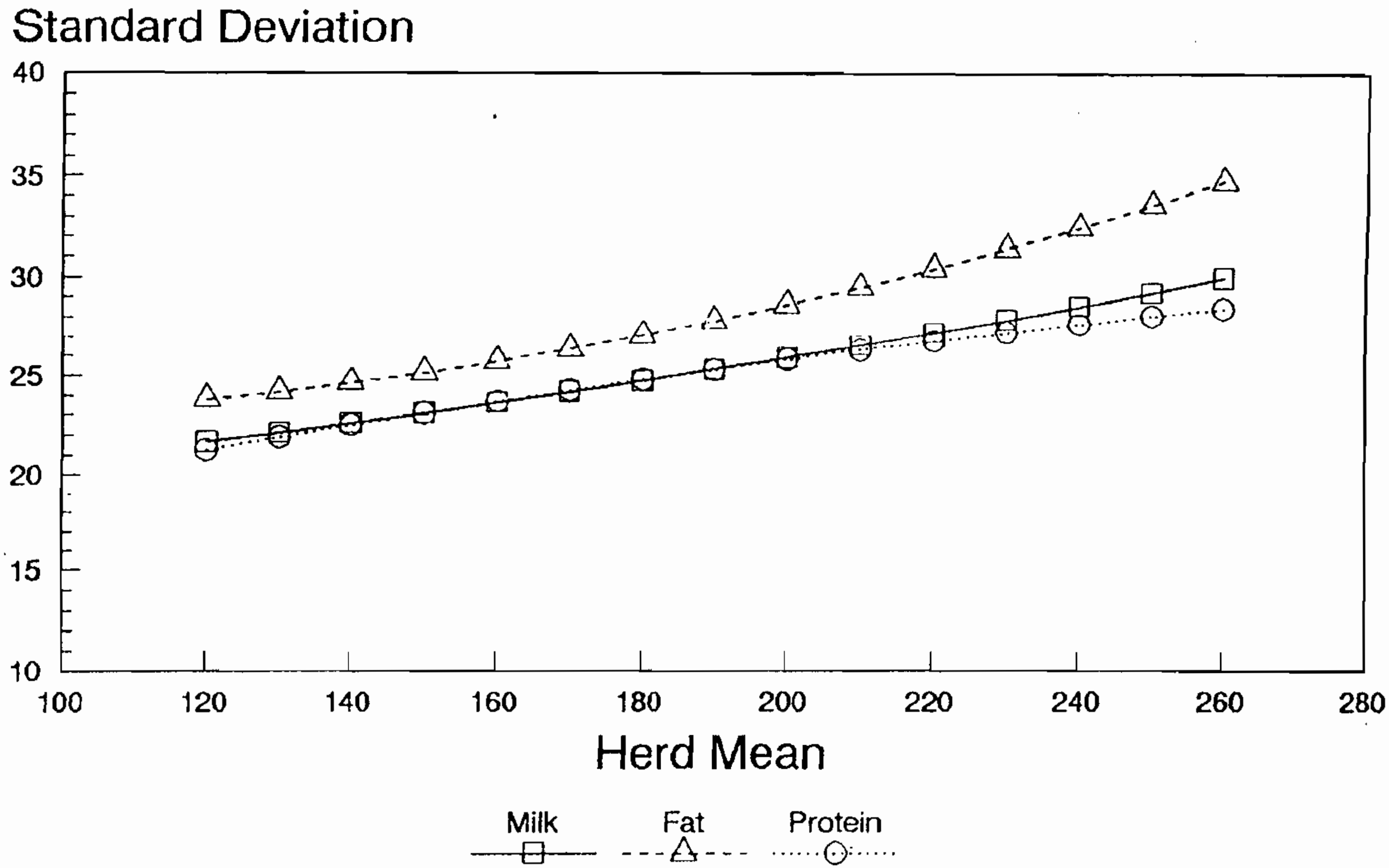
Appendix 1

Arbitrary assignment of provinces to regions, and estimates of regional effect on the standard deviation for final class score.

Region	Province	N	Solution
1	Ontario	509 859	0.2729
2	Quebec	351 310	0.1975
3	New Brunswick Nova Scotia PEI Newfoundland NWT	48 870	0.2692
4	Manitoba Saskatchewan Alberta	95 110	0.3897
5	British Columbia	79 640	0.0000
Overall mean standard deviation for FS		1 084 789	12.4819

