

Polytech'Lille — GIS3 — Projet de Calcul Numérique
Applications de l'Algèbre Linéaire aux Moindres Carrés et à l'Interpolation

Les questions posées dans ce sujet de TP sont indicatives : ce sont davantage des propositions de pistes à étudier que des questions scolaires. En particulier, vous n'êtes pas obligé de traiter chaque question dans son intégralité. Une partie importante du projet doit toutefois consister à programmer des méthodes numériques en FORTRAN. En fin de projet, on vous demande de retourner par courrier électronique, à votre encadrant, une archive contenant vos codes FORTRAN ainsi que votre rapport au format PDF.

L'archive doit être propre : pas de fichier objet, d'exécutable, de copie de sauvegarde de votre éditeur de texte. Les noms des fichiers et des répertoires ne doivent contenir que des caractères ASCII. Si vous réalisez plusieurs variantes du même programme, pour répondre aux différentes questions, placez chaque variante dans l'archive finale. Le correcteur ne doit pas avoir à modifier votre code FORTRAN pour tester votre travail. Les programmes doivent compiler sans *warnings* et permettre de reproduire les résultats donnés dans le rapport. Ils doivent être agréables d'emploi. Pas d'interface compliquée mais une ou deux lignes de texte bien pensées qui expliquent ce qui est attendu en entrée et ce qui est affiché en sortie. Ils doivent être indentés et raisonnablement commentés.

Le rapport doit aider le correcteur à noter votre projet. Il doit mettre en avant les aspects positifs de votre projet, être honnête et expliquer comment reproduire les calculs que vous avez obtenus. Soyez synthétique : ne répétez pas l'énoncé et n'incluez aucun code FORTRAN, sauf en cas de nécessité. Numérotez vos pages. Rédigez une introduction et une conclusion. Pensez à appliquer, avec discernement, la

Règle du « un septième ». Toute texte explicatif doit être précédé d'un résumé approximativement sept fois plus court. Les six septièmes restant ne font que redire ce qui est déjà énoncé dans le résumé en ajoutant des détails et des justifications. Cette règle s'applique récursivement jusqu'à la première phrase.

1 Introduction

Dans beaucoup de problèmes appliqués, on est amené à chercher une courbe reliant une suite de points expérimentaux (interpolation) ou passant « au plus près » de ces points (moindres carrés). Prenons un exemple. Les points expérimentaux sont les $m = 4$ points suivants :

i	1	2	3	4
x_i	1	2	-6	4
y_i	3.01	3.99	-4.07	6

Ces points sont à peu près alignés. On cherche les $n = 2$ coefficients α_1, α_2 d'une droite d'équation

$$y = \alpha_1 x + \alpha_2.$$

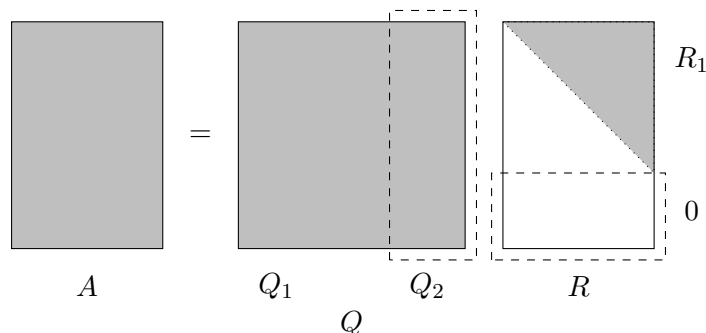
Posons que la droite passe par les points. On obtient le système de m équations en n inconnues suivant (sous forme développée à gauche et matricielle à droite) :

$$\begin{array}{rcl} 3.01 & = & \alpha_1 + \alpha_2, \\ 3.99 & = & 2\alpha_1 + \alpha_2, \\ -4.07 & = & -6\alpha_1 + \alpha_2, \\ 6 & = & 4\alpha_1 + \alpha_2; \end{array} \quad \underbrace{\begin{pmatrix} 1 & 1 \\ 2 & 1 \\ -6 & 1 \\ 4 & 1 \end{pmatrix}}_A \underbrace{\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}}_x = \underbrace{\begin{pmatrix} 3.01 \\ 3.99 \\ -4.07 \\ 6 \end{pmatrix}}_b.$$

Ce système n'a pas de solution (aucune droite ne passe par les quatre points) mais au moins deux méthodes d'algèbre linéaire étudiées en cours permettent de chercher le meilleur vecteur x possible, au sens des moindres carrés [2, chapitre 7].

