



User awareness concerning feedback data and input observations used in reanalysis systems

H. Gregow¹, P. Poli², H. M. Mäkelä¹, K. Jylhä¹, A. K. Kaiser-Weiss³, A. Obregon³, D. G. H. Tan², S. Kekki¹, and F. Kaspar³

¹Finnish Meteorological Institute, P.O. Box 503, 00101 Helsinki, Finland

²European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, RG2 9AX, UK

³Deutscher Wetterdienst, Frankfurter Straße 135, 63067 Offenbach, Germany

Correspondence to: H. Gregow (hilppa.gregow@fmi.fi)

Received: 21 January 2015 – Revised: 25 March 2015 – Accepted: 10 April 2015 – Published: 22 April 2015

Abstract. A web-based survey to assess the benefits and gaps in reanalyses as part of growing climate services was carried out in 2013–2014. The survey elicited responses from about 2500 users of climate information. One of the eleven survey points specifically addressed the observations used in reanalysis, with a multiple-choice question “Have you used reanalysis input observations and feedback data?”. Almost half of the respondents admitted to not knowing what such data were about. Among the others, specific queries asked for these observations to be made available more openly. This paper summarizes the main findings in regard to use of existing reanalyses as well as user awareness and needs in regard to reanalysis feedback data and input observations. In the future, the information obtained via the survey makes it possible to perform various statistically robust analyses addressing different aspects of the use of reanalysis data.

1 Introduction

Users are generally attracted to reanalyses because these are constructed from observations while providing temporally and spatially complete representations of some components of the Earth System, something that observations alone cannot do. Reanalysis input observations and feedback data are relatively well-known within the data assimilation and reanalysis communities. The input observations only tell about the possible information from observations that went into reanalyses. The feedback contains in addition quantitative measures of the agreement between observations and model fields, either before or after assimilation or bias correction. Analysis of the feedback data is routinely carried out to assess the quality of a reanalysis. It is also useful to assess the quality of observations. For the user, feedback data can be of high practical value, e.g. to reveal how well frequency distributions or time series of observed and reanalysed parameters are matching.

To know how aware the users are of reanalysis characteristics, a web-based survey was conducted in winter 2013–2014 as part of the EU FP7 CORE-CLIMAX project. The survey

was advertised by email to over 20 000 registered users of ECMWF reanalysis data and on several websites. The aim of the enquiry was to collect information on the users of reanalyses, their awareness of the maturity and limitations of various reanalysis products as well as their opinions about climate services.

A previous survey about reanalyses had been conducted in 2004–2005 by ECMWF (Hollingsworth and Pfrang, 2005) on ERA-40 users (Uppala et al., 2005). Most of the 127 respondents had given positive feedback about the quality and accessibility of the ERA-40 data. That survey had revealed that users desired increased resolution, longer time spans and more regular extensions of the time series to the present. Based on this feedback, ERA-Interim (Dee et al., 2011) and ERA-20C (Poli et al., 2013) were developed.

Central to the reanalysis process is the generation of quantitative differences between observational data and model-simulated equivalents, together with associated indicators of observation quality. Reanalysis producers are developing ways to make such “observation feedback” information publicly available. The ISPD dataset includes feedback from the

20CR reanalysis (Compo et al., 2011), while the MERRA Gridded Innovations and Observations (GIO) dataset is generated by re-gridding observation-space feedback from the MERRA reanalysis (Rienecker et al., 2011). At ECMWF, observation feedback from the ERA-20C reanalysis is available (Poli et al., 2013) and feedback from the ERA-Interim reanalysis is being prepared. Anecdotal experience has suggested that user uptake of the feedback datasets is growing but awareness remains patchy. We therefore chose to survey user awareness of observation feedback in the CORE-CLIMAX questionnaire. This paper will thus present novel results from the CORE-CLIMAX survey regarding users' awareness of reanalysis feedback data and input observations.

