C. Troupin, S. Watelet,
A. Barth & J.-M. Beckers

GHER-University of Liège

Experiences with using netCDF 4

NetCDF = network Common Data Form (not format!)

NetCDF = network Common Data Form (not format!)

Wrong in these pages:

IPPC:
http://www.ipcc-data.org/help/formats.html
Ocean Color:
https://oceancolor.gsfc.nasa.gov/docs/format/l2nc/
NASA Earth Data:
https://earthdata.nasa.gov/user-resources/acronym-list

NetCDF = network Common Data Form (not format!)

Wrong in these pages:

IPPC:
http://www.ipcc-data.org/help/formats.html
Ocean Color:
https://oceancolor.gsfc.nasa.gov/docs/format/l2nc/
NASA Earth Data:
https://earthdata.nasa.gov/user-resources/acronym-list
$\rightarrow$ "*Good catch Charles! We will make the correction, thank you.*"

NetCDF = network Common Data Form (not format!)

Correct spelling according to Unidata Best practices:

netCDF: Original spelling of the name of the data model, API, and format.
CDF part capitalized in part to pay homage to the NASA "CDF" data model

netcdf: Used in certain file names, such as:
#include <netcdf.h>

NetCDF: Used in titles and at the beginning of sentences, where "netCDF" is awkward or violates style guidelines.

NetCDF = software libraries and self-describing, machine-independent data formats

OGC standards
: netCDF since **2011**
: Climate and Forecast (CF) extension since **2013**

Version 4
: released in ...

What's new?
: HDF5 as a storage layer
: use of groups
: user-defined types
: multiple unlimited dimensions
: compression
: data chunking
: parallel I/O
: ...

NetCDF = software libraries and self-describing, machine-independent data formats

OGC standards
: netCDF since **2011**
: Climate and Forecast (CF) extension since **2013**

Version 4
: released in **2008!**

What's new?
: HDF5 as a storage layer
: use of groups
: user-defined types
: multiple unlimited dimensions
: compression
: data chunking
: parallel I/O
: ...

NetCDF = software libraries and self-describing, machine-independent data formats

compression
data chunking

$\rightarrow$ benchmarks

Benchmark: WMS GetMap requests in Oceanbrowser

Surface concentration of ammonium in the Mediterranean

WMS GetMap is performed
and generate a 512 x 512 PNG image

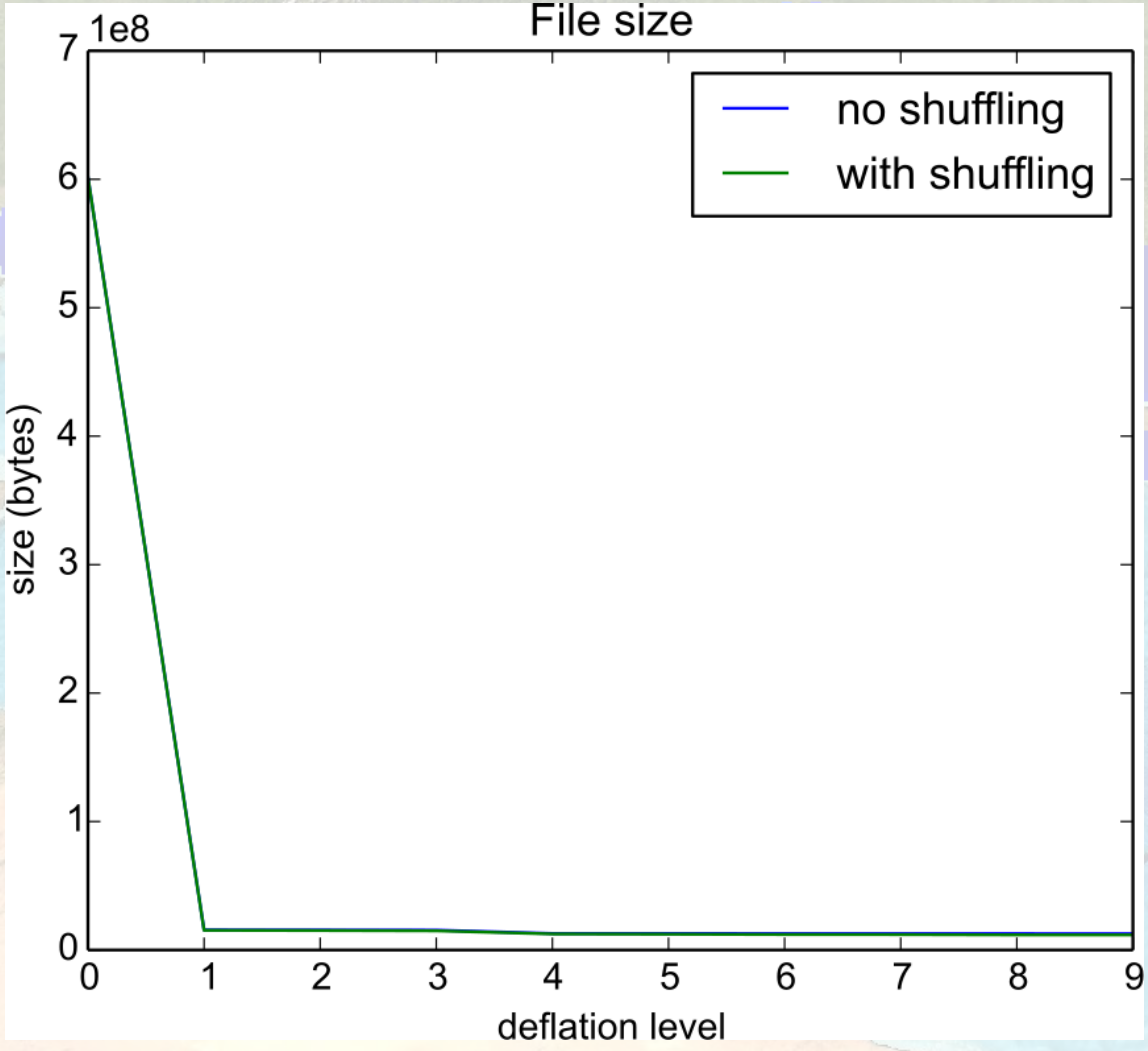
Data are chunked over time
(every time frame is compressed independently)

