

# Chapter 1

## Past, Present and Future Trends in the Use of Computers in Fisheries Research

Bernard A. Megrey and Erlend Moksness

*I think it's fair to say that personal computers have become the most empowering tool we've ever created. They're tools of communication, they're tools of creativity, and they can be shaped by their user.*

Bill Gates, Co-founder, Microsoft Corporation

*Long before Apple, one of our engineers came to me with the suggestion that Intel ought to build a computer for the home. And I asked him, 'What the heck would anyone want a computer for in his home?' It seemed ridiculous!*

Gordon Moore, Past President and CEO, Intel Corporation

### 1.1 Introduction

Twelve years ago in 1996, when we prepared the first edition of *Computers in Fisheries Research*, we began with the claim “The nature of scientific computing has changed dramatically over the past couple of decades”. We believe this statement remains valid even since 1996. As Heraclitus said in the 4th century B.C., “Nothing is permanent, but change!” The appearance of the personal computer in the early 1980s changed forever the landscape of computing. Today’s scientific computing environment is still changing, often at breath-taking speed.

In our earlier edition, we stated that fisheries science as a discipline was slow to adopt personal computers on a wide-scale with use being well behind that in the business world. Pre-1996, computers were scarce and it was common for more than one user to share a machine, which was usually placed in a public area. Today, in many modern fisheries laboratories, it is common for scientists to use multiple computers in their personal offices, a desktop

---

B.A. Megrey (✉)

U.S. Department of Commerce, National Oceanic and Atmospheric Administration,  
National Marine Fisheries Service; Alaska Fisheries Science Center, 7600 Sand Point  
Way NE, BIN C15700, Seattle, WA 98115, USA

personal computer and a portable laptop is often the minimum configuration. Similarly, in many lab offices, there are several computers, each dedicated to a specific computational task such as large scale simulations. We feel that because of improvements in computational performance and advances in portability and miniaturization, the use of computers and computer applications to support fisheries and resource management activities is still rapidly expanding as well as the diversity of research areas in which they are applied. The important role computers play in contemporary fisheries research is unequivocal. The trends we describe, which continue to take place throughout the world-wide fisheries research community, produce significant gains in work productivity, increase our basic understanding of natural systems, help fisheries professionals detect patterns and develop working hypotheses, provide critical tools to rationally manage scarce natural resources, increase our ability to organize, retrieve, and document data and data sources, and in general encourage clearer thinking and more thoughtful analysis of fisheries problems. One can only wonder what advances and discoveries well known theorists and fisheries luminaries such as Ludwig von Bertalanffy, and William Ricker, or Ray Beverton and Sidney Holt would have made if they had had access to a laptop computer.

The objective of this book is to provide a vehicle for fisheries professionals to keep abreast of recent and potential future developments in the application of computers in their specific area of research and to familiarize them with advances in new technology and new application areas. We hope to accomplish this by comparing where we find ourselves today compared to when the first edition was published in 1996. Hopefully, this comparison will help explain why computational tools and hardware are so important for managing our natural resources. As in the previous edition, we hope to achieve the objective by having experts from around the world present overview papers on topic areas that represent current and future trends in the application of computer technology to fisheries research. Our aim is to provide critical reviews on the latest, most significant developments in selected topic areas that are at the cutting edge of the application of computers in fisheries and their application to the conservation and management of aquatic resources. In many cases, these are the same authors who contributed to the first edition, so the decade of perspective they provide is unique and insightful.

Many of the topics in this book cover areas that were predicted in 1989 to be important in the future (Walters 1989) and continue to be at the forefront of applications that drive our science forward: image processing, stock assessment, simulation and games, and networking. The chapters that follow update these areas as well as introduce several new chapter topic areas. While we recognize the challenge of attempting to present up to date information given the rapid pace of change in computers and the long time lines for publishing books, we hope that the chapters in this book taken together, can be valuable where they suggest emerging trends and future directions that impact the role computers are likely to serve in fisheries research.

## 1.2 Hardware Advances

It is difficult not to marvel at how quickly computer technology advances. The current typical desktop or laptop computer, compared to the original monochrome 8 KB random access memory (RAM), 4 MHz 8088 microcomputer or the original Apple II, has improved several orders of magnitude in many areas.

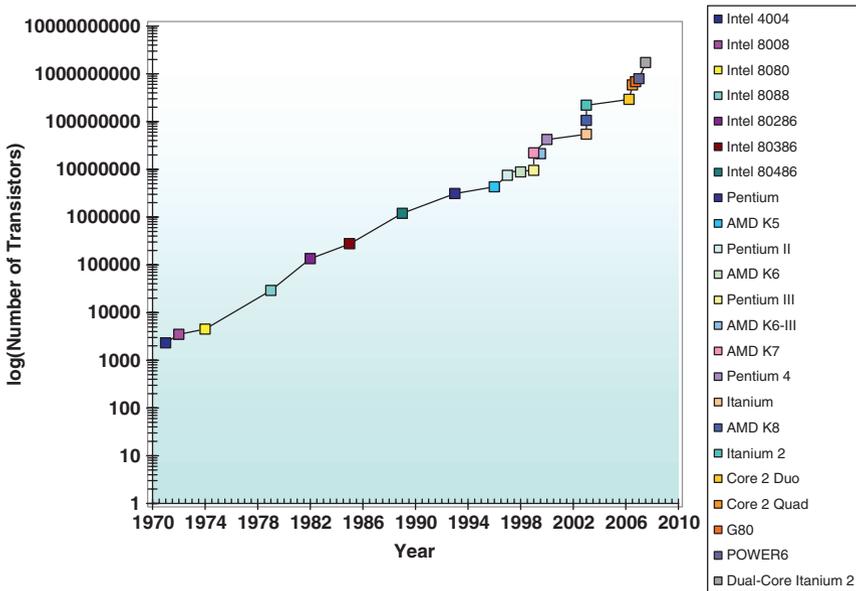
The most notable of these hardware advances are processing capability, color graphics resolution and display technology, hard disk storage, and the amount of RAM. The most remarkable thing is that since 1982, the cost of a high-end microcomputer system has remained in the neighborhood of \$US 3,000. This statement was true in 1982, at the printing of the last edition of this book in 1996, and it holds true today.

### 1.2.1 CPUs and RAM

While we can recognize that computer technology changes quickly, this statement does not seem to adequately describe what sometimes seems to be the breakneck pace of improvements in the heart of any electronic computing engine, the central processing unit (CPU). The transistor, invented at Bell Labs in 1947, is the fundamental electronic component of the CPU chip. Higher performance CPUs require more logic circuitry, and this is reflected in steadily rising transistor densities. Simply put, the number of transistors in a CPU is a rough measure of its computational power which is usually measured in floating point mathematical operations per second (FLOPS). The more transistors there are in the CPU, or silicon engine, the more work it can do.

Trends in transistor density over time, reveal that density typically doubles approximately every year and a half according to a well know axiom known as Moore's Law. This proposition, suggested by Intel co-founder Gordon Moore (Moore 1965), was part observation and part marketing prophesy. In 1965 Moore, then director of R&D at Fairchild Semiconductor, the first large-scale producer of commercial integrated circuits, wrote an internal paper in which he drew a line through five points representing the number of components per integrated circuit for minimum cost for the components developed between 1959 and 1964 (Source: <http://www.computerhistory.org/semiconductor/timeline/1965-Moore.html>, accessed 12 January 2008). The prediction arising from this observation became a self-fulfilling prophecy that emerged as one of the driving principals of the semiconductor industry. As it related to computer CPUs (one type of integrated circuit), Moore's Law states that the number of transistors packed into a CPU doubles every 18–24 months.

Figure 1.1 supports this claim. In 1979, the 8088 CPU had 29,000 transistors. In 1997, the Pentium II had 7.5 million transistors, in 2000 the Pentium 4 had 420 million, and the trend continues so that in 2007, the Dual-Core Itanium 2 processor has 1.7 billion transistors. In addition to transistor density, data



**Fig. 1.1** Trends in the number of transistors placed on various CPU chips. Note the y-axis is on the log scale (Source: <http://download.intel.com/pressroom/kits/IntelProcessorHistory.pdf>, accessed 12 January 2008)

handling capabilities (i.e. progressing from manipulating 8, to 16, to 32, to 64 bits of information per instruction), ever increasing clock speeds (Fig. 1.2), and the number of instructions executed per second, continue to improve.

The remarkable thing is that while the number of transistors per CPU has increased more than 1,000 times over the past 26 years, and another 1,000 times since 1996, performance (measured with millions of instructions per second, MIPS) has increased more than 10,000 times since the introduction of the 8088 (Source: <http://www.jcmit.com/cpu-performance.htm>, accessed 12 January 2008). Scientific analysts, who use large databases, scientific visualization applications, statistics, and simulation modeling need as many MIPS as they can get. The more powerful computing platforms described above will enable us to perform analyses that we could not perform earlier (see Chapters 8, 11 and 12).