2 Methodology

Over a one-year process we devised a concise set of survey points (limited to eleven to encourage respondents to complete the survey in its entirety) (Fig. 1). Through these survey points (accompanied by pre-prepared multiple-choice response options) we sought to ascertain how widely and well used reanalysis products currently are compared to other sources of data, and what problems were to be solved. To obtain as many responses as possible, we advertised the survey via national networks of weather and environmental agencies, WMO regional offices, reanalysis user networks, and two emailing campaigns that turned out to be extremely effective. The web portal enquiry was linked to several websites, including CORE-CLIMAX webpage <http://www.coreclimax.eu/>, the reanalysis web portal <http://reanalysis.org> hosted by NOAA, the Japanese Meteorological Agency's website, the Deutscher Wetterdienst's (DWD) website and the Finnish Meteorological Institute's (FMI) website. Overall, the survey was distributed to regional meteorological offices around the world with the help of WMO and to universities, research institutes and Copernicus User Forum members in Finland. ECMWF contacted 23 957 of their registered reanalysis users twice, prompting around 1000 responses on each occasion. In total 2578 respondents participated in the survey. Thus, users of ECMWF reanalysis products form the bulk of the respondents. This is important to keep in mind when interpreting the results.

Use of reanalysis input observations and feedback data was surveyed by presenting the respondents with 11 ready-made statements, and asking them to "select all that apply". Responses from 2473 respondents were received for this survey point. To analyse the results, the 11 statements were grouped into 5 categories (Table 1) and the responses were consolidated accordingly. Categories 1, 2 and 4 combined responses from two or more of the initial 11 statements. Because respondents could choose more than one statement, these aggregated categories could have included more than one response from each person. However, we took care to re-

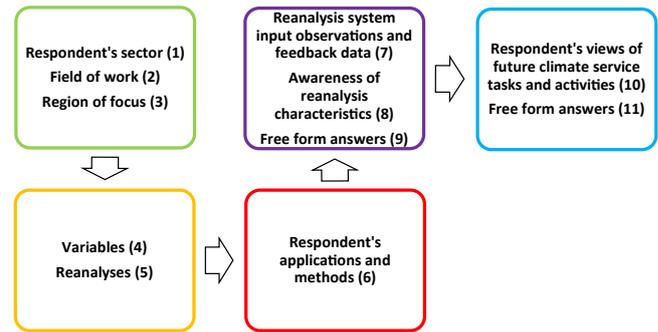


Figure 1. Flowchart of the questionnaire grouped into five main focus areas: respondent's specific background (green), data (yellow), methods (red), awareness and needs concerning reanalysis properties (purple) and future climate service (blue).

move such multiple responses in the consolidation process. To keep the analysis straightforward, we included only those respondents who belonged to only one of the five categories. The resulting number of respondents was 2331 which is 94 % of all the respondents in this section of the survey.

In addition to analyses concerning all respondents, it is of interest to consider smaller subgroups composed, e.g., based on the field/subject of work of the respondents. As an example, we present here responses of four subgroups consisting of those respondents who indicated their field/subject of work to be "fresh water resources and management" (WAT; 172 respondents), "agriculture and food production" (AGR; 139), "forests" (FOR; 113) or "energy" (ENE; 249). Responses of these subgroups were compared with those from all the respondents (ALL; 2331).

3 Results

3.1 Reanalyses used and fields of work of respondents

Among the respondents, the most widely-used reanalysis data sets were the Global ECMWF Interim Reanalysis (ERA-Interim) (opted by 79 % of 2502 respondents), the Global ECMWF 40-year Reanalysis (ERA-40) (51 %), and the Global NCEP/NCAR Reanalysis I (R1) (1948 to present) (39 %) (Fig. 2). Many respondents use more than one reanalysis, resulting in a total of 7597 declared instances of reanalysis use. The share of ERA-Interim was almost half (48 %) of the declared instances. This comes without surprise as about 2000 respondents participated after the emailing campaigns from ECMWF, suggesting these were already users of some ERA products. Compared to the atmospheric reanalyses, the oceanic counterparts were less widely used by the respondents of this survey. The most widely-used oceanic reanalysis data sets were NCEP Global Ocean Data Assimilation System (GODAS) (7 %), and ECMWF Ocean Reanalyses ORA S4 and ORA S3 (5 % each) (Fig. 3). The most common fields of work were climate (opted by 73 % of the