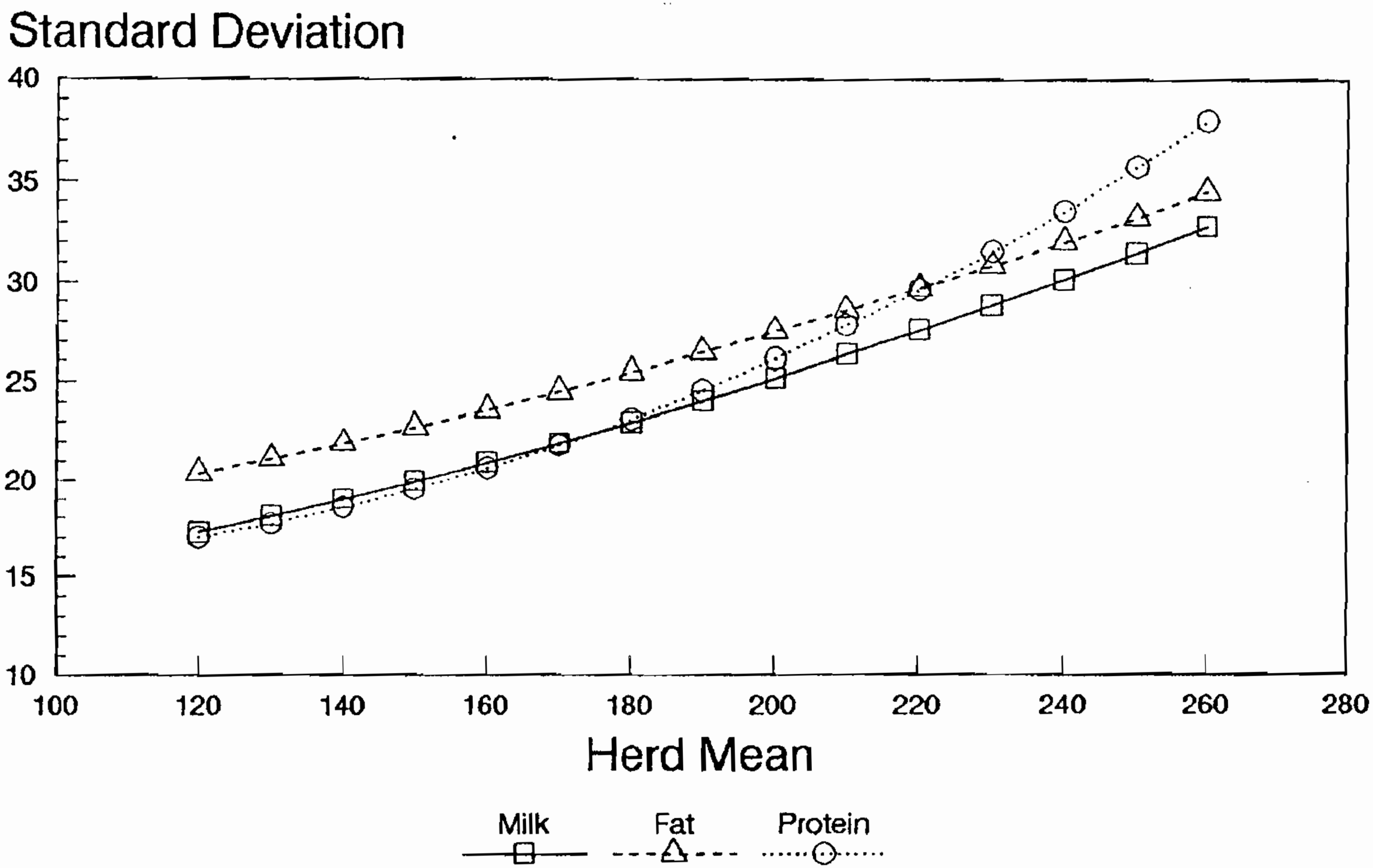
Herd Standard Deviation by Herd Mean

B. C.



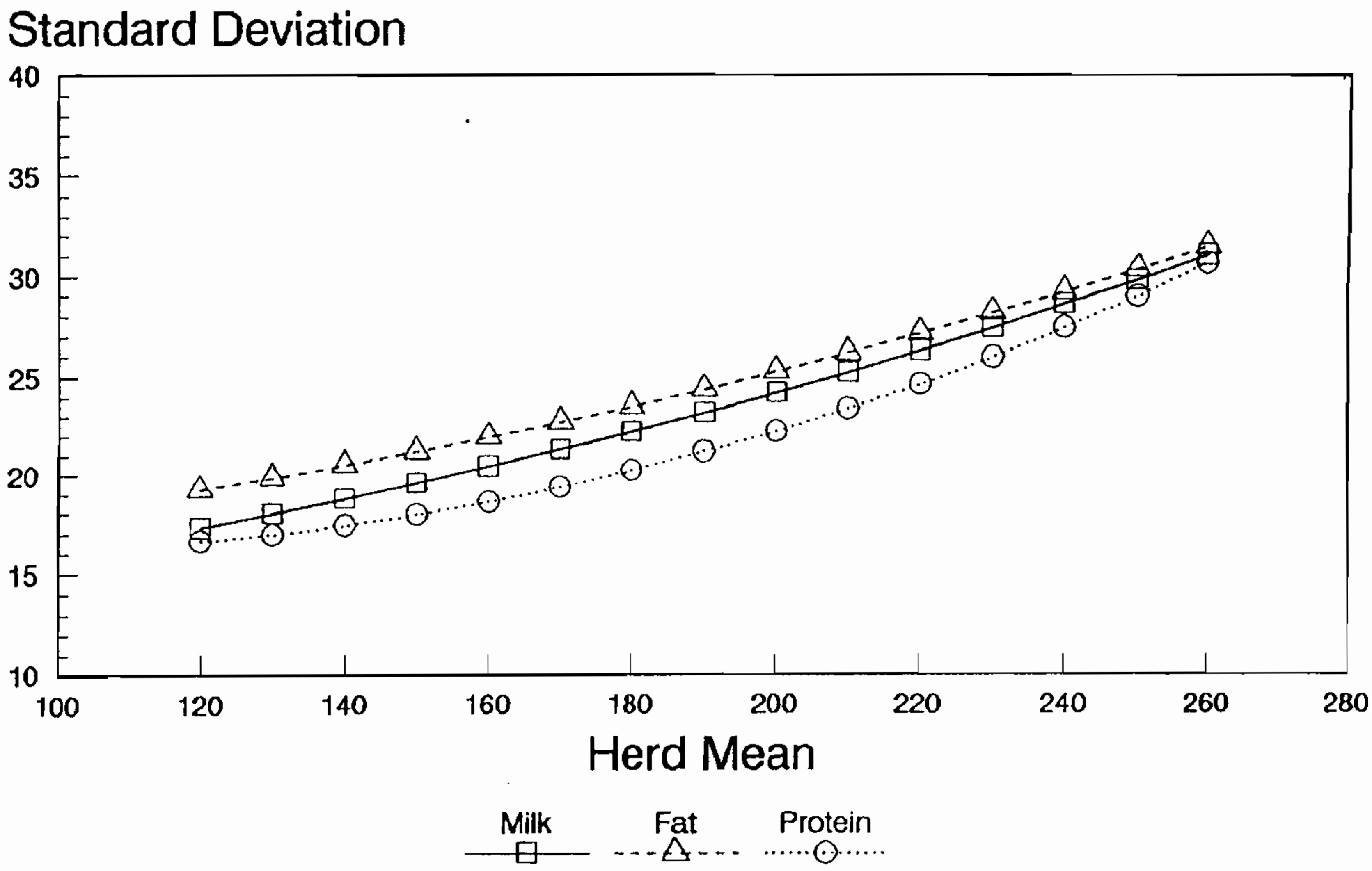