In the original edition we predicted that “Three years from now CPU’s will be four times faster than they are today and multi-processor designs should be commonplace.” This prediction has generally proven to be true. CPU performance has continued to increase according to Moore’s Law for the last 40 years, but this trend may not hold up in the near future. To achieve higher transistor densities requires the manufacturing technology (photolithography) to build the transistor in smaller and smaller physical spaces. The process architecture of

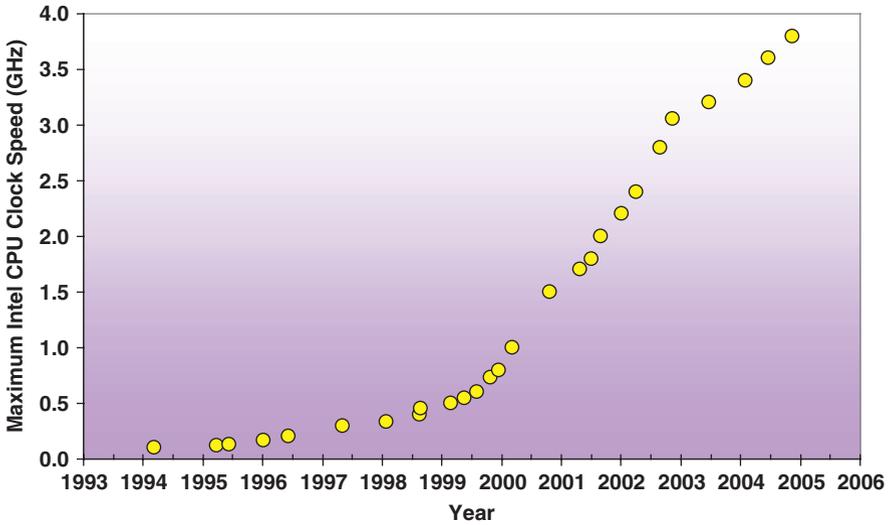


Fig. 1.2 Trends in CPU clock speed (Source: [http://wi-fizzle.com/compsci/cpu\\_speed\\_Page\\_3.png](http://wi-fizzle.com/compsci/cpu_speed_Page_3.png), accessed 12 January 2008)

CPUs in the early 1970s used a 10 micrometer ( $\mu\text{m}$ ,  $10^{-6}\text{m}$ ) photolithography mask. The newest chips use a 45 nanometer ( $\text{nm}$ ,  $10^{-9}\text{m}$ ) mask. As a consequence of these advances, the cost per unit of performance as measured in gigaflops has dramatically declined (Fig. 1.3).

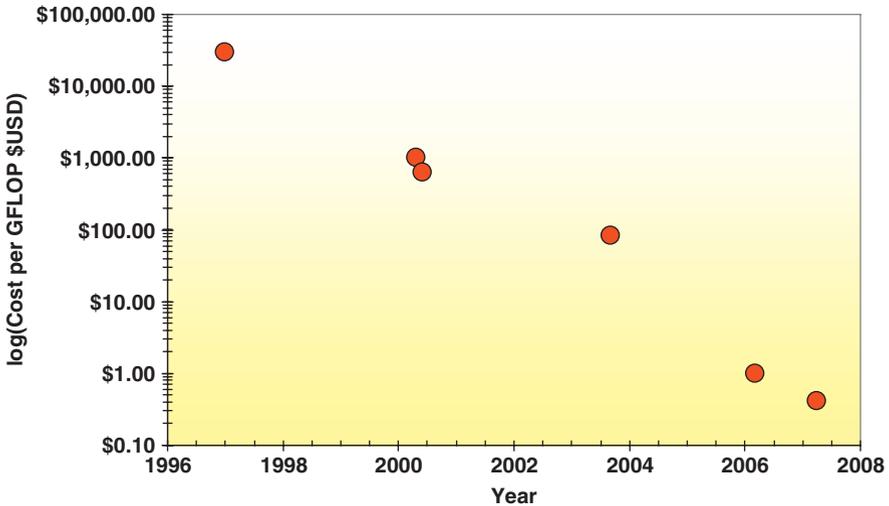


Fig. 1.3 Trends in the cost (\$USD) per gigaflop ( $10^9$  floating point instructions  $\text{s}^{-1}$ ) of CPU performance. Note y-axis is on the log scale (Source: <http://en.wikipedia.org/wiki/Teraflop>, accessed 12 January 2008)

Manufacturing technology appears to be reaching its limits in terms of how dense silicon chips can be manufactured – in other words, how many transistors can fit onto CPU chips and how fast their internal clocks can be run. As stated recently in the BBC News, “The industry now believes that we are approaching the limits of what classical technology – classical being as refined over the last 40 years – can do.” (Source: <http://news.bbc.co.uk/2/hi/science/nature/4449711.stm>, accessed 12 January 2008). There is a problem with making microprocessor circuitry smaller. Power leaks, the unwanted leakage of electricity or electrons between circuits packed ever closer together, take place. Overheating becomes a problem as processor architecture gets ever smaller and clock speeds increase.

Traditional processors have one processing engine on a chip. One method used to increase performance through higher transistor densities, without increasing clock speed, is to put more than one CPU on a chip and to allow them to independently operate on different tasks (called threads). These advanced chips are called multiple-core processors. A dual-core processor squeezes two CPU engines onto a single chip. Quad-core processors have four engines. Multiple-core chips are all 64-bit meaning that they can work through 64 bits of data per instruction. That is twice rate of the current standard 32-bit processor. A dual-core processor theoretically doubles your computing power since a dual-core processor can handle two threads of data simultaneously. The result is there is less waiting for tasks to complete. A quad-core chip can handle four threads of data.

Progress marches on. Intel announced in February 2007 that it had a prototype CPU that contains 80 processor cores and is capable of 1 teraflop ( $10^{12}$  floating point operations per second) of processing capacity. The potential uses of a desktop fingernail-sized 80-core chip with supercomputer-like performance will open unimaginable opportunities (Source: <http://www.intel.com/pressroom/archive/releases/20070204comp.htm>, accessed 12 January 2008).

As if multiple core CPUs were not powerful enough, new products being developed will feature “dynamically scalable” architecture, meaning that virtually every part of the processor – including cores, cache, threads, interfaces, and power – can be dynamically allocated based on performance, power and thermal requirements (Source: <http://www.hardwarecentral.com/hardwarecentral/reports/article.php/3668756>, accessed 12 January 2008). Supercomputers may soon be the same size as a laptop if IBM brings to the market silicon nanophotonics. In this new technology, wires on a chip are replaced with pulses of light on tiny optical fibers for quicker and more power-efficient data transfers between processor cores on a chip. This new technology is about 100 times faster, consumes one-tenth as much power, and generates less heat (Source: [http://www.infoworld.com/article/07/12/06/IBM-researchers-build-supercomputer-on-a-chip\\_1.html](http://www.infoworld.com/article/07/12/06/IBM-researchers-build-supercomputer-on-a-chip_1.html), accessed 12 January 2008).

Multi-core processors pack a lot of power. There is just one problem: most software programs are lagging behind hardware improvements. To get the most out of a 64-bit processor, you need an operating system and application programs that support it. Unfortunately, as of the time of this writing, most

software applications and operating systems are not written to take advantage of the power made available with multiple cores. Slowly this will change. Currently there are 64-bit versions of Linux, Solaris, and Windows XP, and Vista. However, 64-bit versions of most device drivers are not available, so for today's uses, a 64-bit operating system can become frustrating due to a lack of available drivers.

Another current developing trend is building high performance computing environments using computer clusters, which are groups of loosely coupled computers, typically connected together through fast local area networks. A cluster works together so that multiple processors can be used as though they are a single computer. Clusters are usually deployed to improve performance over that provided by a single computer, while typically being much less expensive than single computers of comparable speed or availability.

Beowulf is a design for high-performance parallel computing clusters using inexpensive personal computer hardware. It was originally developed by NASA's Thomas Sterling and Donald Becker. The name comes from the main character in the Old English epic poem Beowulf.

A Beowulf cluster of workstations is a group of usually identical PC computers, configured into a multi-computer architecture, running a Open Source Unix-like operating system, such as BSD (<http://www.freebsd.org/>, accessed 12 January 2008), Linux (<http://www.linux.org/>, accessed 12 January 2008) or Solaris (<http://www.sun.com/software/solaris/index.jsp?cid=921933>, accessed 12 January 2008). They are joined into a small network and have libraries and programs installed that allow processing to be shared among them. The server node controls the whole cluster and serves files to the client nodes. It is also the cluster's console and gateway to the outside world. Large Beowulf machines might have more than one server node, and possibly other nodes dedicated to particular tasks, for example consoles or monitoring stations. Nodes are configured and controlled by the server node, and do only what they are told to do in a disk-less client configuration.

There is no particular piece of software that defines a cluster as a Beowulf. Commonly used parallel processing libraries include Message Passing Interface; (MPI, <http://www-unix.mcs.anl.gov/mpi/>, accessed 12 January 2008) and Parallel Virtual Machine, (PVM, <http://www.csm.ornl.gov/pvm/>, accessed 12 January 2008). Both of these permit the programmer to divide a task among a group of networked computers, and recollect the results of processing. Software must be revised to take advantage of the cluster. Specifically, it must be capable of performing multiple independent parallel operations that can be distributed among the available processors. Microsoft also distributes a Windows Compute Cluster Server 2003 (Source: <http://www.microsoft.com/windowsserver2003/ccs/default.aspx>, accessed 12 January 2008) to facilitate building a high-performance computing resource based on Microsoft's Windows platforms.

One of the main differences between Beowulf and a cluster of workstations is that Beowulf behaves more like a single machine rather than many workstations. In most cases client nodes do not have keyboards or monitors, and are

accessed only via remote login or through remote terminals. Beowulf nodes can be thought of as a CPU + memory package which can be plugged into the cluster, just like a CPU or memory module can be plugged into a motherboard. (Source: [http://en.wikipedia.org/wiki/Beowulf\\_\(computing\)](http://en.wikipedia.org/wiki/Beowulf_(computing)), accessed 12 January 2008). Beowulf systems are now deployed worldwide, chiefly in support of scientific computing and their use in fisheries applications is increasing. Typical configurations consist of multiple machines built on AMD's Opteron 64-bit and/or Athlon X2 64-bit processors.

Memory is the most readily accessible large-volume storage available to the CPU. We expect that standard RAM configurations will continue to increase as operating systems and application software become more full-featured and demanding of RAM. For example, the "recommended" configuration for Windows Vista Home Premium Edition and Apple's new Leopard operating systems is 2 GB of RAM, 1 GB to hold the operating system leaving 1 GB for data and application code. In the previous edition, we predicted that in 3–5 years (1999–2001) 64–256 megabytes (MB) of Dynamic RAM will be available and machines with 64 MB of RAM will be typical. This prediction was incredibly inaccurate. Over the years, advances in semiconductor fabrication technology have made gigabyte memory configurations not only a reality, but commonplace.

