Detecting atmospheric rivers using persistent homology

Kristian Alfsvåg

1 June, 2015

Master's thesis in topology

Department of Mathematics
University of Bergen
Norway

1 Aknowledgements

First of all I want to thank my supervisors Morten Brun and Bjørn Ian Dundas, whithout whose suggestions, ideas and encouragement this thesis would never have seen the day of light. Also I want to thank Thomas Spengler and Clemens Spensberger at the Geophysical Institute for providing valuable insight to the world of geophysics, (and a special thanks to Clemens for the help he gave in my first ventures into the realms of serious computation).

Lastly, I want to thank my wife Inger Kristine and my son Natanael for providing support and motivation throughout all my work with this thesis.

2 Introduction

An atmospheric river is defined to be a long and narrow region of intense, vertically integrated water transport. They often stretch from the humid (sub)tropics to more temperate regions (for example Norway). There are methods in geophysics to detect these atmospheric rivers, but there are those who find these methods slightly ad hoc, and are hoping that there exists better ways of detecting them. This master's thesis is a start on the investigation on whether or not persistent homology can be used as a tool for detecting atmospheric rivers.

Persistent homology is a tool from algebraic topology that can be used to capture topological features of data. It is described in section 3, the section where the mathematical basis for this thesis is described. Its ability to discern shapes in data was the initial reason for the idea of using persistent homology to detect atmospheric rivers being considered.

There are two major computations in this thesis, both of which try to detect atmospheric rivers reaching Bergen: The *bec* (section 5.5) and the *tbec* (Section: 5.6). The *bec* is a number that says if it is possible to go from the tropics to Bergen if one only is allowed to go through regions where the humidity is high. It does so by growing regions around Bergen and equator, respectively, and seeing when these two region merge into one component. It only depends on the humidity values at given time steps and not the water transport, but we see in sections 6.1 and 6.2 that it seemingly does quite a good work of detecting atmospheric rivers.

The *tbec* is a slightly more complicated affair that in many ways does the same as the *bec*, but does so over whole time intervals at once. It not only detects atmospheric rivers, but also so called plumes, which are small regions of high humidity that moves through the atmosphere. The *tbec* is shown in sections 6.3 and 6.4.

In section 5.7 a way to compare the *bec* and the *tbec* is described, which is further discussed in section 6.5.

On the road to making these computations, an implementation of persistent homology was made, which is described in section 4.

This thesis has only been a small start on the problem of seeing whether persistent homology is a good tool to detect atmospheric rivers, and there is much more that can be done. In section 7 a few possible paths for further work on the problem are described.

Contents

3 Basics

In this section the basics for the implementations and computations will be discussed.

Note: When I talk about homology in this thesis, I will always mean homology with $\mathbb{F}_2$ coefficients. This is the simplest case, not the least because it lets me totally disregard orientation of simplices. Also, since I am mainly interested in actually computing things, all complexes will be finite.

3.1 Cellular complexes

Definition 3.2. *A cellular complex is a space X with a filtration*

$$\emptyset \subset X^0 \subset X^1 \subset \ldots \subset X^N = X$$

and maps $\{(\Phi_\alpha, \phi_\alpha) : (D^{n_\alpha}, S^{n_\alpha - 1}) \to (X^{n_\alpha}, X^{n_\alpha - 1})\}$ (where A is some indexing set) such that given n, the following diagram is a pushout diagram:

$$\begin{array}{ccc} \coprod_{\alpha \in A_n} S^{n-1} & \xrightarrow{\sum_\alpha \phi_\alpha} & X^{n-1} \\ \downarrow & & \downarrow \\ \coprod_{\alpha \in A_n} D^n & \xrightarrow{\sum_\alpha \Phi_\alpha} & X^n, \end{array}$$

(where A_n is the subset of A given by $n_\alpha = n$.)

See [4], page 5.

If X and Y are cell complexes, a map of spaces $X \to Y$ is a *map of cell complexes* if for all n-cells Φ of X, there is an n-cell Φ' of Y such that the following diagram commutes:

$$\begin{array}{ccc} D^n & & \\ \downarrow{\scriptstyle \Phi} & \searrow{\scriptstyle \Phi'} & \\ X & \longrightarrow & Y. \end{array}$$

If the map is an inclusion of spaces, we say that X is a subcomplex of Y.

A cellular complex can also be thought of as a space constructed by starting with a discrete set of vertices X^0, and forming X^n by glueing n-discs to X^{n-1} along the boundary of the disc. (That X^0 is discrete follows from $X^{-1} = S^{-1} = \emptyset$.)

In my applications, all the maps $\Phi_\alpha : D^n \to X^n$ are embeddings, and I will call these the n-cells of X. The map ϕ_α is called the attaching map of the cell Φ_α. The 0-cells of X are also called the *vertices* of X. I will denote the set of cells in X by $cells(X)$ (This set is isomorphic to the set A in definition 3.2). It has a partial ordering by saying that $\Phi_\alpha \subseteq \Phi_\beta$ if $\operatorname{im} \Phi_\alpha \subseteq \operatorname{im} \Phi_\beta$.

Definition 3.3. *The standard n-simplex Δ^n is defined to be the subspace*

$$\{(t_0, \ldots, t_n) \mid \sum_{i=0}^{n} t_i = 1, t_i \geq 0 \ \forall i = 0, \ldots, n\} \subseteq \mathbb{R}^{n+1}.$$

Given $i = 0, \ldots, n$, the i'th coface map $d^i : \Delta^{n-1} \to \Delta^n$ is given by

$$d^i(t_0, \ldots, t_{n-1}) = (t_0, \ldots, t_{i-1}, 0, t_i, \ldots, t_{n-1}).$$

The boundary $\partial\Delta^n$ of the n-simplex is defined to be the subset of Δ^n where at least one of the coordinates is zero. The vertices of the n-simplex are all the points where one coordinate is exactly 1 *(and the the rest of the coordinates must be* 0*).*

It should be mentioned that the vertices have a natural ordering given by the order of the coordinates. Note that the coface maps factor through the boundary. Also note that $(\Delta^n, \partial\Delta^n)$ is homeomorphic to (D^n, S^{n-1}). I choose one such homeomorphism and regard it as canonical.

