NCAR/TN-451+STR
NCAR TECHNICAL NOTE

April 2002

RELEASE NOTES FOR PIKAIA 1.2

Paul Charbonneau

HIGH ALTITUDE OBSERVATORY

NATIONAL CENTER FOR ATMOSPHERIC RESEARCH
BOULDER, COLORADO

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

The theory of punctuated equilibrium, a variation on conventional Darwinian evolution theory, states that biological species typically remain genetically (and thus phenotypically) stable for extended periods of (geological) time, and that evolutionary significant change and speciation occur in short "bursts" of evolutionary activity.

I happen to believe that that punctuated equilibrium also adequately characterizes the evolution of computing software.

Version 1.0 of the genetic algorithm-based optimization subroutine `PIKAIA` was released in December 1995, and except for a few minor fixes in the following few months, has remained the only version publicly available. Seven years in the computing software world is arguably equivalent to a full geological era in the physical world, so a punctuated change should be expected. Sure enough, a new software version, `PIKAIA 1.2`, is now ready for public release. Version 1.2 retains the overall format and Numerical Recipes "style" of version 1.0. It only differs internally, in that it incorporates additional genetic operators and algorithmic strategies not included in the original release of `PIKAIA 1.0`.

I am not releasing version 1.2 just for the sake of abiding to my own computing software version of punctuated equilibrium; experience garnered since 1995 indicated that the performance of the original `PIKAIA` could be improved significantly, with relatively minor and user-transparent changes and additions. This is of course the true motivation behind the release of `PIKAIA 1.2`.

These Release Notes are organized as follows: Chapter 1 is really all that a current `PIKAIA` user needs to know to move on to version 1.2. It describes the new operators and strategies included in `PIKAIA 1.2`, their coding implementation, and the new control variable settings that allows these new options to be used. Chapter 2 is my attempt to convince current `PIKAIA` users that switching to `PIKAIA 1.2` is worth the (small) effort. It is a performance study of various algorithmic versions of `PIKAIA`, tested again a suite of maximization problems I have been developing and using over the years whenever I experimented with new strategies and operators. The third and final chapter discusses three topics that I thought could be of interest to some current or potential `PIKAIA` users out there, although they have nothing really to do with `PIKAIA 1.2` specifically: combinatorial optimization with `PIKAIA`; a parallel implementation of `PIKAIA`; and `PIKAIA` in computing languages other than plain old dumb tough FORTRAN-77.

I wish to thank all of my colleagues, at NCAR and elsewhere and too numerous to list here, whose questions and queries over `PIKAIA` and genetic algorithms continue to sharpen my thinking on the topic. I do wish to thank explicitly Adrian Webster, Richard Boivin, Luciano Mantegazza, and Andreas Bobinger, for sending in bug reports following the original release of `PIKAIA`; Steve Tomczyk for promptly picking out and reporting a number of typos in the original `PIKAIA` User's Guide; Sarah Gibson, Scott McIntosh, and Alan Miller, for kindly making publicly available their versions of `PIKAIA` in computing languages other than FORTRAN-77.

Finally, I take this opportunity to thank once again my friend and former colleague Barry Knapp, now at the University of Colorado at Boulder, without whom `PIKAIA 1.0` would have never been the user-friendly software that so many people evidently have found it to be. `PIKAIA` is as much his brainchild as it is mine, and I hope that in preparing `PIKAIA 1.2` I have managed to live up to the good programming standards he tried so hard to imprint upon my thick skull back then in the previous millennium.

Paul Charbonneau

April 2002, Boulder

1. INTRODUCING PIKAIA, VERSION 1.2

This chapter contains everything needed by a user of `PIKAIA 1.0` to get going with the new version `PIKAIA 1.2`. It assumes that the reader is reasonably familiar with the basic ideas underlying a Genetic Algorithm (GA), as well as with the operation and overall structure of the `PIKAIA` subroutine. Whenever in doubt, consult the User's Guide to `PIKAIA 1.0` (Charbonneau & Knapp 1995, NCAR Technical Note 418-IA; hereafter "PUG"). For good, general introductions to genetic algorithms, see Goldberg (1989), Davis (1991), Bäck (1996), or Mitchell (1996); and for `PIKAIA`-specific introductions, Charbonneau (1995, 2002).

1.1 `PIKAIA` then and now

`PIKAIA` is a general-purpose genetic algorithm-based numerical optimization subroutine, written in ANSI-standard FORTRAN-77. Version 1.0 was released in December 1995. It's main selling point, as compared to other genetic algorithm packages available commercially or in the public domain, was and remains its ease of use and "Numerical-Recipes" style. At the time of its public release, `PIKAIA 1.0` was viewed by its designers primarily as a learning tool, although it had been used successfully already on a number of research applications. As of March 2002, an informal and undoubtedly incomplete tally reveals 71 past or present users, and 30 papers published in refereed journals where `PIKAIA` was used as part of the work. Perhaps because `PIKAIA` was first described in the pages of *The Astrophysical Journal*, the majority of research applications remain in the space, solar and astrophysical sciences, although it has also been used in a number of applications in atomic physics, engineering design, geoseismic inversion, biophysical modeling, aerodynamics, acoustics, and game theory.

In the course of the past seven years it has become clear that `PIKAIA 1.0` suffers from a few simple shortcomings that can easily be remedied. Starting already in 1998, a few users reporting convergence problems were supplied with a private-release "improved" version of the original subroutine. Recurrent surfacing of these problems, and additional intermittent twiddling and testing of `PIKAIA` "improvements" by the present author, has made it clear that relatively minor modifications and additions to `PIKAIA 1.0` could lead to significantly better performance. Thus evolved `PIKAIA 1.2`.

1.2 Description of new features

`PIKAIA 1.2` retains the overall structure, I/O, and calling sequence of version `1.0`. Internally, it includes three new features: two-point crossover, creep mutation and distance-based adjustment of the mutation rate. The following gives brief descriptions of these features, along with the occasional difficulties encountered with `PIKAIA 1.0` that motivated their inclusions in this new software release. Full listings for new code elements can be found in the Appendix.

1.2.1 Two-point crossover

The uniform one-point crossover included in `PIKAIA 1.0` (see PUG, §3.6), suffers from "end-point" bias; consider the following strings resulting from the decimal encoding of a set of floating-point parameter values defining the "model" being optimized:

```
123456789012345678901234567890123
```

Suppose now, for the sake of the argument, that the two substrings "123" at the two ends of the string, when decoded, turn out to be advantageous, in that their give their bearer above-average fitness. This advantageous combination is basically impossible to copy intact into a single offspring string following the application of uniform one-point crossover. This would be much less of a problem for the two other 123 substrings more centrally located, where the disruption probability would here be $\simeq 0.4$.

Two-point crossover bypasses this difficulty by selecting *two* splicing points along the string, and exchanging the string segments located in between these splicing points in a manner otherwise identical to one-point crossover. For example:

```
123456789012345678901234567890123
987654321098765432109876544321098
.....|..........|................


    12345|67890123456|78901234567890123
    98765|43210987654|32109876544321098


        12345|43210987654|78901234567890123
        98765|67890123456|32109876544321098


            123454321098765478901234567890123
            987656789012345632109876544321098
```

A moment of reflection will reveal that the endpoints of the string are now much more likely to end up in the same offspring string. In fact, we have overcorrected our problem: string endpoints are now *too likely* to end up on the same offspring. The simplest way out of this dilemma is to introduce an additional probabilistic test to the crossover operation, whereby either one-point or two-point crossover is chosen, with equal probability. `PIKAIA 1.2` does precisely this, as shown in the following fragment of the new crossover subroutine `pcross`:

```
      ispl=int(urand()*n*nd)+1      [ first crossover point         ]
      if (urand().lt.0.5) then      [ pick 1-pt vs 2-pt crossover   ]
         ispl2=n*nd                 [ this for one-point crossover  ]
      else
         ispl2=int(urand()*n*nd)+1  [ second crossover point        ]
         if (ispl2.lt.ispl) then    [ ensure islp<=ispl2            ]
            tmp=ispl2
            ispl2=ispl
            ispl=tmp
         endif
      endif
      do 1 i=ispl,ispl2             [ now swap from ispl to ispl2   ]
         t=gn2(i)
         gn2(i)=gn1(i)
         gn1(i)=t
    1 continue
```

Note that two-point crossover is *not* included in most of the "improved" versions of `PIKAIA 1.0` privately distributed since 1998.

1.2.2 Creep mutation

In conjunction with the decimal encoding/decoding scheme, the uniform one-point mutation included in `PIKAIA 1.0` (see PUG, §3.7.1), is liable to getting stuck at "Hamming Walls". This problem was already noted in the context of the non-linear least-squares fit discussed in the example chapter of the PUG (§5.3), and most often occurs in later evolutionary phases, when a population of pretty good trial solutions is being refined primarily by the action of mutation.

Consider the portion of a string `...4251...` encoding a floating-point parameter $a = 0.4251$. Suppose now that the optimal setting for this parameter happens to be $a^* = 0.4110$. Suppose also that the fitness landscape is smooth enough that the fitness increases gradually as a^* is approached from above or below. Evidently, many possible sequences of single-digit replacements can transform the current substring to its "target" optimum. For example,

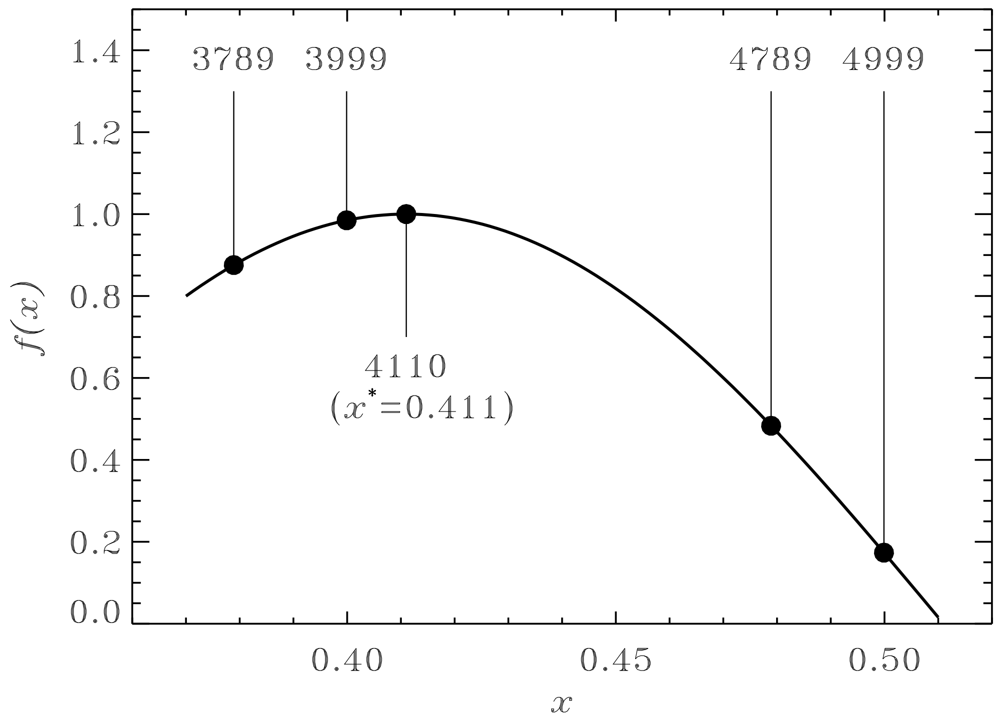

Figure 1: A Hamming Wall in a decimal encoding scheme. One-point mutation cannot move an decimally-encoded parameter value $x = 0.3789$ to the optimum $x^* = 0.4110$ along an evolutionary favored sequence of intermediate, i.e., a sequence of gradually increasing fitness $f(x)$ (see text).