Not all RAM performs equally. Newer types, called double data rate RAM (DDR) decrease the time it takes for the CPU to communicate with memory, thus speeding up computer execution. DDR comes in several flavors. DDR has been around since 2000 and is sometimes called DDR1. DDR2 was introduced in 2003. It took a while for DDR2 to reach widespread use, but you can find it in most new computers today. DDR3 began appearing in mid-2007. RAM simply holds data for the processor. However, there is a cache between the processor and the RAM: the L2 cache. The processor sends data to this cache. When the cache overflows, data are sent to the RAM. The RAM sends data back to the L2 cache when the processor needs it. DDR RAM transfers data twice per clock cycle. The clock rate, measured in cycles per second, or hertz, is the rate at which operations are performed. DDR clock speeds range between 200 MHz (DDR-200) and 400 MHz (DDR-400). DDR-200 transfers 1,600 megabits per second ( $\text{Mb s}^{-1}: 10^6 \text{ bits s}^{-1}$ ), while DDR-400 transfers 3,200  $\text{MB s}^{-1}$ . DDR2 RAM is twice as fast as DDR RAM. The bus carrying data to DDR2 memory is twice as fast. That means twice as much data are carried to the module for each clock cycle. DDR2 RAM also consumes less power than DDR RAM. DDR2 speeds range between 400 MHz (DDR2-400) and 800 MHz (DDR2-800). DDR2-400 transfers 3,200  $\text{MB s}^{-1}$ . DDR2-800 transfers 6,400  $\text{MB s}^{-1}$ . DDR3 RAM is twice as fast as DDR2 RAM, at least in theory. DDR3 RAM is more power-efficient than DDR2 RAM. DDR3 speeds range between 800 MHz (DDR3-800) and 1,600 MHz (DDR3-1600). DDR3-800 transfers 6,400  $\text{MB s}^{-1}$ ; DDR3-1600 transfers 12,800  $\text{MB s}^{-1}$ .

As processors increased in performance, the addressable memory space also increased as the chips evolved from 8-bit to 64-bit. Bytes of data readily

accessible to the processor are identified by a memory address, which by convention starts at zero and ranges to the upper limit addressable by the processor. A 32-bit processor typically uses memory addresses that are 32 bits wide. The 32-bit wide address allows the processor to address  $2^{32}$  bytes (B) of memory, which is exactly 4,294,967,296 B, or 4 GB. Desktop machines with a gigabyte of memory are common, and boxes configured with 4 GB of physical memory are easily available. While 4 GB may seem like a lot of memory, many scientific databases have indices that are larger. A 64-bit wide address theoretically allows 18 million terabytes of addressable memory ( $1.8 \cdot 10^{19}$  B). Realistically 64-bit systems will typically access approximately 64 GB of memory in the next 5 years.

### ***1.2.2 Hard Disks and Other Storage Media***

Improvements in hard disk storage, since our last edition, have advanced as well. One of the most amazing things about hard disks is that they both change and don't change more than most other components. The basic design of today's hard disks is not very different from the original 5¼" 10 MB hard disk that was installed in the first IBM PC/XTs in the early 1980s. However, in terms of capacity, storage, reliability and other characteristics, hard drives have substantially improved, perhaps more than any other PC component behind the CPU. Seagate, a major hard drive manufacturer, estimates that drive capacity increases by roughly 60% per year (Source: <http://news.zdnet.co.uk/communications/0,100,0000085,2067661,00.htm>, accessed 12 January 2008).

Some of the trends in various important hard disk characteristics (Source: <http://www.PCGuide.com>, accessed 12 January 2008) are described below. The areal density of data on hard disk platters continues to increase at an amazing rate even exceeding some of the optimistic predictions of a few years ago. Densities are now approaching 100 Gbits in<sup>-2</sup>, and modern disks are now packing as much as 75 GB of data onto a single 3.5 in platter (Source: <http://www.fujitsu.com/downloads/MAG/vol42-1/paper08.pdf>, accessed 12 January 2008). Hard disk capacity continues to not only increase, but increase at an accelerating rate. The rate of technology development, measured in data areal density growth is about twice that of Moore's law for semiconductor transistor density (Source: <http://www.tomcoughlin.com/Techpapers/head&medium.pdf>, accessed 12 January 2008).

The trend towards larger and larger capacity drives will continue for both desktops and laptops. We have progressed from 10 MB in 1981 to well over 10 GB in 2000. Multiple terabyte (1,000 GB) drives are already available. Today the standard for most off the shelf laptops is around 120–160 GB. There is also a move to faster and faster spindle speeds. Since increasing the spindle speed improves both random-access and sequential performance, this is likely to continue. Once the domain of high-end SCSI drives (Small Computer System Interface), 7,200 RPM spindles are now standard on mainstream desktop and

notebook hard drives, and a 10,000 and 15,000 RPM models are beginning to appear. The trend in size or form factor is downward: to smaller and smaller drives. 5.25 in drives have now all but disappeared from the mainstream PC market, with 3.5 in drives dominating the desktop and server segment. In the mobile world, 2.5 in drives are the standard with smaller sizes becoming more prevalent. IBM in 1999 announced its *Microdrive* which is a tiny 1 GB or device only an inch in diameter and less than 0.25 in thick. It can hold the equivalent of 700 floppy disks in a package as small as 24.2 mm in diameter. Desktop and server drives have transitioned to the 2.5 in form factor as well, where they are used widely in network devices such as storage hubs and routers, blade servers, small form factor network servers and RAID (Redundant Arrays of Inexpensive Disks) subsystems. Small 2.5 in form factor (i.e. “portable”) high performance hard disks, with capacities around 250 GB, and using the USB 2.0 interface are becoming common and easily affordable. The primary reasons for this “shrinking trend” include the enhanced rigidity of smaller platters. Reduction in platter mass enables faster spin speeds and improved reliability due to enhanced ease of manufacturing. Both positioning and transfer performance factors are improving. The speed with which data can be pulled from the disk is increasing more rapidly than positioning performance is improving, suggesting that over the next few years addressing seek time and latency will be the areas of greatest attention to hard disk engineers. The reliability of hard disks is improving slowly as manufacturers refine their processes and add new reliability-enhancing features, but this characteristic is not changing nearly as rapidly as the others above. One reason is that the technology is constantly changing, and the performance envelope is constantly being pushed; it’s much harder to improve the reliability of a product when it is changing rapidly.

Once the province of high-end servers, the use of multiple disk arrays (RAIDs) to improve performance and reliability is becoming increasingly common, and multiple hard disks configured as an array are now frequently seen in consumer desktop machines. Finally, the interface used to deliver data from a hard disk has improved as well. Despite the introduction to the PC world of new interfaces such as IEEE-1394 (FireWire) and USB (universal serial bus) the mainstream interfaces in the PC world are the same as they were through the 1990s: IDE/ATA/SATA and SCSI. These interfaces are all going through improvements. A new external SATA interface (eSATA) is capable of transfer rates of 1.5–3.0 Gbits  $s^{-1}$ . USB transfers data at 480 Mbits  $s^{-1}$  and Firewire is available in 400 and 800 Mbits  $s^{-1}$ . USB 3.0 has been announced and it will offer speeds up to 4.8 Gbits  $s^{-1}$ . Firewire will also improve to increases in the range of 3.2 Gbits  $s^{-1}$ . The interfaces will continue to create new and improved standards with higher data transfer rates to match the increase in performance of the hard disks themselves.

In summary, since 1996, faster spindle speeds, smaller form factors, multiple double-sided platters coated with higher density magnetic coatings, and improved recording and data interface technologies, have substantially increased hard disk storage and performance. At the same time, the price per

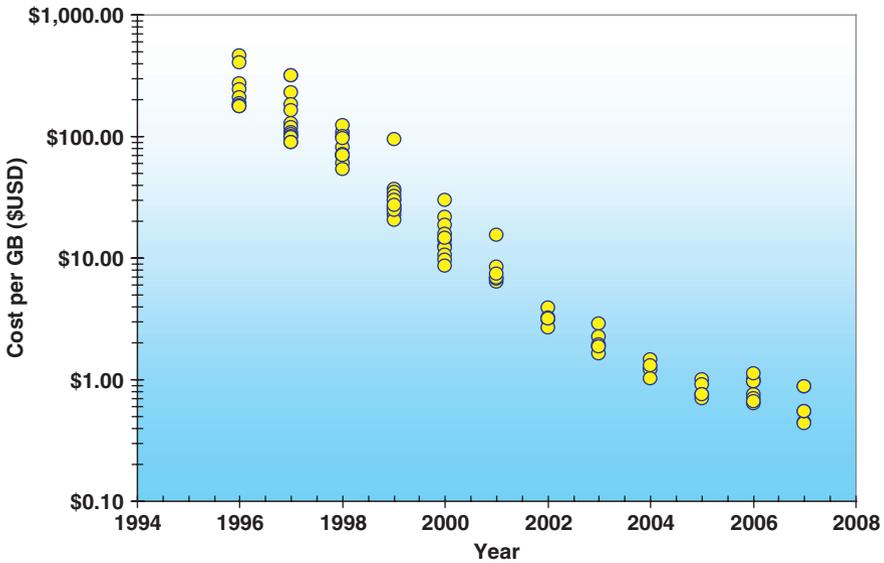


Fig. 1.4 Trends in cost per GB of hard disk storage (\$USD) (Source: <http://www.mattscopmputertrends.com/harddiskdata.html>, accessed 12 January 2008)

unit of storage has decreased (Fig. 1.4). In 1990, a typical gigabyte of storage cost about \$US 20,000 (Kessler 2007). Today it is less than \$US 1. The total hard disk capacity shipped as of 2003 (Fig. 1.5) indicates exponentially

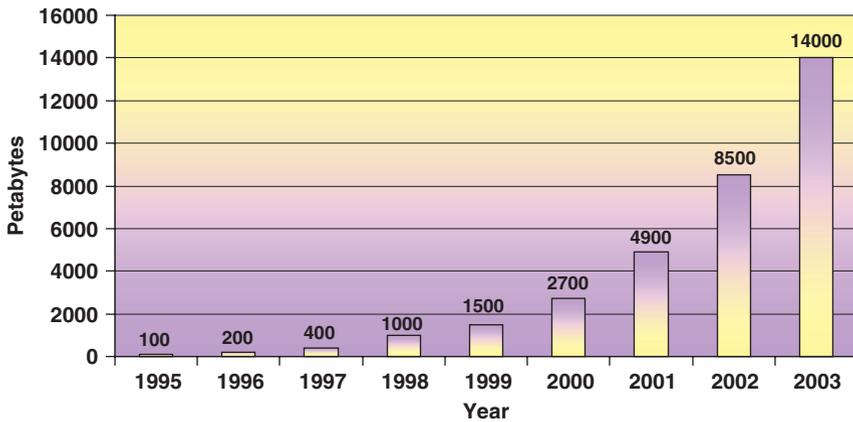


Fig.1.5 PC hard disk capacity (in petabytes- $10^{15}$  B or 1,000 TB) shipped as of 2003 (Data from 1999 Winchester Disk Drive Market Forecast and Review Table C5, International Data Corporation) (Source: <http://www2.sims.berkeley.edu/research/projects/how-muc+h-info/charts/charts.html>, accessed 12 January 2008)

increasing capacity through time. Today 2.5" 250 GB hard disks are common and multiple terabyte hard disks collected together in RAID configurations provide unprecedented storage capacity. The trends continue as recently Seagate announced research into nanotube-lubricated hard disks with capacities of several terabits per square inch, making possible a 7.5 TB 3.5 in hard disk (Source: <http://www.dailytech.com/article.aspx?newsid=3122&ref=y>, accessed 12 January 2008).

