

Compactação de dados Cartográficos Digitais

*Nilo Cesar Coelho da Silva*¹*

*José Duarte Correia*¹*

*Cláudio Márcio Rodrigues de Jesus*²*

Luiz Felipe Ferreira²

RESUMO

Os processos de aquisição, manipulação e arquivamento de dados cartográficos vêm se modificando de modo extraordinário nos últimos anos, em função do grande avanço tecnológico de todas as ciências, principalmente da informática. Um dos maiores objetivos a ser alcançado, conseqüentemente, é a digitalização das bases cartográficas nacionais, essenciais para a maioria das aplicações em Cartografia e ciências afins. Esta operação envolve grandes volumes de dados, tornando críticas as etapas de armazenamento e transmissão. O presente trabalho compara diversos programas de compactação de dados digitais, aplicados a amostras representativas do universo cartográfico.

INTRODUÇÃO

O progresso da informática registrado nestes últimos anos, no que concerne a programas e máquinas, tem contribuído muito para a evolução da Cartografia Automatizada. Por conseguinte, a demanda de dados cartográficos digitais vem aumentando consideravelmente.

Compactadores são programas computacionais que para um dado arquivo, reestruturam seus dados e eliminam re-

dundâncias de tal forma que resulte um novo arquivo de tamanho sensivelmente menor do que o original. Como possibilitam, também, a descompactação, isto é, o

* Alunos do Curso de Pós-Graduação em Engenharia Cartográfica no Instituto Militar de Engenharia.

¹ Fundação Instituto Brasileiro de Geografia e Estatística, Diretoria de Geociências

Av. Brasil, 15.671, Parada de Lucas, Rio de Janeiro, RJ, CEP 21241-051
Tel.: (021) 391-8217; Fax: (021) 391-7070;
e-mail: nilo@datum.ime.br; duarte@datum.ime.br

² Instituto Militar de Engenharia, Departamento de Cartografia
Praça General Tibúrcio, 80, Rio de Janeiro, RJ, CEP 22290-270
Tel.: (021) 542-3598; Fax: (021) 275-9047;

e-mail: marcio@datum.ime.br; felipe@aquarius.ime.br

retorno integral do arquivo original de forma segura, esses programas são importante ferramenta no armazenamento de dados.

Ao se buscar economia de espaço de armazenamento, não se considera apenas o aspecto físico, mas também a redução de tempo na transmissão de dados.

O objetivo deste trabalho é comparar algum(ns) programa(s) de compactação, disponíveis no mercado, mais adequado(s) ao armazenamento dos dados cartográficos.

ARQUIVOS DE DADOS

Existe uma grande diversidade de documentos cartográficos, cartas, imagens, etc., todos com correspondentes digitais. Optou-se neste trabalho, pela seleção de arquivos que representam temas, altimetria, rodovia, hidrografia e canevá, usuais de uma mapoteca digital.

Para melhor desenvolvimento foi adotada a divisão dos arquivos em cinco áreas, cada uma com características específicas, conforme a seguir:

a) Carta Militar de Orientação

Folha de ITABORAÍ, na escala de 1:15.000, utilizada em Competições de Orientação, representando a área 1 com dimensões de 2'00" x 2'30" (latitude e longitude).

b) Mapeamento Sistemático Brasileiro

Foram selecionadas quatro áreas de dimensões 2'30" x 2'30" (Figuras 1 e 2) correspondendo cada uma, a 3% da área total de uma folha topográfica na escala de 1:50.000, com relevos diferenciados, já que na região com relevo acidentado o volume de dados é consideravelmente maior:

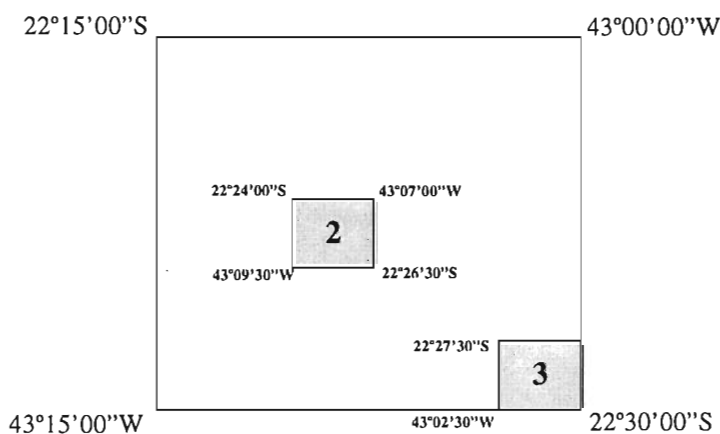


Figura 1 – Folha topográfica ITAIPAVA (RJ) – Áreas (2 e 3), de Relevo Acidentado.