Herd Standard Deviation by Herd Mean

Maritimes



Herd Standard Deviation by Herd Mean

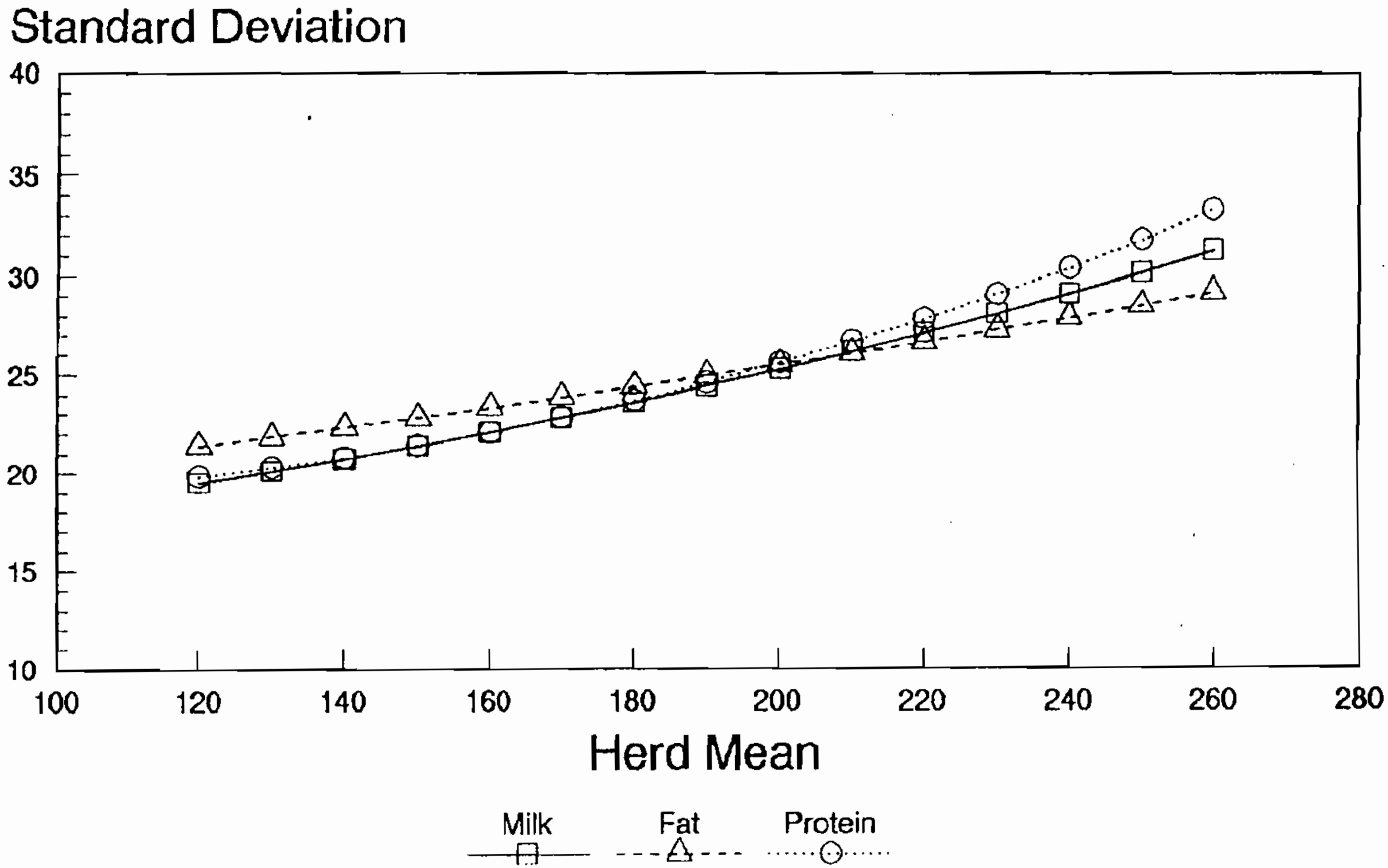
Ontario



))))