La méthode historique. Elle consiste à résoudre le système $A^T A x = A^T b$. La matrice $A^T A$ est symétrique. Sauf cas dégénéré, elle est définie positive [2, chapitre 7, Proposition 14, page 71].

La méthode moderne. Elle consiste à calculer une factorisation QR réduite¹ de A puis à résoudre $R_1 x = Q_1^T b$ avec :



2 Codage FORTRAN

Dans cette section, on fournit un code complet (fichier `projet.f`) qui résout complètement l'exemple avec les deux méthodes. Ce code s'appuie sur la bibliothèque numérique mondialement connue LAPACK ainsi que sur une bibliothèque « pédagogique » `libmutils`².

```
PROGRAM PROJET
* La matrice A et le vecteur b sont surdimensionnés
  INTEGER MMAX, NMAX
  PARAMETER (MMAX = 100)
  PARAMETER (NMAX = 10)
  DOUBLE PRECISION A(MMAX,NMAX), B(NMAX)
* Les dimensions mathématiques de A (m x n) et b (m) sont lues sur l'entrée standard
  INTEGER M, N
```

1. Calculer la factorisation réduite de A revient à ne calculer que Q_1 et R_1 . C'est bien une factorisation de A puisque $A = Q_1 R_1$. Dans le cadre des problèmes de moindres carrés, seules les matrices Q_1 et R_1 sont utiles.

2. Les fonctions de la bibliothèque pédagogique se reconnaissent au fait que leur identificateur contient un blanc souligné.

```

*
* Lecture de A et b.
    CALL DREAD_MPL (M, N, A, MMAX)
    CALL DREAD_MPL (M, 1, B, MMAX)
* Affichage de A et b
    CALL DPRINT_MPL ('A', M, N, A, MMAX)
    CALL DPRINT_MPL ('b', M, 1, B, MMAX)
* Résolution par l'algorithme de Cholesky
    CALL CHOLESKY_SOLVE (M, N, A, MMAX, B)
* Résolution par la factorisation QR
    CALL FACT_QR_SOLVE (M, N, A, MMAX, B)
    END PROGRAM

    SUBROUTINE CHOLESKY_SOLVE (M, N, A, LDA, B)
* Les paramètres formels (LDA = Leading Dimension of A)
    INTEGER M, N, LDA
    DOUBLE PRECISION A(LDA,*), B(*)
* Variables locales
    DOUBLE PRECISION ATA(N,N), ATB(N)
    INTEGER INFO
*
    WRITE (*,*) 'Résolution par la méthode du Cdt Cholesky'
* ATA = Transpose (A) . A (DGEMM est une BLAS pour le pdt de matrices)
    CALL DGEMM ('Transpose', 'No Transpose',
    $           N, N, M, 1D0, A, LDA, A, LDA, 0D0, ATA, N)
    CALL DPRINT_MPL ('A**T . A', N, N, ATA, N)
* ATB = Transpose (A) . b (DGEMV est une BLAS pour le pdt matrice . vecteur)
    CALL DGEMV ('Transpose', M, N, 1D0, A, LDA, B, 1, 0D0, ATB, 1)
    CALL DPRINT_MPL ('A**T . b', N, 1, ATB, N)
* Factorisation de Cholesky : ATA = L . Transpose (L)
* DPOTRF est une fonction LAPACK pour la méthode de Cholesky
    CALL DPOTRF ('Lower', N, ATA, N, INFO)
    IF (INFO .NE. 0) STOP 'Erreur DPOTRF'
    CALL DPRINT_MPL ('Après DPOTRF, A', N, N, ATA, N)
* Résolution L . y = ATB avec résultat dans ATB.
* DTRSV est une BLAS pour la substitution avant/arrière
    CALL DTRSV ('Lower', 'No Transpose', 'Not Unit Triangular',
    $           N, ATA, N, ATB, 1)
* Résolution Transpose (L) . x = y (résultat dans ATB)
    CALL DTRSV ('Lower', 'Transpose', 'Not Unit Triangular',
    $           N, ATA, N, ATB, 1)
    CALL DPRINT_MPL ('solution', N, 1, ATB, N)
    END SUBROUTINE

    SUBROUTINE FACT_QR_SOLVE (M, N, A, LDA, B)
* Les paramètres formels (LDA = Leading Dimension of A)
    INTEGER M, N, LDA
    DOUBLE PRECISION A(LDA,*), B(*)
* Variables locales
    DOUBLE PRECISION TAU(N), WORK(N), QTB(M)
    INTEGER INFO
*
    WRITE (*,*) 'Résolution par la factorisation QR'
* Factorisation A = Q . R
* DGEQRF est une fonction LAPACK pour la factorisation QR (Householder)
    CALL DGEQRF (M, N, A, LDA, TAU, WORK, N, INFO)