Table 1. The five main categories and the different statements surveyed in regard to reanalysis input observations and feedback data (“it” refers to reanalysis input observations and feedback data).

| | |
|----|---|
| 1. | I have used “it” to ... (combined the following four original statements:) |
| | a. assess the reanalysis data using observations as a reference; |
| | b. as above but the other way around; |
| | c. merge the observations and reanalysis data together to create an improved product; |
| | d. understand how the observations had been used by reanalysis. |
| 2. | I have not used “it” because ... (combined the following three original statements:) |
| | a. the data files are too big; |
| | b. the data formats are too complicated; |
| | c. there is no easy interface to get these data. |
| 3. | I could not find “it”. |
| 4. | I have had no time or resources or interest to look into “it” (combined two original statements). |
| 5. | I do not know what “it” is about. |

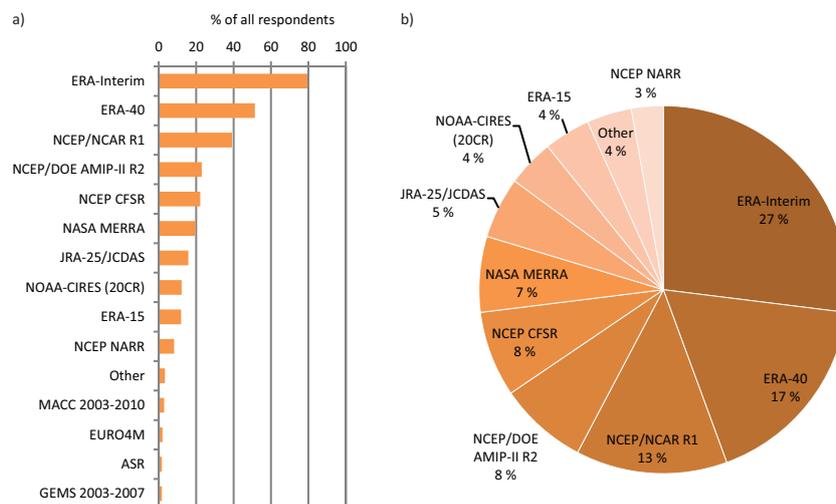


Figure 2. Atmospheric reanalysis data sets that the respondents most often used. **(a)** Percentages of all respondents (2502). **(b)** Percentages of all declared instances of atmospheric reanalysis use (7597) given by respondents. The sector “Other” includes the four least used reanalysis data sets as well as the option “other”. It was asked to choose all that apply.

respondents), weather (47 %), and oceans and seas (25 %). The share of the three above-mentioned fields was more than half (56 %) of all the votes given for altogether 28 different choices.

3.2 User awareness of feedback data and input observations with field-wise examples

Almost half of the respondents declared not to know what is meant by reanalysis input observations and feedback data that are used in the reanalysis systems (Fig. 4). In relative terms, users in the subgroups FOR and WAT were the most acquainted with the issue, but even in those groups the share of them not knowing what it is about was around 40 %. In ALL and ENE the share was roughly 50 %. Category 1 en-

compasses those responses that declared use of reanalysis input observations and feedback data for one or more purposes. In percentage terms, Category 1 instances were largest for AGR (37 % of the respondents) and the smallest for ALL (24 %). The most common use of reanalysis feedback data was to assess reanalysis data using observations as a reference (72 % of those in ALL choosing at least one of the options a–d in Category 1 (598 respondents)). Percentages for the other options in Category 1 were as follows: 38 % used the data to assess observations using reanalysis as a reference, 26 % used it to merge observations and reanalysis data together to create an improved product, and 24 % used it to understand how the observations had been used by reanalysis. The percentages varied somewhat for the four subgroups but the order of the options was the same as for ALL.