Hard disks are not the only available storage media. Floppy disks, formerly a mainstay of portable storage, have become a thing of the past. Today computers are rarely shipped with floppy disk drives. At one time, Iomega's portable ZIP drives looked promising as a portable device to store about 200 MB of data. In 1996, we predicted that "Newer storage media such as read-write capable CD-ROM's and WORM's (write once read many times) will eventually displace floppy disks as the storage medium of choice". This has taken place and today even the CD-ROM, which in the past held promise for large capacity storage (~700 MB) has been replaced with the ubiquitous "thumb drive" memory sticks. These marvels of miniaturization can accommodate 8–16 GB of data, use very fast USB 2.0 transfer interfaces, easily connect to any computer with a USB port, and are unusually inexpensive. As of the time of this writing a 4 GB USB 2.0 memory stick costs around \$US 40. Double-sided rewritable DVD media are increasingly being used to easily store data in the 4–6 GB range.

### ***1.2.3 Graphics and Display Technology***

In 1996, we predicted that in 3–5 years (1999–2001), support for 24-b color, full 3-D acceleration, broadcast quality video, and full-motion near-lifelike virtual-reality capabilities would be commonplace. This forecast has proven to be true.

The very first video card, released with the first IBM PC, was developed by IBM in 1981. The MDA (monochrome display adapter) only worked in text mode representing 25×80 lines in the screen. It had a 4 KB video memory and just one color. Today's graphic cards offer radically improved capabilities. Modern video cards have two important components. The first is the GPU (graphics processing unit). This dedicated microprocessor, separate from the main CPU, is responsible for resolution and image quality. It is optimized for floating point calculations, which are fundamental to 3D graphics rendering. The GPU also controls many graphic primitive functions such as drawing lines, rectangles, filled rectangles, polygons and the rendering of the graphic images. Ultimately, the GPU determines how well the video card performs. The second important component is the video RAM (or vRAM). In older graphics cards, system RAM was used to store images and textures. But with a dedicated video card, built-in vRAM takes over this role, freeing up system RAM and the main CPU for other tasks. When it comes to vRAM, there are a variety of options. If

you're just doing simple tasks, 64 MB is adequate. If you're editing video, 128 MB should be the minimum, with larger amounts up to 512 MB – 1 GB available for more demanding tasks. Also as a general rule, the more powerful the GPU, the more vRAM it will require.

Modern video cards also incorporate high speed communication channels that allow large amounts of graphic data to pass quickly through the system bus. Today's video cards also contain multiple output options including S-video, super VGA (SVGA), Digital Video Interface (DVI), and High Definition Multimedia Interface (HDMI) connections are common as well as options for up to 32-b and 64-b colors, and resolutions approaching  $2,560 \times 1,600$  at very fast refresh rates in the range of 85 Hz. The very newest cards include television tuners, some even offering the newly emerging High-definition standard. This feature is mainly relevant to home computer systems for those who want to turn their computer into a personal video recorder. We are convinced a scientific application for this feature will become useful in the years to come.

The ability to produce graphics is just one piece of the graphics system, with the other being the display device. Old large and power hungry analog monitors are slowly being replaced by digital Liquid Crystal Display (LCD) panels, the latter appearing sometimes in large (19–22 in) formats. LCD monitors are sleeker than bulky cathode-ray tube models and they are more energy efficient. Some LCD monitors consume  $\frac{1}{2}$  to  $\frac{2}{3}$  the energy of traditional monitors.

Since Windows XP was released, with its expanded desktop feature, dual LCD monitor desktop computers have become more common. The increased popularity of multi-display systems has to do with advances in technology as well as economics. Though Windows 98 first allowed for dual display configurations, the bulky analog CRTs that sat on most desks and workspaces simply could not accommodate more than one monitor. Flat-panel displays solved the space problem. Originally expensive they were considered a luxury, with prices often exceeding \$US 1000. Resolution increased along with the ability to pack more and more transistors into the LCD panel, and today's monitors, by contrast, are just a fraction of original cost. Today a good quality 22 in LCD monitor costs around \$US 300. That means adding a second or third monitor is comparable to the cost of some of the original models. Research shows that there is a productivity benefit that is almost immediate. Numerous studies estimate productivity increases of anywhere from 10 to 45% (Russel and Wong 2005; Source: <http://www.hp.com/sbso/solutions/finance/expert-insights/dual-monitor.html>, accessed 12 January 2008). Efficiency experts suggest that using two LCD monitors improves efficiency by up to 35% and researchers at Microsoft also found similar results, reporting that workers increased their productivity 9–50% by adding a second or third monitor. (Source: <http://www.komando.com/columns/index.aspx?id=1488> accessed, 12 January 2008).

### ***1.2.4 Portable Computing***

Another recent trend is the appearance of powerful portable computer systems. The first portable computer systems (i.e. “luggables”) were large, heavy, and often portability came at a cost of reduced performance. Current laptop, notebook, and subnotebook designs are often comparable to desktop systems in terms of their processing power, hard disk, and RAM storage and graphic display capabilities. In 1996, we observed that “It is not unusual, when attending a scientific or working group meeting, to see most participants arrive with their own portable computers loaded with data and scientific software applications.” Today, it is unusual to see scientists attending technical meetings arrive *without* a portable computer.

Since 1996, the performance and cost gap between notebooks and desktops capable of performing scientific calculations has continued to narrow, so much so, that the unit growth rate of notebook computers is now faster than for desktops. With the performance gap between notebooks and desktop systems narrowing, commercial users and consumers alike are beginning to use the notebooks more and more as a desktop replacement since the distinction between the two as far as what work can be accomplished is becoming more and more blurred. Moreover, the emergence of notebook “docking stations” allows the opportunity to plug notebooks into laboratory network resources when scientists are in their office and then unplug the notebook at the end of the day to take it home or on the road, all the while maintaining one primary location for important data, software, working documents, literature references, email archives, and internet bookmarks.

We have seen that miniaturization of large capacity hard disk storage, memory sticks, printers, and universal access to email made available via ubiquitous Internet connectivity (see below) all contribute to a portable computing environment, making the virtual office a reality.

## **1.3 Coping with Oceans of Data**

The information explosion is well documented. Information stored on hard disks, paper, film, magnetic, and optical media doubled from 2000 to 2003, expanding by roughly 5 EB (exabytes: over 5 billion gigabytes) each year or about 800 MB per person per year (Lyman and Varian 2003). These authors present, as of 2003, an intriguing look into the volume of digital information produced worldwide, where it originates and interesting trends through time. For example, in the United States we send, on average, 5 billion instant messages and 31 billion emails each day (Nielsen 2006).

The trend is clear for scientific pursuits; the growth of data is one of the biggest challenges facing scientists today. As computer software and hardware improve, the more sensors we place into the biosphere, the more satellites we

put into orbit, the more model runs we perform, the more data that can be – and is being – captured. In fact, Carlson (2006) tells us, “Dealing with the ‘data deluge,’ as some researchers have called it, along with applying tested methods for controlling, organizing and documenting data will be among the great challenges for science in the 21st century.” Unfortunately, having more data does not mean we are able to conduct better science. In fact, massive volumes of data can often become detrimental to scientific pursuits. Data originating from different sources can sometimes be conflicting and certainly require ever increasing resources, hardware and maintenance. Someone once said: “We are drowning in data, but starving of information”. We feel this is particularly true for fisheries data. In addition to the ever increasing quantity of data we add the vexing problems of harmonizing heterogeneous data collected on different spatial and temporal scales and the ever present problem of inappropriate use of online data because the metadata are missing (Chapter 5). Documenting data by writing metadata is a task scientists are reluctant to undertake, but a necessary step that will allow efficient data discovery as volumes of data continue to grow. Scientists have been struggling with this issue for years and metadata software solutions are scarce and often inadequate. Metadata will be a major issue in the coming decade.

Below we present some current examples of the increasing amounts of data we are required to accommodate.

Current generation general ocean circulation models using the Regional Ocean Modeling System (ROMS; Source: <https://www.myroms.org/>, accessed 24 December 2007) linked to lower trophic level (NPZ) ecosystem models (see Chapter 10) using typical grid spacing (3 km horizontal and 30 vertical levels, giving  $462 \times 462$  horizontal gridpoints) over a typical ocean domain such as the central Gulf of Alaska (Hermann, in press, Chapter 10) generates 484 MB of output a day (where all the physical and biological variables are saved at every horizontal/vertical gridpoint). Hence a full model year of daily output from this model generates up  $484 \text{ MB} \times 365 = 176 \text{ GB}$  (Albert Hermann, pers. comm., NOAA, Pacific Marine Environmental Laboratory). If a relatively short time series of model simulations (say 10 years) were permanently archived, it would require almost 2 TB (TB, 1,000 GB) of storage.