```

```

        IF (INFO .NE. 0) STOP 'Erreur DGEQRF'
        CALL DPRINT_MPL ('Après DGEQRF, A', M, N, A, LDA)
* QTB = b (DCOPY est une BLAS qui copie un vecteur dans un autre)
        CALL DCOPY (M, B, 1, QTB, 1)
* QTB = Transpose (Q) . b
* DORMQR est une fonction LAPACK pour multiplier par Q après DGEQRF
        CALL DORMQR ('Left', 'Transpose',
$           M, 1, N, A, LDA, TAU, QTB, M, WORK, N, INFO)
        IF (INFO .NE. 0) STOP 'Erreur DORMQR'
        CALL DPRINT_MPL ('QTB', N, 1, QTB, N)
* Résolution R . x = QTB (résultat dans QTB)
        CALL DTRSV ('Upper', 'No Transpose', 'Not Unit Triangular',
$           N, A, LDA, QTB, 1)
        CALL DPRINT_MPL ('solution', N, 1, QTB, N)
END SUBROUTINE

```

3 Exécution

On suppose l'archive `cnum.tgz` téléchargée et décompressée dans le répertoire principal. On commence par compiler la bibliothèque `libmutils` présente dans l'archive.

```

boulrier@connell:~$ cd cnum/libmutils/
boulrier@connell:~/cnum/libmutils$ make
gfortran -Wall -fimplicit-none -g -c -o maple.o maple.f
gfortran -Wall -fimplicit-none -g -c -o mutils.o mutils.f
gfortran -Wall -fimplicit-none -g -c -o dlatmr.o dlatmr.f
gfortran -Wall -fimplicit-none -g -c -o dlatm3.o dlatm3.f
gfortran -Wall -fimplicit-none -g -c -o dlatm2.o dlatm2.f
gfortran -Wall -fimplicit-none -g -c -o dlatm1.o dlatm1.f
gfortran -Wall -fimplicit-none -g -c -o dlaran.o dlaran.f
gfortran -Wall -fimplicit-none -g -c -o dlarnd.o dlarnd.f
ar cru libmutils.a maple.o mutils.o dlatmr.o dlatm3.o dlatm2.o dlatm1.o dlaran.o dlarnd.o

```

On se déplace dans le répertoire qui contient le code FORTRAN. On le compile avec l'utilitaire `gfor`. Cet utilitaire affiche la commande de compilation qu'il exécute.

```

boulrier@connell:~/cnum/libmutils$ cd ../projet/
boulrier@connell:~/cnum/projet$ ../gfor projet.f
gfortran -Wall -fimplicit-none -L/home/boulrier/cnum/libmutils projet.f -lmutils -llapack -lblas

```

Le fichier `droite.mat` contient une description de la matrice A et du vecteur b de l'exemple.

```

boulrier@connell:~/cnum/projet$ cat droite.mat
A := << 1 | 1>,
      < 2 | 1>,
      <-6 | 1>,
      < 4 | 1>>;
b := <3.01, 3.99, -4.07, 6>;

```

On exécute le programme. L'entrée standard est redirigée sur le fichier `droite.mat`.

```

boulrier@connell:~/cnum/projet$ ./a.out < droite.mat
A :=
<< 1.000000000    | 1.000000000    >,
< 2.000000000    | 1.000000000    >,