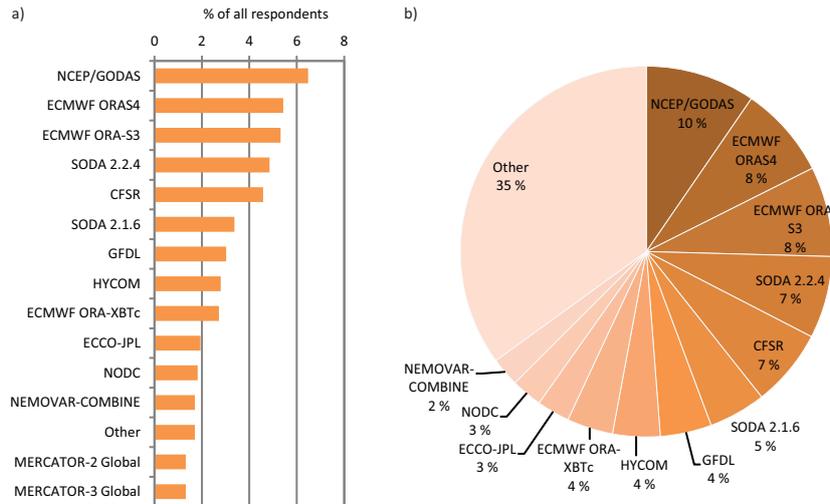


Figure 3. Oceanic reanalysis data sets that the respondents most often used. (a) Percentages of all respondents (2502). Only the 15 most used reanalyses are shown. (b) Percentages of all declared instances of oceanic reanalysis use (1746) given by respondents. The sector “Other” consists of over 40 oceanic reanalysis data sets, whose share from all given votes was less than 2% each. It was asked to choose all that apply.

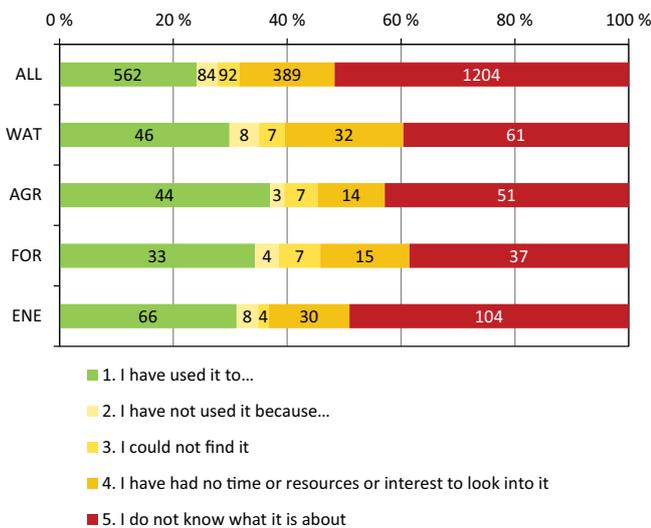


Figure 4. Use of reanalysis input observations and feedback data by all respondents and by the four subgroups. Categories 1–5 are explained in more detail in Table 1. The numbers show the number of the respondents in each category.

Given that provision of reanalysis input observations and reanalysis feedback data is still relatively new, the number of responses falling into Category 1 is somewhat higher than we had expected. In retrospect, we believe that respondents may have interpreted the term “reanalysis input observations” as meaning “any observational dataset”. We therefore allow for the possibility that feedback uptake is actually substantially lower than 24–37% (perhaps by factors of 2 or more).

Roughly 5% of the respondents reported that they had not used the data for one reason or another (Category 2). The most reported factor limiting the use of the data was that the respondents felt that there is no easy interface to get the data (75% of those in ALL choosing at least one of the options a–c in Category 2 (157 respondents)). Users also reported that data formats are too complicated (36%) and that the files are too big (26%). The share of those who did not find the data at all (Category 3) varied from 2% (ENE) to 7% (FOR). The proportion of those who simply have had no time, resources or interest to look into it (Category 4) ranged from 12% (AGR) to 21% (WAT).