Definition 3.4. *An n-cell $\Phi_\alpha : D^n \to X^n$ of a cellular complex X is a* simplicial cell *(or just a simplex) if there, given a coface map d of Δ^n, exists an $(n-1)$-cell $\Phi_\beta : D^{n-1} \to X^{n-1}$ such that the following diagram commutes:*

$$\begin{array}{ccccc} \Delta^{n-1} & \xrightarrow{\cong} & D^{n-1} & & \\ \downarrow{\scriptstyle d} & & & \searrow^{\Phi_\beta} & \\ \partial\Delta^n & \xrightarrow{\cong} & S^n & \xrightarrow{\phi_\alpha} & X^{n-1}. \end{array}$$

Since a point in the interior of an $(n-1)$-cell can only be contained in that single $(n-1)$-cell, and these Φ_β cover the whole image of ϕ_α, these Φ_β are indeed the only $(n-1)$ cells contained in Φ_α.

From now on the chosen isomorphism between the n-disc and the standard n-simplex will be suppressed, and simplicial cells regarded as maps $\Delta^n \to X^n$.

The following definitions for cubes and cubical cells are completely analogous to the simplicial case above.

Definition 3.5. *The* standard n-cube *is defined as the n-fold cartesian product $I^n = [0,1]^n \subset \mathbb{R}^n$. The* coface maps $d^i_k : I^{n-1} \to I^n$ *of the n-cube are given by*

$$d^i_k(t_1, \ldots, t_{n-1}) = (t_1, \ldots, t_{i-1}, k, t_i, \ldots, t_{n-1})$$

for i in $\{1, \ldots, n\}$ and k in $\{0, 1\}$. The boundary ∂I^n of the n-cube is the subset of I^n where at least one of the coordinates is zero or one. The vertices of the cube are all the points where all the coordinates are either 0 *or* 1*.*

Note that the coface maps factor through the boundary ∂I^n. Also note that $(I^n, \partial I^n)$ is homeomorphic to (D^n, S^{n-1}). As above, I choose one such homeomorphism and regard it as canonical.

Definition 3.6. *An n-cell $\Phi_\alpha : D^n \to X^n$ of a cellular complex X is a* cubical cell *(or just a cube) if there, given a coface map d of I^n, exists an $(n-1)$-cell $\Phi_\beta : D^{n-1} \to X^{n-1}$ such that the following diagram commutes:*

$$\begin{array}{ccccc} I^{n-1} & \xrightarrow{\cong} & D^{n-1} & & \\ \downarrow{\scriptstyle d} & & & \searrow^{\Phi_\beta} & \\ \partial I^n & \xrightarrow{\cong} & S^n & \xrightarrow{\phi_\alpha} & X^{n-1}. \end{array}$$

By the same argument as for simplicial cells, these Φ_β are the only $(n-1)$-cells contained in Φ_α.

From now on the chosen isomorphism between the n-discs and the standard n-cube will be suppressed, and cubical cells regarded as maps $I^n \to X^n$.

Note that since $(\Delta^1, \Delta^0) \cong (I, I^0)$, there is no difference between cubical and simplicial cells when we are in dimensions 0 and 1.

Definition 3.7. *The chain complex $C(K) = C(K; \mathbb{F}_2)$ of a cellular complex K is the vector space over $\mathbb{F}_2$ on the set of cells in K. This is a graded vector space by giving an n-cell the degree n. The boundary map $\partial : C(K) \to C(K)$ of degree -1 is given by $\partial(\Phi_\alpha) = \sum_{\beta \in A_{n_\alpha - 1}} d_{\alpha\beta}\Phi_\beta$. (See [4], section 2.2 (page 140) for how to find the coefficients $d_{\alpha\beta}$ in general.) In the special case where c is a simplex or a cube, $\partial(c) = \sum_d \phi \circ d$ where ϕ is the attaching map of c and d runs over the set of coface maps of the simplex or cube (respectively) of the right dimension.*

We call the elements of $Z(K) = \ker \partial$ the *cycles* in $C(K)$ and the elements in $B(K) = \operatorname{im} \partial$ the *boundaries* in $C(K)$. The homology of $C(K)$ is the quotient group $H(K) = Z(K)/B(K)$.

In my thesis, all cell complexes will satisfy the following:

- All cells are either simplicial or cubical.
- All cells are uniquely given by their restriction to the vertices of the standard simplex/cube.
- All cells are embeddings.

This gives that all the cells can be given purely combinatorically. A simplicial cell s of X is just an embedding of the n-simplex into X, and it can be given as an injective map $\{0, \dots, n\} \to X^0$ obtained by restricting the characteristic map to the vertices of Δ^n and precomposing with the natural ordering on these vertices. Somewhat more compactly, I will write this as an ordered list $(x_0, \dots, x_n)$ of distinct vertices of X. By abuse of notation, I will write $s = (x_0, \dots, x_n)$. It should be noted that if you permute this list (non-trivially), you will get a different simplex. This is not needed in any of my examples, but it is slightly more efficient to implement than the alternative of saying that simplices are equal if their sets of vertices are equal.

So, in this notation, if $s = (x_0, \dots, x_n)$ is a simplicial cell, then

$$\partial s = \sum_i (x_0, \dots, \hat{x_i}, \dots, x_n). \tag{1}$$

A cubical cell $c : I^n \to X^n$, on the other hand, can be given as an injective map $\{0,1\}^n \to X^0$, obtained by restricting the characteristic map $I^n \to X$ to the vertices of I^n. By abusing notation once again, I will write $c : \{0,1\}^n \to X^0$.

If $c : \{0,1\}^n \to X^0$ is a cubical cell, then

$$\partial c = \sum_{i,k} c \circ d^i_k, \tag{2}$$

where $d^i_k : \{0,1\}^{n-1} \to \{0,1\}^n$ is the restriction of the coface map $d^i_k : I^{n-1} \to I^n$ (and is given by inserting a $k \in \{0,1\}$ between term number $i-1$ and i).

Example 3.8. From a grid $G = \prod_{i=1}^{N} \{0, 1, \ldots, n_i - 1\}$, one can construct a cellular complex $K(G)$ with G as vertices as follows: For all $m = (m_i)_i \in G$ where $0 \leq m_i < n_i - 1$, let $c_m : \{0,1\}^N \to G$ be given by $c_m(k) = k + m$ (where $k \in \{0,1\}^N$ and addition is term-wise). The map c_m can be regarded as an N-cube in a cell complex with G as vertices, and including all the faces of these cubes will give a cellular complex $K(G)$, which as a topological space is just the convex hull of $G \subset \mathbb{R}^N$. See Figure 1.