The image is generated 1000 times
and the median time is computed.

Deactivated WMS tile cache
(designed to optimise the delivery of images)

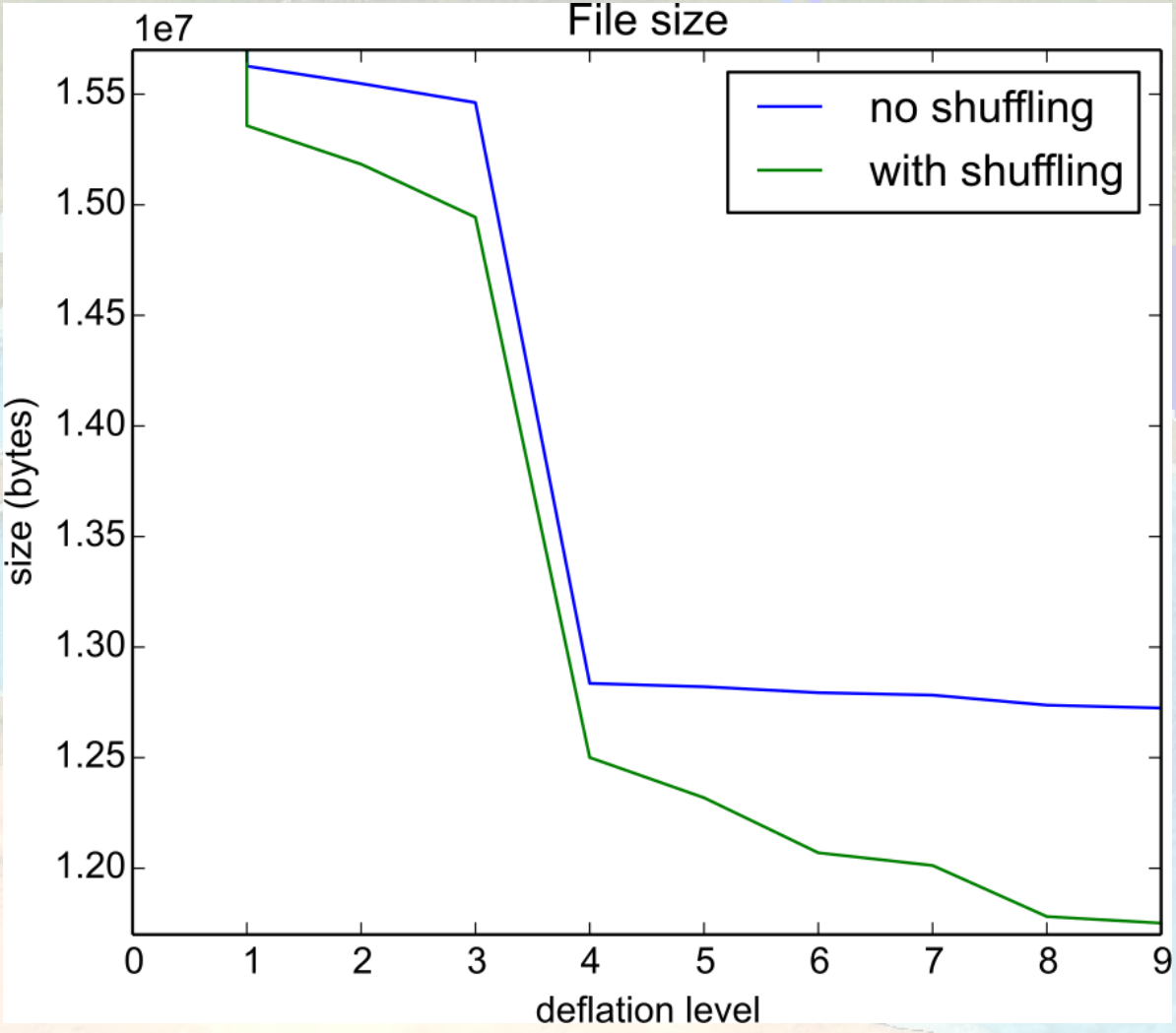
Benchmark: WMS GetMap requests in Oceanbrowser