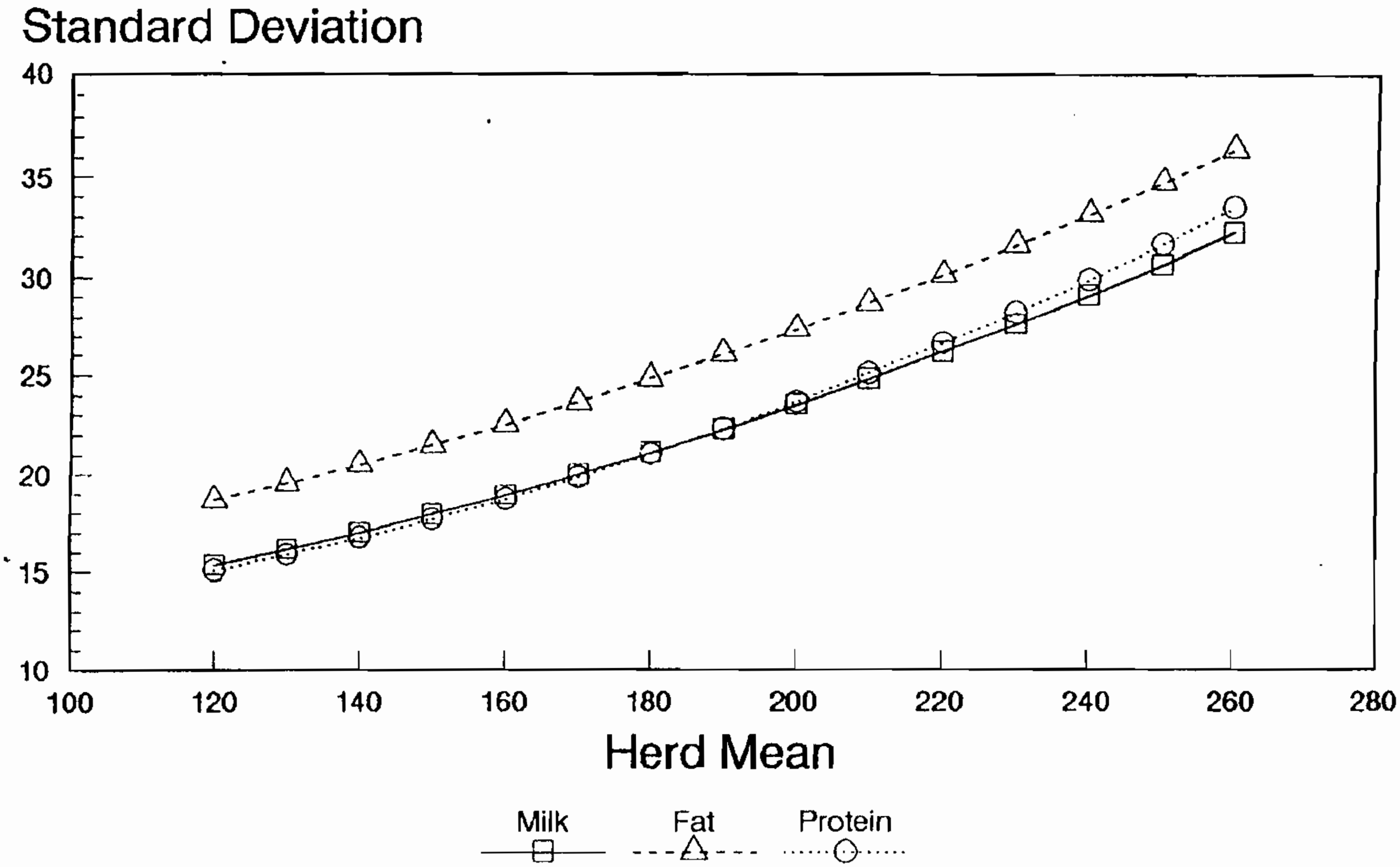
Herd Standard Deviation by Herd Mean

Prairies



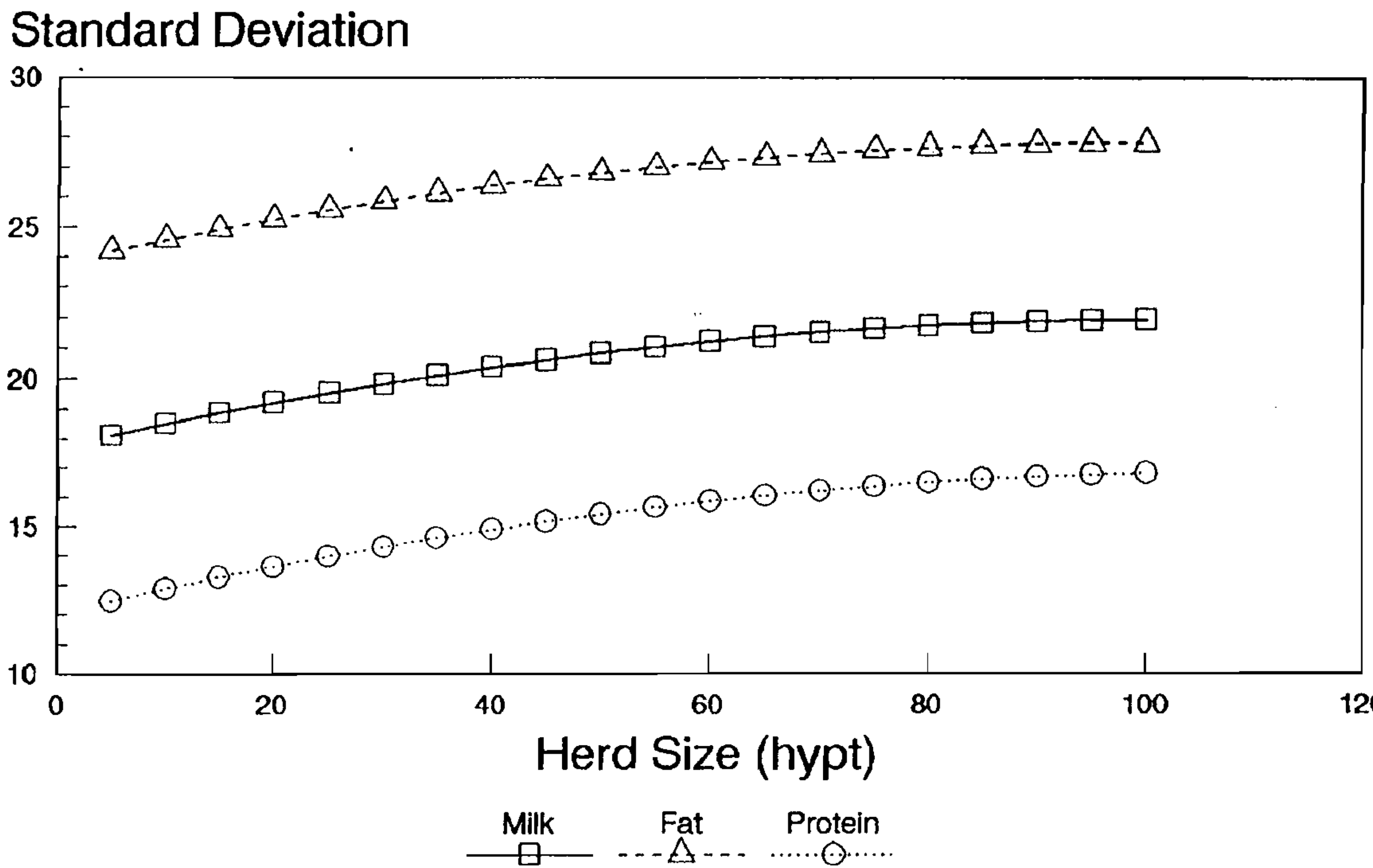
Herd Standard Deviation by Herd Mean