Data collection rates for a typical shipboard acoustic echosounder system (see Chapter 5), such as the Simrad EK60 using 3 frequencies (3 frequencies at 1 ms pulse to 250 m contains 3,131 pings; 1 frequency = 16.7 MB) generates about 50 MB of data. A hypothetical acoustic mooring designed to measure down to 250 m will generate about  $4 \text{ MB h}^{-1}$  or about  $50 \text{ MB day}^{-1}$ . In the case of a typical groundfish survey, the echosounder will generate about  $95 \text{ MB-1.2 GB h}^{-1}$ , depending on the ping rate (Alex deRoberts, pers. comm. NOAA, Alaska Fisheries Science Center). Finally, newer multibeam systems, such as the Simrad ME70 will collect  $10\text{--}15 \text{ GB h}^{-1}$  for typical applications (see e.g. Ona et al. 2006).

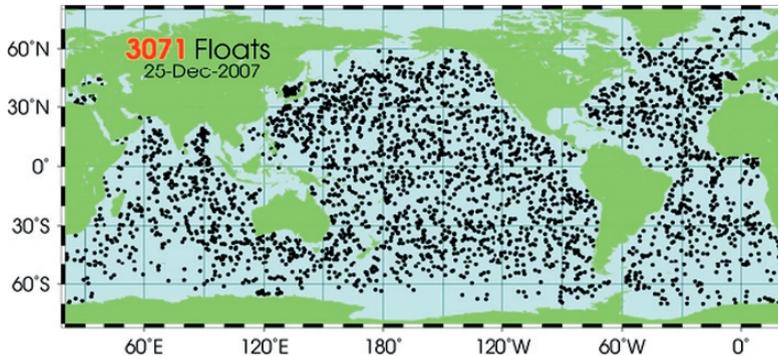
Many of our “standard” field data collection devices (i.e. measuring boards, scales, and net sampling equipment) are now digital and, interact with other

on-board ship sensors (i.e. GPS) providing large amounts of additional quality controlled information. For example, our old paradigm of net measurement (i.e. spread and height) has improved over the last decade with the use of depth sensors and bottom contact sensors. The amount of potential additional net information is about to explode with the use of automated net mensuration systems capable of providing height, spread, bottom contact, temperature, depth, net symmetry, speed, geometry of the codend, fish density, distance and angle of the net relative to the boat and even net damage reporting. In addition, there is a rapidly expanding flow of information from sounders and enhanced sonar devices capable of providing many data streams regarding the sea bottom condition and hardness, currents and other sea states. This means that even traditional data sources have the potential to rapidly expand in quantity and metadata requirements.

Cabled ocean observing systems, such as VENUS (Victoria Experimental Network Under the Sea) (Source: <http://www.martlet.ca/view.php?aid=38715>, accessed 12 January 2008) and NEPTUNE (North-East Pacific Time-Series Undersea Networked Experiments) (Source: [http://www.neptunecanada.ca/documents/NC\\_Newsletter\\_2007Aug31F.pdf](http://www.neptunecanada.ca/documents/NC_Newsletter_2007Aug31F.pdf), accessed 12 January 2008), off the North American west coast, are some of the world's first regional permanent ocean observatories. The observing system consists of undersea nodes to regulate and distribute power and provide high bandwidth communications ( $4 \text{ GBs}^{-1}$ ) through fiber-optic cable, connecting more than 200 instruments and sensors, such as video cameras, a 400 m vertical profiler (to gather data at various ocean depths) and a remotely operated vehicle, as they collect data and imagery from the ocean surface to beneath the seafloor. The existing VENUS node is similarly configured and collecting data at a rate of 4 GB per day. John Dower (pers. comm.), affiliated with the NEPTUNE and VENUS cabled observing system, characterized the problems associated with coping with the vast amounts of data being delivered from "always on" data streams such as these new cabled system as trying to take a "Drink from a fire hose".

The Intergovernmental Panel on Climate Change (IPCC) coordinated scientists at 17 major climate modeling centers throughout the world to run a series of climate models under various standard prescribed climate scenarios to examine the anticipated affect of factors contributing to climate change. They then prepared climate assessment reports, the most recent being the Fourth Assessment Report or AR4 (IPCC 2007). The massive output files are archived at the Lawrence Livermore National Laboratory (Source: <http://www-pcmdi.llnl.gov/>, accessed 12 January 2008) and are made available to the scientific community for analysis. These data consist of 221 output files from different "experiment scenario/model" combinations and the data volume totals approximately 3 TB.

Remote sensing equipment such as the ARGO system is a global array of about 3,000 free-drifting profiling ARGO floats (Fig. 1.6) that measures the temperature and salinity of the upper 2,000 m of the ocean. The floats send their data in real-time via satellites to ARGO Global Data Acquisition Centers (GADC). Data from 380,472 individual profiles are instantly available at the



**Fig. 1.6** Location of 3071 active ARGO floats which have delivered data within the past 30 days, as of 25 December 2007) (Source: <http://www.argo.ucsd.edu/Acindex.html>, accessed 25 December 2007)

GDACs including 168,589 high quality profiles provided by the delayed mode quality control process. Google Earth can be used to track individual floats in real-time (Source: <http://w3.jcommops.org/FTPRoot/Argo/Status/>, accessed 12 January 2008). This amazing resource allows, for the first time, continuous monitoring of the temperature, salinity, and velocity of the upper ocean, with all data being relayed and made publicly available within hours after collection (Source: <http://www.argo.ucsd.edu/Acindex.html>, accessed 12 January 2008).

Satellites offer another example of broadband high capacity data delivery systems. The Advanced Very High Resolution Radiometer (AVHRR) data set is comprised of data collected by the AVHRR sensor and held in the archives of the U.S. Geological Survey's EROS Data Center. AVHRR sensors, carried aboard the Polar Orbiting Environmental Satellite series, consist of a 4- or 5-channel broad-band scanning radiometer, sensing in the visible, near-infrared, and thermal infrared portions of the electromagnetic spectrum (Source: <http://edc.usgs.gov/guides/avhrr.html>, accessed 12 January 2008). The AVHRR sensor provides for global (pole to pole) on board collection of data from all spectral channels. Each pass of the satellite provides a 2,399 km (1,491 mi) wide swath. The satellite orbits the Earth 14 times each day from 833 km (517 mi) above its surface. The objective of the AVHRR instrument is to provide radiance data for investigation of clouds, land-water boundaries, snow and ice extent, ice or snow melt inception, day and night cloud distribution, temperatures of radiating surfaces, and sea surface temperature. Typical data file sizes are approximately 64 MB per a 12 min (in latitude-longitude coordinates) sampling swath per orbit.

The Sea-viewing Wide Field-of-view Sensor (SeaWiFS) is another example of a satellite system designed to provide quantitative data on global ocean bio-optical properties to the Earth science community (Source: <http://oceancolor.gsfc.nasa.gov/SeaWiFS/>, accessed 12 January 2008). Subtle changes in ocean color, and

in particular surface irradiance on every band, signify various types and quantities of marine phytoplankton (microscopic marine plants), the knowledge of which has both scientific and practical applications. Since an orbiting sensor can view every square kilometer of cloud-free ocean every 48 h, satellite-acquired ocean color data provide a valuable tool for determining the abundance of ocean biota on a global scale and can be used to assess the ocean's role in the global carbon cycle and the exchange of other critical elements and gases between the atmosphere and the ocean. The concentration of phytoplankton can be derived from satellite observations of surface irradiance and quantification of ocean color. This is because the color in most of the world's oceans in the visible light region, (wavelengths of 400–700 nm) varies with the concentration of chlorophyll and other plant pigments present in the water (i.e. the more phytoplankton present, the greater the concentration of plant pigments and the greener the water). A typical SeaWiFS SST file of sea surface temperature for 1 day from the MODIS sensor can be as large as 290 MB.

On-line databases of compiled and quality controlled data are another source of large quantities of information. Examples include biological databases such as a comprehensive database of information about fish (FishBase) that includes information on 29,400 species (Source: <http://www.fishbase.org/>, accessed 12 January 2008), a database on all living cephalopods (octopus, squid, cuttlefish and nautilus) Cephbase (Source: <http://www.cephbase.utmb.edu/>, accessed 12 January 2008), Dr. Ransom Myer's Stock Recruitment Database consists of maps, plots, and numerical data from over 600 fish populations (over 100 species) from all over the world (Source: <http://www.mscs.dal.ca/~myers/welcome.html>, accessed 12 January 2008), Global Information System about fish larvae (LarvalBase) (Source: <http://www.larvalbase.org/>, accessed 28 December 2007), the FAO Statistical Database consists of a multilingual database currently containing over 1 million time-series records from over 210 countries (Source: [http://www.fao.org/waicent/portal/statistics\\_en.asp](http://www.fao.org/waicent/portal/statistics_en.asp), accessed 12 January 2008), not to mention the numerous catch and food habits databases, often consisting of tens of millions of records.

Even given the sometimes overwhelming quantity of data, one trend that has definitely happened in the last decade is the movement of data from flat ASCII files and small ad-hoc databases (i.e. EXCEL spreadsheets) into relational databases with designs based on actual data relationships and collection methodology. This has been a very important and powerful step towards control of data quality.

Hopefully, the problems mentioned at the beginning of this section can be addressed with the tremendous advancements in hardware mentioned above as well as software advances covered in the next section.

## 1.4 Powerful Software

At the time of the last writing of this book, application software was only available from commercial sources. Since 1996, a remarkable development has taken place – Open Source software (free source code) is widely available

for almost any purpose and for almost any CPU platform. Open Source software is developed by an interested community of developers and users. As Schnute et al. (2007) eloquently put it, “Open source software may or may not be free of charge, but it is not produced without cost. Free software is a matter of liberty, not price. To understand the concept, one should think of free as in free speech, not as in free beer”. To our knowledge, no one has attempted to estimate the true cost of Open Source software.