`4251` $\rightarrow$ `4221` $\rightarrow$ `4121` $\rightarrow$ `4111` $\rightarrow$ `4110`
`..1.` $\rightarrow$ `...0`

Because each intermediate step represents an increase in fitness, the above transition is evolutionary favored. But what if the starting string had been `...3789...`? The sequence

`3789` $\rightarrow$ `4789` $\rightarrow$ `4189` $\rightarrow$ `4119`, etc
`..1.`

while possible in principle, will *not* be evolutionary favorable. The problem, illustrated on Figure 1, is simply that the second-step individual with $a = 0.4789$ has much smaller fitness than the original $a = 0.3789$. On the other hand, the following sequence would be evolutionary favorable:

`3789` $\rightarrow$ `3799` $\rightarrow$ `3899` $\rightarrow$ `3999`

```
.9..
```

and now we're stuck! As can be seen on Figure 1, the one-point mutation transition `3999` $\rightarrow$ `4999` is even less evolutionary advantageous than our earlier possibility `3789` $\rightarrow$ `4789`. At least two well-coordinated and simultaneous mutations must occur for the transition to take place with a corresponding increase in fitness, e.g., something like

`3999` $\rightarrow$ `4099`

Because such a chance event is very unlikely to take place in a timely manner, the population will cluster at $a = 0.3999$, without further improvement in fitness. By all appearances, the solution has converged... but unfortunately, not on the global optimum; just flat against a Hamming Wall located in its vicinity.

There are two ways out of this quandary. The first is to devise an encoding scheme where successive single-digit changes at the level of the string map into smooth variations of the decoded floating-point parameter across its whole range. Gray binary coding (see, e.g., Press et al. 1992, §20.2) is a well-known instance of this approach. However, this can rapidly become cumbersome as the string length increases. Alternately, one can simply devise mutation operators that can "carry over the one".

The creep mutation operator included in `PIKAIA 1.2` achieves this in the following way. Once a digit has been targeted for mutation, the corresponding digit is either *incremented* or *decremented* (with equal probabilities). In the case of a unit increment, if the digit happened to be a "`9`", it becomes "`0`" and that located to its left is also increment by unity (and the whole process repeats as many time as needed if the left neighbouring digit also happened to be a "`9`"). The corresponding sequences of digit operation associated with unit decrement should now be obvious to anyone having survived first-grade arithmetics.

In the case of the above example, once stuck at `3999` an evolutionary favored transition could be produced by creep mutation as follows:

`3999` $\rightarrow$ `3909` $\rightarrow$ `3009` $\rightarrow$ `4009`, etc

```
..+.
```

Note that the above sequence of steps represents the action of a single creep mutation event beginning at the third digit of the substring. In `PIKAIA 1.2`, creep mutation is coded up as follows in subroutine `mutate`:

```
   do 1 i=1,n                             [loop over substrings     ]
    do 2 j=1,nd                           [loop over digits         ]
     if (urand().lt.pmut) then            [mutation probability test]
      loc=(i-1)*nd+j                      [location on string       ]
      inc=nint ( urand() )*2-1            [pick mutation increment  ]
      ist=(i-1)*nd+1                      [substring left boundary  ]
      gn(loc)=gn(loc)+inc                 [creep mutation           ]
      if(inc.lt.0.and.gn(loc).lt.0)then[creep decrement (-1) case]
       if(j.eq.1)then                     [at substring boundary:   ]
        gn(loc)=0                         [no creep allowed!        ]
       else
        do 3 k=loc,ist+1,-1               [creep back up substring  ]
         gn(k)=9                          [turn 0 to 9              ]
         gn(k-1)=gn(k-1)-1                [carry over the one       ]
         if(gn(k-1).ge.0)goto 4           [exit creep process       ]
3       continue
        if( gn(ist).lt.0.)then            [we popped under 0.0000...]
         do 5 l=ist,loc                   [...so we fix it up       ]
          gn(l)=0
5        continue
        endif
4       continue                          [done with decrement creep]
       endif
      endif
      if(inc.gt.0.and.gn(loc).gt.9)then[creep increment (+1) case]
       if(j.eq.1)then                     [at substring boundary:   ]
        gn(loc)=9                         [no creep allowed!        ]
       else
        do 6 k=loc,ist+1,-1               [creep back down substring]
         gn(k)=0                          [turn 9 to 0              ]
         gn(k-1)=gn(k-1)+1                [carry over the one       ]
         if( gn(k-1).le.9 )goto 7         [exit creep process       ]
6       continue
        if( gn(ist).gt.9 )then            [we popped over 9.9999....]
         do 8 l=ist,loc                   [...so we fix it up       ]
          gn(l)=9
8        continue
        endif
7       continue                          [done with increment creep]
       endif
```

```
        endif
       endif
    2   continue                                   [done with this digit     ]
    1 continue                                     [done with this substring ]
```

The coding ends up looking a tad intricate, because ANSI-standard FORTRAN-77 does not have recursion capabilities. As you examine the code, notice how extra precautions are needed to prevent the operator from creeping across from one parameter-defining substring to another: creep mutation should only operate within a substring that decodes to a single parameter, otherwise other parameters might get corrupted by "runaway creep". The large number of short `DO` loops and `IF` statements might also raise concerns regarding execution time; this is a good place to remind ourselves that in most real research GA applications, most of the CPU time is spent evaluating the fitness function, so extra work within the GA usually has little consequence on total CPU or wallclock execution time.

Used in conjunction with strong selection pressure, creep mutation operates as a local hill climber: by design, it always generates small steps away from a current good solution (the largest jump generated in parameter space is caused by a ± 1 change in a string digit decoding into a leading digit in the corresponding floating-point parameter). An automatic consequence is then that creep mutation cannot cause large jumps in parameter space, the way one-point mutation does when it happens to operate on a digit that decodes into a leading digit in the decoded parameter. Because such large jumps are useful for the exploration of parameter space, it turns out to be advantageous to use either one-point or creep mutation, with equal probabilities. This is achieved by an additional probability test upon entering subroutine `mutate` (see full code listing in the Appendix).

1.2.3 Distance-based adjustment of the mutation rate

`PIKAIA 1.0` includes a very basic form of self-adaptation of the mutation probability p_m ($\equiv$`pmut=ctrl(6)`). At the end of each generational iteration, the degree of clustering of the population is computed using as a measure the normalized fitness difference between the best and median individuals (according to fitness-based rank; see PUG, §3.7.2):

$$\Delta_f = \frac{f^{\rm max} - f^{\rm med}}{f^{\rm max} + f^{\rm med}} \tag{1}$$

where $f^{\rm max} \equiv f(x^{\rm max})$ and $f^{\rm med} \equiv f(x^{\rm med})$, and $x^{\rm max}$, $x^{\rm med}$ are the parameter sets defining the best and median individuals, respectively. When $\Delta_f \to 0$, the population is strongly clustered, while if $\Delta_f \to 1$ the population is scattered across parameter space. The idea is to increase p_m in the former case, and decrease it in the latter. Doing so helps preventing premature convergence, since high p_m

favors displacement away from secondary extrema, while ensuring that efficient exploration of parameter space by the crossover operator (requiring $p_m \ll 1$ to avoid excessive disruption of the building blocks being assembled by crossover) takes place whenever the population is dispersed in parameter space.

Evidently, eq. (1) is just one of many possible measures of population clustering. Another obvious possibility would be to use a measure of metric distance between best and median, for example:

$$\Delta_d = \frac{1}{n}\left(\sum_{j=1}^{n}(x^{\max} - x^{\text{med}})^2\right)^{1/2} , \tag{2}$$

where n is the number of parameters in a solution vector $\mathbf{x}$. This latter measure of population clustering turns out to be preferable to eq. (1) in certain cases[1]. Figure 2 illustrates the idea. A fitness-based criterion such as eq. (1) behaves as wanted for high-contrast fitness landscapes (panels A and B), but in a low-contrast case (panel C) indicates clustering ($\Delta_f \to 0$) when the best and median individuals are in fact far apart from one another. A distance-based criterion such as eq. (2) behaves appropriately in this latter case, but in high-contrast situation can indicate clustering prematurely ($\Delta_d \to 0$, as on panel B).

`PIKAIA 1.2` now includes the possibility of using either a fitness-based (eq. (1)) or distance-based (eq. (2)) criterion to monitor population clustering. The choice is made by appropriate setting of the mutation mode control variable `imut`$\equiv$`ctrl(5)`, a detailed in Table II. Unless one knows for a fact that a problem is characterized by low-fitness contrast across the parameter space, fitness-based mutation rate adjustment should be tried first[2]

Independently of which adjustment criterion is invoked, `PIKAIA 1.2` varies the

---

[1] The 1/`n` normalization factor in eq. (2) means that Δ_d is not a true measure of metric distance. However, numerical experiments I carried out with the suite of test problems described in the next chapter indicate that such a normalization factor yields a better mutation rate adjustment.

[2] One can of course design criteria that incorporate both fitness-based and distance-based clustering measures. For example, one could deem the population clustered only IF Δ_f AND Δ_d are both small, and spread out IF Δ_f OR Δ_d are large. I played a bit with criteria of this type, and surprisingly (to me anyway), on many of the test problems described in chapter 3 below they often *degraded* performance as compared to either criterion used in isolation.