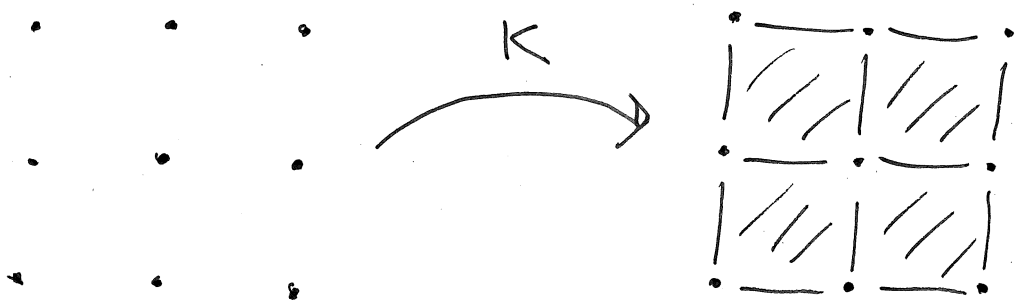


Figure 1: An example of a grid G and $K(G)$.

Example 3.9. One example of a cell complex that has both simplicial and cubical cells is the following: Take $K(G)$ of a grid G, and add simplices as shown in Figure 2.

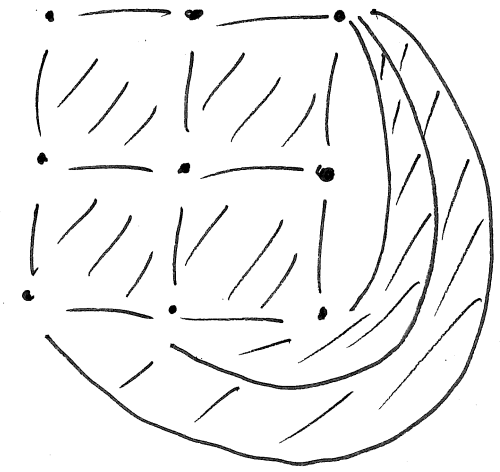


Figure 2: A simple example of a cellular complex.

3.10 Persistent homology

Definition 3.11. *A* cellular map *on a cell complex K is a map $g : cells(K) \to \mathbb{R}$ such that $c_0 \subseteq c_1$ implies $g(c_0) \leq g(c_1)$. An* opposite *cellular map on K is defined in the same way, except that it is a map $g : cells(K) \to \mathbb{R}^{op}$. (I.e. $\mathbb{R}$ with the reversed order.)*

By the isomorphism $-1 : \mathbb{R} \to \mathbb{R}^{op}$ of ordered sets, an opposite cellular map can be given by a cellular map, and vice versa.

Definition 3.12. *A* filtered complex *is a cellular complex K together with a filtration of subcomplexes*

$$\emptyset = K^0 \subset K^1 \subset \cdots \subset K^m = K. \tag{3}$$

A filtration of a complex K can also be given by a cellular map on K in the following way: From a cellular map, make a filtration of K by setting K^r to be the union of the images of the cells in $g^{-1}(-\infty, r]$. Since K is finite there will only be finitely many steps where anything happens, and reindexing those with integers gives a filtration on the form (3). Going the other way around, starting with a filtration, you can make a cellular map g by setting $g(c) = i$ if c is in K^i but not in K^{i-1}. We say that $g(c)$ is the filtration value of the cell c (or the filtration degree if all the values are integers).

A filtered complex gives a filtration of chain complexes, and taking the k'th homology of this gives a sequence

$$H_k(K^0) \to H_k(K^1) \to \cdots \to H_k(K).$$

This diagram is called the *k'th persistent homology* of the filtered complex (3). I will write H_k^i as shorthand for $H_k(K^i)$.

The intuitive way to think of this is as follows: the filtered complex is a space evolving over time, getting more and more cells added. The persistent homology captures the homology of this space, and how this evolves over time.

Definition 3.13. *Given a homology class γ in H_k^b:*

- *We say that γ is* born *in K^b if it is not contained in* $\mathrm{im}\{H_k^{b-1} \to H_k^b\}$.
- *Furthermore, if γ is born in K^b we say that it has died in K^j if there is a $\gamma' \in H_k^{b-1}$ such that γ and γ' are mapped to the same class in H_k^j. If no such j exists, we say that γ dies at infinity, and else we say that γ dies entering K^d where d is the minimum of all such j's. This is called the* elder rule *because when two classes merge, the oldest one lives.*
- *We call the interval $[b, d)$ the* persistence interval *associated to γ. The list of all these pairs is the* persistence diagram *of the filtered complex. The difference $d - b$ is called the* persistence *of γ.*

See [3], chapter VII for more details on persistent homology.

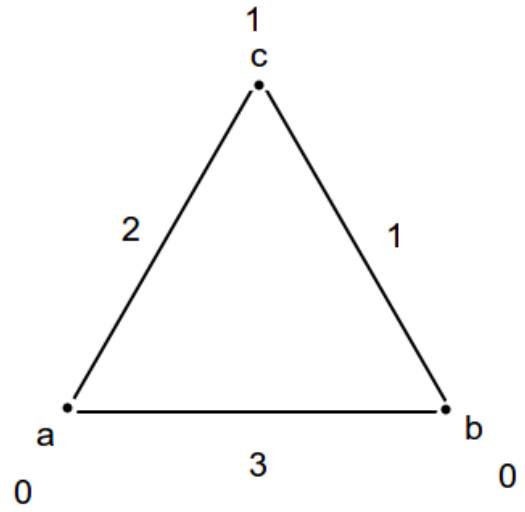
The filtered complex (3) gives rise to a graded $F[t]$-module $C = \oplus_i C(K^i)$ where multiplication by t on summand $C(K^i)$ is induced by the inclusion $K^i \subseteq K^{i+1}$ (the grading is given by filtration degree). Taking the k'th homology of this module gives $H_k(\oplus_i C(K^i)) = \oplus_i H_k^i$.

When F is a field, $F[t]$ is a principal ideal domain, and the structure theorem for finitely generated modules over a principal ideal domain gives that

$$\oplus_i H_k^i \cong \bigoplus_j t^{b_j} F[t] \oplus \bigoplus_l t^{b_l} F[t]/t^{d_l} F[t].$$

So the persistent homology is uniquely given (up to isomorphism) by the b_j's, b_l's and d_l's. Or, in other words, the persistence intervals $[b_j, \infty)$ and $[b_l, d_l)$, as the b's correspond to filtration values where homology classes are born, and the d's correspond to filtration values where homology classes die.