PROGRAMAS DE COMPACTAÇÃO

Os programas utilizados para a compactação dos arquivos mencionados no item anterior estão relacionados na Tabela 2.

Alguns destes programas oferecem várias opções: maior taxa de compactação; maior rapidez; entre outras. Pelas razões expostas na Introdução, a escolha foi pela maior taxa de compactação.

Tabela 2 – Programas de Compactação utilizados [FOR95]

COMPACTADOR	VERSÃO	EXTENSÃO	DESCRIÇÃO
ARJ	2.41A	.ARJ	Formato para DOS, autor: Robert K. Jung
GZIP	1.2.4	.Z	Formato derivado do compress do UNIX; utiliza o método de Lempel-Ziv
LHArc	2.13	.LZH	Originário do PC, autor: Haruyasu Yoshizaki (Yoshi)
LHArc	2.55b	.LZH	Idem, em outra versão
PKPAK	3.61	.ARC	Formato ARC
PKZIP	1.10	.ZIP	Formato ZIP
PKZIP	2.04g	.ZIP	Idem, última versão
SQueueze	1.08.3	.SQZ	Antigo formato do PC e do CP/M, autor: Jonas I. Hammarberg
ZOO	2.1	.ZOO	Também usa o algoritmo de Lempel-Ziv

PROCESSAMENTO

Para avaliação dos compactadores foram realizados 148 testes, envolvendo todos os arquivos selecionados, cujos resultados são apresentados nas tabelas 3 e 4.

Elaborou-se um programa, em linguagem C, que fornece automaticamente os valores de tempo de processamento e taxa de compactação.

O ambiente computacional empregado na consecução dos diversos testes foi um PENTIUM 133, com 16 Mb de RAM.

Os comandos utilizados foram os seguintes:

Seja *A* – arquivo(s) de dado(s) a compactar;
seja *C* – arquivo compactado.

ARJ: arj a -jm C A (*)
GZIP: gzip -9 -S (nom) A (*) (nom) é uma extensão a escolher
LHArc 2.13: lha a C A
LHArc 2.55: lha a C A
PKPAK: pkpak -a C A
PKZIP 1.10: pkzip C A
PKZIP 2.04: pkzip -aex C A (*)
SQueue: sqz a -q0 C A (*)
ZOO: zoo ah C A (*)

Observações:

a) Os comandos assinalados com (*) são os que oferecem opção de maior taxa de compactação mas com um acentuado consumo de tempo. Por exemplo, a altimetria da área 3 (1,5 MB), com o emprego do programa SQueue passou de 10,4 s, na compactação normal, para 46,8 s na máxima, com ganho de apenas 0,4% (6 Kb).

b) Excetuando o programa GZIP, todos os demais admitem *A* como um ou mais arquivos, por exemplo:

A = arquivo1 arquivo2 ...(vários arquivos)

A =nome *.* (vários arquivos dados por sua extensão e complementos de nome)

O programa GZIP é o que apresenta maior grau de dificuldade para uso, pois ignora a extensão do arquivo *A* gerando o arquivo compactado com o mesmo nome de *A* e sufixo escolhido no comando de compactação. Por esse motivo é o único caso em que não aparece *C* no comando. Após o comando GZIP o arquivo original é automaticamente substituído pelo compactado.

A seguir são apresentadas as Tabelas 3 e 4, com os resultados da compactação dos dados.

No caso das folhas do mapeamento sistemático, os valores apresentados correspondem à média aritmética das taxas obtidas nas duas áreas de cada folha (ver Figuras 1 e 2).

As versões 2.13 e 2.55b do programa LHArc apresentaram os mesmos resultados, razão pela qual só foi incluído um deles nas tabelas.