This analysis of the use of reanalysis feedback data examined those respondents who belonged to one category only. However, 4% (ALL) to 8% (FOR and ENE) of the respondents belonged to two categories. Typically, these respondents had chosen two of the statements falling into the Categories 2–5 indicating that they do not know the issue and/or they have not used or found the data for a reason or another. The share of respondents belonging to three or more categories was mostly less than one percent for each study group.

4 Conclusions

This work has provided quantitative indications that many users of climate information remain unaware of the availability of reanalysis feedback data and input observations, and that uptake of feedback data is currently rather low. It was found that the most important factors limiting the use of the feedback data and input observations, presuming that one knew what it was about, was that the respondents felt that there is no easy interface to get the data or they did not

find it at all. This is indeed very true of many reanalyses but not so of others like MERRA or 20CR. In any case, this is a very important message to the reanalysis community. The EU FP7 ERA-CLIM project has contributed to improve this with the development of an Observation Feedback Archive at ECMWF, but substantial content such as ERA-Interim or ERA-20C observation feedback was not yet available online at the time of the survey. Assuming this content becomes available in the near-future, there remains the issue that there is a significant learning curve for users to take up these new products. To resolve this, we encourage the relevant communities to pool and invest resources to develop tools and provide training that will bridge the gap between current capabilities and comprehensive exploitation of reanalysis observation input and feedback.

Because of the success of obtaining so many responses (2578 check-box answerers and approximately 1000 free comments) further work is needed to identify in more detail the benefits of different reanalyses and their characteristics. Additionally, depending on the amount of field-wise and sectorial responses (the examples shown here are for the fields of fresh water resources and management, agriculture and food production, forests, and energy), we may be able to make statistically robust analyses of user needs in other sectors/fields, and/or in distinct geographical regions. These are planned to be the focus of our future research.

Author contributions. Hilppa Gregow (PI of WP5 in FP7 CORE-CLIMAX), Kirsti Jylhä, Satu Kekki, Andrea Kaiser-Weiss, Andre Obregon, Frank Kaspar, Paul Poli, David Tan designed, prepared and tested the survey and launched it in November 2013. Analyses in the manuscript were conducted by Hilppa Gregow, Paul Poli, Hanna Mäkelä, and Kirsti Jylhä. Hilppa Gregow prepared the manuscript with contribution from all co-authors.

Acknowledgements. This work was financially supported by the EU FP7 CORE-CLIMAX project. We thank Carsten Maass from ECMWF for arranging emailing about 24 000 ECMWF users, Karolin Eichler from WMO for helping distribute the survey to the WMO regional offices, Gil Compo from CIRES for advertising the survey in the United States, and Shinya Kobayashi for advertising the survey in Japan. We are very grateful to all of the respondents for giving their time to take part in the survey.

Edited by: E. Bazile

Reviewed by: B. Kurnik and G. Verver

References

- Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.: The Twentieth Century Reanalysis Project, *Q. J. Roy. Meteor. Soc.*, 137, 1–28, doi:10.1002/qj.776, 2011.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. Roy. Meteor. Soc.*, 137, 553–597, 2011.
- Hollingsworth, A. and Pfrang, C.: A preliminary survey of ERA-40 users developing applications of relevance to GEO (Group on Earth Observations), *ECMWF Newsletter*, 104, 5–9, 2005.
- Poli, P., Hersbach, H., Tan, D., Dee, D., Thépaut, J.-N., Simmons, A., Peubey, C., Laloyaux, P., Komori, T., Berrisford, P., Dragani, R., Trémolet, Y., Holm, E., Bonavita, M., Isaksen, L., and Fisher, M.: The data assimilation system and initial performance evaluation of the ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C) ERA Report Series 14 available from ECMWF, Shinfield Park, Reading RG2 9AX, UK, available at: <http://www.ecmwf.int/en/research/publications>, last access: 16 April 2015, 2013.
- Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, *J. Climate*, 24, 3624–3648, doi:10.1175/JCLI-D-11-00015.1, 2011.
- Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, *Q. J. Roy. Meteor. Soc.*, 131, 2961–3012, 2005.