```

```

    <-6.000000000    | 1.000000000    >,
    < 4.000000000    | 1.000000000    >>;
b :=
< 3.010000000      , 3.990000000      , -4.070000000      , 6.000000000      >;
Résolution par la méthode du Cdt Cholesky
A**T . A :=
<< 57.00000000    | 1.000000000    >,
    < 1.000000000    | 4.000000000    >>;
A**T . b :=
< 59.41000000      , 8.930000000      >;
Après DPOTRF, A :=
<< 7.549834435     | 0.1324532357    >,
    < 1.000000000    | 1.995609215     >>;
solution :=
< 1.007533040      , 1.980616740      >;
Résolution par la factorisation QR
Après DGEQRF, A :=
<<-7.549834435     | -0.1324532357   >,
    < 0.2339226584    | -1.995609215    >,
    < -0.7017679752   | 0.6572371959    >,
    < 0.4678453168    | 0.1721853681    >>;
QTB :=
<-7.869046733      , -3.952537019     >;
solution :=
< 1.007533040      , 1.980616740      >;

```

4 Travail de Programmation

Il s'agit de remplacer les fonctions LAPACK du fichier `projet.f` par vos propres fonctions FORTRAN. Les fonctions à reprogrammer sont, par difficulté croissante : DTRSV, DPOTRF, DGEQRF et DORMQR. Des codes FORTRAN au niveau scalaire sont disponibles dans votre cours et dans [2].

5 Exercice

Il consiste à résoudre le problème suivant [3, Chapitre 4, exemple 6.2]. Pour étudier le phénomène de la thermo-électricité, on fait l'expérience suivante. On soude un fil de cuivre avec un fil de constantan de manière à obtenir une boucle fermée. Un point de soudure est maintenu à température fixe ($T_0 \simeq 24$ degrés Celsius) alors que l'on fait varier la température de l'autre. Ceci génère une tension U , qu'on mesure en fonction de T . On suppose que cette dépendance obéit à la loi :

$$U = a + bT + cT^2.$$

On cherche les paramètres a, b, c , à partir des mesures données dans le tableau ci-dessous :

i	T	U	i	T	U	i	T	U
1	0	-0.89	8	35	0.42	15	70	1.88
2	5	-0.69	9	40	0.61	16	75	2.10
3	10	-0.53	10	45	0.82	17	80	2.31
4	15	-0.34	11	50	1.03	18	85	2.54
5	20	-0.15	12	55	1.22	19	90	2.78
6	25	0.02	13	60	1.45	20	95	3.00
7	30	0.20	14	65	1.68	21	100	3.22

6 Questions Ouvertes

Les graphiques donnés sur les pages suivantes sont extraits de [1] et ont probablement été tracés à la main. Les méthodes étudiées dans ce projet permettent-elles de les reproduire de façon satisfaisante ?

Références

- [1] L. C. Birch and H. G. Andrewartha. *The Distribution and Abundance of Animals*. The University of Chicago Press, 1954.
- [2] François Boulier. Calcul Numérique. Algèbre Linéaire. Technical report, Université Lille 1, 59655, Villeneuve d'Ascq, France, 2013. Support de cours de GIS3, accessible sur <http://cristal.univ-lille.fr/~boulier>.
- [3] Ernst Hairer. Polycopié du cours “Analyse Numérique”. Accessible sur <http://www.unige.ch/~hairer>, octobre 2001.

the average numbers in the area might be lower. One might expect similar phenotypical differences to occur between populations from different climatic zones, but this may be complicated by the existence of genetic differences in such populations. Hart (1952) compared the lethal temperatures of freshwater fish from different localities in North America. He studied 14 species. Geographic differences in the upper lethal temperature were found only in three species—*Notropis cornutus*, *Gambusia affinis*, and *Micropterus salmoides*, each