Tabela 3 – ARQUIVOS VETORIAIS – Tempo e Taxa de Compactação

ARQUIVO		COMPACTADORES – tempo (s) e taxas (%)															
Folha Topog.	Tema	ARJ		GZIP		LHArc 2.55b		PKPAK		PKZIP 1.10		PKZIP 2.04G		SQueueze		ZOO	
Itaboraí		temp	taxa	temp	taxa	temp	taxa	temp	taxa	temp	taxa	temp	taxa	temp	taxa	temp	taxa
		6,8	36,45	9,0	36,98	5,1	36,26	2,6	14,17	4,1	31,03	5,5	36,46	9,3	36,61	7,7	36,24
Itaipava	Alt.	2,5	48,98	3,5	49,29	1,2	48,60	0,6	36,45	1,4	44,89	1,5	49,20	2,8	49,21	1,5	48,48
	Can	0,6	75,64	0,8	76,29	0,4	73,49	0,4	49,31	0,4	-69,39	0,4	75,84	0,7	76,01	0,5	73,08
	Hid.	0,4	70,93	0,5	71,75	0,4	71,42	0,4	63,00	0,4	69,52	0,3	711,11	0,5	71,52	0,4	70,56
	Rod	0,6	72,12	0,6	72,76	0,4	72,19	0,4	63,52	0,4	70,12	0,4	72,24	0,6	72,62	0,4	71,27
	Todos	3,1	58,02	— ¹	1,8	57,16	0,9	42,81	1,8	53,58	1,8	58,15	3,7	58,23	2,2	57,03	
Ituverava	Alt.	0,5	66,06	0,5	66,76	0,4	66,50	0,3	57,95	0,4	64,13	0,3	66,27	0,5	66,73	0,4	65,59
	Can.	0,6	68,62	0,6	69,32	0,4	67,62	0,3	55,80	0,4	65,70	0,4	68,90	0,6	69,20	0,4	67,00
	Hid.	0,4	72,67	0,4	73,58	0,4	73,25	0,3	64,26	0,4	71,48	0,3	72,81	0,4	73,37	0,3	72,14
	Rod.	0,6	70,52	0,8	70,97	0,5	70,34	0,4	58,65	0,4	68,39	0,4	70,74	0,8	70,70	0,5	69,90
	Todos	1,2	69,16	— ¹	0,9	68,62	0,5	57,59	0,9	66,68	0,6	69,27	1,5	69,46	1,0	68,32	

¹ GZIP não compacta vários arquivos em um só.


 → ocorrência mínima de tempo de processamento e máxima de taxa de compactação numa linha da tabela.

Tabela 4 – ARQUIVOS MATRICIAIS – Tempo e Taxa de Compactação

ARQUIVO		COMPACTADORES – tempo (s) e taxas (%)															
Folha Topog.	Tema	ARJ		GZIP		LHArc 2.55b		PKPAK		PKZIP 1.10		PKZIP 2.04G		SQueueze		ZOO	
		temp	taxa	temp	taxa	temp	taxa	temp	taxa	temp	taxa	temp	taxa	temp	taxa	temp	taxa
Itaipava	Alt.	32,0	61,06	36,0	61,08	7,8	60,60	2,5	49,77	7,9	57,43	9,4	61,07	46,8	61,23	11,5	60,59
	Hid.	1,7	51,51	2,0	50,97	1,0	51,98	0,7	37,81	1,2	47,56	0,9	50,84	2,2	50,84	1,3	51,85
	Rod.	1,8	59,99	2,0	59,98	0,9	60,31	0,6	46,71	1,1	55,00	0,9	59,64	2,5	59,44	1,2	60,12
	Todos	34,8	59,75	— ¹		9,8	59,43	3,2	48,04	9,6	56,10	10,8	59,69	51,2	59,81	13,5	59,41
Ituverava	Alt.	2,9	55,45	3,3	55,14	1,4	55,64	0,7	41,22	1,5	52,26	1,4	55,00	3,8	55,01	1,8	55,55
	Hid.	1,5	58,20	1,7	58,38	0,9	58,98	0,4	42,08	1,2	53,04	0,7	57,96	1,9	57,86	1,0	58,81
	Rod.	2,2	48,86	2,7	48,75	1,3	49,12	0,8	36,34	1,3	45,39	1,5	48,70	3,0	48,67	1,7	49,03
	Todos	6,0	53,49	— ¹		3,1	53,80	1,5	39,51	3,7	49,77	3,2	53,19	8,2	53,16	4,0	53,74

¹ GZIP não compacta vários arquivos em um só.



→ ocorrência mínima de tempo de processamento e máxima de taxa de compactação numa linha da tabela.