This decomposition gives rise to an algorithm for computing persistent homology with field coefficients by matrix reduction. One assumes inductively a homogeneous basis for the cycles $Z_k \subseteq C_k$ (sorted from highest to lowest filtration degree) and a matrix giving the map $\partial : C_{k+1} \to Z_k$. Column reduce this matrix to lower diagonal form. From the reduced matrix one can find a basis for Z_{k+1} by observing which columns only contains zero, and it is also easy to obtain the matrix giving the map $\partial : C_{k+2} \to Z_{k+1}$ in this new basis (just delete the right rows in the matrix in the previous basis for C_{k+1}). The top non-zero element in the pivot columns will remain unchanged if the matrix is further row reduced to diagonal form, so the persistence of the cycles in the basis can be read off from the degree of these elements.



$$\begin{array}{c} \\ c \\ b \\ a \end{array}\!\!\begin{array}{c} \begin{array}{ccc} bc & ac & ab \end{array} \\ \begin{pmatrix} 1 & t & \\ t & & t^3 \\ & t^2 & t^3 \end{pmatrix} \end{array} \sim \begin{array}{c} \\ c \\ b \\ a \end{array}\!\!\begin{array}{c} \begin{array}{ccc} bc & ac+tbc & ab+tac+t^2bc \end{array} \\ \begin{pmatrix} 1 & & 0 \\ t & t^2 & 0 \\ & t^2 & 0 \end{pmatrix} \end{array}$$

Figure 3: Illustrating the algorithm for computing persistent homology by reduction of the matrix to lower diagonal form.

Figure 3 gives an illustration on how the algorithm works. The filtered complex is given by specifying the filtration degrees of the simplices. (The filtration degrees are the numbers floating by the simplices.) The matrix is corresponding to the map $\partial : C_1 \to Z_0$. The first diagonal entry in the reduced matrix says that the cycle born by c is killed immediately (since the element 1 has degree 0). The second diagonal entry (t^2) says that the cycle born by b lives for two filtration steps. The third diagonal entry (0) says that the cycle born by a is added lives for ever. When one in addition remembers that both a and b have filtration degree 0, this means that the persistence intervals in the persistent 0-homology are $[0, 2)$ and $[0, \infty)$. In addition, there is one column with only zeros, which gives that the chain $ab + t \cdot ac + t^2 \cdot bc$

generates the 1-cycles. This cycle has filtration degree 3 and does not die (since there are no 2-cells). Thus the persistence diagram for the 1-homology is $[3, \infty)$. This example reappears in a different form in appendix A. See [8] for more details on the algorithm.

Note: The algorithm assumes that the cells have unique filtration values, i.e. only one cell is added at the time. So for cells with the same filtration value, some arbitrary ordering is made (arbitrary except that a cell must appear after its boundary). This means that it makes sense to talk about the cell causing a cycle's birth (the cycle's birth cell) or the cell killing it (its death cell).

Also note: It is perhaps not immediate that the pairings of births and deaths above correspond to the pairings of b_l and d_l in the decomposition (i.e. the decomposition respects the elder rule). It is clear that some class dies at the filtration steps d_l, but not necessarily the youngest possible (as the elder rule demands). Keeping track of the basis in the algorithm one sees that if a cell c is added killing a cycle, the cycle getting killed will have the filtration degree corresponding to the maximal degree of the cells in ∂c, i.e. when a cycle dies, it is the youngest possible that dies.

One way to visualize persistence diagrams is the *barcode*, where one draws all the persistence intervals above each other (see Figure 4).

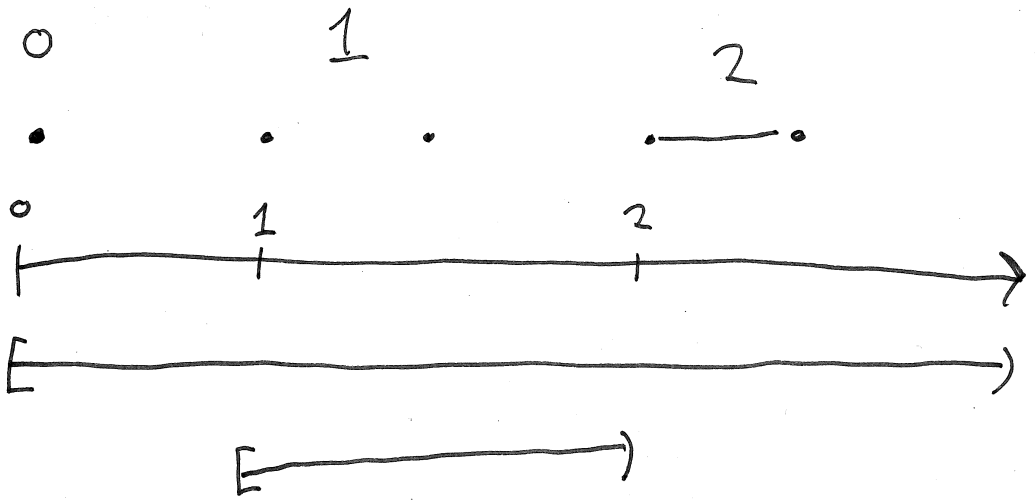


Figure 4: The barcode of a simple filtered complex with persistence intervals $[0, \infty)$ and $[1, 2)$.

3.14 Constructing filtered complexes from functions on grid points

Starting with a function g on a grid G, one can construct a filtration of the complex $K(G)$ described in example 3.9: Start by letting the filtration value on the 0-cells be the function value on it as a grid point. Then, for higher dimensional cells, let the filtration value be given inductively by the maximum of the filtration values of cells in its boundary. This will give a cellular map $K(f)$ on $K(G)$, which, as I noted, is equivalent to giving a filtration. Call this the *bottom-up* filtration $B(f)$ induced by the function f. Doing the same in $\mathbb{R}^{op}$ instead of $\mathbb{R}$ (effectively exchanging the word 'maximum' with 'minimum') gives an opposite cellular map on K, which results in what I call the *top-down* filtration $T(f)$ induced by f.

These two filtrations can be pictured as follows: Regard the $K(f)$ as a height function on a rectangle, and imagine that you fill water in the resulting landscape. B^t in the bottom-up filtration will correspond to the surface of the water when it is at heigh t. This also explains the name 'bottom-up' as it starts by including the points on the bottom and goes up. On the other hand, if you start with water above the highest "mountain tops" and let it sink, T^t in the top-down filtration will correspond to the mountain peaks that are above the water surface at height t. So you start by adding the points on the top, and go downwards. See Figure 6.