Some notable examples of Open Source or no cost software includes operating systems such as Fedora Linux (Source: <http://fedoraproject.org/>, accessed 12 January 2008); web sever software by Apache (Source: <http://www.apache.org/>, accessed 12 January 2008); high level numerical computing software such as Octave (Source: <http://www.gnu.org/software/octave/>, accessed 12 January 2008) and SciLab (Source: <http://www.scilab.org/>, accessed 12 January 2008) which is similar to MATLAB; statistical software such as R (source: <http://www.r-project.org/>, accessed 12 January 2008) and WinBUGS; (Lunn et al. 2000; Source: <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>, accessed 12 January 2008) for implementing Bayesian statistics; compilers such as the GNU Family of Compilers for C (Source: <http://gcc.gnu.org/>, accessed 12 January 2008) and FORTRAN (Source: <http://www.gnu.org/software/fortran/fortran.html>, accessed 12 January 2008); plotting software, also from the GNU development team (Source: <http://www.gnu.org/software/fortran/fortran.html>, accessed 12 January 2008 and [www.gnuplot.info/](http://www.gnuplot.info/) accessed 12 January 2008); database software such as MySQL (Source: <http://www.mysql.com/>, accessed 12 January 2008); software business productivity programs such as OpenOffice (Source: <http://www.openoffice.org/>, accessed 12 January 2008); ecosystem modeling software such as Ecopath with Ecosim (Source: <http://www.ecopath.org/>, accessed 12 January 2008) and the newly released fisheries library in R (FLR, Kell et al. 2007) (Source: <http://www.flr-project.org/>, accessed 12 January 2008). Many other offerings can be located at the Free Software Foundation (Source: <http://www.fsf.org/>, accessed 12 January 2008).

Similar to our previous observation, we still see software functionality and growing feature sets advancing in lockstep with improvements in computer hardware performance and expanded hardware capability. Today’s application software packages are extremely powerful. Scientific data visualization tools and sophisticated multidimensional graphing applications facilitate exploratory analysis of large complex multidimensional data sets and allow scientists to investigate and undercover systematic patterns and associations in their data that were difficult to examine several years ago. This trend enables users to focus their attention on interpretation and hypothesis testing rather on the mechanics of the analysis. Software that permits the analysis of the spatial characteristics of fisheries data are becoming more common. Programs to implement geostatistical algorithms (see Chapter 7) and Geographic Information System (GIS) software (see Chapter 4) have made significant advances that offer the fisheries biologist the ability to consider this most important aspect of natural

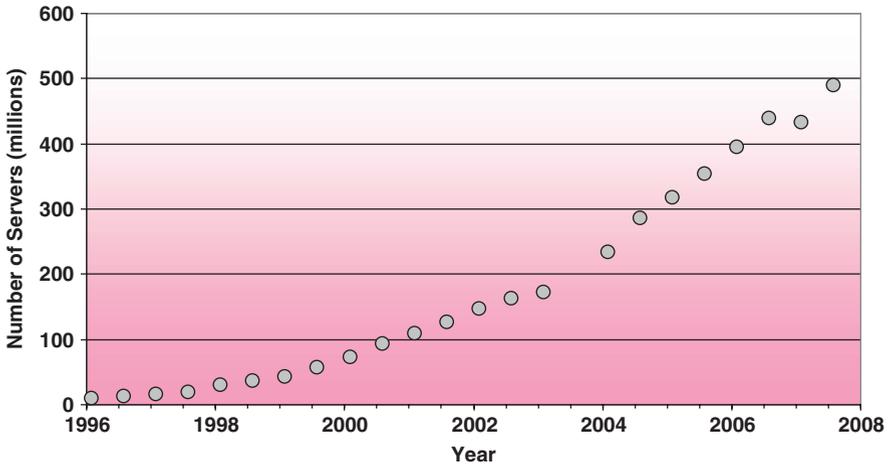
populations in both marine and freshwater ecosystems. Image analysis software (see Chapter 9) also offers promise in the areas of pattern recognition related to fisheries science such as identification of taxa, fish age determination and growth rate estimation, as well as identifying species from echo sounder sonar records.

Highly specialized software such as neural networks and expert systems (see Chapter 3), which in the past have received limited application to fisheries problems, are now becoming commonplace. Very advanced data visualization tools (see Chapter 10) offer exciting new research opportunities heretofore unavailable to fisheries scientists. Whole ecosystem analysis tools (see Chapter 8) allow the simultaneous consideration of the entirety of the biological components that make up these dynamic systems. The marriage of powerful computer systems to remote sensing apparatus and other electronic instrumentation continues to be an area of active research and development (see Chapter 5). The area of population dynamics, fisheries management, stock assessment and statistical methodology software (see Chapters 11 and 12), long a mainstay of computer use in fisheries, continues to receive much attention.

## 1.5 Better Connectivity

No aspect of our scientific lives remains untouched by the World Wide Web and Internet connectivity. The explosive growth of the Internet over the last decade has led to an ever increasing demand for high-speed, ubiquitous Internet access. The Internet is the fastest growing communication conduit and has risen in importance as the information medium of first resort for scientific users, basically achieving the prominence of a unique, irreplaceable and essential utility. How did we do our jobs without it?

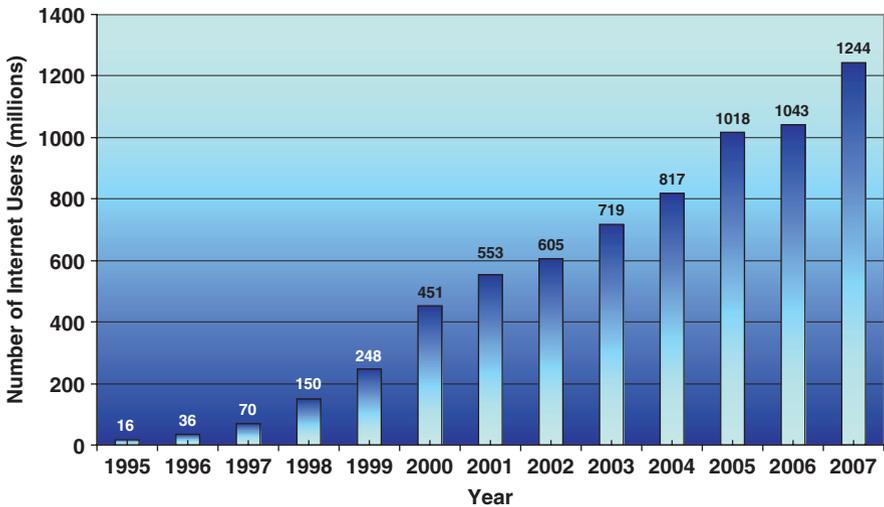
In 1996, we predicted that “the Internet, other network, and Wide-Area-Network connectivity resources held great promise to deliver global access to a vast and interactive knowledge base. In addition, the Internet would provide to the user a transparent connection to networks of information and more importantly people.” This has largely proven to be true and compared to today’s Internet resources, it may seem as a bit of an understatement. Compared to 12 years ago, access has improved, speed has increased, content has exploded providing significantly more resources available over the web. We feel it is true to say that the Internet is considered the method of choice for communication in scientific circles. O’Neill et al. (2003) present a nice summary of trends in the growth of the web, current as of 2003. Figure 1.7 depicts the steadily increasing trend in the number of Internet host servers on line (Source: <http://www.isc.org/index.pl?/ops/ds/host-count-history.php>, accessed 12 January 2008) and the large and growing community of users (Fig. 1.8) (Source: <http://www.esnips.com/doc/f3f45dae-33fa-4f1f-a780-6cfbce8be558/Internet-Users>, accessed 12 January 2008; 2007 statistics from: <http://www.internetworldstats.com/stats.htm>, accessed 12 January 2008). These data show that the number of hosts



**Fig. 1.7** Trends in the number of servers that make up the World Wide Web (Source: <http://www.isc.org/index.pl?/ops/ds/host-count-history.php>, accessed 12 January 2008)

and users have increased 5.071% and 3.356%, respectively since 1996. Lyman and Varian (2003) estimate that the web accounted for 25–50 TB of information.

Most electronic communication flows through four main channels: radio and television broadcasting, telephone calls and the Internet. The Internet, the



**Fig. 1.8** Trends in the number of internet users (millions) (Source: <http://www.esnips.com/doc/f3f45dae-33fa-4f1f-a780-6cfbce8be558/Internet-Users>, accessed 26 November 2007; 2007 statistics from: <http://www.internetworldstats.com/stats.htm>, accessed 12 January 2008)

newest electronic information medium, has proven to be capable of subsuming all three other communication channels. Digital TV stations are broadcast over IPTV (Internet Protocol Television is a system where digital television service is delivered by using Internet Protocol over a network infrastructure; See Section 1.6). We believe this relatively quick transition is a prelude of what opportunities we can expect in the near future.

Today, there are three basic options available to access the Internet: The oldest method is dial-up access. Back in 1996, dial-up was one of the few options for many scientists to connect to the Internet, mainly through commercial Internet providers such as America On Line. During this era, typical fisheries laboratories did not have broadband connections or significant within-lab Internet resources (i.e. application servers, distributed electronic databases, web pages, etc.). Even email connectivity was slow by today's standards. As the Internet grew, with its emphasis on visual and multimedia delivery, it presented a problem for the dial-up community. So much so that web pages offered "image free" versions of their content with the intention to speed up access to the core of their content. Ironically, in today's telecommunication environment, this is the exact situation for the fledgling web-enabled cell phone and Personal Digital Assistant (PDA) customers.

Eventually, the speed of dial-up modems simply could not accommodate the abundant digital content desired by web users. The progression of modem speed was impressive at first: 1,200 baud, 3,600, 9,600 baud. No end was in sight. But the expanding and band-width consuming Internet content continued to push the envelope. Eventually, even the speedy 56,000 baud modem was too slow. It was simply not fast enough to carry multimedia, such as sound and video except in low quality. In modern digital societies, dial-up is the method of last resort. If you are using dial-up, it is clear that either broadband access is not available, or that broadband access is too expensive.

We suggest, that today, the main communication link to the web and the Internet for fisheries and resource scientists within their offices is via high-speed connections such as T1 or T3 lines (often connected via fiber optic cable) or access to proven high-speed technologies such as cable modems and digital subscriber lines (DSL). While it is true that individual situations vary, we feel confident saying that within-laboratory internet connectivity has come a long way compared to 12 years ago and we expect it will continue to improve with alarming speed.

