



ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

Towards Harmony between CANARI and SURFEX

E. Kurzeneva

Joint 29th ALADIN Workshop &
HIRLAM All Staff Meeting
2019, 1-5 April 2019, Madrid

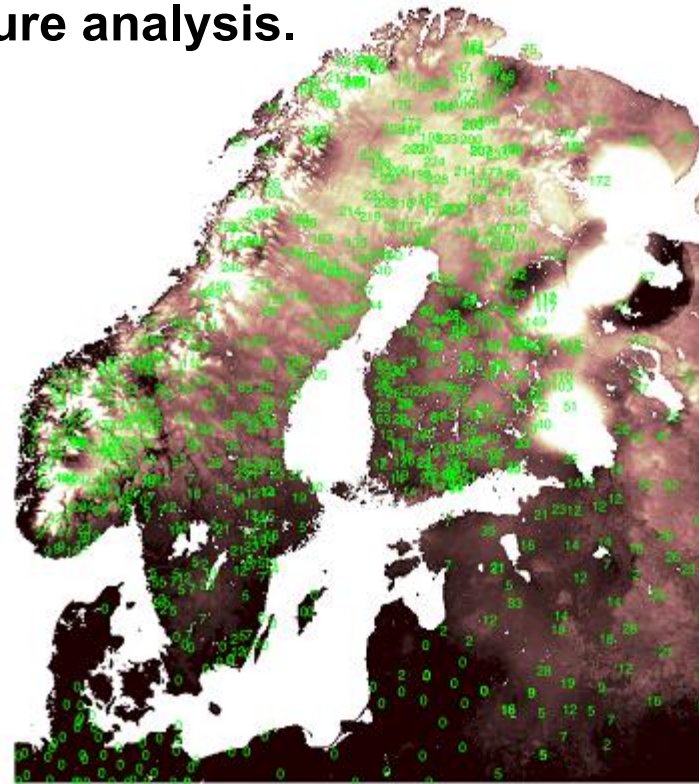
2.4.2019



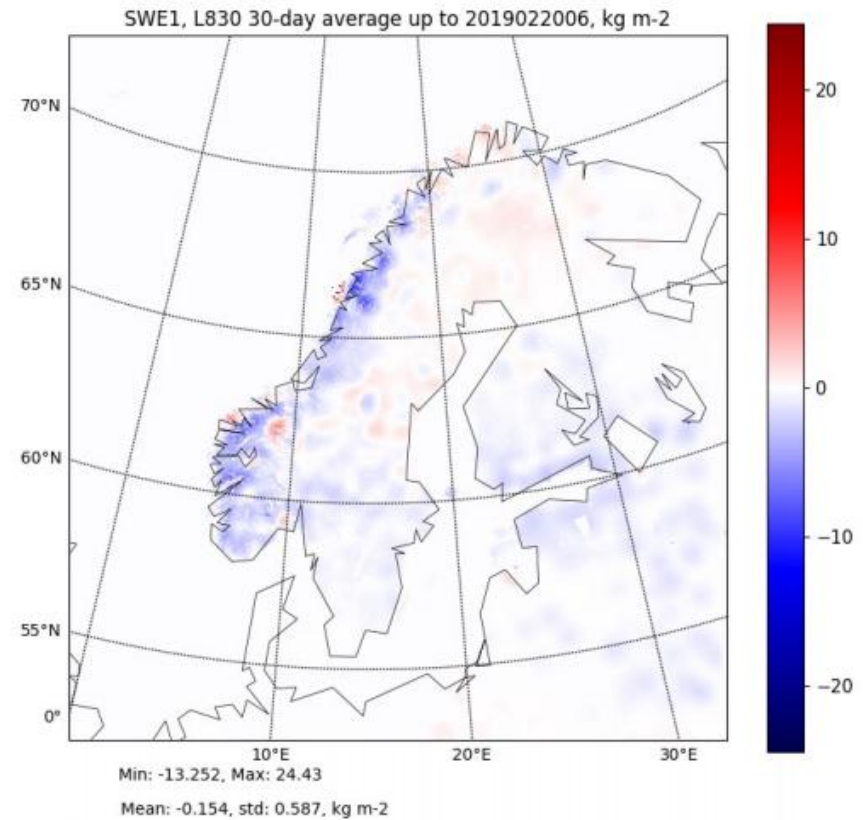
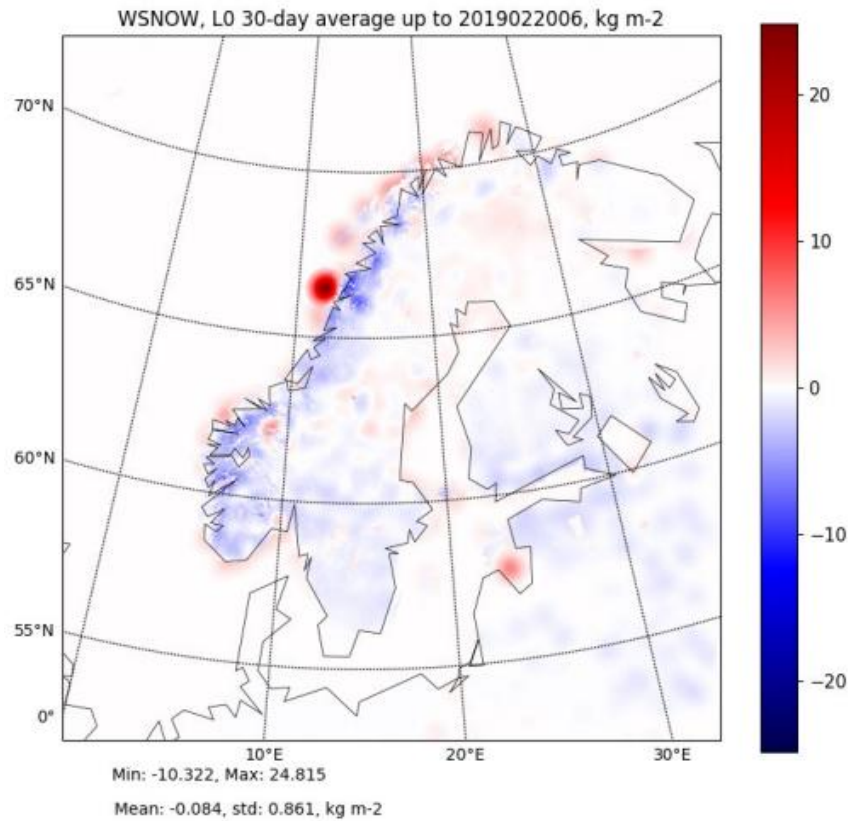
Motivation

Practical: the unrealistic features noticed in the operational runs of FMI before 2016, in MetCoOp operational runs and in experimental runs over MetCoOp domain in the SWE and deep soil temperature analysis.

MetCoOp (cy40h1.1) operational SWE analysis field (the background color) and the SWE for the SYNOP stations calculated from the snow depth observations using the climatological values of the snow density (numbers in green), kg/m², for 19.02.2016, 06 UTC