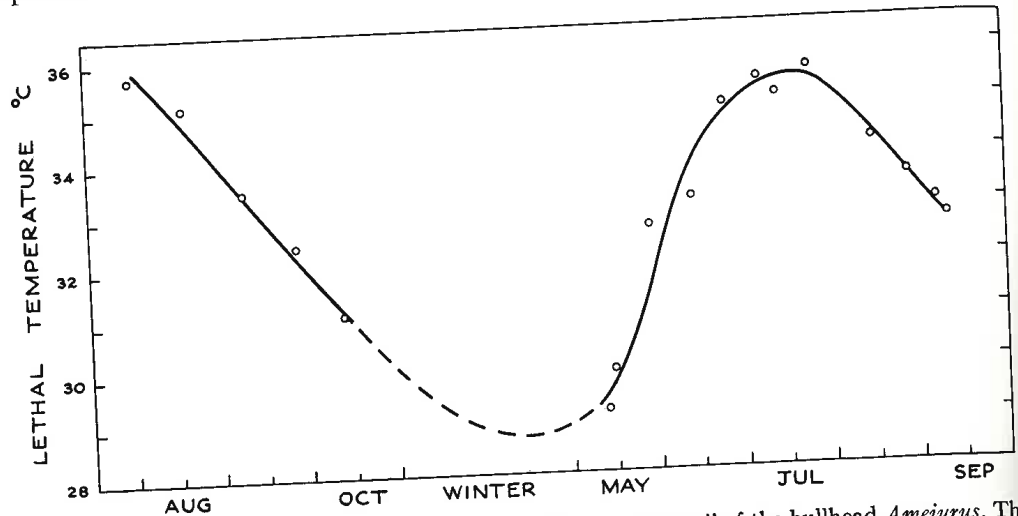


FIG. 6.30.—Illustrating the seasonal trend in the "heat resistance" of the bullhead *Ameiurus*. The ordinate gives the lowest temperature at which just half the sample of fish were dead after an exposure of 12 hours. The fish have clearly become acclimatized to the prevailing temperature of each season. (After Brett, 1944.)

from southern Canada and Tennessee. In each case he was able to find morphological differences between the populations from the different regions, and he recognized them as subspecies. Hart suggested that the similarity in the upper lethal temperatures for the other 11 species may have been due to their great capacities for acclimatization; or perhaps it was due to the relatively low temperatures prevailing in the places where these fishes lived. Since none of them ran any risk of being killed by high temperature, there was little likelihood of any of them being selected for this character.

6.323 DELAY IN THE MANIFESTATION OF THE LETHAL INFLUENCE OF HIGH TEMPERATURE

We mentioned in section 6.31 that it is often not possible to tell merely by looking at an animal the precise moment when it has had a fatal "dose" of high or low temperature. Sometimes the harm that has been done cannot be detected until much later in the life-cycle. This is well illustrated by Darby and Kapp's (1933) experiments with maggots of the fruit fly *Anastrepha*. In nature the maggots of *Anastrepha* live surrounded by the juicy contents of the fruit, so it is possible to measure the influence of high temperature on survival.

found to be correlated with temperature, the correlations with evaporation and saturation deficit will be of the same order. This follows from the simple rule in statistics which states that if two (or more) variables x and z are closely correlated positively and if a third variable y is correlated with x , then the correlation between y and z will be of the same order and sign as that between y and x . This does not preclude the possibility of a relationship between y and z which is independent of that between y and z and vice versa; but more compli-