The most recent major change in Internet interconnectivity developments since 1996 involves the almost ever-present wireless Internet access. Trends in wireless communications today are vast and exciting and accelerating at the high speeds they employ. Twelve years ago, we felt privileged as scientists, if we had the opportunity to attend a meeting or working group venue that had wireless connectivity to the Internet. At the time, these services were supplied by visionary venue hosts and often our laptops were not even capable of accessing wireless signals without external adapters connected to USB or PCMCIA ports.

Most laptops manufactured today come with Wi-Fi cards already built-in. Once the laptop's Wi-Fi capability is turned on, software usually can detect an access point's SSID – or “service set identifier” – automatically, allowing the laptop to connect to the signal without the user having to intervene. Today's portable laptops, almost without exception, utilize some variation of the Intel Centrino CPU, which has a wireless mobile chipset, wireless interface, and Wi-Fi adapter embedded directly onto the computational CPU. Thus, Wi-Fi capability is already built in. Wireless extends beyond Internet connectivity. We have wireless computers; wireless Internet, WANs, and LANs; wireless keyboards and mice; pagers and PDAs; and wireless printers, scanners, cameras, and hubs. The potential is very real for our children to say, “What is a cable?”

Today, we don't feel it is an over exaggeration to say that scientists expect SOME level of Internet connectivity when they are outside of their laboratory environment. This includes meeting locations, workshop venues, alternate work locations, not including public access points located at places such as airports, local hot spots, hotels. A minimum is access to email – and the expectation is that both hardware and software tools will be available to accomplish communication with the parent laboratories or distant colleagues. Even better would be to have access, via wired or wireless connections to files at the home lab for interactions or demonstrations for such things as examination of virtual databases, access to simulation model animations, or the ability to instantaneously access working documents, PDF publications, or library resources.

Since the last edition of this book, not only has wireless connectivity become commonplace, it has gone through several iterations of improvements. The Institute of Electrical and Electronic Engineers (IEEE) standard or protocol, known as 802.11, began with the 802.11B version (a data transfer rate of around 11 Mbits s<sup>-1</sup> (Mbps) using the 2.4 GHz band with a range of 38 m), then progressed to 802.11G (a data transfer rate of 54 Mbps using the same 2.4 GHz band as 802.11B with similar range) and now the emerging standard is 802.11N (over 248 Mbps using the 5 GHz and 2.4 GHz spectrum bands and a wider range of 70 m). This is just about as fast as can be experienced over a hard-wired network. With each new iteration of the Wi-Fi standard, transmission speed and range are generally improved.

Lyman and Varian (2003) report the number of users who connect wirelessly has doubled from 2002. They estimate that 4% or roughly 1.4 million users now access the Internet without wires. Just 1 year later, the estimate was updated to be over 40,000 hot spots catering to over 20 million users. Hot spots are Wi-Fi locations setup to provide Internet access through a wireless network to nearby computers.

This extraordinary explosion of access points and users is a testimony to the utility and demand for Wi-Fi Internet access. Current estimates suggest that there are 100,000 Wi-Fi hot spots worldwide (Source: <http://www.jiwire.com/about/announcements/press-100k-hotspots.htm>, accessed 12 January 2008). Data indicate that Europe has the fastest annual growth (239%: Source: <http://ipass.com/>

pressroom/pressroom\_wifi.html, accessed 12 January 2008), with most of that annual growth (255%) taking place in hotel venues (Source: [http://ipass.com/pressroom/pressroom\\_wifi.html](http://ipass.com/pressroom/pressroom_wifi.html), accessed 12 January 2008).

The wide adoption of Wi-Fi and its rapid growth help scientists gradually become familiar with this new means of Internet access. We typically search for Wi-Fi hot spots while we are away from the home or office. Many airports, coffee bars, hotels and motels now routinely provide these services, some for a fee and some for free. According to projections, over 90 million laptops and personal digital assistants (PDAs) will have ready to access Wi-Fi LANs and local hot spots.

This technology is not without problems. The main problem with wired broadband access is that it requires a cost-based subscription and it does not reach all areas. The main problem with Wi-Fi access is that hot spots are physically small, so coverage is sparse and spatially very localized. WiMAX (Worldwide Interoperability for Microwave Access) was designed to solve these problems. WiMAX, which is based on the IEEE 802.16 standard, is a new telecommunications technology designed to provide wireless data over long distances. The expectation is that WiMAX will be considered an alternative to wired broadband like cable and DSL since it will provide universal wireless access, where almost everywhere is a “hot spot”. In practical terms, WiMAX would operate similar to Wi-Fi but at higher speeds, over greater distances and for a greater number of users. WiMAX can provide broadband wireless access up to 30 mi (50 km) for fixed stations, and 3–10 mi (5–15 km) for mobile stations. In contrast, the Wi-Fi/802.11 wireless local area network standard is limited in most cases to only 100–300 feet (30–100 m). WiMAX could potentially erase the suburban and rural blackout areas that currently have no broadband Internet access because phone and cable companies have not yet run the necessary wires to those remote locations.

Another trend that is beginning to impact scientific activity is the growth of smaller devices to connect to the Internet and the convergence of cellular phones, other portable devices, and computers. Small tablets, pocket PCs, smart phones, and even GPS devices are now capable of tapping into the web, further advancing the realization of true mobile computing. Output from real time data streams from Ocean Observing Systems such as the Alaska Ocean Observing System (Source: <http://ak.aos.org/op/data.php?region=AK>, accessed 12 January 2008) can now be viewed on web-enabled hand-held mobile devices.

### ***1.5.1 Security***

Increased levels of unwanted and malevolent computer intrusions are a regrettable outcome of better connectivity. They have grown so rapidly in recent years that they are no longer just an aggravation. A recent report estimates that in 2000, hacking attacks and computer viruses cost global businesses around 40,000 human years in lost productivity equaling about \$US1.6 trillion dollars

(Source: <http://www.vnunet.com/vnunet/news/2113080/hackers-viruses-cost-business-6tn>, accessed 12 January 2008).

The Internet was not designed with security in mind. It follows the democratic philosophy of open sharing. Because everything is interconnected, everything is vulnerable. For example, SPAM (unsolicited bulk email) is considered a threat because SPAM often contains malicious attachments. If the attachment is opened, it may unleash a virus. Because of overly aggressive SPAM filters placed on our email systems to protect us from malicious SPAM, email has become an unreliable form of communication. All too many times legitimate emails go undelivered or unread because they are incorrectly identified as SPAM. According to figures reported in Lyman and Varian (2003), unsolicited bulk email make up 40% of all email traveling over the Internet.

In addition there are other serious attacks and threats delivered via email such as viruses, worms (software that attach to and take control of a normal program then spread to other computers). Very aggressive viruses can spread quickly. For example, the Sapphire/Slammer worm, released in February 2003, required roughly 10 min to spread worldwide. In the early stages of infection, the number of compromised hosts doubled every 8.5 s. At its peak infection rate, achieved approximately 3 min after it was released, the virus scanned over 55 million IP address per second and infected 1 in 12 emails on the Internet (Source: <http://www.caida.org/research/security/sapphire/>, accessed 12 January 2008). Other threats include, phishing (using email to search for valuable information such as a credit card number), trojan horses (software that appearing to be useful but is in reality malicious), and system monitors (software that tracks everything the user does, then emails it back to the author of the malicious software). Adware (software to display unwanted adds) and Spyware (software to secretly monitor and record what a user types) can also be serious as they tend to degrade system performance. Many of the hacker tools that required in-depth knowledge a few years ago to implement these threats have been automated and are much easier to use.

A consequence to fisheries computer users is the unanticipated constraints or restrictions information technology administrators place on our use of computers because of real or perceived threats to system security. We need increased security to protect us, our computers and the computer network resources we rely on. Typically, computer security efforts focus on keeping “outsiders” out, through physical and technical measures such as gates, guards, passwords, locks and firewalls. In today’s computing environment, it is absolutely necessity to use virus protection software with current virus definition files. We are now often required to use “strong” passwords that have to be changed on a frequent schedule, use Virtual Private Network (VPN) software to remotely gain secure access to computer network resources contained behind firewalls, or encrypt entire laptop computer hard disks. As unfortunate as this situation appears, the reality is that serious computer hackers are in a cat-and-mouse dance with security professionals. We expect security issues to remain a part of the computing landscape for many years to come.

## 1.6 Collaboration and Communication

The old paradigm of scientific communication consisted of publishing research results in peer-reviewed printed publication outlets such as journals and books. Collaborations took place mainly through face-to-face meetings and attending scientific workshops and conferences. Today, these reliable and still pervasive outlets remain important, but we have many other alternatives for scientific communication and collaboration. It is not uncommon today, some would say required, for a scientist to prepare their research results in several different mediums and formats such a written paper for publication in a published journal or one of the newer open access digital journals, a PowerPoint presentation published on the web, and a personal or laboratory web page.

Email is currently one of the most widespread and pervasive methods of communication. Extrapolated from data supplied by the International Data Corporation, and reported by Lyman and Varian (2003), it is estimated that the current worldwide email volume amounts to about 60 billion emails a day or the equivalent of 1,829 TB daily or 3.35 PB (PB:  $10^{15}$  B) annually. Lyman and Varian (2003) offer more intriguing statistics that were valid 4 years ago: email ranks second behind the telephone for the largest information flow; 60% of workers with email access receive 10 or fewer messages on an average day, 23% receive more than 20, and 6% more than 50. They estimate that workers spend an hour or less per day on email. They also report that 78% of workers send 10 or fewer messages on an average day and that 11% send more than 20. In our 1996 edition, we mentioned the benefits of using electronic communication and the benefits of the Internet to help us organize the book, especially email. By our current thinking and expectations, it would be unthinkable to contemplate organizing and planning an endeavor such as the second edition of this book without the benefits of email and the Internet.

The decreasing cost of video capture and display technology along with widely available high speed Internet connectivity has fostered the increased use of personal video teleconference systems based on webcams, personal computer systems, and software compression. Use of video conferencing saves valuable time and reduces the cost of collaborations since often it removes or reduces the need to travel. The hardware used for this technology continues to improve in quality, prices have dropped dramatically, and the availability of freeware (often as part of chat programs) has made software based videoconferencing accessible to many.