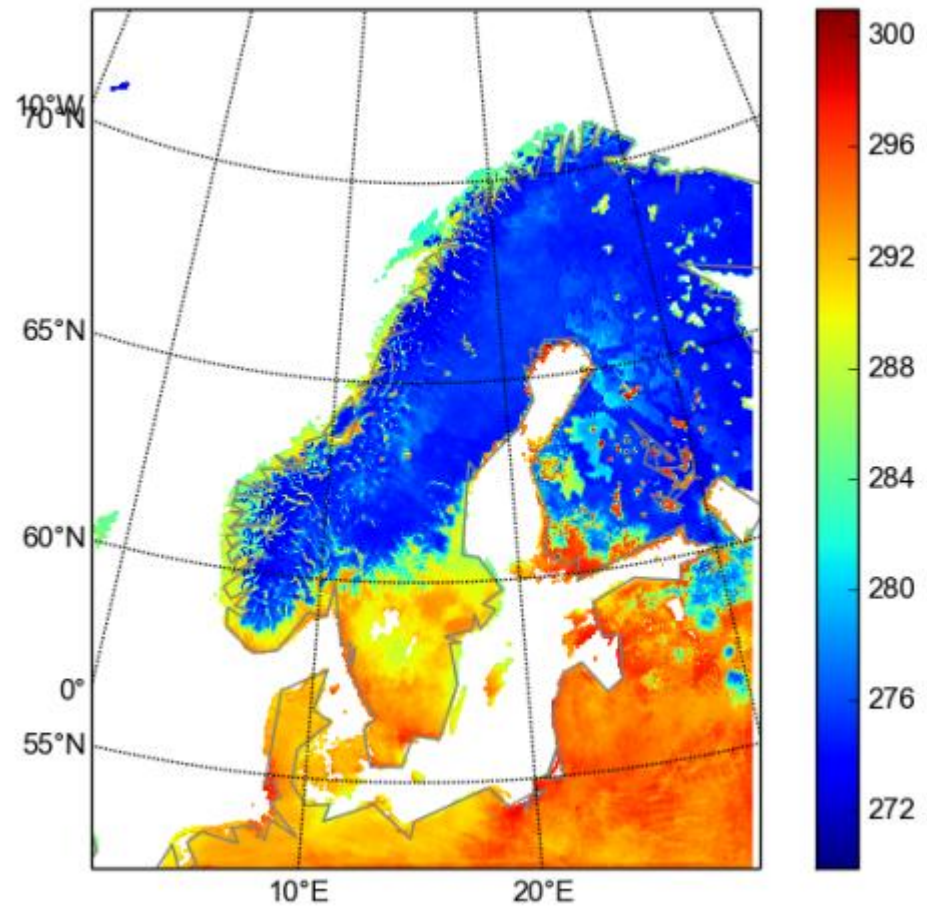
by courtesy of M. Quenon



Increments from the snow analysis for the operational MetCoOp (cy40h1.2) averaged for the period of 20.01.2019-20.02.2019, from CANARI (right) and used by SURFEX (left).

by courtesy of C. Fortelius

The deep soil temperature analysis field, K, in the research experiment of Harmonie cy40h1.1 over the MetCoOp domain, for 01.06.2016, 00 UTC. The experiment used the standard Harmonie configuration and started at 15.04.2016



experiment run by P. Samuelsson

Motivation

General:

Need of further development of the surface analysis:

- * the snow analysis
with different kinds of satellite data
- * the lake analysis
- * the analysis for urban tile?

- * the full use of the MESCAN options
(at the moment, after a lot of tests,
we are not sure ...)



**Tromsø, June 2018, strong opinion:
CANARI is a terrible code.**

What to do with CANARI?

Why CANARI is so terrible?

Why?

For the HIRLAM community, it is *someone else'* code:

- Contains parts of IFS code, ARPEGE and ALADIN code
- The main expert, F. Taillefer, passed away ... whom to ask?
- Documentation exists, but distributed in different locations and difficult to find, contains many errors and corrections
- Sometimes the authors (from ECMWF, Météo-France) are not easy to contact, because they have another versions of the system
- Initially was a multi-variate upper-air DA code
- Not externalized, very dependent on the model itself
- Contains many obsolete features and unused code
- Just very large

Consequences:

- Coding reminds a detective work or struggling with monsters



... however:

- All these problems do not mean that the code itself is bad!
- The code has a complicated but good structure
- The difficulties are psychological rather than scientific or technical.

For us, it is Legacy code



THREEPANELSOUL.com matthew boyd - ian mcconville



The screenshot shows the Wikipedia page for "Legacy code". The browser address bar displays "https://en.wikipedia.org/wiki/Legacy_code". The page title is "Legacy code" with the subtitle "From Wikipedia, the free encyclopedia". A prominent yellow warning box at the top states: "This article **needs additional citations for verification**. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. Find sources: "Legacy code" – news · newspapers · books · scholar · JSTOR (February 2019) (Learn how and when to remove this template message)".