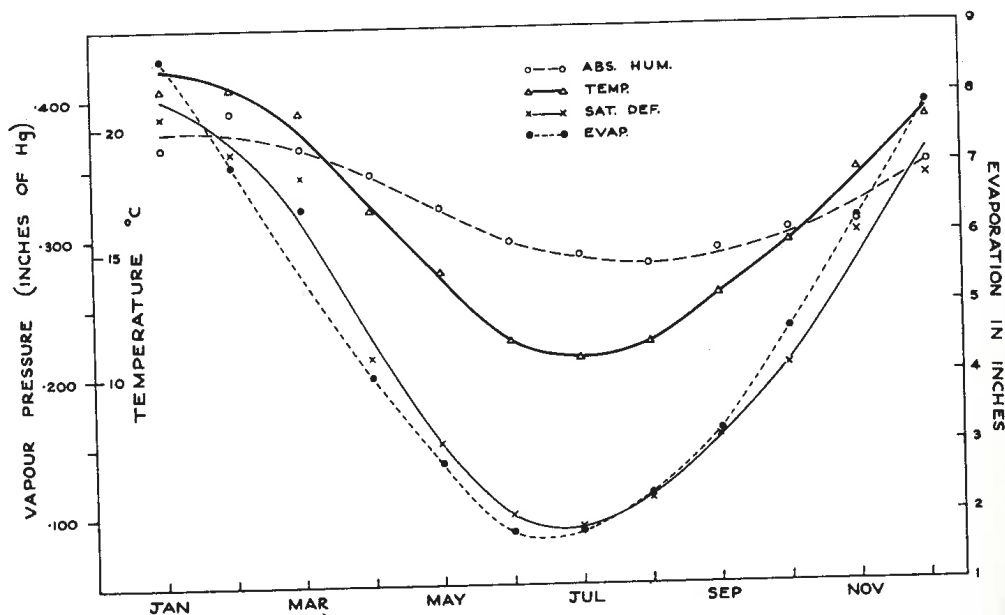


FIG. 6.04.—Showing the secular trends of atmospheric temperature and humidity at the Waite Institute, Adelaide. The curves are based on mean daily records for 23 years. Note that changes in absolute humidity are slight and gradual compared to the more abrupt changes in saturation deficit and evaporation, which are closely related to temperature as well as to absolute humidity.

cated techniques are needed to evaluate this. The student who wishes to assimilate the "philosophy" behind this argument should read the section on *partial regression* in a good modern textbook on statistical methods. The immediate practical outcome is that causal relationships may be overlooked by concentrating on an obvious simple correlation; but the descriptive value of such calculations may remain unimpaired. Field observations are often related to temperature because it is more easily measured: Clark (1947a) got an adequate estimate of the temperature associated with the behavior of locusts by simply placing a thermometer with a blackened bulb among the insects in the situations that they occupied. And Wellington (1949a) painted the bimetallic strip of a thermograph black and exposed it in the foliage of spruce trees. Such calculations are valuable for descriptive purposes, provided that we remember that in this method temperature is being allowed to "speak for" humidity and any other variable in the environment which is correlated with temperature

production had fallen from an initial rate of more than 8 per day to about 3 per day. On the 101st day the insects were placed at 18° C. and kept at this temperature for 8 days. On being returned to 27° C. on the 109th day, the insects began producing eggs more rapidly, quickly reaching a peak of about 5 eggs per day, which was nearly double the rate for the period just before they had been placed at low temperature. At a constant temperature of 27° C. the rate began to decline again but required about 70 days to decrease to the origi-

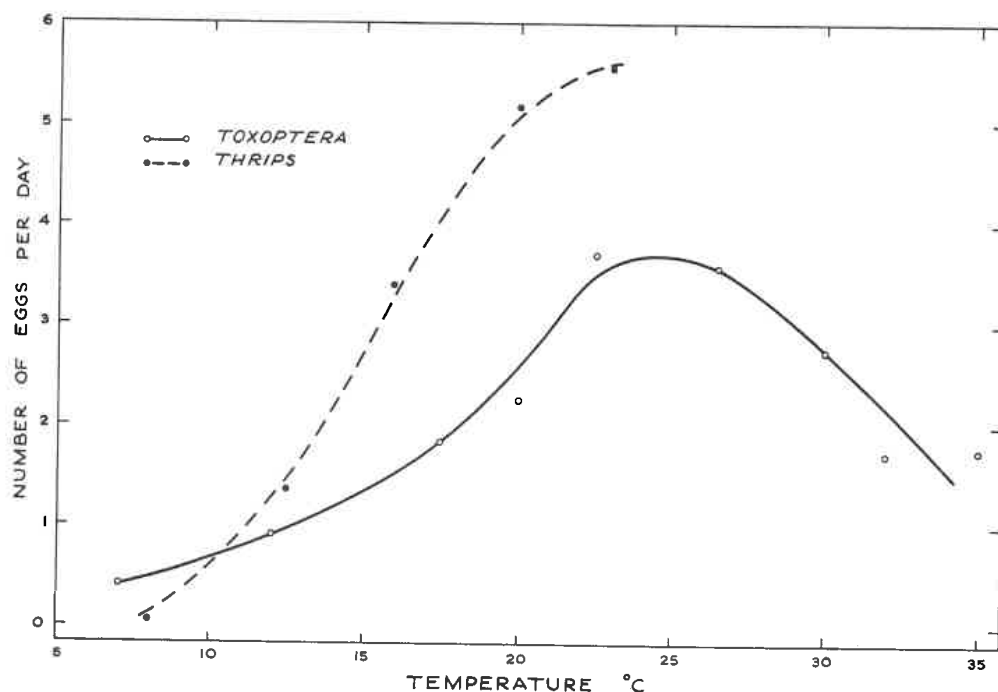


FIG. 6.20—Showing the relationship between temperature and speed of egg-production for *Thrips imaginis* and *Toxoptera graminum*. (Data from Wadley, 1931; and Andrewartha, 1935.)