File size vs. deflation level



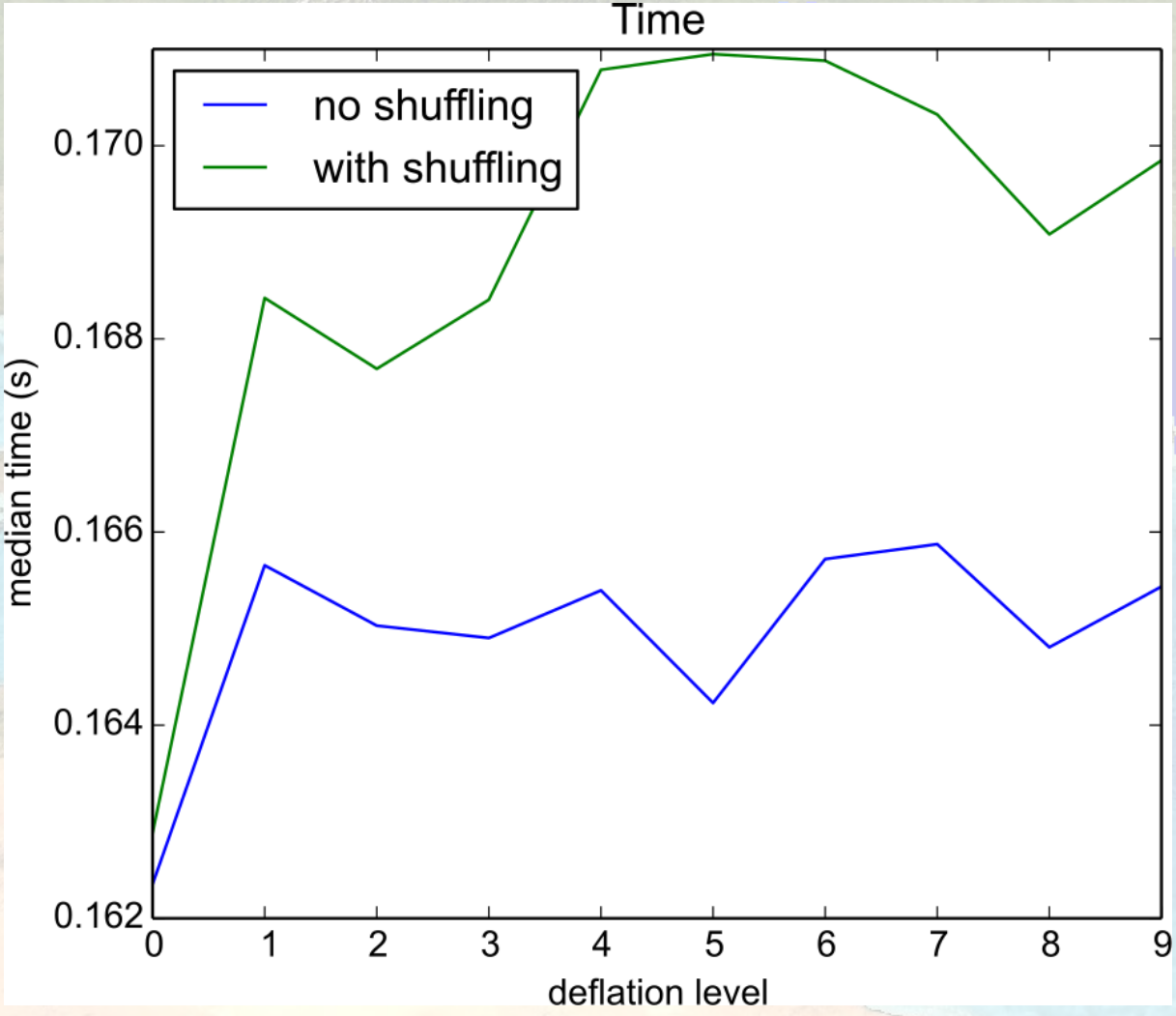
Benchmark: WMS GetMap requests in Oceanbrowser