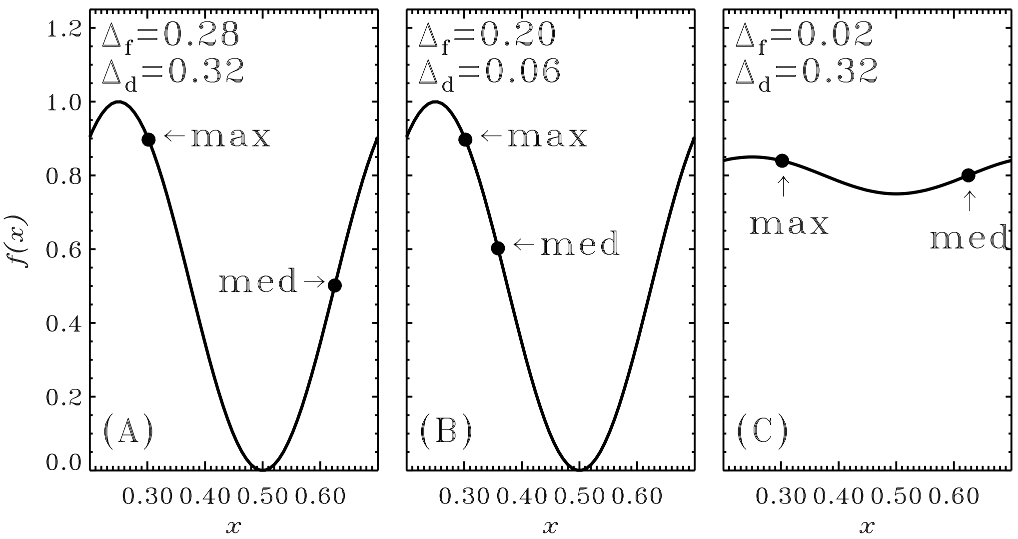

Figure 2: Two measures of population clustering in three distinct cases. In high fitness contrast situations (panel B), a fitness-based criterion is preferable, while in low-contrast cases (panel C) a distance-based criterion is better behaved. There are of course many situations where either criterion behaves more-or-less equally well (panel A).

mutation rate according to the "recipe" used in `PIKAIA 1.0` (see PUG, §3.7.2):

$$p_m \rightarrow \begin{cases} p_m \times \delta & \text{if } \Delta \leq \Delta^{\text{low}}, \\ p_m & \text{if } \Delta^{\text{low}} < \Delta < \Delta^{\text{high}}, \\ p_m/\delta & \text{if } \Delta \geq \Delta^{\text{high}}, \end{cases} \tag{3}$$

where $\delta = 1.5$, $\Delta^{\text{low}} = 0.05$, $\Delta^{\text{high}} = 0.25$ are hardwired via a `PARAMETER` statement in subroutine `adjmut`, and Δ stands for either Δ_f or Δ_d, as the case may be. Again as in `PIKAIA 1.0`, the absolute lower and upper bounds allowed on the mutation probability p_m are set by the control variables `pmutmn`≡`ctrl(7)` and `pmutmx`≡`ctrl(8)` (see PUG, §4.5). For implementation details, see the source code for the new subroutine `adjmut` listed in the Appendix of these Release Notes.

1.2.4 New control parameter settings for `PIKAIA 1.2`

To facilitate upward compatibility, the new strategies and operators described above have been incorporated into `PIKAIA 1.2` in a manner such that the calling sequence and input control vector (`ctrl`) length of `PIKAIA 1.0` remained unchanged.

A 50/50 mixture of one-point and two-point crossover is now hardwired in `PIKAIA 1.2`. Quantitative support for this decision can be found in §3.3 below.

The crossover probability `pcross` (≡`ctrl(4)`) is then the probability per breeding event of either one-point or two-point crossover to take place. The user wishing to enforce exclusive use of one-point crossover must enforce `ispl2=n*nd` in subroutine `cross`, by un-commenting the appropriate line of source code (see full listing of the new subroutine `cross` in the Appendix).

The remaining two additional operators and strategies included in `PIKAIA 1.2`, namely creep mutation and distance-based adjustment, can be invoked by setting the value of the control variable `imut` (≡`ctrl(5)`) to the values listed in the following Table:

Table I

New valid values for control variable `imut`

`imut`	Adjustable rate	Based on	Creep	Note
1	No	N/A	No	1,2
2	Yes	Fitness	No	1,3
3	Yes	Distance	No	
4	No	N/A	Yes	2
5	Yes	Fitness	Yes	
6	Yes	Distance	Yes	

Notes:
1= Only allowed settings in `PIKAIA 1.0`
2= Setting not recommended; see main text.
3= Default value in both `PIKAIA 1.0` and `PIKAIA 1.2`

All other control variable settings remain the same as in `PIKAIA 1.0`, and are specified by the user as elements of the 12-dimensional input control vector `ctrl` (see PUG, Table I). Note that for completeness, the possibility of using 50/50 one-point and creep mutation *without* dynamical adjustment of the mutation rate is allowed (`imut`=4). However, this setting, like pure one-point mutation without rate adjustment, yields a much less-than-optimal global algorithm on most problems. Therefore, neither `imut`=1 or 4 are recommended settings for `PIKAIA 1.2`.

1.3 Compatibility with PIKAIA 1.0

`PIKAIA 1.2` is operationally compatible with `PIKAIA 1.0`, in the sense that the new subroutine can be substituted for the old one without any changes in the source code for the calling program, user-supplied fitness function `ff(n,x)`, or settings of the control variable input array `ctrl`. In particular, all default settings in `PIKAIA 1.2` are the same as in `PIKAIA 1.0`. However, even if starting from the exact same initial random population, different evolutionary paths can and will occur, for two reasons:

(1) `PIKAIA 1.2` uses two-point crossover for 50% of the breeding events which pass the initial crossover probability test upon entering subroutine `cross`.
(2) Even if the user enforces one-point crossover (by forcing `ispl2=n*nd` in subroutine `cross`), comparison of the version `1.0` and `1.2` source codes for subroutines `cross` and `mutate` will reveal three additional calls to function `urand()`. *This implies that even if using the same random number generator seed, from one generation to the next the two sequences of random deviates will become increasingly out of sync between otherwise identical version* `1.0` *and* `1.2` *runs.* This will lead to distinct evolutionary paths to the global optimum.

However, everything else being equal `PIKAIA 1.2` should be expected to be performing as well as version `1.0` (and in fact probably better, as argued in the following chapter).

1.4 Obtaining a copy of PIKAIA 1.2

Source codes for `PIKAIA 1.2`, including the subroutine itself as well as driver programs and fitness functions for the examples discussed in chapter 5 of the PUG, can be obtained from NCAR's High Altitude Observatory via the Web or anonymous ftp. For direct Web access point your browser to

```
http://download.hao.ucar.edu/archive/pikaia
```

For download via anonymous ftp, use the address

```
ftp.hao.ucar.edu 122
```

and log in using the username `anonymous` and your e-mail address as the password (you are now on a UNIX system). Type the following command:

```
cd /archive/pikaia
```

The command `ls` can then be used to display the directory's content.

The file `README` is self-explanatory; please do read it, as it will most likely includes additional (and possibly important) information which did not make it in these release notes. The file `userguide.ps` is a standard postscript file containing the `PIKAIA 1.0` User's Guide (a.k.a. PUG in these release notes). Nearly everything in that User's Guide remains relevant to `PIKAIA 1.2`, so the novice

PIKAIA user should definitely grab a copy. The file `pikaia.f` contains not only the genetic subroutine `PIKAIA 1.2` itself, but also a driver code, fitness function and other required routines, including a random number generator. The file is a completely self-contained source code which can be used for installation check (see PUG, §2.2). The original version `PIKAIA 1.0` remains available in the file `pikaia1.0.f`. The directory `examples` contains drivers, fitness functions and synthetic datasets for the examples discussed in chapter 5 of the PUG. They can be used with either versions `1.0` or `1.2` of `PIKAIA`.

1.5 The PIKAIA Web Page

The `PIKAIA` Web page has been around since 1996, and can be found at the (lengthy) URL:

`http://www.hao.ucar.edu/public/research/si/pikaia/pikaia.html`

It is the place to look for Bug Reports, tutorial material, lists of known past and present users, and lists of `PIKAIA`-related publications. And of course, links are set up there to the most up-to-date, corrected versions of the `PIKAIA` User's Guide, the present Release Notes, and source code for all the software. You can also see there what a real *Pikaia* actually looks like.

1.6 Reference and credits

The GA-based optimization subroutine `PIKAIA` was first described in the following paper, published in *The Astrophysical Journal (Supplements)* in December 1995:

Charbonneau, P. 1995, ApJS, 101, 309.

and described in greater details in the `PIKAIA` User's Guide:

Charbonneau, P., & Knapp, B. 1996, *A User's Guide to PIKAIA 1.0*, NCAR Technical Note 418+IA (Boulder: National Center for Atmospheric Research)

The above two publications remain the primary reference for both `PIKAIA`. The present document shall serve as the "official" specific reference to version 1.2:

Charbonneau, P. 2002, *Release Notes for PIKAIA 1.2*, NCAR Technical Note 451+STR (Boulder: National Center for Atmospheric Research)

Both versions of `PIKAIA` remain public domain software, and, consequently, no restrictions are imposed on further distribution for research and education purposes.

2. A PERFORMANCE STUDY

The seasoned `PIKAIA` user satisfied with version `1.0` is entitled to ask: why bother switching to a new version? This is a very legitimate question[3], and this chapter makes an attempt at answering it.

2.1 Defining performance

In the computer science literature on search algorithms (which includes GAs), a distinction is often made between "off-line" and "on-line" performance. In the context of a GA, the "off-line" performance is the fitness of the very best solution found in the course of the whole evolutionary run. The "on-line" performance is a measure of how quick a "good enough" solution can be found. These two measures of performance are obviously related, and each can be more or less relevant depending on the problem under consideration. In the present case of global, multimodal numerical optimization using a GA, it is most important to focus on an aspect of off-line performance sometimes dubbed "global performance" or "success rate": given N runs of the GA on the same problem but starting with different initial populations, how many will succeed in finding the global maximum?

2.2 A suite of test functions

Readers with even just a little experience in global optimization will be well aware of the fact that global performance is highly problem-dependent. Consequently, the first thing that needs to be done is to assemble a suite of test problems that cover a broad range of optimization (mis)behaviors. Here all such test problems will take the form of maximization problems for functions of many (real) variables, i.e., finding the D-dimensional parameter set $\mathbf{x}^*$ such that the function

$$f(\mathbf{x}) \ , \qquad \mathbf{x} \equiv \{x_1, x_2, ..., x_D\}$$

is maximized. Following Bäck (1996), preference will be given to functions that are readily generalizable to arbitrary large D. In addition, we will require that

$$\max(f) \equiv f(x^*) = 1.0 \ , \qquad x_j \in [0.0, 1.0] \ , \quad j = 1, ..., D.$$

[3] I personally rank software upgrades very high on my most-hated list, up there with lawnmowing, suppertime telemarketers, and my neighbour's barking rat.