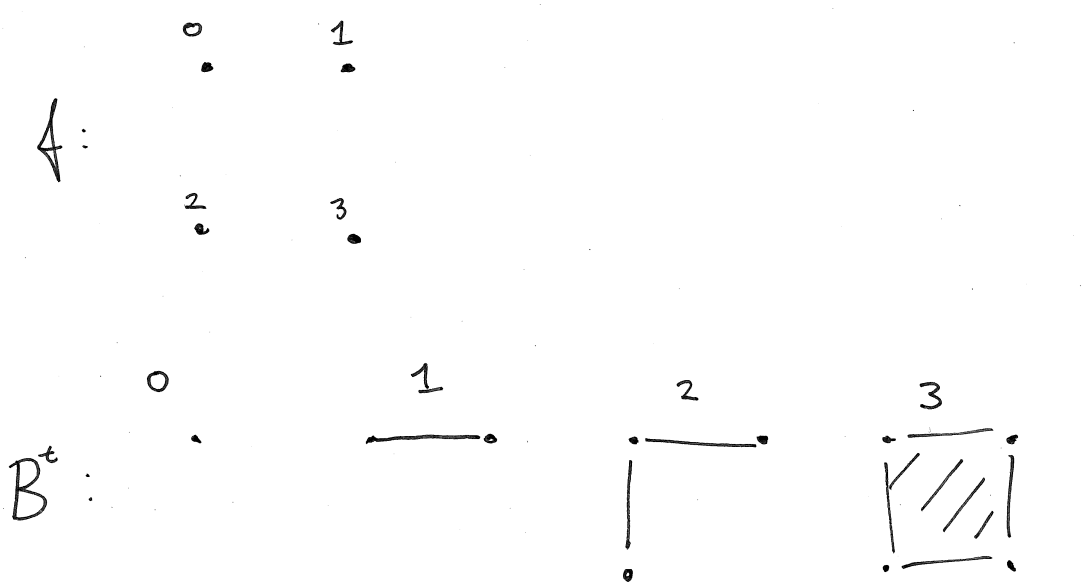


Figure 5: A map f on a grid, and the bottom-up filtration $B(f)^t$ induced by f.

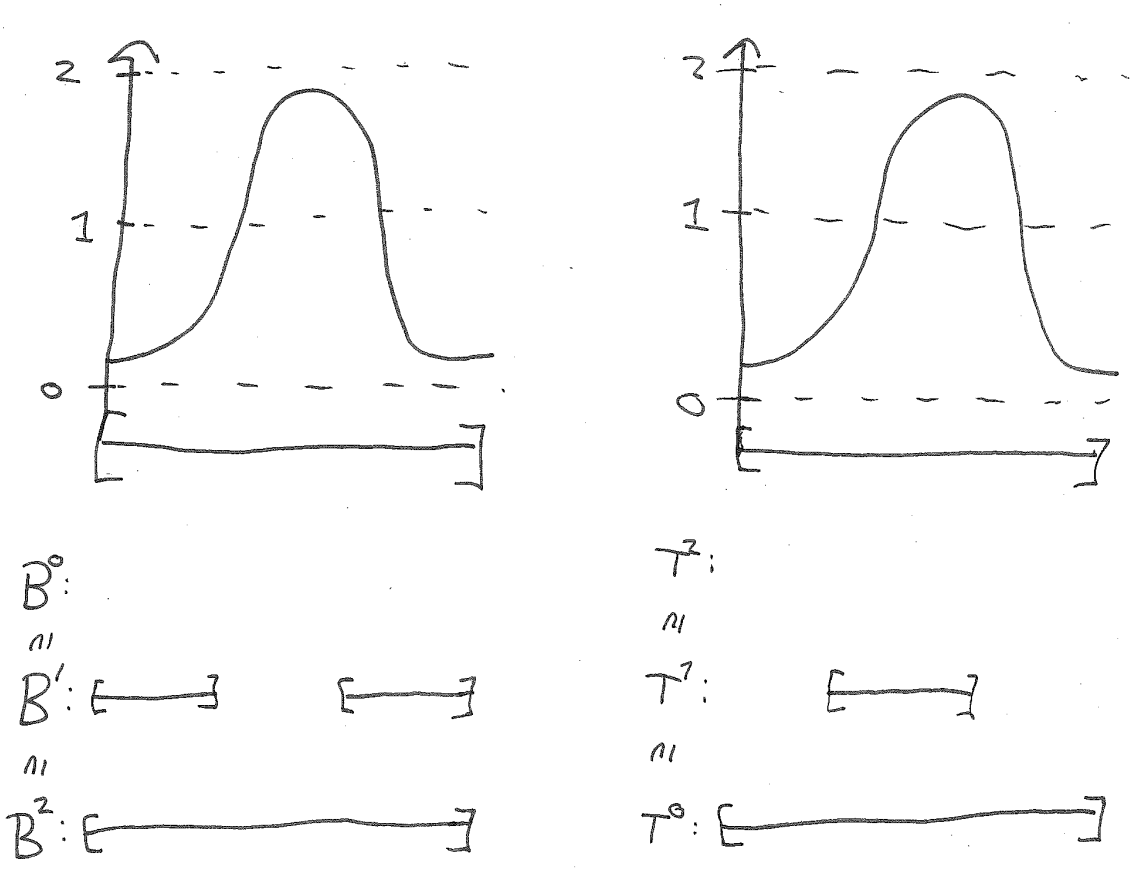


Figure 6: Illustrating the bottom-up and top-down filtrations given by a map on the unit interval.

4 The implementation of persistent homology

When I started working on this thesis, I didn't think that I would be using programming that much. I expected to just install the persistent homology package, run a few scripts, and then have ample time to write my thesis on the results. About a thousand lines of code later, and after learning Python almost from scratch, it can be safely said that this was slightly naive.

So programming ended up being a significant part of this thesis.

Initially, I used JavaPlex ([7]) to get acquainted with computing persistent homology. A problem with this was that the methods to retrieve the data were written in Python, and so we had a slight dilemma: Should we find some way to get the data over to Java or MatLab, where JavaPlex had its natural domain, or should we start using the slightly less user-friendly Dionysus [5] instead, (a C++ based package, but with Python bindings enabling it to be used within Python). We landed on using Dionysus.