Close-up view



Benchmark: WMS GetMap requests in Oceanbrowser

Time vs. deflation level



Benchmark: results

1. File size reduced by a factor of 38
   (from 574M to 15M)
   with deflation level 1
2. File size reduced by 20%
   at deflation level 4
3. *Shuffling* reduces the file size even more
4. Compression slightly increases WMS map generation time
   with shuffling: $<5\%$
   without shuffling: $< 2\%$

Checksum

Files getting bigger, need to be sure the file integrity

Need to assess how performance is affected

NetCDF

in languages and software tools

Programing languages: netCDF4 is there!

Language	Package/module installation
python	https://github.com/Unidata/netcdf4-python
Fortran	https://github.com/Unidata/netcdf-fortran
C	https://github.com/Unidata/netcdf-c
Java	https://github.com/Unidata/thredds
JS	https://www.npmjs.com/package/netcdf4
GNU Octave	https://github.com/Alexander-Barth/octave-netcdf
julia	https://github.com/meggart/NetCDF.jl
Matlab	Native support since R2010b

Visualisation and analysis tools: comparison

Tools

IDL	ODV	IDV	(arc)GIS	DIVA	divand
Diva-on-web	SURFER	Ferret	GNU R	Panoply	

Visualisation and analysis tools: comparison

Error estimation

IDL	ODV	IDV	(arc)GIS	DIVA	divand
No	Yes	No	No	Yes	Yes

Diva-on-web	SURFER	Ferret	GNU R	Panoply	
Yes	No	No	No	No	

Visualisation and analysis tools: comparison

Interpolation or analysis?

IDL	ODV	IDV	(arc)GIS	DIVA	divand
Int./Anl.	Int./Anl.	Int.	Int./Anl.	Anl.	Anl.
Diva-on-web	SURFER	Ferret	GNU R	Panoply	
Int./Anl.	Int./Anl.	Int.	Int./Anl.	Anl.	

Visualisation and analysis tools: comparison

Maximum dimensions

IDL	ODV	IDV	(arc)GIS	DIVA	divand
2D (3D if inv. dist.)	2D	2D	2D	2D	nD

Diva-on-web	SURFER	Ferret	GNU R	Panoply
2D	2D	4D	2D	2D

Visualisation and analysis tools: comparison

netCDF as input

IDL	ODV	IDV	(arc)GIS	DIVA	divand
Yes	Yes	Yes	Yes	Yes	No

Diva-on-web	SURFER	Ferret	GNU R	Panoply	
No	Yes	Yes	Yes	Yes	

Visualisation and analysis tools: comparison

netCDF as output

IDL	ODV	IDV	(arc)GIS	DIVA	divand
Yes	Yes	No	Yes	Yes	Yes

Diva-on-web	SURFER	Ferret	GNU R	Panoply	
Yes	Yes	Yes	Yes	Yes	

Visualisation and analysis tools: comparison

OGC compliance

IDL	ODV	IDV	(arc)GIS	DIVA	divand
Yes ?	?	No ?	Yes	?	?

Diva-on-web	SURFER	Ferret	GNU R	Panoply	
Yes	?	?	?	No?	

Visualisation and analysis tools: summary

Table available at

https://github.com/gher-ulg/ODIP/blob/master/netCDFtools.md

Gridding tools	IDL	ODV	IDV	(arc)GIS	DIVA	divand
Gridding techniques	Nearest neighbours, inverse distance, bilinear, polynomial, spline, natural neighbours, kriging	Inverse distance, Variational Inverse Method (VIM)	Inverse distance	Inverse distance, polynomial, spline, natural neighbours, kriging	Variational Inverse Method	Variational Inverse Method

Visualisation and analysis tools: summary

1. Many possibilities for quick visualisation of NetCDF
2. Not easy to assess OGC compliance
3. Most of these software tools can work with simple text files
4. The majority can deal with netCDF4 (import/export)
5. 8 out of 11 software tools can also import netCDF via OPeNDAP

Conclusions

Reasonable trade-off: level-5 compression without shuffling

Implication: users downloading directly the netCDF files need to have the netCDF4 and HDF5 libraries with compression enabled in order to be able to read them

Final message:

push for the use of netCDF-4