The first test function, plotted in its $D = 2$ version on Figure 3A, is a simple, smooth unimodal function defined as a Gaussian centered in the middle of the spatial domain:

$$f_1(\mathbf{x}; D) = \exp(-r^2/\sigma^2), \tag{4a}$$

$$r^2 = \sum_{j=1}^{D} (x_j - 0.5)^2 \ , \tag{4b}$$

with $\sigma^2 = 0.15$. The maximum is at $x_j^* = 0.5\,\forall j$. This smooth, unimodal function is a very easy maximization problem, on which almost any local hill-climbing technique would be far more efficient than PIKAIA. It is nonetheless useful to have such a function in the test suite, in order to be able to test PIKAIA's local hill-climbing capabilities under various setting of its internal control parameters.

The second test function, plotted on Figure 3B again for $D = 2$, is unimodal like f_1, but step-wise discontinuous:

$$f_2(\mathbf{x}; D) = \frac{1}{D} \sum_{j=1}^{D} \frac{\mathrm{int}(10x_j - \epsilon)}{9} \ , \tag{5}$$

with $\epsilon = 10^{-6}$. The maximum is at $x_j \geq 0.9\,\forall j$. This is a variation on one of the test functions in the famous test suite developed by DeJong (1975) in his pioneering study of GA performance. This function is a useful test on two grounds: (1) search for a corner maximum is hard for the crossover operator; (2) only relatively large mutation can lead to fitness improvements. For small D, this function is easy for the GA, even though it would readily defeat most local gradient-based hill-climbing schemes, including the usually robust and pseudo-global downhill simplex method (see, e.g., Press *et al.* 1992, §10.4). For large D, it becomes pretty hard even for a GA.

The third test function is the D-dimensional generalization of the multimodal 2D function distributed with PIKAIA as installation check. A plot of its $D = 2$ version is shown on Fig. 3C. Its D-dimensional form is defined as:

$$f_3(\mathbf{x}; D) = \cos^2(9\pi r) \exp(-r^2/\sigma^2), \tag{6a}$$

$$r^2 = \sum_{j=1}^{D} (x_j - 0.5)^2 \ , \tag{6b}$$

with again $\sigma^2 = 0.15$. The maximum is at $x_j = 0.5\,\forall j$, but is surrounded by concentric rings of progressively higher secondary maxima centered on the central

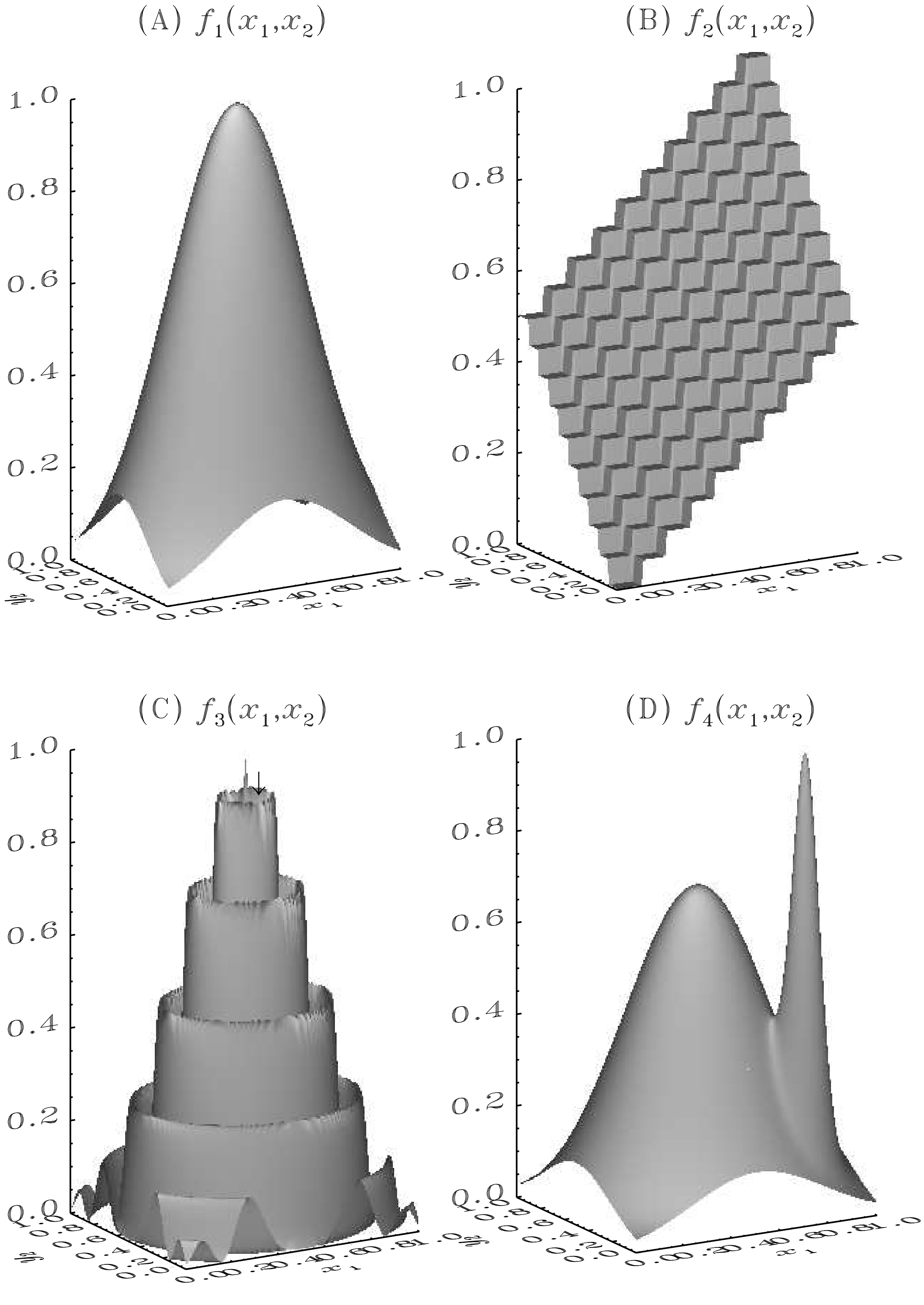

Figure 3: A suite of test functions. The plots show the two dimensional ($D = 2$) versions of the four functions that make up `PIKAIA`'s test suite. The fourth test function has been rotated by 90° about the vertical, to facilitate viewing.

maximum. This is a very hard test function by any standards, and it rapidly gets a lot harder as D increases.

The fourth and final test function is the sum of two D-dimensional Gaussians, as shown on Figure 3D (for $D = 2$ as usual). It's general form is:

$$f_4(\mathbf{x}; D) = A_1 \exp(-r_1^2/\sigma_1^2) + A_2 \exp(-r_2^2/\sigma_2^2) \ , \tag{7a}$$

$$r_1^2 = \sum_{j=1}^{D} (x_j - 0.5)^2 \ , \qquad r_2^2 = \sum_{j=1}^{D} (x_j - 0.2)^2 \ , \tag{7b}$$

with

$$A_1 = 0.7, \qquad A_2 = 1 - 0.7 \exp(-r_1^2/\sigma_1^2) \ , \tag{7c}$$

and $\sigma_1^2 = 0.15$, $\sigma_2^2 = 0.005$. This is a *very* hard problem, and increasingly so the larger D gets. Almost invariably, early in the evolutionary run the central, broad lower Gaussian draws towards it most trial solutions. The challenge is then for the prematurely converged population to find and move onto the taller, narrower Gaussian which is the true, global maximum.

2.3 Experimental design

The idea here is to examine the improvement (or degradation) in performance associated with the new operators and strategies included in `PIKAIA 1.2`, namely distance-based mutation rate adjustment, two-point crossover, and creep mutation. This is in fact impossible to establish in a robust, problem-independent manner, since the settings of the various GA control variables interact with each other in complex ways, which results in very non-linear effects on performance. In other words, examining performance as one control parameter at a time is varied is useful, but care is warranted in making grand sweeping conclusions on the basis of such experiments. With this important caveat firmly in mind, we proceed nonetheless.

We will test six distinct implementations of `PIKAIA 1.2`, labeled very originally PK1 through PK6. All implementations use a population size `np`=`ctrl(1)`= 50, 5-digit encoding `nd`=`ctrl(3)`= 5, crossover probability `pcross`=`ctrl(4)`= 0.85, full-generational-replacement `irep`=`ctrl(10)`= 1, elitism `ielite`=`ctrl(11)`= 1, full selection pressure `fdif`=`ctrl(9)`= 1.0, and variable mutation rate in the range $0.0005 \leq p_m \leq 0.25$, initially set at $p_m = 0.005$ (as specified by `ctrl(6)`, `ctrl(7)`, and `ctrl(8)`). The six implementations differ in whether or not they use two-point crossover, creep mutation, and whether they adjust the mutation rate according to the fitness-based or distance-based criterion (eq. (1) or (2) herein). Note that in this context PK1 corresponds to `PIKAIA 1.0` operating under its the default settings, except for population size.

2.4 Off-line performance Results

To get interesting and meaningful numbers in a reasonable amount of CPU time, we'll pick $D = 10$ and 15 for f_1 and f_2, but $D = 3$ and 4 for the much harder f_3 and f_4. Likewise, we'll only run out to 100 generations with the first two, but up to 2500 generation with the latter pair. To get representative statistics, each `PIKAIA` implementation listed in Table II is run 100 times, each time with a different initial random seed, and thus with a distinct initial random population.

The results are listed in Table II. The top part of the Table details which operators and adjustment strategy are used in which of the six `PIKAIA` implementation. Remember that a "yes" on two-point crossover or creep mutation means that these are used 50% of the time, with one-point crossover and uniform one-point mutation used for the other 50%. For the unimodal f_1 and f_2, the numerical entries correspond to the 100-run average of the best individual found at the end of each run. On the other hand, for f_3 and f_4 the entries give the success rate, e.g., 58/100 means that the global maximum was found in 58 out of 100 independent trials. This is by far the most relevant performance measure for these hard multimodal functions.