After a while, Dionysus started to feel a little restrictive to use. It could only handle simplicial complexes, and it was difficult to keep track of the connection between the homology classes and the points in the complex. So, to get a bit more control over the situation, I decided to implement persistent homology myself. This gave me a much better understanding of how things were computed, and more control on how it was done.

The aim has been to keep the implementation as simple and straightforward as possible. It is based on the algorithm for computing persistent homology described in [8]. The structure is inspired by JavaPlex and Dionysus, but kept a bit simpler. It has also been important to try to document it well and in general keep it accessible to others, so that it can be an asset for anyone continuing the work on this problem.

The source code and full documentation of this implementation can be found at

`http://org.uib.no/mi/master/kal045/main.html`.

In addition to this, a more simplified presentation of the classes in the implementation is given here:

First it should be mentioned that a vertex can be any (immutable) object. (In the calculations, the vertices have typically been tuples corresponding to points in a grid.)

Cell	
Data	vertices filtrationValue
Methods	equals(other) boundary() getDim()

The class Cell (an abstraction of the classes Simplex and CubicalCell) contains the following data: Vertices and filtration value. In the Simplex class, the vertices are stored in an ordered list, and in the CubicalCell class, the vertices are stored in an array of shape 2^n (effectively a map from $\{0,1\}^n$). The equals method says that two simplices are equal if their ordered lists of vertices are equal, and two cubical cells are equal if their arrays of vertices are equal. Furthermore, in dimensions 0 and 1 a simplex and cubical cell are equal if their vertices correspond. Note that the equals method does not depend on the filtration value. The boundary method returns a list of cells according to the formulas (1) and (2) in section 3.1. (A chain can be represented by a list of cells since we work with $\mathbb{F}_2$-coefficients.) The getDim method returns the dimension of the cell.

PersistenceCycle	
Data	birthCell deathCell
Methods	equals(other) getBirth() getDeath() getDim() persistence()

The class PersistenceCycle represents persistence intervals. The cell birthCell is the cell that causes the birth of the cycle. If the cycle dies, the cell deathCell is the cell causing the death. (This makes sense since the algorithm for computing persistent homology assumes that only one cell is added at the time, as mentioned in section 3.10.) The method getBirth returns the filtration value of birthCell, and getDeath returns the filtration value of deathCell if it exists, and infinity otherwise. The method persistence returns the difference between these two. The method getDim returns the dimension of the cycle, which is given as the dimension of the birth cell. The equals method say that one cycle is equal to another if they are of the same dimension, and are born and die at the same filtration steps.

PersistenceDiagram	
Data	cycles
Methods	getBettiNumbers()

The class PersistenceDiagram represents persistence diagrams. It is basically a list of persistence cycles, but with the method getBettiNumbers, which returns two lists: one list of all filtration values where the homology changes, and one list of the ranks of the homology (the betti numbers) at theses filtration values.

FilteredComplex	
Data	cells
Methods	computePersistentHomology(maxDim) getDiagram(dim)

The class FilteredComplex is more or less the backbone of the implementation, as this is where the computation of persistent homology takes place. It contains a list of cells which form the filtered complex. The method computePersistentHomology(maxDim) computes the persistent homology in dimensions up to maxDim (inclusive). The method getDiagram(dim) returns the persistence diagram of the given dimension if this is computed (and an error otherwise).

One notable difference between my implementation and JavaPlex/Dionysus is that the latter two have their own classes to compute persistent homology, whereas in my implementation this task is assigned to the FilteredComplex itself. One advantage of my choice is that there is one less layer of abstraction between the data and the results, making it easier to compare them. On the other hand, it is less flexible, as it only allows for one way to compute the persistent homology.

There are a few possible pitfalls when using this implementation. Notably, constructing a cell with non-distinct vertices will give unexpected results without any error message. This is noted in the documentation. More direct steps to ensure this could be taken, but that would

increase computation time. It is also possible to construct a FilteredComplex which does not satisfy the conditions for a filtered complex (e.g. a cell appearing before its boundary, or that its boundary is not contained in the complex at all). This will result in an error message when trying to compute the persistent homology.

In appendix A there is a small example of how this implementation can be used to compute persistent homology.

5 Computations

The data that has been used in this thesis comes from the ERA-Interim project, which provides an array of meteorological data from 1979 and until today [2]. It uses data assimilation to fit observed data unto a Gaussian grid (i.e. coordinates are given by longitude and latitude). I have accessed the data using software made by Clemens Spensberger at the Geophysical Institute. Since the data sets I would work over were quite large, I was allowed access to the power of skd-cyclone, a quite capable computer belonging to the Geophysical Institute. I have mostly been working on the total water column, but also the integrated water vapour transport, both of which I define here for easy reference:

Definition 5.1. *The* total water column *(abbreviated tcw[1]) at a point on the earth's surface is defined to be the integral of the water density (given in kg/m^3) along the line from the point and directly upwards to the top of the atmosphere. It is given in kg/m^2.*

Definition 5.2. *The* integrated water vapour transport *(IVT) is defined to be the vertical integral of the vector field (water density)·(horizontal wind velocity). This is a vector field on the earth's surface, with coefficients given in $kg/m^2 \cdot m/s$.*

The data I have used is given on a grid with half-degree intervals, covering the earth, every six hours. In other words, given a year, it is given on the grid

$$\{-180°E, -179,5°E, \dots, 179,5°E\} \times \{-90°N, -89,5°N, \dots, 90°N\} \times \{t_0, t_1, \dots, t_{M-1}\}.$$

The first two coordinates are longitude and latitude, and the last coordinate is given by $t_i = 6 \cdot i$ hours since the year started. Normally, $M = 1460$, but in leap years, $M = 1464$.