Voice over Internet Protocol (VOIP) is a common and widely used protocol developed for the transmission of voice though the Internet. Software products have been built around this new technology allowing instantaneous voice communications between two or more people using computers connected to the Internet. Typically these computers are portable laptops with a built in or externally attached digital video camera. There are many capable software programs to facilitation communication and collaboration via VOIP. Video

conferencing services that use VOIP are common today on most free instant message clients, such as Yahoo! Messenger (Source: <http://messenger.yahoo.com/>, accessed 12 January 2008) and MSN Messenger (Source: <http://join.msn.com/messenger/overview2000>, accessed 12 January 2008) or the newer Windows Live Messenger (Source: <http://get.live.com/messenger/features>, accessed 12 January 2008). Skype is another widely used free software program using VOIP that allows users to make telephone calls from their computer to other Skype users free of charge, or to landlines and cell phones for a fee (Source: <http://www.skype.com/>, accessed 12 January 2008). All tools provide very similar features sets including video conferencing, instant messaging, text messaging, PC-mobile messaging, file transfer, and an ability to circumvent firewalls.

One new developing medium to facilitate communication are blogs or on-line diaries where commentaries or news on a particular subject are placed for users to read. Communication is the key focus and readers can leave comments in an interactive format. A typical blog combines text, images, and links to other blogs, web pages, and other media related to its topic. As of December 2007, blog search engine Technorati was tracking more than 112 million blogs (Source: <http://en.wikipedia.org/wiki/Technorati>, accessed 12 January 2008). Blogs related to the aquatic sciences exist as well. Some examples of marine and freshwater blogs are: "Thoughts from Kansas" (Source: <http://scienceblogs.com/tfk>, accessed 12 January 2008), MarineBio Blog (Source: <http://marinebio.org/blog/?cat=2>, accessed 12 January 2008), blogs about fisheries (Source: <http://www.blogtoplist.com/rss/fisheries.html>, accessed 12 January 2008), What is Your Ecotype (Source: <http://whatsyourecotype.blogspot.com/>, accessed 12 January 2008), and John's Marine and Environmental Blog (Source: <http://jmcarroll-marinebio.blogspot.com/2007/10/letter-to-national-marine-fisheries.html>, accessed 12 January 2008).

LISTSERV, a pervasive mailing list manager, is also a newer medium for communication. Recent estimates suggest that LISTSERV's sends approximately 30 million messages a day using approximately 300,000 mailing lists (Lyman and Varian 2003).

Finally, we remind readers of the pervasiveness of PowerPoint or similar presentation software as a medium of communication. In the past, presentations at scientific meetings were made predominately using 35 mm slides and overhead transparencies. Now, it is rare to see these approaches used. Almost without exception, scientific results are communicated using presentation software such as PowerPoint, running on a computer connected to a digital computer projector. On many university campuses, open lectures are delivered using PowerPoint presentations. Often PowerPoint presentations given at scientific meetings are published to the web as a record of the meeting and a resource for those unable to attend the meeting. For example, the North Pacific Marine Science Organization (PICES) routinely performs this service (Source: <http://www.pices.int/publications/presentations/default.aspx>, accessed 12 January 2008).

## 1.7 Summary

We will see further miniaturization of the microcomputer, increased capacity both in terms of processing speed, memory and storage capacity, and relatively stable prices. Although we eagerly anticipate these changes, we probably will never be satisfied. As performance levels markedly improve so will the complexity of our work, the size of the data sets we are expected to handle, and our expectations of “minimum system configuration” and “adequate performance”. The desire for faster machines, better graphics, and larger hard disks will not go away in the near future.

Computer software will continue to provide expanded analysis capabilities to fisheries professionals. We now have microcomputer operating systems that surpass what were commonly available on large mainframe computers many years ago, communication tools that make world-wide communication not only possible but fun, great suites of consistent, powerful productivity applications, personal database applications with sophisticated development and data management tools, and wider availability of highly specialized fisheries analysis programs.

Better connectivity allows fisheries scientists to more easily communicate with colleagues within the scientific community who they might not otherwise have the opportunity to meet as well as external clients. Increased microcomputer connectivity holds great promise for research collaboration and synthesis activities. Opportunities abound for scientists to effectively communicate the value and relevance of their research findings to all audiences. The Internet is a proven technology, already traveled by billions of users around the world, and it will continue to open new frontiers that are difficult to imagine today. Moreover, it will provide expanded capabilities to share databases, computer programs, and experience. Carl Walters realized the benefits of an expansion of networked microcomputers and predicted this outcome back in 1989 (Walters 1989).

Storage and sharing of fisheries database and historic information products is encouraging the systematic analysis of past fisheries information and facilitating inter-regional comparison of fisheries experience. Very often data sets from a scientific perspective become valuable when they become longer in length and one is able to put the data into perspective. Environmental data sets are most valuable after they are long enough to sample the target phenomenon multiple times. It is only when we start collecting data more regularly and routinely that we gain immense understanding into various processes. We see opportunities for scientists to ask increasingly broader questions as they focus their attention on processing and analyzing large volumes of data. Our ability to understand the fast paced and considerable changes we observe gives us a distinct advantage when dealing with complex multidisciplinary issues. Improvements in hardware and software performance will help investigators sift through increasing amounts of data in a much more powerful way or

execute computationally intensive statistical models such as the 550 parameter stock assessment model that includes a Bayesian Markov Chain Monte Carlo (MCMC) analysis (Iannelli et al. 2007). Through these activities, a deeper understanding of the data will improve the design of studies that look for associations between components of the biosphere on spatial and temporal scales relevant to biological and ecological systems. New and upcoming technologies will soon permit the ability to built truly integrated fisheries ecosystem models. For example, fine-scale biogeochemical-physical models dynamically coupled with spatially-explicit multi-trophic level ecosystem models will soon be commonplace, thus allowing holistic examination of the ecosystem and the response of its species-to-species interrelationships to management or climate change scenarios.

Common patterns and trends may emerge as data sets are shared, made available for quick retrieval and analysis, and accompanied by companion data sets such as vast oceanographic and meteorological data libraries. One of the ongoing challenges will be the development of tools to provide this capacity to process large volumes of raw data and to make it available in real time over the Internet.

Computers, fisheries research, and the resulting synergism are indeed exciting. We think that it is true that in any branch of science computers are essential. You can often set yourself apart as a valuable scientist and researcher by learning and using advanced computational tools (i.e. going beyond using basic computer applications). A necessary and valuable skill of future scientists will be to have some knowledge of how to solve problems and being able to look at a problem and come up with a computer solution. We hope the chapters in this book contribute to the development of this skill set. The following collection of papers, while not intended to be comprehensive, do characterize the breadth and sophistication of the application of computers to modern fisheries analysis, and documents the progression in advances in technology and their application to fisheries and resource management problems. We believe that the topics covered here are a prelude to new future opportunities and we anticipate with enthusiasm the challenges ahead.

## References

- Carlson S (2006) Lost in a sea of science data. *The Chronicle of Higher Education*, Information Technology Section 52(42):A35 (Source: <http://chronicle.com/free/v52/i42/42a03501.htm>, accessed 12 January 2008).
- Hermann AJ, Hinckley S, Dobbins EL, Haidvogel DB, Mordy C (in press) Quantifying cross-shelf and vertical nutrient flux in the Gulf of Alaska with a spatially nested, coupled biophysical model. *Progress in Oceanography*.
- Iannelli JN, Barbeaux S, Honkalehto T, Kotwicki S, Aydin K, Williamson N (2007) Eastern bering sea pollock. In: National Marine Fisheries Service Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Region in 2007. North Pacific Fishery Management Council, Anchorage, AK.

- IPCC (2007) Climate Change (2007) Synthesis Report. Intergovernmental Panel on Climate Change.
- Kell LT, Mosqueira I, Grosjean P, Fromentin J-M, Garcia D, Hillary R, Jardim E, Mardle S, Pastoors MA, Poos JJ, Scott F, Scott RD (2007) FLR: an open-source framework for the evaluation and development of management strategies. *ICES Journal of Marine Science* 64:640–646.
- Kessler M (2007) Days of officially drowning in data almost upon us. *USA Today, Technology News*, March, 05, 2007 (Source: [http://www.usatoday.com/tech/news/2007-03-05-data\\_N.htm](http://www.usatoday.com/tech/news/2007-03-05-data_N.htm) accessed 12 January 2008).
- Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000) WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 10:325–337.
- Lyman P, Varian HR (2003) How much information? School of Information Management and Systems, University of California at Berkeley (Source: <http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/>, accessed 12 January 2008).
- Moore GE (1965) Cramming more components onto integrated circuits. *Electronics Magazine* 38(8) (April 19, 1965).
- Nielsen JL (2006) Thoughts from Kansas. President's hook. *Fisheries* 31(10):480, 514–515.
- Ona E, Dalen J, Knudsen HP, Patel R, Andersen LN, Berg S (2006) First data from sea trials with the new MS70 multibeam sonar. *Journal Acoustic Society of America* 120:3017–3018.
- O'Neill ET, Lavoie BF, Bennett R (2003) Trends in the evolution of the public web. *D-Lib Magazine*, 9(4), April 2003 (Source: <http://www.dlib.org/dlib/april03/lavoie/04lavoie.html>, accessed 12 January 2008). DOI: 10.1045/april2003-lavoie.
- Russell SE, Wong K (2005) Dual-Screen monitors: a qualitative analysis of their use in an academic library. *The Journal of Academic Librarianship* 31(6):574–577.
- Schnute JT, Maunder MN, Ianelli JN (2007) Designing tools to evaluate fishery management strategies: can the scientific community deliver? *ICES Journal of Marine Science* 64:1077–1084.
- Walters CJ (1989) Development of microcomputer use in fisheries research and management. In: Edwards EF, Megrey BA (eds.) *Mathematical Analysis of Fish Stock Dynamics*. American Fisheries Society Symposium 6:3–7.