nal figure of 3 eggs per day. There can be no doubt that in this case a short exposure to 18° after the insects had been living at 27° for 100 days caused a marked increase in the rate of egg-production at 27° and that this increase was sustained for at least 60 days.

During a 10-year study of the European cornborer *Pyrausta nubilalis*, in the field in Ontario, Stirret (1938) accumulated a great volume of quantitative data on the flight activity, fecundity, and oviposition-rate of this species. The results of this study of a natural population confirmed the results of the laboratory experiments discussed in this section. The fecundity was estimated by counting the numbers of eggs laid each night on a sample plot of 100 plants of corn, and at the same time the number of moths flying in the plot was estimated independently. The average number of eggs per moth varied from 2.5 in 1928 to 15.2 in 1930. There was an apparent relationship with temperature,

in moist air; the time required to kill 50 per cent was the same for all three lots. Robinson (1928) tried to acclimatize adults of *Calandra* by exposing them to a series of descending temperatures: 72 hours at 10° , 65 hours at 7.2° , 48 hours at 4.4° , and 36 hours at 1.6° C. These beetles were then exposed to -1.1° alongside controls which came to the low temperature directly from 23° . The time required to kill about 50 per cent of the controls was 90 hours, compared to 28 hours for those that had been "acclimatized." This result might have been

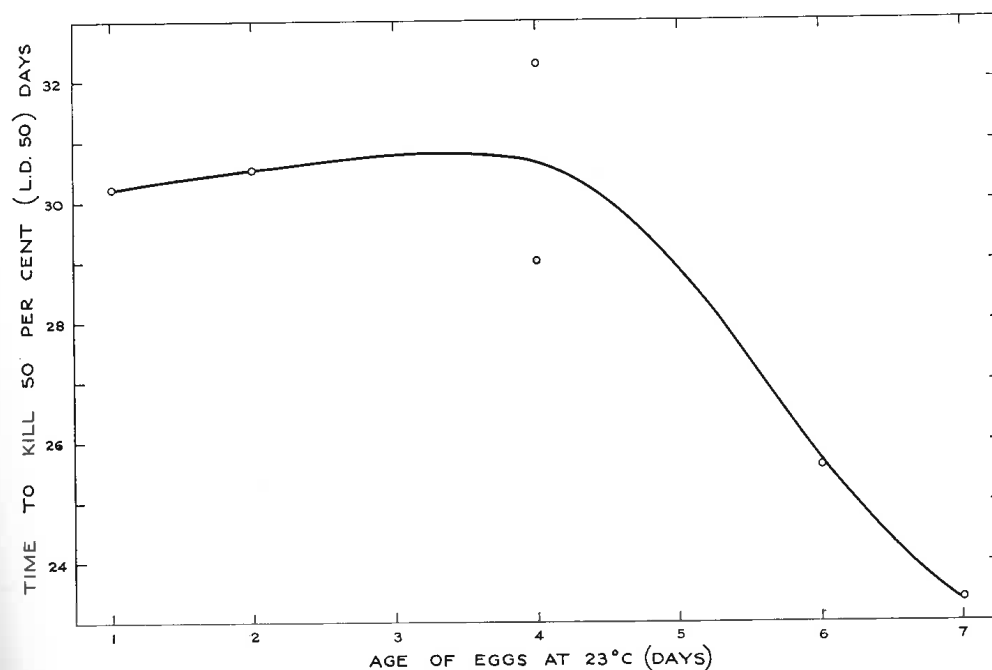


FIG. 6.24.—Showing the way in which the stage of the developing embryo influenced the survival time for eggs of *Cimex* held at 7.7° C. and 90 per cent relative humidity. (Data from Johnson, 1940.)