The main text defines legacy code as source code that relates to a no longer supported or manufactured operating system or other computer technology. It notes that the term was first used by computer scientist George Olivetti to describe code maintained by an administrator that did not develop the code. The text explains that the term can also mean code inserted into modern software for the purpose of maintaining an older or previously supported feature – for example supporting a serial interface even though many modern systems do not have a serial port. It may also be in the form of supporting older file formats that may have been encoding in non-ASCII characters.

In practice, most source code has some dependency on the platform for which it is designed – even if a programmer uses a platform-independent programming language like Java, it is hard to write a large, useful program that is totally independent of its environment. When the manufacturer upgrades a platform (or the platform is superseded), the code may no longer work without changes, and becomes legacy code. A large part of the task of a software engineer is to continually alter code to prevent this.

While the term usually refers to source code, it can also apply to executable code that no longer runs on a later version of a system, or requires a compatibility layer to do so. An example would be a classic Macintosh application which will not run natively on Mac OS X, but runs inside the Classic environment, or a Win16 application running on Windows XP using the Windows on Windows feature in XP.

Modern interpretations [edit]

More recently, the software engineering community has developed other interpretations for the term *legacy code*. Among the most prevalent are *source code inherited from someone else* and *source code inherited from an older version of the software*. Michael Feathers^[2] introduced a definition of *legacy code* as *code without tests*, which reflects the perspective of legacy code being difficult to work with in part due to a lack of automated regression tests. He also defined *characterization tests* to start putting legacy code under test.

**Just recognizing of the problem may help;
also by management.**

Merits and issues of CANARI

- Effectively parallelized
- Uses ODB to work with different types of observations
- Support of IFS cycles

- Uses its own physiography, which is old, coarse and inconsistent with SURFEX (the same with the upper air analysis)
- Does not use the tiling approach
- Observation operator is assumed to depend only on the grid geometry. It is applied for the obs report. It is not variable-specific
- In the observation operator, only the land-sea mask is used
- In the observation operator, CANARI runs it's own diagnostic (of T2m and RH2m)