Table II

Off-line Performance results

	PK1	PK2	PK3	PK4	PK5	PK6
2-pt. cross.	No	Yes	No	Yes	Yes	Yes
Creep	No	No	No	No	Yes	Yes
p_m-adj.	fitn.	fitn.	dist.	dist.	fitn.	dist.
$f_1(D=10)$	0.9921	0.9929	0.8558	0.8717	0.9893	0.9232
$f_1(D=15)$	0.9668	0.9714	0.5118	0.5401	0.9656	0.7408
$f_2(D=10)$	0.9906	0.9912	0.9929	0.9949	0.9983	0.9984
$f_2(D=15)$	0.9586	0.9629	0.9542	0.9544	0.9813	0.9822
$f_3(D=3)$	43/100	44/100	64/100	66/100	91/100	100/100
$f_3(D=4)$	37/100	25/100	7/100	10/100	27/100	27/100
$f_4(D=3)$	93/100	93/100	70/100	100/100	67/100	58/100
$f_4(D=4)$	2/100	3/100	14/100	14/100	2/100	2/100

There is a lot of information, trends, and pseudo-trends that can be extracted from Table II. The most obvious trend is that performance rapidly degrades as problem dimensionality (D) is increased (for constant population size and number of generations). Perhaps the most noteworthy other general observation is that:

(1) Two-point crossover often helps, and very rarely hurts (cf. PK1 vs PK2, and PK3 vs PK4, on all f's)

Other trends in Table II are harder to interpret, or just plain confusing or apparently inconsistent. As a small selection, consider that:

(2) Creep mutation helps a lot on some problems (PK2 vs PK5 and PK3 vs PK6 on f_3), but degrades it on others (cf. f_4)
(3) Creep mutation degrades performance on f_1 when used with fitness-based mutation rate adjustment (PK2 vs PK5), but improves it when used with distance-based adjustment (PK4 vs PK6).
(4) PK3 did worst on $f_3(D = 4)$, but best (ex aequo with PK4) on $f_4(D = 4)$
(5) On f_3, PK1 did worst at $D = 3$, but degraded the least —and did the best— at $D = 4$.

The sources of these apparent inconsistencies are of course that (1) performance is highly problem-dependent, and (2) The GA's internal strategies and operator interact in very complex, "nonlinear" ways.

It is on the basis of these —and other— numerical experiments that the decision was made to hardwire 50% two-point crossover in `PIKAIA 1.2`, but to leave the choice of creep mutation and distance- vs fitness-based mutation rate adjustment in the hand of the user, through the setting of the control variable `imut` (≡`ctrl(5)`), as described in Table I herein.

2.5 On-line performance Results

On-line performance, i.e., how fast is an "acceptable" solution found, is mostly an issue in applications relating to real-time control, and is rarely critical in the types of numerical global optimization modeling applications that most `PIKAIA` users are likely to find themselves involved with. Nonetheless, a quick look at on-line performance on our suite of test functions is useful to appreciate various behavioral differences in the six P̂IKAIA versions used to generate the off-line performance results of table II.

Figure 4 depicts average convergence curves for two of the test problems of Table II, namely $f_1(D = 15)$ and $f_2(D = 10)$. Each curve corresponds to the average of 100 "best fitness vs generation" convergence curves, and so can be expected to be statistically representative.

Consider first Figure 4A, for the unimodal f_1 test function. Initially (first 15 generations, say), all `PIKAIA` versions show more or less the same rapid increase

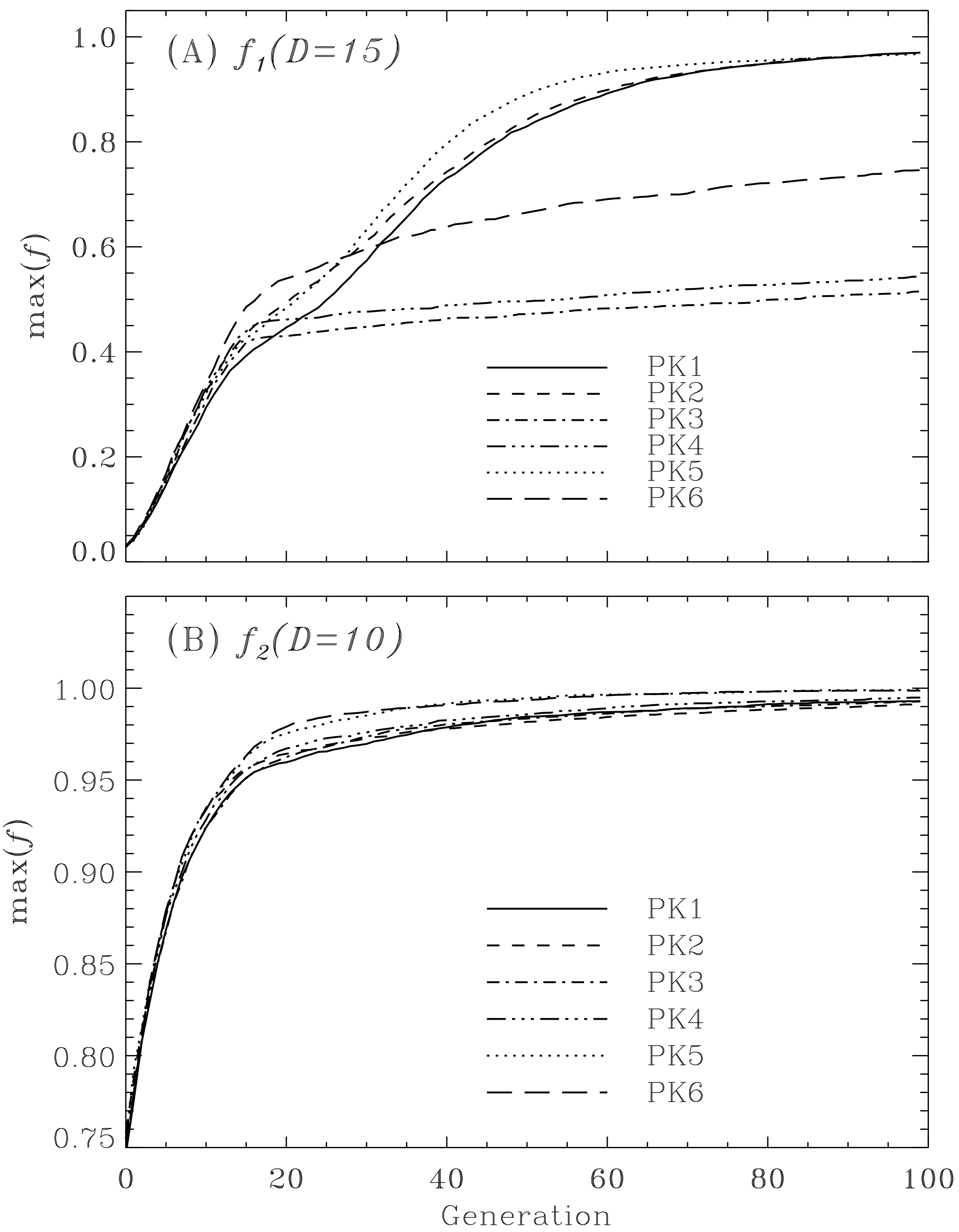

Figure 4: On-line performance results on $f_1(D = 15)$ and $f_2(D = 10)$. Note the different vertical scales on each panel.

in fitness, which is here primarily due to the exploratory action of the crossover operator. Then a marked divergence can be seen between the algorithms using fitness-based adjustment of the mutation probability (PK1, PK2, and PK5) and those using distance-based adjustment (PK2, PK4, and PK6). This reflects the fact that f_1 is a sharply-peaked function, for which a fitness-based criterion (eq. (1)) provides a measure of population clustering superior to that computed using a distance-based criterion. Note how the use of creep mutation (PK2 vs PK5) makes little difference in the former case, but a significant one in the latter (PK4 vs PK6). Here the use of two-point crossover seems to make little difference in either case (cf. PK1 vs PK2, and PK3 vs PK4).

Results for test function f_2, shown on Fig. 4B, show a totally different pattern, with now PK5 and PK6 coming out slightly ahead, and all four other implementations doing more-or-less the same. This indicates a total lack of dependence on the criterion used to measure population clustering and control the mutation rate. This is a natural consequence of the fact that the unimodal function f_2 is globally linear in each parameter (see Fig. 3B), so that there exists a linear relationship between fitness difference and metric distance. The superior on-line performance of PK5 and PK6 indicates the enhanced local hill-climbing capabilities of creep mutation, over uniform one-point mutation.

On-line results for the multimodal f_3 and f_4 (not shown) show no obvious trends. All versions do about equally well over the first 50–100 generations, following which the convergence curves mostly reflect the success rate at locating the true extrema. On $f_3(D = 3)$, PK4 does marginally better early on, with PK5 and PK6 taking a slight lead from generation 100 onward, and ultimately achieving the best off-line performance (see Table II).

3. SELECTED SPECIAL TOPICS

This chapter discusses three special topics that, although not specifically related to `PIKAIA 1.2`, might be of interest to the `PIKAIA` user community.

3.1 `PIKAIA` for combinatorial optimization

This section describes the use of `PIKAIA` for a combinatorial optimization problem. It is based on a chapter of the PhD thesis of Scott McIntosh, then a graduate student at the University of Glasgow, Scotland.

The physical problem Scott was working on involved inferring physical properties of the solar outer atmosphere, using measurements of the intensities I_m of various spectral lines observed by space-borne solar telescopes (see McIntosh *et al.* 2000). Mathematically, the problem is defined by an integral equation of the form

$$I_m = \int_T K_m(T)\xi(T)\mathrm{d}T \ , \qquad m = 1, ..., M \ . \tag{8}$$

Here the quantity $\xi(T)$, known as the *differential emission measure*, is the "source function" to be inverted from the M observed lines and is assumed to depend only on the temperature T of the solar atmospheric plasma. The Kernel $K_m(T)$ is a known quantity involving a lot of atomic physics, as well as properties of the observing instrument. Discretizing K_m and ξ on a temperature mesh of J discretization points, and introducing a suitable quadrature formula to evaluate the integral, turns eq. (8) into the matrix system:

$$[K]\{\xi\} = \{I\} \ , \tag{9}$$

where the matrix $[K]$ is of size $M \times J$. In practice, even if one chooses $J = M$ the linear system described by eq. (9) is ill-conditioned. This occurs because of the smoothing properties of the integral operator in eq. (8), and is compounded by the fact that the Kernels K_m are not localized enough in T to avoid overlap (see Craig & Brown 1986 for an accessible introduction to these matters).

The technique of choice for solving eq. (9) is of course Singular Value Decomposition (see Press *et al.* 1992 for a lucid, no-nonsense introduction). However, Scott was interested in a method that would allow him to simply select a subset of lines I_m that would yield the least ill-conditioned version of eq. (9), while allowing additional, physically-based constraints to be imposed on line choices. The reduction in dataset size was also appealing to him, because of the massive amount of data produced by current solar spectral imaging instruments (one spectrum per pixel at a high temporal cadence, for over 10^6 pixels!). Already familiar with `PIKAIA` from a prior modeling study (McIntosh *et al.* 1998), Scott decided to modify it to tackle his line selection problem.

3.1.1 Modifying PIKAIA's encoding scheme

What Scott was facing is a combinatorial optimization problem than involves identifying the subset of N *distinct* labels from a search list of M ($> N$) possible choices, so as to minimize some computable quantity, in his case the Condition Number of the matrix $[K]$ in eq. (9). An important simplifying aspect of this combinatorial problem is that the *ordering* of selected elements within the optimal subset is unimportant, unlike, e.g., in the Travelling Salesman problem.