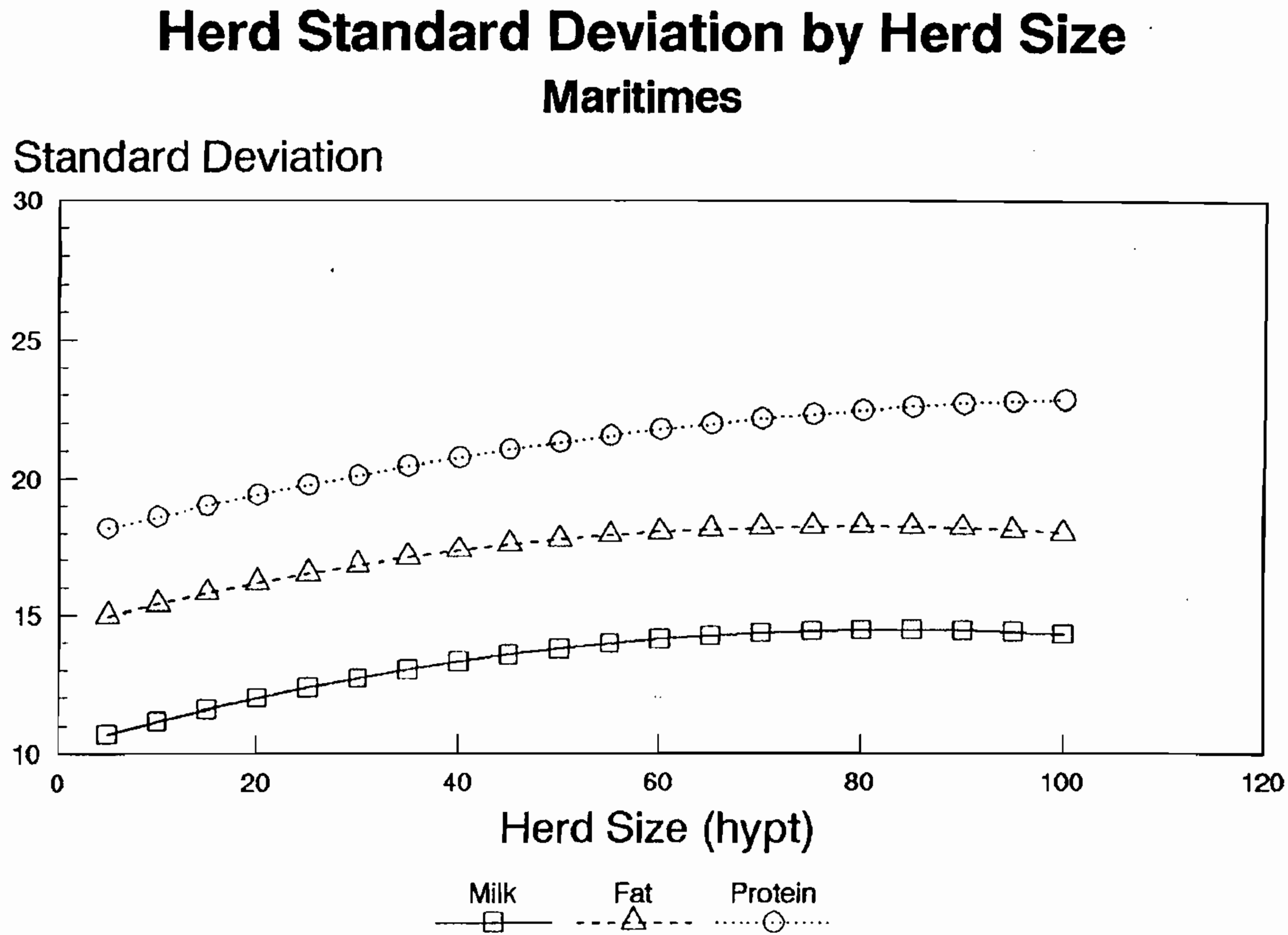
Quebec



Herd Standard Deviation by Herd Size

B. C.

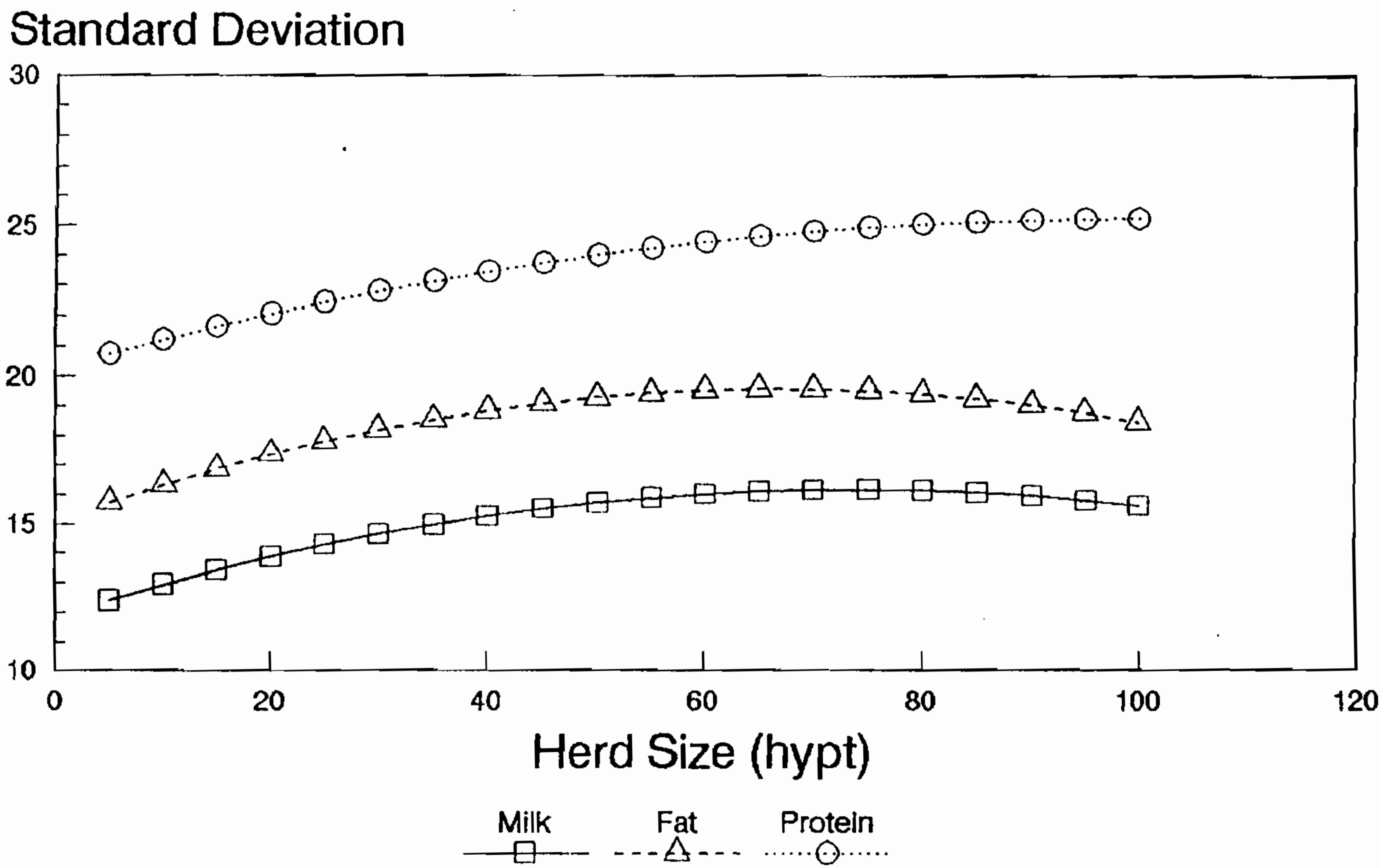


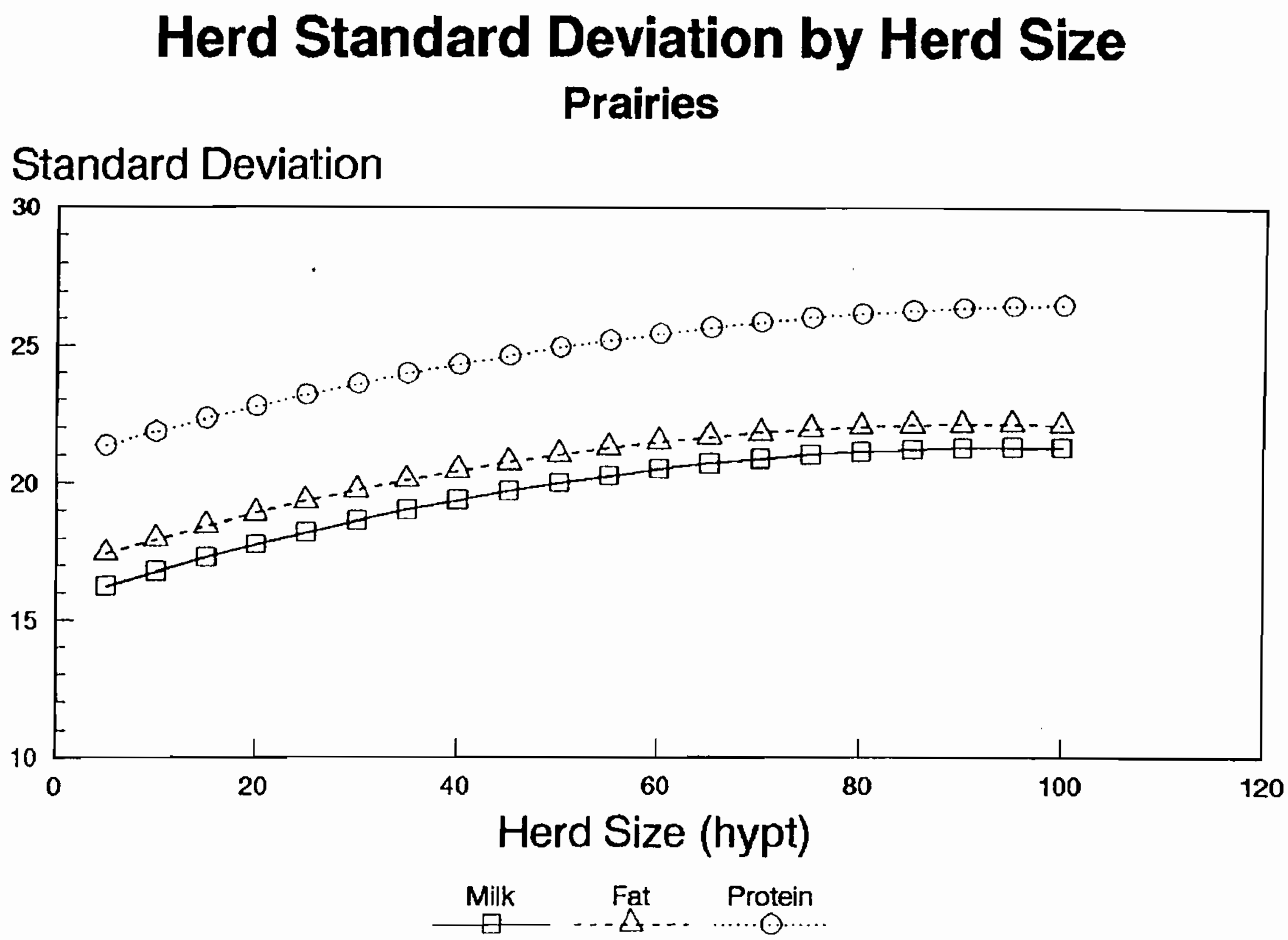


