Geoscientific
Model Development
Discussions

Open Access

EGU

Interactive comment on "Flex_extract v7.1 – A software to retrieve and prepare ECMWF data for use in FLEXPART" *by* Anne Philipp et al.

Anonymous Referee #3

Received and published: 16 June 2020

This manuscript describes the newest version, 7.1, of the flex_extract software. This piece of open source software is employed by a broad community of users of both FLEXPART and FLEXTRA in order to retrieve ECMWF meteorological fields required to run these models. It is very encouraging to see that in parallel to an adaptation and enhancement of functionalities of flex_extract, professional software quality assurance methods have been implemented in the most recent version of the tool. In addition to the unit and regression testing, modern code metrics have been used to assess notable improvements of the newest release with respect to the previous one, due to a reduced complexity. Cumbersome as it may seem, this is certainly the way the scientific software development needs to undertake in the times of continually growing codes and increased collaborative effort required. The tool itself is very versatile, in terms of

different mode of data accessibility, streams of input ECMWF data, including backward compatibility and anticipating future releases, grid-spacing and temporal frequency, which can be configured. I could not emphasise enough the importance of an early transition towards python3. Often, in the scientific community, we find ourselves with legacy issues trailing far too long with respect to the contemporary work environment. In case of flex_extract these issues are addressed in a timely manner and I would like to congratulate software developers on it.

I must admit that when I downloaded the manuscript for the first time and saw 62 pages, I sighed. But the manuscript is not as voluminous as the number of pages may indicate as the Authors included many diagrams and tables to support the content. The manuscript, effectively containing about one thousand lines of text, is very pleasant to read, the content and the logical flow of information keeps a reader focussed and is presented in smooth English. The paper is very well structured, informative, broader context of this work is described, some historical aspects are addressed and their connection to legacy issues explained. A new way of disaggregating precipitation fields is a noteworthy scientific development. Information on the sources of the ECMWF meteorological fields and advice on how to intertwine forecast and analysis fields as well as on the importance of maintaining consistency between a target application and horizontal and temporal resolution are very valuable, especially for for new end users. The content of Tab.1, Tab.2 and Tab.3 is a treasure. Respect to the Authors for gathering this information. Included are clear instructions to follow for software installation and usage. The tool as well as the instructions make an impression of being user-friendly. There are illustrative diagrams supporting the information described in the text.

I recommend the manuscript for publication after the following issues and typos have been addressed:

l. 5 I suggest replacing 'therefore' with 'consequently' to avoid a repetition with the previous sentence

l. 22 should be 'Integrated Forecast System'

l. 41 I believe you want ECMWF here

l. 58 should read '1990s'

l. 62 You probably want 'intersect' and not 'interSect.'

l. 99 Should read US National Centers

l. 107/108 – the two reference should be in brackets

l. 116 reference should be in brackets

l. 146 reference should be in brackets

l. 155 suggest skipping 'it' or better replacing it with 'the'

l. 169 'implied; instead of 'inplied', I believe

l. 181 missing 'n' in 'ecessary'

l. 184 - 187 and then also l. 206 – 209: are 'access mode' and 'application mode' two different things? Are 'ECMWF Member States Linux servers', 'Member State Gateway server' and 'local host' rather access modes than application modes? And based on those three access modes, four application modes are derived? Maybe it is then clearer in conjunction with the header of section 3.1, where 'Remote', 'Gateway' and 'Local' are referred to as application modes. And then in conjunction with user mode it results in four user application modes.

l. 187 results

l. 227 – depends which spelling you follow; if it is BE then 'licence'; but I noticed indeed that Copernicus Publications follow AmE spelling of this word, namely 'license'

l. 240 – I may be missing something important in the software structure but I do not understand why sending a script to the ECMWF batch queue in step 1 contrasts in

steps to and 3; in particular do not understand 'or' here; I would imagine that that job sent to the batch queue also retrieves data from MARS and post-processes them to obtain FLEXPART input fields. From l. 249-252 I understand that flex_extract proceeds with steps 2 and 3 locally; is it correct? It is a small thing but it would be good to clarify

l. 268 – you want to use 'which' once

l. 283 – should read 'for a correct setting'

l. 284/285 – you probably want to use the word 'combination' only once

l. 355/356 – these papers should be referenced in brackets

l. 363 – does it stem from private communication with Paul James?

l. 373/374 – what is the position of 1 and 2 with respect to a,b,c,d

l. 416 should read 'pressure'

l. 504 – what is a pure forecast? Do you mean deterministic forecast? 'Pure forecast'also later appears in l.527

l. 513/514 – not sure what this sentence mean? Is there just 'is' missing in this sentence? Or is there more to add?

l. 524 - is YYMMDDHH in this file name the analysis hour?

l. 529 would 'base time' be better than 'starting forecast time'

l. 535 – I may be missing something here but why the file names selected for the ensemble members do not account for FORECAST_STEP?

l. 548 – should read 'makes'

l. 581 – I would skip the coma after e.g.

l. 615 – do you mean 'paths'?

l. 624 – should read 'except'

l. 687 – better 'indicates a lower complexity'

l. 720 'Lagrangian' should be one word

l. 721 – 1990s

l. 940 – skip 'an'

l. 942 and 943 are too tightly formatted in vertical (also true for l. 848 and 849 and l. 888 and 889 then for lines 966, 967 and 968; subsequently lines 982, 983 and 984 have different vertical formatting than the rest of the manuscript; l. 1062 and 1063 are too close in vertical as well)

l. 954 you may want to insert a blank space after the bracket

In the caption of Tab.2 'resolution' instead of 'resoltion'

In the body of Tab.6 – should be 'Public users have to register for obtaining an account'

In the caption of Tab. 18 you need to insert a blank space after etadot

In the caption of Tab. 24 it is difficult to understand the first sentence starting from: 'Python code'; could you, please, re-phrase. I would also suggest putting sections v7.0.4 and v7.1 of this table side by side and not one on top of the other, if possible (I am aware there is no strict correspondence between theses two sections)