The only non-trivial modification to `PIKAIA` involved the encoding scheme. Instead of `PIKAIA`'s floating point array `x(1:n)`, for Scott's problem an individual is defined as a list of N integers (the labels) with values $1 \leq n \leq M$ defining which of the M available spectral lines measurements are to be used to carry out the inversion. The user-supplied fitness function `ff(n,x)` must then accept as input one such integer array, and compute the Condition Number of the resulting matrix system (now of size $N \times J$) given by eq. (9). The primary difficulty was to design an encoding scheme that always ensures that the action of crossover and mutation produces offspring lists of labels that always contain *distinct* entries.

3.1.2 Standard Ordinal Representation

The construction of a list of distinct labels from a reference list with possible non-distinct entries is readily carried out using the *ordinal representation* scheme (Michalewicz 1996, chap. 10). The task is to extract N distinct labels from a master list $\mathbf{S}$ of M possible values, where $S(m) = m$, $m = 1, ..., M$. An *ordinal vector* $\mathbf{e}$ is made up of N elements $e(n)$ with values in the n-dependent range $1 \leq e(n) \leq M - n + 1$. The corresponding *label vector* $\mathbf{E}$ is constructed according to the following iterative procedure:

do $n = 1, N$
  $E(n) := S(e(n))$
  do $k = n, M - n - 1$
    $S(k) := S(k + 1)$
  enddo
enddo

Note that the second algorithmic step has the consequence that $S(e(n))$ is removed from the list, and the list size is reduced by one at each iteration. This scheme has some attractive characteristics. It is quite simple to implement, and having the $e(n)$'s uniformly distributed in their allowed bounds results in a uniform distribution of $E(n)$'s. However, one can easily verify that when pairs of $\mathbf{e}$'s are acted upon by the uniform one-point crossover operator, the $e(n)$'s located right of the splicing point can decode into $E(n)$'s not originally coded by the parent $\mathbf{e}$'s. Consider for example a situation where $N = 6$ distinct elements must be extracted from a reference list of $M = 20$ possible values. The following two ordinal vectors:

$$\mathbf{e}_1 = (4, 5, 1, 16, 11, 5) \tag{10a}$$

$$\mathbf{e}_2 = (3, 12, 10, 1, 6, 13) \tag{10b}$$

decode into the two lists

$$\mathbf{E}_1 = (4, 6, 1, 19, 14, 8) \tag{11a}$$

$$\mathbf{E}_2 = (3, 13, 11, 1, 8, 18) \tag{11b}$$

The six iterative steps needed to convert $\mathbf{e}_1$ to $\mathbf{E}_1$ are detailed on Figure 5. Now carry out crossover of the two ordinal vectors (10a), (10b), with a splicing point at the third element, yielding the offspring ordinal vectors:

$$\mathbf{e}_3 = (4, 5, 10, 1, 6, 13) \tag{12a}$$

$$\mathbf{e}_4 = (3, 12, 1, 16, 11, 5) \tag{12b}$$

These decode into the two offspring lists:

$$\mathbf{E}_3 = (4, 6, 12, 1, 9, 18) \tag{13a}$$

$$\mathbf{E}_4 = (3, 13, 1, 19, 14, 7) \tag{13b}$$

Note that the two offspring lists jointly contain three labels —7,9,12— that were not present in either of the parent lists (11a), (11b). This difficulty is a direct consequence of the leftward shifting associated with the encoding procedure. This is conceptually at variance with the expected behavior of crossover, which should lead to exchange of intact subsets of the $E(n)$'s. The only tolerable exception is when crossover introduces additional duplicate entries in the pair of offspring $\mathbf{e}$'s resulting from the operation. Likewise, under ordinal representation, uniform one-point mutation can potentially alter *all* elements of $\mathbf{E}$. This made the standard ordinal representation unsuitable for Scott's application.

3.1.3 Ranked Ordinal Representation

Scott and I came up with a simple modification of standard ordinal representation, which we dubbed Ranked Ordinal Representation (ROR), and that can bypass the

$n = 1:$
$\mathbf{S} = [1, 2, 3, \mathbf{4}, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]$
$e(1) = 4 \rightarrow E(1) = 4$

$n = 2:$
$\mathbf{S} = [1, 2, 3, 5, \mathbf{6}, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]$
$e(2) = 5 \rightarrow E(2) = 6$

$n = 3:$
$\mathbf{S} = [\mathbf{1}, 2, 3, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]$
$e(3) = 1 \rightarrow E(3) = 1$

$n = 4:$
$\mathbf{S} = [2, 3, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, \mathbf{19}, 20]$
$e(4) = 16 \rightarrow E(4) = 19$

$n = 5:$
$\mathbf{S} = [2, 3, 5, 7, 8, 9, 10, 11, 12, 13, \mathbf{14}, 15, 16, 17, 18, 20]$
$e(5) = 11 \rightarrow E(5) = 14$

$n = 6:$
$\mathbf{S} = [2, 3, 5, 7, \mathbf{8}, 9, 10, 11, 12, 13, 15, 16, 17, 18, 20]$
$e(6) = 5 \rightarrow E(6) = 8$

Figure 5: Construction of a label list using standard ordinal representation, for $M = 20$ and $N = 6$. The label $E(n)$ removed from $\mathbf{S}$ at each iteration (n) is indicated in boldface.

above problem. It consists in *sorting* the ordinal vector in decreasing order, so that $e(n+1) \leq e(n)\ \forall n$ prior to applying the ordinal algorithm. Under this scheme the ordinal vector (10a) now decodes into

$$\mathbf{E}_1 = (16, 11, 5, 6, 4, 1) \tag{14}$$

Clearly, if $\mathbf{e}$ does not contain duplicate entries then $\mathbf{E} = \mathbf{e}$ (with $\mathbf{e}$ sorted). If however entry $e(n) = e(n+1)$ for some n $(< N)$, then $E(n) = e(n)$ and $E(n+1) = e(n) + 1$. This is the case here with the duplicate "5" in $\mathbf{e}_1$.

This modified scheme behaves far better than standard ordinal representation under one-point crossover. Under ROR, the two parent ordinal vectors $\mathbf{e}_1, \mathbf{e}_2$

(eqs. (10a,b) above) now decode into:

$$\mathbf{E}_1 = (16, 11, 5, 6, 4, 1) \tag{15a}$$

$$\mathbf{E}_2 = (13, 12, 10, 6, 3, 1) \tag{15b}$$

and the two offspring ordinal vectors $\mathbf{e}_3, \mathbf{e}_4$ (eqs. (12a,b) above) resulting from crossover starting at the third element then decode into:

$$\mathbf{E}_3 = (13, 10, 6, 5, 4, 1) \tag{16a}$$

$$\mathbf{E}_4 = (16, 12, 11, 5, 3, 1) \tag{16b}$$

Clearly the crossover operation has had the desired effect, namely the interchange of intact subsets of $E(n)$'s from each parent $\mathbf{E}$ (remember that element ordering within each $\mathbf{E}$ is irrelevant here)[4]. Similarly, one-point mutation has now only a local effect.

To ensure that a statistically uniform distribution of $e(n)$'s translates into a similarly uniform distribution of $E(n)$'s, it is necessary to let $1 \le e(n) \le M, \forall n$. On the other hand, the ordinal decoding algorithm requires that $S(e(n)) \le M - n + 1$ at every iterative step; to reconcile these conflicting requirement, we modify the algorithm in the following way:

do $n = 1, N$
  $E(n) := \min(\ M - n + 1, S(e(n))\ )$
  do $k = n, M - n - 1$
    $S(k) := S(k + 1)$
  enddo
enddo

This actually introduces a slight bias toward high values of $E(n)$, but for relatively small population sizes ($N_p \le 100$, say) it remains statistically insignificant as compared to the realization noise ($\propto \sqrt{N_p}$).

A more serious problem arises when crossover is repeatedly applied to *randomly* selected members of an initially randomly distributed population of $\mathbf{e}$'s, when N_p is small. Because the occurrence of duplicates $[e(n), e(n)]$ in newly generated offspring always translates into a pair $[E(n), E(n+1)]$, in the absence of

[4] Note that one of the "6" label present in the parents has been turned into a "5", a consequence of the existence of duplicate "5" in $\mathbf{e}_1$, combined with leftward shift by the ordinal algorithm.

mutation, any inhomogeneity in the original distribution of $e(n)$'s across the population (inevitable for finite population size) will tend to grow until the whole population is clustered around N allowed values for the search list.

This is illustrated in Figure 6, for the $M = 133$, $N = 30$ cases originally considered by Scott. Consider first panel A; the thick histogram shows the summed count distribution of $E(n)$'s corresponding to a population of $N_p = 50$ individuals, each with labels chosen randomly in the allowed range $[1, M]$. The dotted line shows the expected mean count, and the horizontal dashed lines the one-σ deviation range expected from Poisson statistics. As expected (assuming also a good random number generator), the counts in most bins are within the $\pm$one-σ range. The thin histogram shows the equivalent distribution having arisen after 10 genetic algorithm-like generational iterations, with random selection of parents bred without mutation but with crossover probability $p_m = 1$. Most bins are now well outside the expected one-σ range, evidence for the aforementioned clustering instability. Panel B is in the same format, but now mutation (with probability of occurrence of $p_m = 0.02$ for each $e(n)$) has been restored to the breeding step. Mutation reduces, but does not eliminate, the clustering instability.

3.2 Parallel `PIKAIA`

It is often said that genetic algorithms readily lend themselves to a parallel implementation, because fitness evaluation, usually the most computationally demanding step, can be carried out independently for each individual in the population. This may seem obvious, but to develop an actual working parallel genetic algorithm implementation is quite another matter. In this section I give an overview of one specific parallel application of `PIKAIA` which I found both elegant and impressive.