Current status

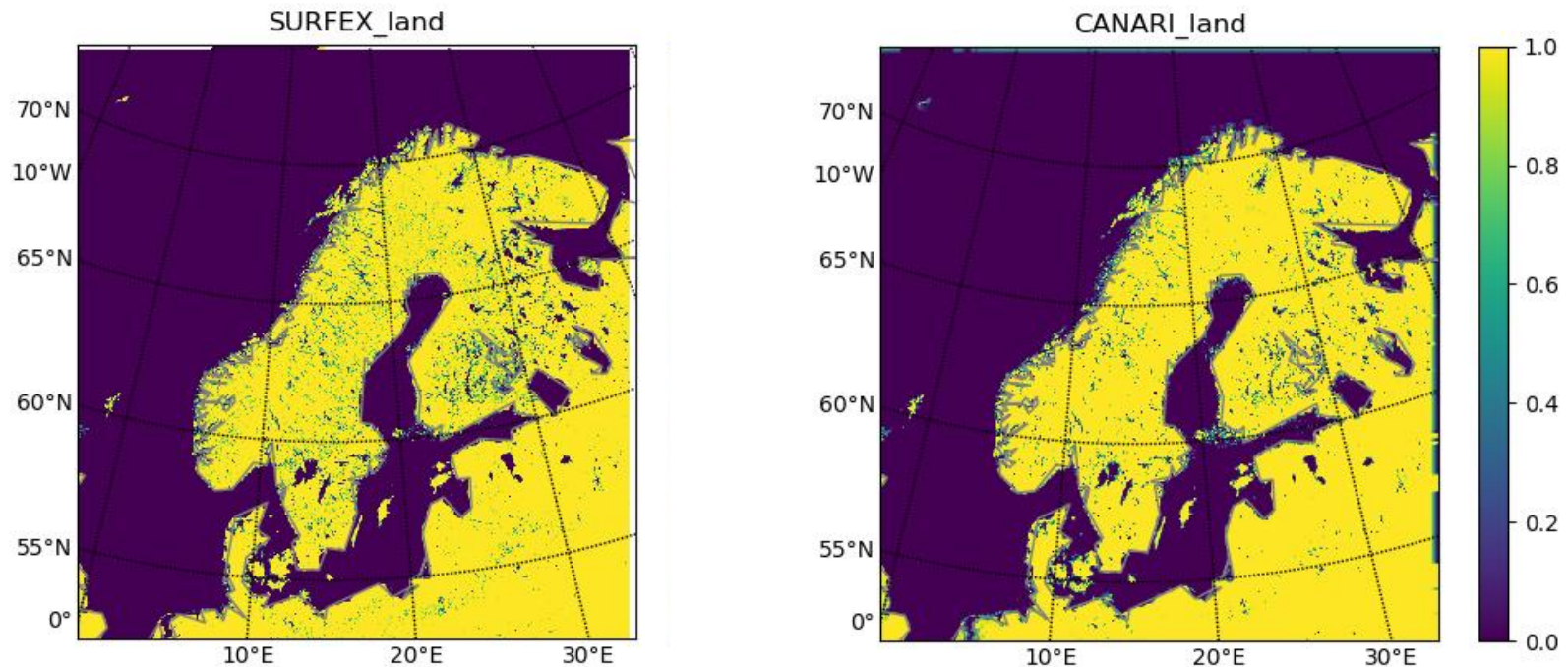
To get rid of the unrealistic features in the snow analysis, currently in MetCoOp we blacklist some coastal stations. These are:

- **22525, RAZNAVOLOK** – the Russian station on the White sea coast, contains wrong geographical coordinates, which is not captured by the quality control for some reasons,
- **22805, VALAAM** – the Russian station, on the island in Lake Ladoga,
- **26226, KIHNU** – the Latvian station, on the island in Gulf of Riga,
- **02750, HANKO TVARMINNE** – the Finnish station, of the island in Gulf of Finland,
- **02976, KOTKA RANKKI** – the Finnish station, on the coast of Gulf of Finland

How serious the problems are?

Inconsistent physiography:

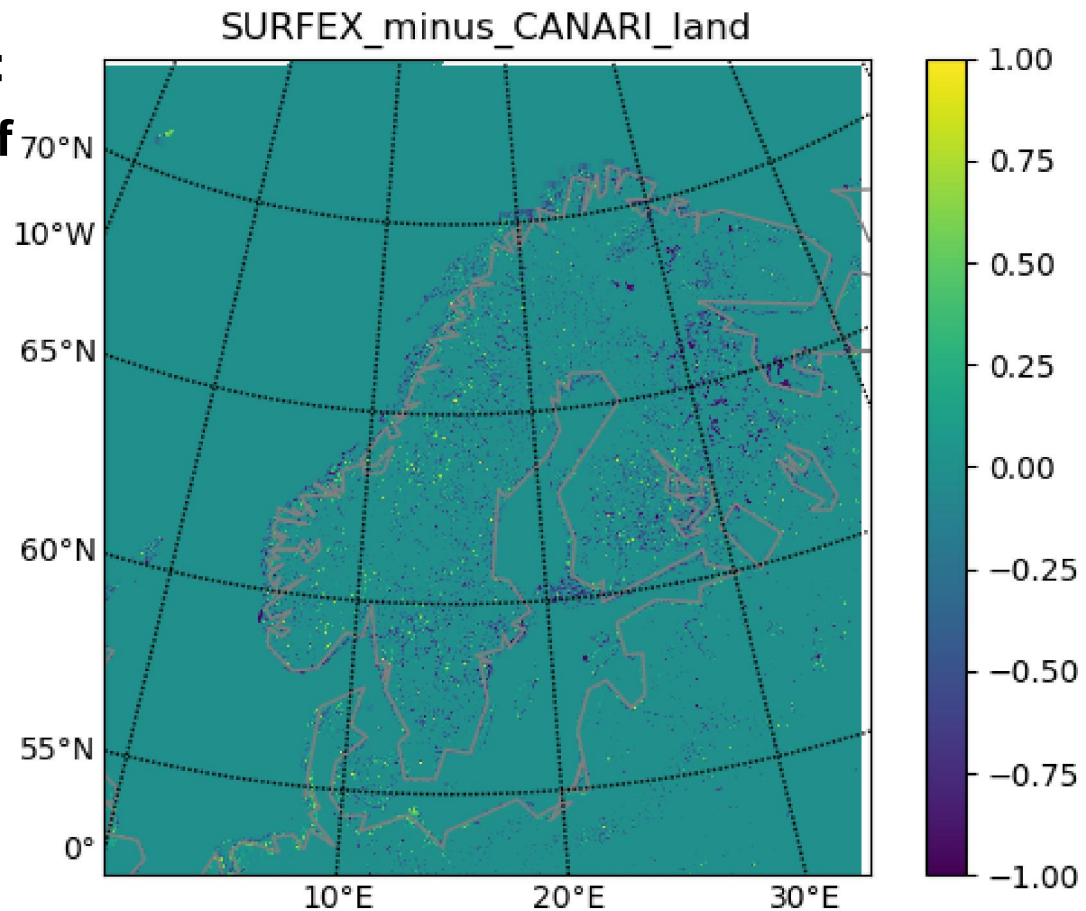
The fraction of land in SURFEX (left) and in old ALADIN (right).
Note that in CANARI the land fraction is used as the land-sea mask.