anticipated, for the "acclimatization" temperatures were below the incipient lethal low temperature for *Calandra*. During their so-called "acclimatization" the beetles were actually experiencing lethal temperatures and were accumulating a sublethal dose of cold; they died in a shorter time at -1.1° C. because they had already been "part-killed." This experiment is often quoted as proof that *Calandra* (and perhaps other similar animals) cannot be acclimatized. The temperatures at which it would be reasonable to acclimatize *Calandra* and other animals lacking the capacity for dormancy are those which lie just inside the favorable zone for development. So far as we know, this experiment has not been done, and we can not yet say whether this sort of animal does or does not possess the capacity to become acclimatized.

Most of the common insect pests of stored products are usually considered to have originated in tropical or subtropical zones. All those that have been studied appear to lack the capacity to become dormant in any stage of the life-

of exposure. In determining the influence of temperature on rate of development, the influence of moisture is therefore most likely to be kept constant throughout the range of different temperatures, when the conditions are such that the animals lose the same amount of water during the experiments at each temperature. Birch (1944a) designed an experiment with eggs of *Calandra oryzae* to test this hypothesis. The interaction between temperature and

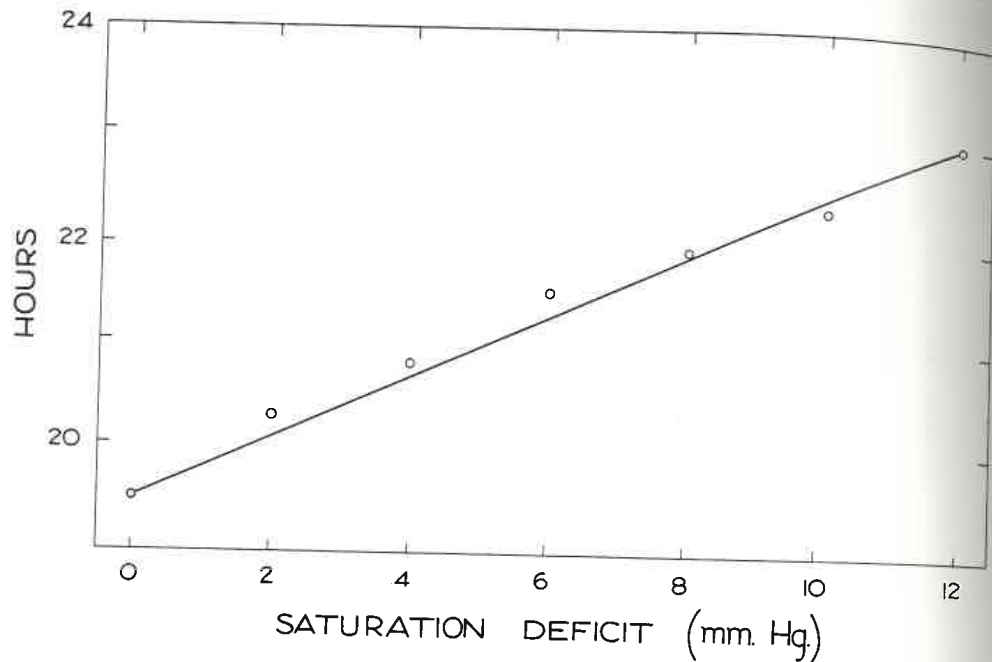


FIG. 7.41.—The relation between saturation deficit and the time taken for eggs of *Lucilia sericata* to hatch at 27° C. (After Evans, 1934.)

“moisture” (defined in this special way) was small, and he concluded that the hypothesis had been confirmed for the eggs of *C. oryzae*.

7.7 SUMMARY

We have seen that animals may respond to moisture in a way which takes them to places where the moisture gives them a good chance to survive and multiply. Some animals move away from moisture to dry places, while others move from dryness to wetter places. But whether they are in dry or wet places, animals are still faced with the necessity for maintaining the water content of their bodies at a relatively constant level. When their water content falls below a certain minimal amount, they die. There is also some evidence, though it is less well substantiated, that animals also die when their water content increases above “normal,” as it may in very wet places. There is thus a zone of lethal wetness and a zone of lethal dryness, with a tolerable zone of favorable