Travis Metcalfe, then a graduate student working on his PhD thesis at the University of Texas at Austin, was facing a tough modelling problem in stellar structure theory: learning something about the physics of white dwarf interiors from astronomical observations of their pulsation frequencies (for a great introduction to the physics of white dwarf, see Fontaine *et al.* 2001). White dwarfs are the ultimate enpoints of stellar evolution for the vast majority of stars, including our Sun. They have exhausted all their nuclear fuel, and gradually cool and fade to black by radiating their thermal energy content into space. Their interior structure is a historical record of their life as thermonuclear furnaces. Travis was trying to obtain a best-fit to pulsation frequencies of one particularly well-observed white dwarfs, matriculated GD358. To do so he had to (1) pick a set of global stellar parameters such as mass, internal chemical composition, etc.; (2) evolve a stellar model from the beginning of the white dwarf phase to the observed surface tem-

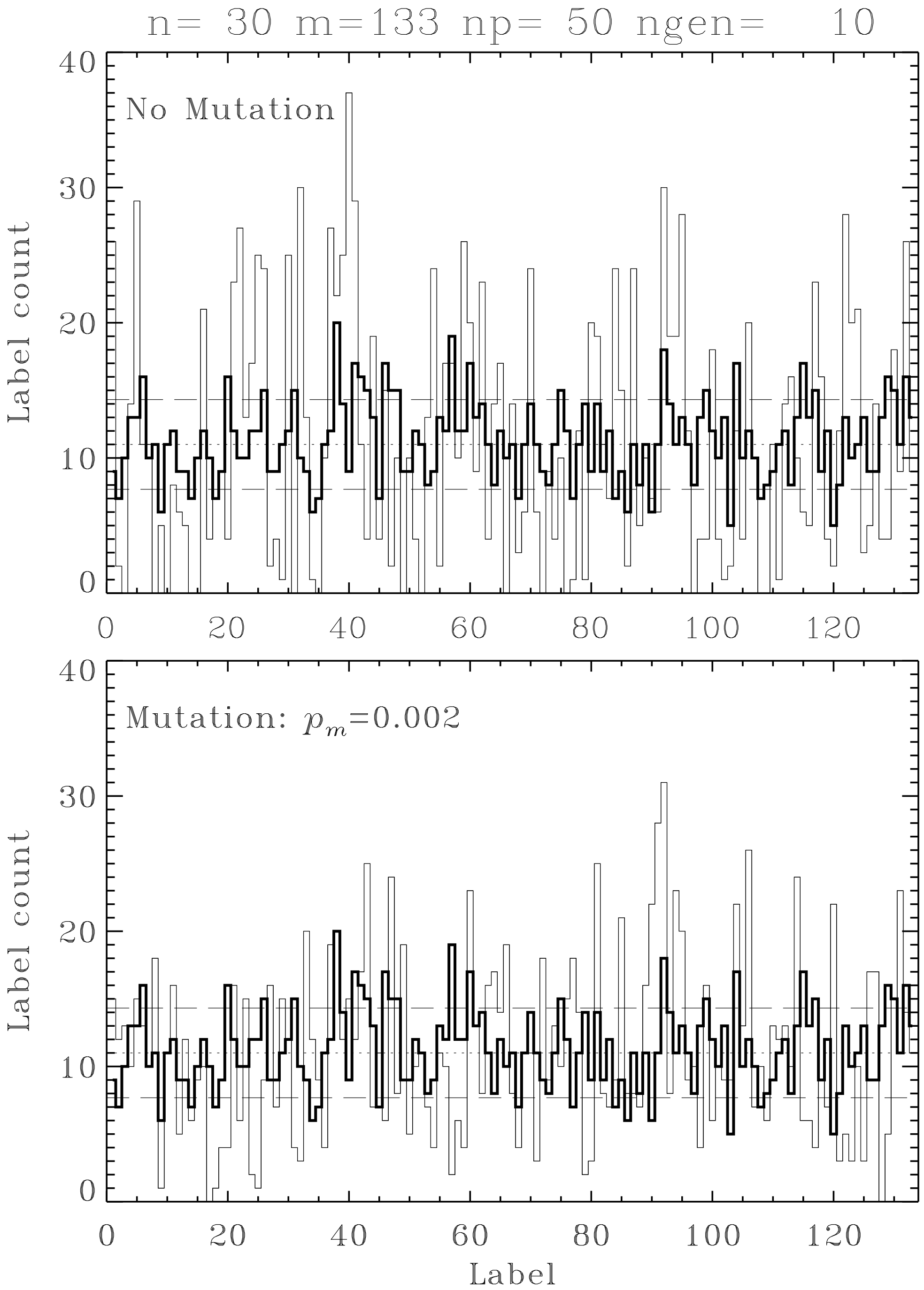

Figure 6: Clustering instability for ROR without selection pressure. The thick-line histograms show the frequency distribution of selected lines in the initial random population, and the thin histograms that having arisen after application of the crossover operator with random selection for ten generations, with (panel B) and without (panel A) use of the mutation operator. The horizontal dotted line is the expected mean count for purely random selection, with the dashed lines indicating the one-σ range expected from Poisson counting statistics.

perature of GD358; (3) solve an eigenvalue problem for the pulsation frequency of his cooled model; and finally (4) compute the root-mean-square residual between these theoretical frequencies and the observed ones. Finding the set of stellar parameters that minimizes this residual defined an optimization problem. For a given parameter set, the above sequence of modeling steps added up to about a minute of CPU time on a typical fast workstation processor. Faced with a vast and likely multimodal parameter space, it rapidly became clear to Travis that he needed an efficient global optimization scheme that could be parallelized. He wisely opted for genetic algorithms, and settled on `PIKAIA` as a specific software.

Travis' next judicious move was to decide that if he was going to go parallel, he might as well go all the way. He managed to obtain no less than 64 minimal "slave" PCs, all of which he connected to a "master" PC via a local network operating under LINUX (see Metcalfe & Nather 1999). With executables of the stellar evolution and pulsation codes as well as observed frequency data copied onto each of the slave PC's, Travis ran a slightly modified version[5] of an early pre-release incarnation of `PIKAIA 1.2` on the master PC, using the higher-level PVM language to manage the communications between master and slaves. All the crucial little details are included in his PhD thesis, which is available online at

`http://ceti.as.utexas.edu/metcalfe/`

(see in particular his Chapter 3 and Appendix C). Travis went on to reap a rich scientific harvest from his customized hardware/software system (see Metcalfe *et al.* 2000; Metcalfe *et al.* 2001; Metcalfe & Charbonneau 2002). From a computational point of view, one truly remarkable thing was the speedup factor achieved using his fully distributed hardware and software design. Figure 7 shows a scalability plot going from one to 64 slave processors. The speedup factor reaches 52.7 at 64 processors. And perhaps even more remarkably, the scaling curve shows little sign of leveling off. And this is for a fixed problem size, i.e., the population size is $N_p = 128$ in all cases. Scaling up a parallel software very rarely looks that good. Figure 7 stands as a testimony of the truly robust parallelization properties of genetic algorithms.

---

[5] This version included distance-based mutation rate adjustment and creep mutation. The former led to significant performance improvement for Travis' modeling problem, but the latter did not. Here distance-based mutation rate adjustment helped a lot because Travis was dealing with a low contrast parameter space; his best solutions did not have a fitness value (inverse residual) differing a whole lot from moderately good or even so-so solutions. The difference was however physically meaningful, given the high accuracy of the pulsation frequencies observed in white dwarfs.

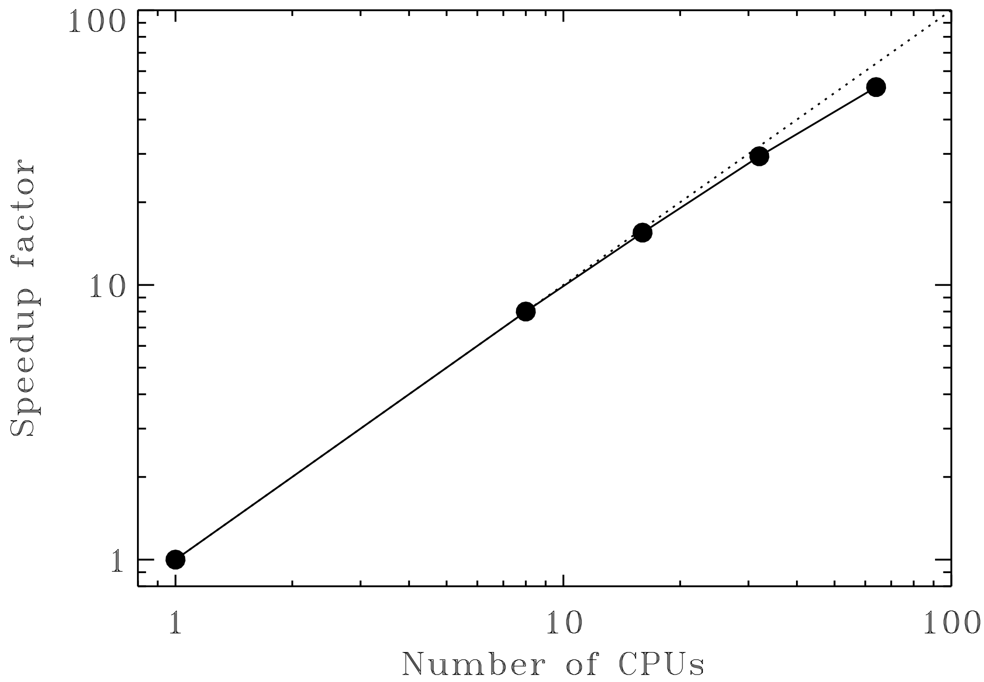

Figure 7: Speedup factor for parallel PIKAIA. The curve shows the speedup factor (inverse wallclock time) achieved by Travis' parallel implementation of PIKAIA on his fully distributed hardware architecture, as a function of the number of processor used in the calculation. The problem size, i.e., population size and generational length, are the same for all runs. The theoretical maximum, speedup $\propto$ number of processors, is shown by the dotted line.

As far as PIKAIA itself is concerned, Travis actually had very little to do, once he decided to restrict his reproduction plan to full-generational-replacement (`irep`$\equiv$`ctrl(10)=1`). In PIKAIA running under full-generational-replacement, at each generation the offspring solutions are bred and stored in the array `newph`, and only after the breeding loop terminates are the fitness calculated in a **do** loop within subroutine `newpop` (see PUG, §§3.2 and A.1). Travis pulled out the call to the user-supplied fitness function `ff(n,x)` from the `do 2 i=1,np` loop in `newpop`, and inserted a single call to a new subroutine that evaluate the fitnesses of population members in parallel by dispatching individual trial solutions to individual slave processors through calls to appropriate PVM routines (a similar modification was made to the `do 1 ip=1,np` loop in the main subroutine, to compute the fitnesses across the initial random population).

Travis also had to introduce various safeguards to ensure that the master processor would not wait forever for results from crashed slave processors, but would instead re-dispatch the failed job to the next available slave processor. Such nitty-gritty details are essential for a succesful parallel implementation.

3.3 `PIKAIA` in other computing languages

Many `PIKAIA` users have inquired (and continue to enquire) about the availability of `PIKAIA` in computing languages other than plain good'ol dumb never-goes-away FORTRAN-77. Some users have gone further and produced their own versions in their favorite computing languages, and kindly made them available to the general `PIKAIA` user's community.

The following is a list of people to contact and URL's to access for such `PIKAIA` versions. While I have not tested these to the same extent than I have the original FORTRAN-77 `PIKAIA`, they certainly seem to work[6].