How serious the problems are?

Inconsistent physiography:

The difference in fraction of land between SURFEX and old ALADIN.



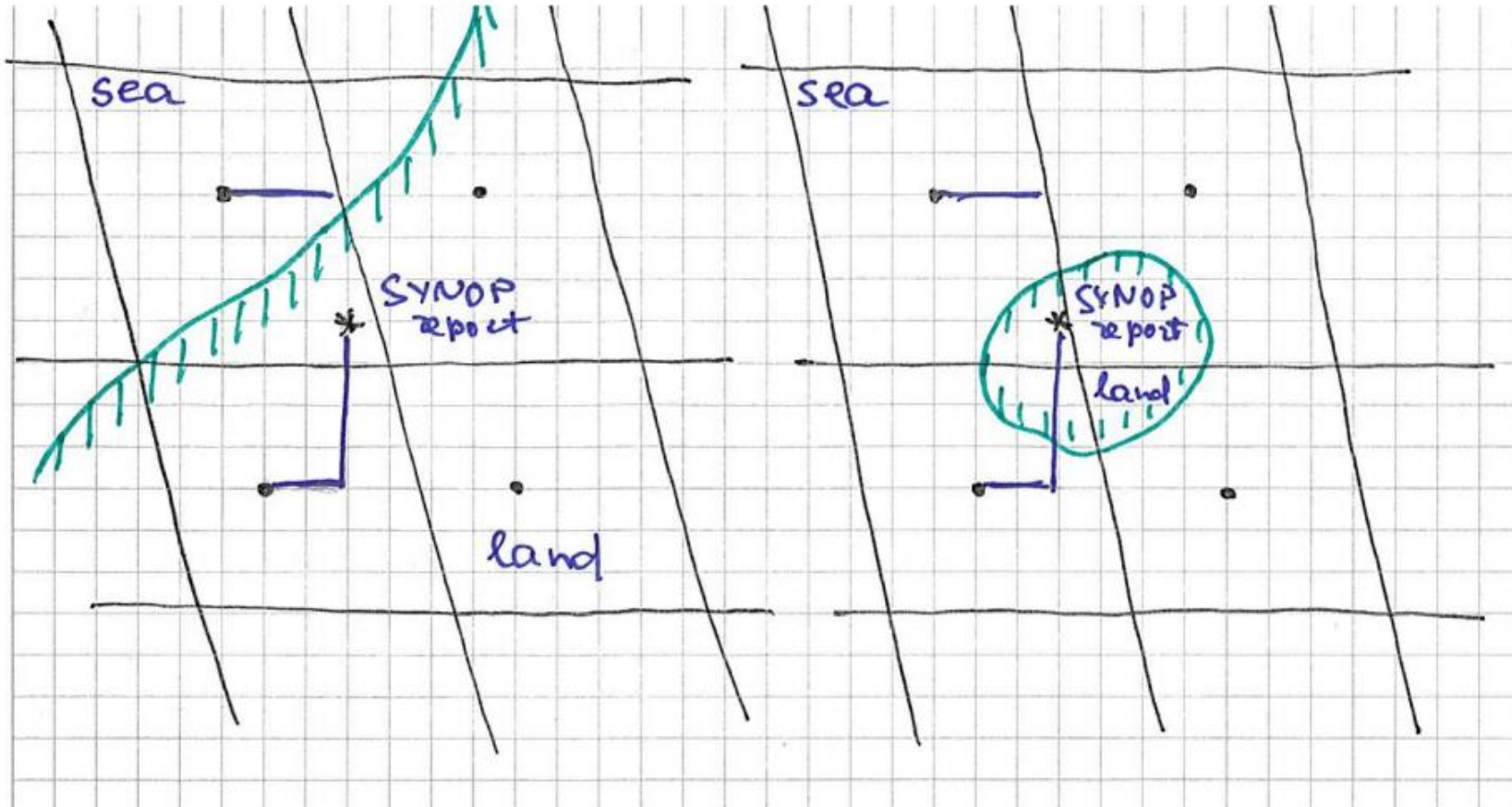
How serious the problems are?