ANÁLISE DOS RESULTADOS

Na análise, os dados foram agrupados de acordo com o seu tipo, vetorial ou matricial, sendo que os valores assinalados nas tabelas 3 e 4 serviram como referência para comparação entre os resultados obtidos na compactação.

a) *Taxa de Compactação*

a.1) **Arquivos vetoriais:**

O programa GZIP foi o que apresentou o melhor resultado para esse tipo de arquivo. A área 1, Carta de Orientação Militar, apresentou taxa de compactação bem menor do que as demais (Mapeamento Sistemático Brasileiro), o que sugere que seus dados podem ter sofrido algum tratamento congênere ao executado nesse trabalho.

Relacionando por ordem decrescente de taxa de compactação, a tendência geral mantida por toda a tabela, em relação as outras áreas, foi:

Variação em relação ao GZIP (%)

1 – GZIP	-x-
2 – SQueue	0,2
3 – PKZIP 2.04g	0,4
4 – ARJ	0,6
5 – LHA	1,0
6 – ZOO	1,2
7 – PKZIP 1.10	5,0
8 – PKPAK	10,0

O PKPAK foi, sistematicamente, o que apresentou os piores resultados.

a.2) **Arquivos Matriciais:**

A exceção das áreas com altimetria densa, a tendência geral em ordem decrescente de taxa de compactação foi:

Variação em relação ao LHArc (%)

1 – LHArc	-x-
2 – ZOO	0,1
3 – ARJ e GZIP	0,6
5 – PKZIP 2,04g e SQueue	0,8
6 – PKZIP 1.10	5,0
7 – PKPAK	14,0

Especificamente para as áreas com altimetria densa, a taxa de compactação em ordem decrescente foi:

Varição em relação ao SQueue (%)

1 – SQueue	-x-
2 – GZIP, PKZIP 2,04g, ARJ	0,2
3 – LHArc e ZOO	0,6
4 – PKZIP 1,10	4,0
5 – PKPAK	11,0

O gráfico a seguir ilustra a taxa de compactação, para os itens “Todos os Temas”, do programa SQueue.

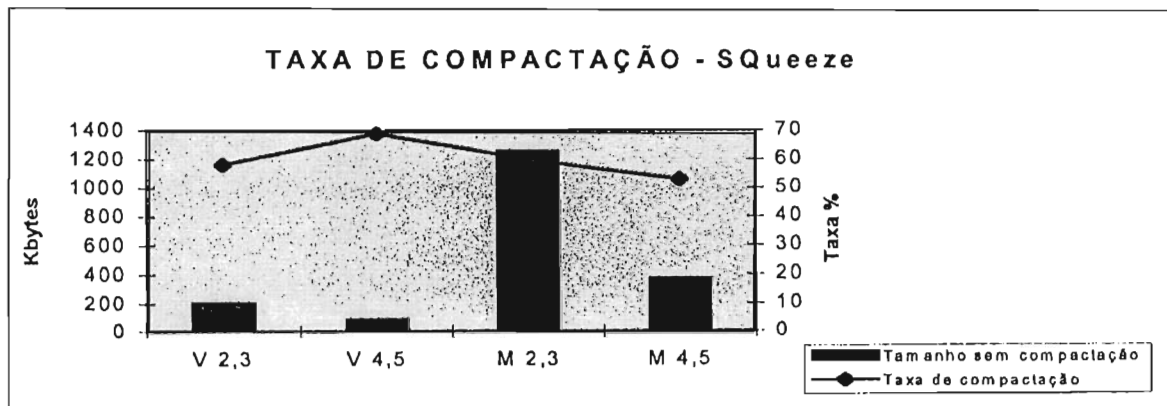


Figura 3 – V 2,3 – média de valores das áreas 2 e 3, arquivos vetoriais
 V 4,5 – média de valores das áreas 4 e 5, arquivos vetoriais
 M 2,3 – média de valores das áreas 2 e 3, arquivos matriciais
 M 4,5 – média de valores das áreas 4 e 5, arquivos matriciais

b) Tempo de Compactação

O programa PKPAK apresentou os melhores resultados de tempo em praticamente todos os testes realizados, superando os demais programas com boa margem de folga, tendo, porém, apresentado baixas taxas de compactação.

O gráfico a seguir mostra o tempo de processamento e a taxa de compactação do programa PKPAK, para os dados correspondentes aos itens “Todos os Temas” das tabelas 3 e 4.