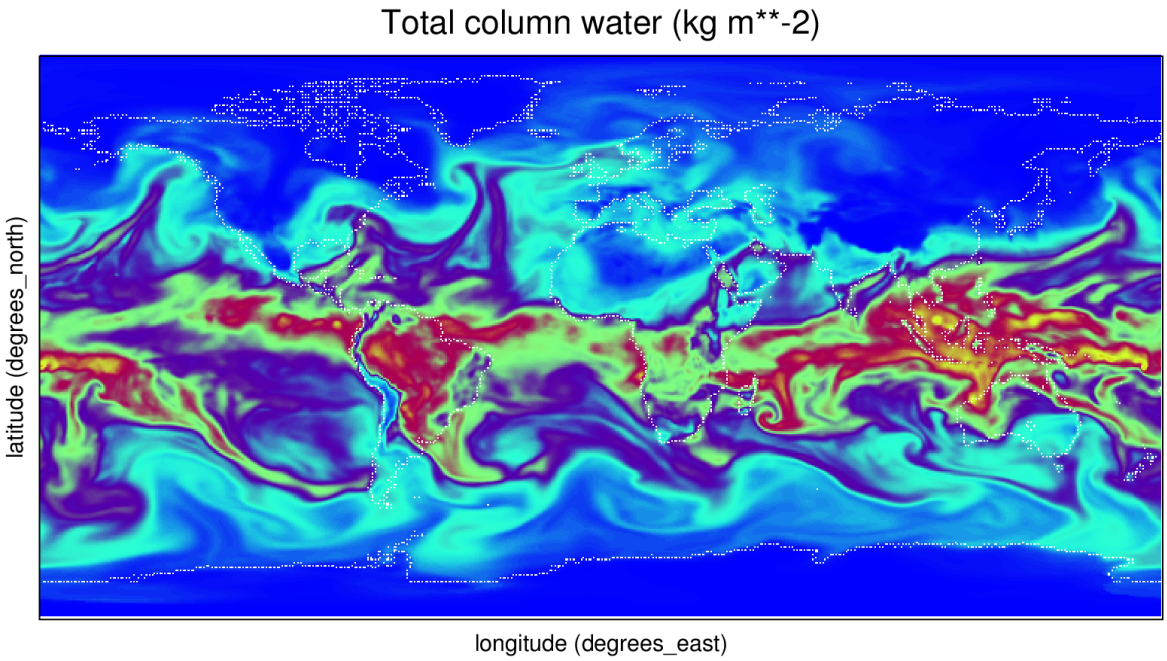


Figure 7: The *tcw* at January 1st, 2012.

[1]It is perhaps a bit peculiar that it is abbreviated *tcw* and not *twc*, but that is how it is done in the data, and I decided to stick with it.

The goal of this thesis is detecting atmospheric rivers. It is not not completely well defined what an atmospheric river is, but loosely speaking it is defined as a long and narrow region of intense, integrated water vapour transport (See [6]).

5.3 What I did

The following example was one of the very first things I did to get acquainted with the data and the computing of persistent homology:

Example 5.4. The *tcw* at a given time step gives a map on the grid $\{-180°E, \dots, 179°E\} \times \{-90°N, \dots, 90°N\}$. Take the bottom-up filtration induced by this grid map (see example 3.9) and compute its persistent 0-homology. This filtered complex corresponds to taking the world map (not the globe) and filter it by first including the driest points, and then include the next-to-driest points until one at the end adds the most humid points and end up with the whole world map. See Figure 8 for the bar code of this persistent homology.

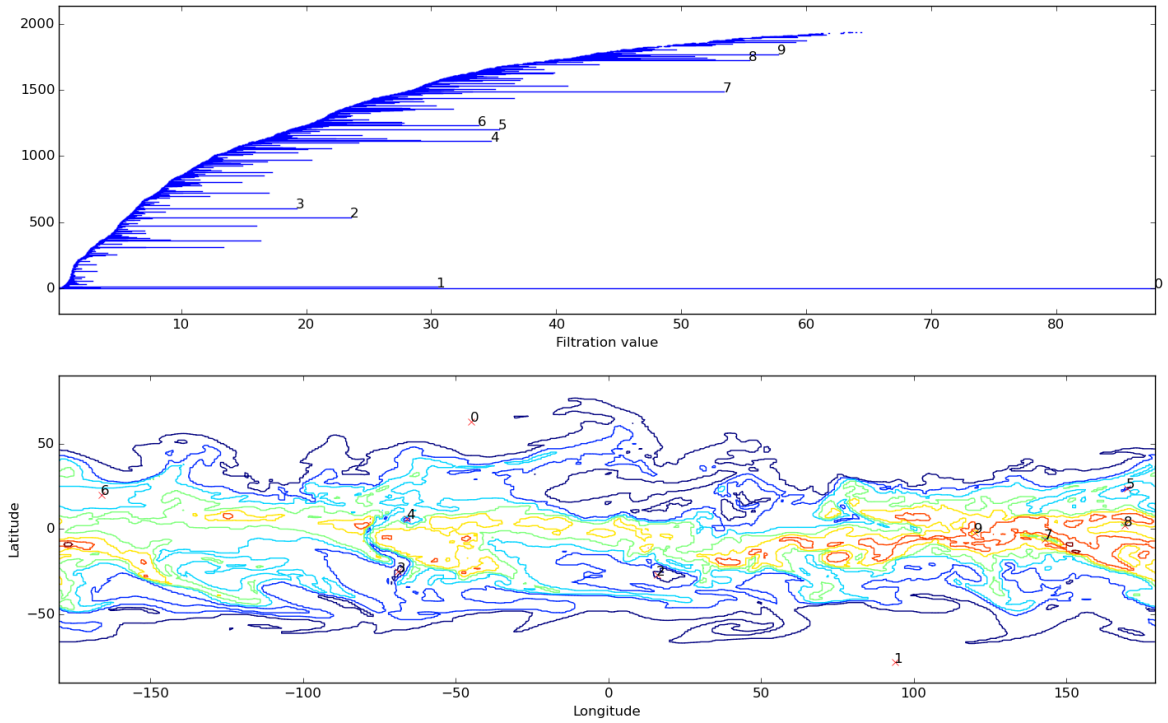


Figure 8: Upper half: Persistence diagram of bottom-up filtration given by *tcw* on the whole world at 1 January 2012. The 10 longest-living classes are numbered.
Lower half: Contour lines of *tcw*. The points where the numbered classes first appear are marked. Compare this with figure 7.

After this example, I started looking for atmospheric rivers. To narrow the problem down, I have only looked for atmospheric rivers reaching Bergen. Looking at the graph of *tcw* (like in Figure 7), it seemed to me that these rivers were characterized by a long, thin 'river-shape' reaching up from the tropics and to Bergen. The *bec* (see section 5.5) is a way to look for these shapes.