Problem with observation operator

- In the observation operator, CANARI uses only the land-sea mask.
- Interpolation weights are first calculated depending on the distances but then modified according to the land-sea mask.
- Separation between “land” and “sea” obs is applied for whole observation reports, for all the observed variables. Interpolations themselves are also variable-independent. Thus, the snow depth is treated in the same way as the T2m and RH2m.
- If around a land observation point there are only sea points in the background (or vice versa), the weights are not modified, which means that the problem is just ignored. This is done, because otherwise the system crashes in the calculation of T2m and RH2m (a stub in the code).

How serious the problems are?

Problem with observation operator



How serious the problems are?

Problem with observation operator

Printing from slint_canari.F90:

for the MetCoOpC domain,

(the run at 15.12.2017, 03UTC, after cold start at 00UTC),

there are

137 SYNOP stations for which all the influencing grid-boxes are water, and
10 SHIP reports for which all the influencing grid boxes bare land.

All these situations are potentially dangerous.

Applying of MESCAN may produce a mess.

Possible plans and perspectives for CANARI

Short term:

- To implement the fractional land cover in CANARI instead of the land-sea mask (SURFPROP.TERRE instead of SURFIND.TERREMER). The field already exist in mXX files but not used by CANARI - done
- To improve the observation operator as simple as possible - in progress

Medium term:

- To replace the CANARI physiography by the SURFEX physiography.

Open questions:

- What is the best way to do it technically?
- What else this change will influence?
perhaps, the ARPEGE and AROME models without SURFEX



If we are going to develop CANARI or any land surface DA ...

- **First, better theoretical understanding how to treat tiles.
We may use experience from HIRLAM (*span*).**
- **Elaboration of observation operators.**
- **Elaboration of the quality control: to implement the first very rough checks depending on physiography.**
- **Harmonization with SODA (for CANARI)**
- **Cleaning of CANARI?**



ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

Thank you for your attention!



2.4.2019

**Joint 29th ALADIN Workshop & HIRLAM All Staff Meeting
2019, 1-5 April 2019, Madrid**