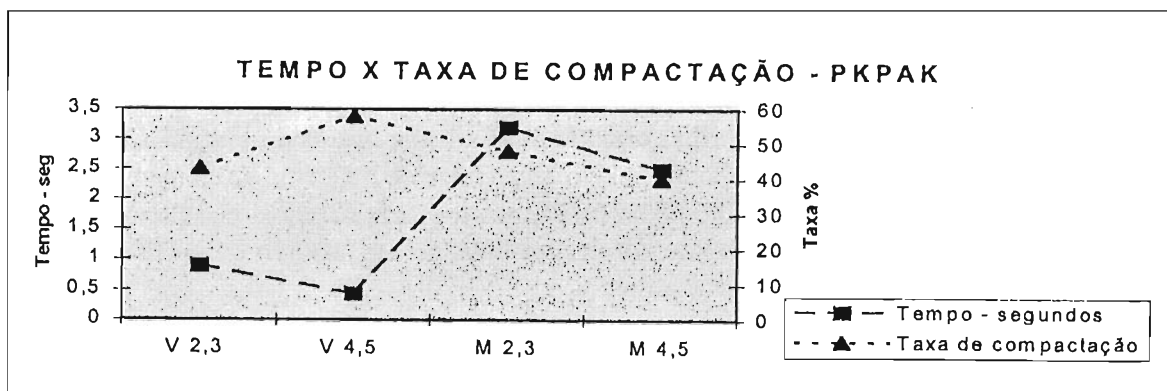


Figura 4 – V 2,3 – média de valores das áreas 2 e 3, arquivos vetoriais
 V 4,5 – média de valores das áreas 4 e 5, arquivos vetoriais
 M 2,3 – média de valores das áreas 2 e 3, arquivos matriciais
 M 4,5 – média de valores das áreas 4 e 5, arquivos matriciais

CONCLUSÕES

O programa PKPAK, foi o que apresentou maior eficiência em relação ao tempo de compactação.

O resultado da análise da eficiência em taxa de compactação dos programas testados foi:

- A Carta de Orientação Militar, aparentemente, apresenta em sua forma original dados de forma mais compactados do que os arquivos de folhas do Mapeamento Sistemático Brasileiro.
- Para os arquivos vetoriais o programa GZIP foi o que se mostrou mais eficiente, sendo que a exceção do PKZIP 1.10 e do PKPAK os outros programas apresentaram discrepância máxima de 1%, na taxa de compactação, em relação a obtida com o GZIP.
- Para os arquivos matriciais das áreas com altimetria densa o programa Squeeze foi o que se mostrou mais eficiente. Já para os outros arquivos matriciais (relevo suave) o mais eficiente foi o programa LHArc, sendo que também nesses casos a exceção dos programas PKZIP 1.10 e PKPAK, todos os outros apresentaram discrepância máxima de 1%, na taxa de compactação, em relação as obtidas pelos Squeeze e LHArc.
- Os programas PKZIP 1.10 e PKPAK geraram, sistematicamente em todos os testes, taxas de compactação discrepantes em mais de 4% em relação aos programas GZIP, Squeeze e LHArc.
- O programa PKZIP 2.04g aparenta ser uma boa solução geral, quando comparado com os programas mais eficientes, por ter apresentado tempos bem menores e taxas bem próximas. Entretanto os valores das tabelas 3 e 4 referem-se exclusivamente

aos comandos com taxa máxima de compressão, sendo que os programas GZIP, Squeeze e LHArc proporcionam tempos bem inferiores, próximos ao PKZIP 2.04g, na opção normal de compactação, sem perderem muito em termos de taxa.

- Para efeitos práticos, de forma geral, pode-se afirmar que os programas ARJ, GZIP, LHArc, PKZIP 2.04g, Squeeze e Zoo apresentaram a mesma eficiência (variação máxima de 1% na taxa de compactação) nos testes realizados, ressaltando-se que com o GZIP não se conseguiu compactar mais de um arquivo simultaneamente e que o arquivo original é destruído.

SUGESTÕES PARA TRABALHOS FUTUROS

Execução de testes em arquivos de maior volume de dados (folha completa do Mapeamento Sistemático Brasileiro, Imagem de Satélite, Fotografias Aéreas e outros);
Elaboração de curvas de taxa de compactação e tempo de processamento, que permitam interpolação e extrapolação dos resultados.

BIBLIOGRAFIA

[FOR95] – Fortes, L.P.S.; Oliveira, L. C. de; Ferreira L. F. Ensaio de Compactação de Dados Para a RBMC, 1995.

[PER94] – Perna, M.A.L., Módulo de Compactação de Imagens Discretas. Tese de Mestrado em Sistemas e Computação apresentada ao Instituto Militar de Engenharia, Rio de Janeiro, 1994.