3.3.1 PIKAIA in IDL

The Interactive Data Language (IDL) has become a very popular software, as it combines numerical computation capabilities and easy data I/O with outstanding graphics and visualization tools, all in a modular and flexible programming environment. Two IDL versions of `PIKAIA` are currently available, courtesy of Sarah Gibson and Scott McIntosh. Sarah's IDL version is a direct transliteration of the original FORTRAN-77 code, and makes no use of IDL's vector operation and string manipulation capabilities. On the application she was working on at the time, her IDL version was about ten times slower than the original FORTRAN-77 code (both running on UNIX-based Sun or SGI platforms). Her code is available in the form of an IDL procedure, on the HAO/NCAR ftp site:

`http://download.hao.ucar.edu/archive/pikaia/pikaia.pro`

Scott's version is a true re-write in IDL of the original `pikaia.f`, and he estimates it to be slower than the original by a factor of about three. It is available at:

`http://zeus.nascom.nasa.gov/~scott/ga.html`

3.3.2 PIKAIA in FORTRAN-90

Over fifteen years in the making and officially released around 1991, FORTRAN-90 is the designated successor to FORTRAN-77. Alan Miller, an Honorary Research Fellow at the CSIRO/Mathematical and Information Sciences in Australia,

---

[6] I will post links on the `PIKAIA` Web Page to similar versions of `PIKAIA 1.2` as they become available and are called to my attention.

has produced a FORTRAN-90 version of `PIKAIA 1.0`, which he kindly makes available on one of his Web sites. Point your browser to:

`http://users.bigpond.net.au/amiller`

scroll down to “Miscellaneous code” and look for “`pikaia.f90`”.

APPENDIX: SOURCE CODE

A.1 New subroutines for PIKAIA 1.2

The following are reduced listing of the three subroutines in `PIKAIA 1.2` that present significant changes from `PIKAIA 1.0`. As with all `PIKAIA 1.0` code listings included in the PUG, the listings are "reduced" in that the corresponding subroutines in `PIKAIA 1.2` many more explanatory comments lines than the listing given here; all operational lines are included as they appear in the source code; and all `REAL`, `INTEGER`, etc, statements have been regrouped independently of their local/input/output status.

```
      subroutine cross(n,nd,pcross,gn1,gn2)
      implicit none
      integer       n, nd
      real          pcross
c=====================================================================
c     breeds two parent chromosomes into two offspring chromosomes
c     Either uniform one-point or two-point crossover is used,
c     chosen anew with equal probability at each breeding event
c=====================================================================
c     USES: urand
      integer       i, ispl, ispl2, itmp, t
      real          urand
      external      urand
c     Use crossover probability to decide whether a crossover occurs
      if (urand().lt.pcross) then
c        Compute first crossover point
         ispl=int(urand()*n*nd)+1
c        Choose between one-point and two-point crossover
         if (urand().lt.0.5) then
            ispl2=n*nd
         else
            ispl2=int(urand()*n*nd)+1
c           un-comment following line to enforce one-point crossover
c           ispl2=n*nd
            if (ispl2.lt.ispl) then
               itmp=ispl2
               ispl2=ispl
               ispl=itmp
            endif
         endif
c        Swap genes from ispl to ispl2
         do 10 i=ispl,ispl2
            t=gn2(i)
            gn2(i)=gn1(i)
            gn1(i)=t
   10    continue
      endif
      return
      end
c*********************************************************************
      subroutine mutate(n,nd,pmut,gn,imut)
      implicit none
      integer       n, nd, imut, gn(n*nd)
      real          pmut
c=====================================================================
c     Generalized uniform one-point and creep mutation operator
```

```
c     Mutations occur at rate pmut at all gene loci
c        imut=1    Uniform mutation, constant rate
c        imut=2    Uniform mutation, variable rate based on fitness
c        imut=3    Uniform mutation, variable rate based on distance
c        imut=4    Uniform or creep mutation, constant rate
c        imut=5    Uniform or creep mutation, variable rate based on
c                  fitness
c        imut=6    Uniform or creep mutation, variable rate based on
c                  distance
c=====================================================================
c     USES: urand
      integer        i,j,k,l,ist,inc,loc,kk
      real           urand
      external       urand
c     Decide which type of mutation is to occur
      if(imut.ge.4.and.urand().le.0.5)then
c     CREEP MUTATION OPERATOR
c     Subject each locus to random +/- 1 increment at the rate pmut
         do 1 i=1,n
            do 2 j=1,nd
               if (urand().lt.pmut) then
c     Infer location on string, substring boundary, and pick +1 or -1
                  loc=(i-1)*nd+j
                  inc=nint ( urand() )*2-1
                  ist=(i-1)*nd+1
                  gn(loc)=gn(loc)+inc
c                 Now we carry over the one, if needed
c                 Decrement case
                  if(inc.lt.0 .and. gn(loc).lt.0)then
                     if(j.eq.1)then
                        gn(loc)=0
                     else
                        do 3 k=loc,ist+1,-1
                           gn(k)=9
                           gn(k-1)=gn(k-1)-1
                        if( gn(k-1).ge.0 )goto 4
    3                   continue
                        if( gn(ist).lt.0.)then
                           do 5 l=ist,loc
                              gn(l)=0
    5                      continue
                        endif
    4                   continue
                     endif
                  endif
c                 Increment case
```

```
                  if(inc.gt.0 .and. gn(loc).gt.9)then
                     if(j.eq.1)then
                        gn(loc)=9
                     else
                        do 6 k=loc,ist+1,-1
                           gn(k)=0
                           gn(k-1)=gn(k-1)+1
                           if( gn(k-1).le.9 )goto 7
    6                   continue
                        if( gn(ist).gt.9 )then
                           do 8 l=ist,loc
                              gn(l)=9
    8                      continue
                        endif
    7                   continue
                     endif
                  endif
               endif
    2       continue
    1    continue
      else
c     UNIFORM MUTATION OPERATOR
c     Subject each locus to random mutation at the rate pmut
         do 10 i=1,n*nd
            if (urand().lt.pmut) then
               gn(i)=int(urand()*10.)
            endif
   10    continue
      endif
      return
      end
c**********************************************************************
      subroutine adjmut(ndim,n,np,oldph,fitns,ifit,pmutmn,pmutmx,
     +                  pmut,imut)
      implicit none
      integer        n, ndim, np, ifit(np), imut
      real           oldph(ndim,np), fitns(np), pmutmn, pmutmx, pmut
c======================================================================
c     dynamical adjustment of mutation rate;
c     imut=2 or imut=5 : adjustment based on fitness differential
c                        between best and median individuals
c     imut=3 or imut=6 : adjustment based on normalized metric distance
c                        between best and median individuals
c======================================================================
      integer        i
      real           rdif, rdiflo, rdifhi, delta
```

```
      parameter      (rdiflo=0.05, rdifhi=0.25, delta=1.5)
      if(imut.eq.2.or.imut.eq.5)then
c     Adjustment based on fitness differential
         rdif=abs(fitns(ifit(np))-fitns(ifit(np/2)))/
     +           (fitns(ifit(np))+fitns(ifit(np/2)))
      else if(imut.eq.3.or.imut.eq.6)then
c     Adjustment based on normalized metric distance
         rdif=0.
         do 1 i=1,n
            rdif=rdif+( oldph(i,ifit(np))-oldph(i,ifit(np/2)) )**2
    1    continue
         rdif=sqrt( rdif ) / float(n)
      endif
      if(rdif.le.rdiflo)then
         pmut=min(pmutmx,pmut*delta)
      else if(rdif.ge.rdifhi)then
         pmut=max(pmutmn,pmut/delta)
      endif
      return
      end
```

A.2 Other minor coding changes in PIKAIA 1.2

This section described other minor coding changes made throughout PIKAIA 1.2 to accommodate the new operators and strategies described in these release notes. The only changes occur in the main program pikaia, and in subroutine setctl.

A.2.1 main program

Because subroutine mutate must know if creep mutation is to be used, and subroutine adjmut whether fitness-based or distance-based adjustment is invoked, the control variable imut is now passed through both subroutine's argument lists. Likewise, to compute metric distance subroutine adjmut must be passed the population array oldph. The corresponding calls in subroutine pikaia now look like:

```
      call mutate(n,nd,pmut,gn1,imut)
      call mutate(n,nd,pmut,gn2,imut)
                     .
                     .
                     .
      if (imut.eq.2.or.imut.eq.3.or.imut.eq.5.or.imut.eq.6)
     +   call adjmut(NMAX,n,np,oldph,fitns,ifit,pmutmn,pmutmx,
     +               pmut,imut)
```

A.2.2 subroutine setctl

The safety checks on the value of imut carried out in subroutine setctl now reflects the new valid numerical values for this control variable:

```
      if(imut.ne.1.and.imut.ne.2.and.imut.ne.3.and.imut.ne.4
     +    .and.imut.ne.5.and.imut.ne.6) then
         write(*,10)
         status = 5
      endif
   10 format(' ERROR: illegal value for imut (ctrl(5))')
```

And that is all!

BIBLIOGRAPHY

Bäck, T. 1996, *Evolutionary Algorithms in Theory and Practice* (Oxford: Oxford University Press).

Charbonneau, P. 1995, *The Astrophysical Journal Supplement*, **101**, 309

Charbonneau, P. 2002, *An introduction to genetic algorithms for numerical optimization*, NCAR Technical Note 450+IA (Boulder: National Center for Atmospheric Research)

Charbonneau, P., & Knapp, B. 1995, *A User's Guide to PIKAIA 1.0*, NCAR Technical Note 418+IA (Boulder: National Center for Atmospheric Research (PUG))

Craig, I.J.D., & Brown, J.C. 1986, *Inverse Problems in Astronomy* (Bristol, UK: Adam Hilger)

Davis, L. 1991, *Handbook of Genetic Algorithms* (New York: Van Nostrand Reinhold)

Fontaine, G., Brassard, P., & Bergeron, P. 2001, *Publication of the Astronomical Society of the Pacific*, **113**, 409

Goldberg, D.E. 1989, *Genetic Algorithms in Search, Optimization & Machine Learning* (Reading: Addison-Wesley)

McIntosh, S.W., Diver, D.A., Judge, P.G., Charbonneau, P., Ireland, J., & Brown, J.C. 1998, *Astronomy and Astrophysics Supplement*, **132**, 145

McIntosh, S.W., Charbonneau, P., & Brown, J.C. 2000, *The Astrophysical Journal*, **529**, 1115

Metcalfe, T., & Charbonneau, P. 2002, *Journal of Computational Physics*, submitted

Metcalfe, T., & Nather, R. 1999, *Linux Journal*, **65**, 58

Metcalfe, T., Nather, R., & Winget, D. 2000, *The Astrophysical Journal*, **545**, 974

Metcalfe, T., Winget, D., & Charbonneau, P. 2001, *The Astrophysical Journal*, **557**, 1021

Michalewicz, Z. 1996, *Genetic Algorithms + Data Structures = Evolution Programs*, third ed. (New York: Springer)

Press, W.H., Teukolsky, S.A., Vetterling, W.T., & Flannery, B.P. 1992, *Numerical Recipes*, Second Edition (Cambridge: Cambridge University Press)