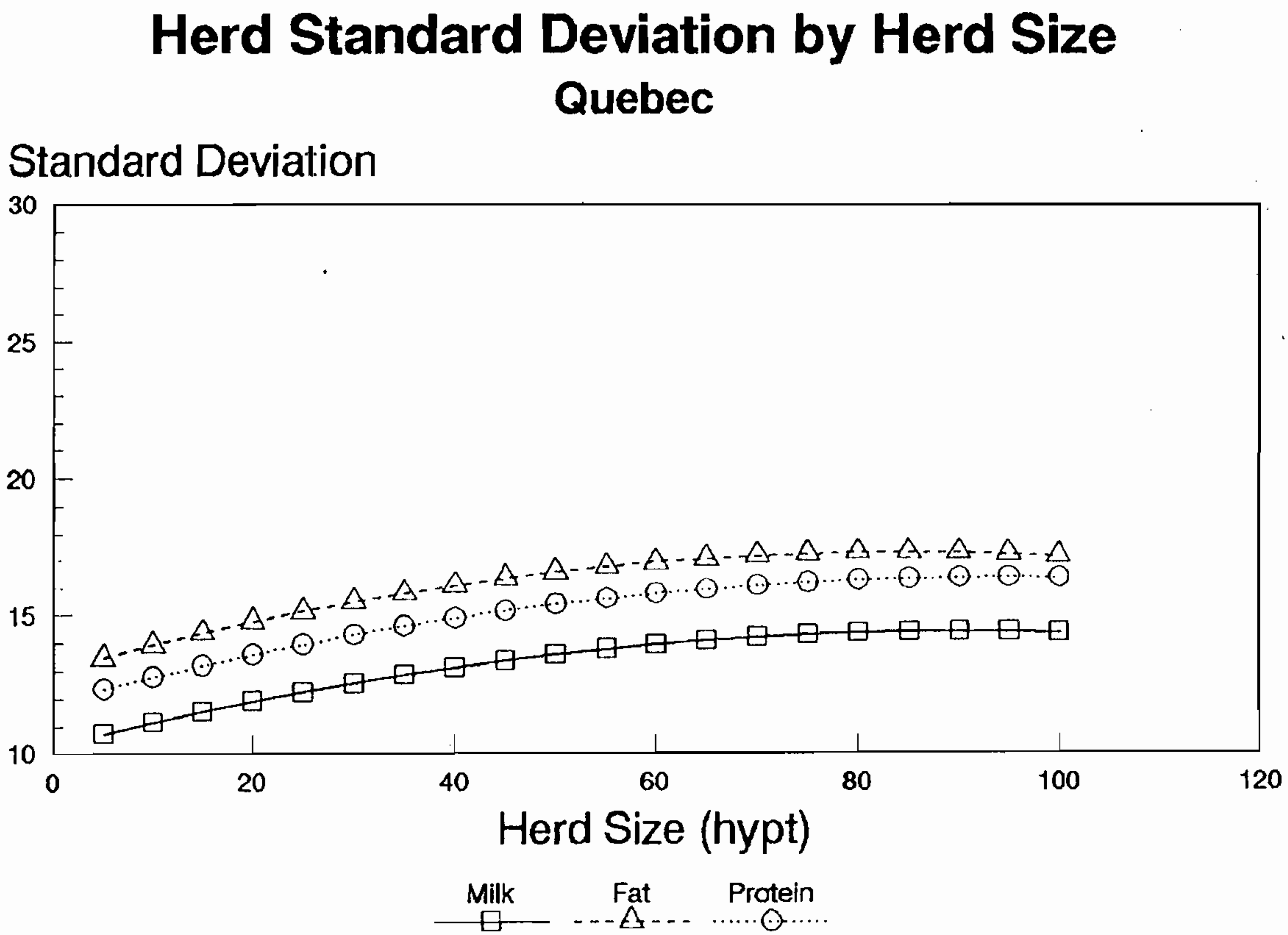
Herd Standard Deviation by Herd Size
Maritimes
Standard Deviation
30
25
20
15
10
0
20
40
60
80
100
120
Herd Size (hypt)
Milk
Fat
Protein

Herd Standard Deviation by Herd Size

Ontario




Herd Standard Deviation by Herd Size
Prairies
Standard Deviation
30
25
20
15
10
0
20
40
60
80
100
120
Herd Size (hypt)
Milk
Fat
Protein



Herd Standard Deviation by Herd Size
Quebec
Standard Deviation
30
25
20
15
10
0
20
40
60
80
100
120
Herd Size (hypt)
Milk
Fat
Protein