This *bec*-value, interesting as it may be, has (at least) one drawback when trying to detect atmospheric rivers. An atmospheric river is dynamic in nature, being defined as large *transport* of water. The *bec*, on the other hand, is more static in nature as it only depends on the humidity levels at a given time. I considered two strategies from here: Either adding dynamics to the equation by somehow applying wind velocities, or increasing the dimension of the complexes by adding the time dimension. Postponing the dynamics for later, I decided to do the time strategy first, as it felt more mathematically interesting. For example, here it was necessary to apply H_1, whereas the *bec* only depended on H_0, and a larger part of the persistence diagram is considered relevant (for computing the *bec*, only the two oldest classes have anything to say). Also, the geophysicists regarded it as a novel idea. This computation is described in section 5.6.

Unfortunately, there would not be enough time to also include the dynamics strategy in my thesis, except for some basic thoughts (see section 5.8.)

5.5 The bec

In this section I will discuss the value I've computed which I have decided to call the *bec*. (Short for Bergen Equator Connected). The idea comes from looking at pictures of the *tcw* values in the world, and observing that there often would be these long "arms" of humidity emerging from the tropics and reaching up to the more temperate north. The *bec* is a calculation that tries to capture these "arms". Because we restricted our interest to rivers reaching Bergen, we only worked on the section of the map from equator to 65 degrees north, and from -95 to 15 deg east (see Figure 9). Bergen is at 60 degrees north and 5 degrees east. This makes Bergen lie in the north-east corner of the section, a choice justified by water transport to Bergen predominately coming from the south-west. (From the south because there is more humidity further south, and from the west partly because of the Coriolis effect and partly because to the west there is the humid ocean, while to the east there are dry landmasses.)

Figure 9: figure
Part of the world I've been working over. (Map retrieved from Google Maps.)

The *bec* is computed as follows:

- At a given time, take the *tcw* values on the grid

$$W = \{-95°E, -94°E, \ldots, 15°E\} \times \{0°N, 1°N, \ldots, 65°N\}.$$

  (This grid is slightly coarser than what was available, to save some computing time. It has not been checked how refining the grid would affect the results.)

- Tweak this grid map by setting the value on the equator to be slightly larger than the maximum value on the grid above. This will make the equator appear in the very beginning of the filtration, and always be a connected region.

- Construct the top-down filtration induced by this map (a filtration of $K(W)$, where the most humid points are added first).

- Add an extra vertex and a line from this to Bergen's vertex, and give the extra vertex an even (slightly) larger value than the equator. (The line gets the same filtration value

as the vertex corresponding to Bergen, making the extra vertex connect to Bergen as soon in the filtration as possible.) See Figure 10.

- Compute the persistent 0-homology of this filtered complex.
- Find the next-to-oldest class in the persistent 0-homology. Because of the tweaking above, this class must represent the equator (the oldest represents the extra point adjacent to Bergen.) We are using the top-down complex, so the highest values correspond the the oldest classes.
- The oldest class lives for ever, representing the single connected component of the complex at the end of the filtration. Since everything is connected by the end, the equator's class must die at a certain filtration value, which will be the value of the filtration step where it merges together with an older class. Since there is only one older class, we know that this is the filtration step where the equator's component and the extra vertex's component merge. This value is the *bec* at the given time.

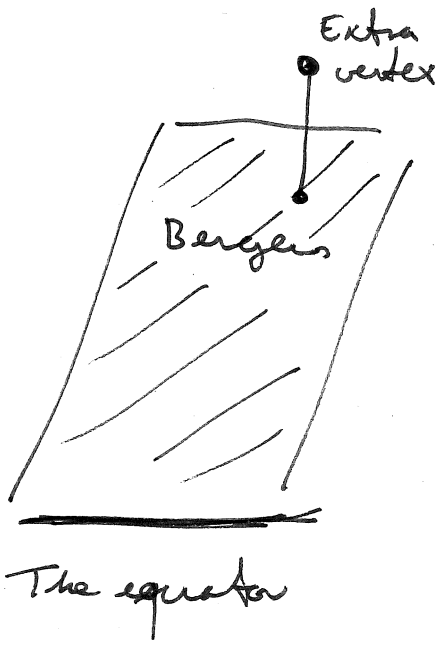


Figure 10: The complex constructed when computing the *bec*.

That the equator's component and the extra vertex's component merge means exactly that there is a path between the equator and Bergen. So the *bec* is characterized by being the maximal *tcw* value q where there is a path between Bergen and the equator through points with value $\geq q$. The reason for adding the extra vertex instead of just changing the filtration value of Bergen is basically that the latter forgets the value of the *tcw* in Bergen, and the former does not. The latter would effectively be looking for paths to a point adjacent to Bergen instead. This is not necessarily a bad thing, but not what was wanted of this calculation.

To describe the *bec* in more a descriptive (but less precise) way: Pretend that the elevation of the ground is given by the *tcw* value, and that there is water filled up to the highest mountain tops. Let the water drain, and stop the draining at the first level where you can walk on dry land from the equator to Bergen. This level is the *bec* value.

The *bec* doesn't say anything directly about water transport, but it does say when there are paths with potential of having large water transport (as the humidity is high along this path). In section 6.1 you can see that, in 2012 at least, the most notable *bec* values all correspond to atmospheric rivers reaching Bergen.

See figure 11 for the *bec* values for 2012. It lies at a higher level in summer. This is because (on the northern hemisphere) it is warmer then, and thus higher humidities. (If one wants a more detailed view of the *bec*, its values are plotted one month at the time in appendix B.)

Since the humidity in general decreases when going further away from the equator, we had a slight fear that this *bec* would be nothing more than a measure of how wet it is in Bergen. There *is* a strong correlation between *bec* and the *tcw* value in Bergen. In 2012, they are equal at 1049 of 1464 time steps. See figure 12 for the relative difference between *bec* and *tcw* in Bergen. Looking at typical contour lines of *tcw* at high *bec* and high *tcw* in Bergen suggests that there is some qualitative difference. When the *bec* is high, the contour lines typically form a narrow river-shape up to Bergen, and the integrated water transport is high and runs along this shape. When the *tcw* is high in Bergen, but *bec* is relatively low, these shapes appear less frequently and the water transport towards Bergen is less pronounced. When it still is high water transport towards Bergen, this seems to be in the form of a plume (a disconnected region of humidity) and not a river. This is illustrated by the figures in sections 6.1 and 6.2. Here I have chosen time steps where the *bec* and relative difference are high, respectively, and plotted the IVT (integrated water transport, see Definition 5.2) and the contour lines of the *tcw*. The graphs of the *tcw* and *bec* around this time step are also included to complete the picture.

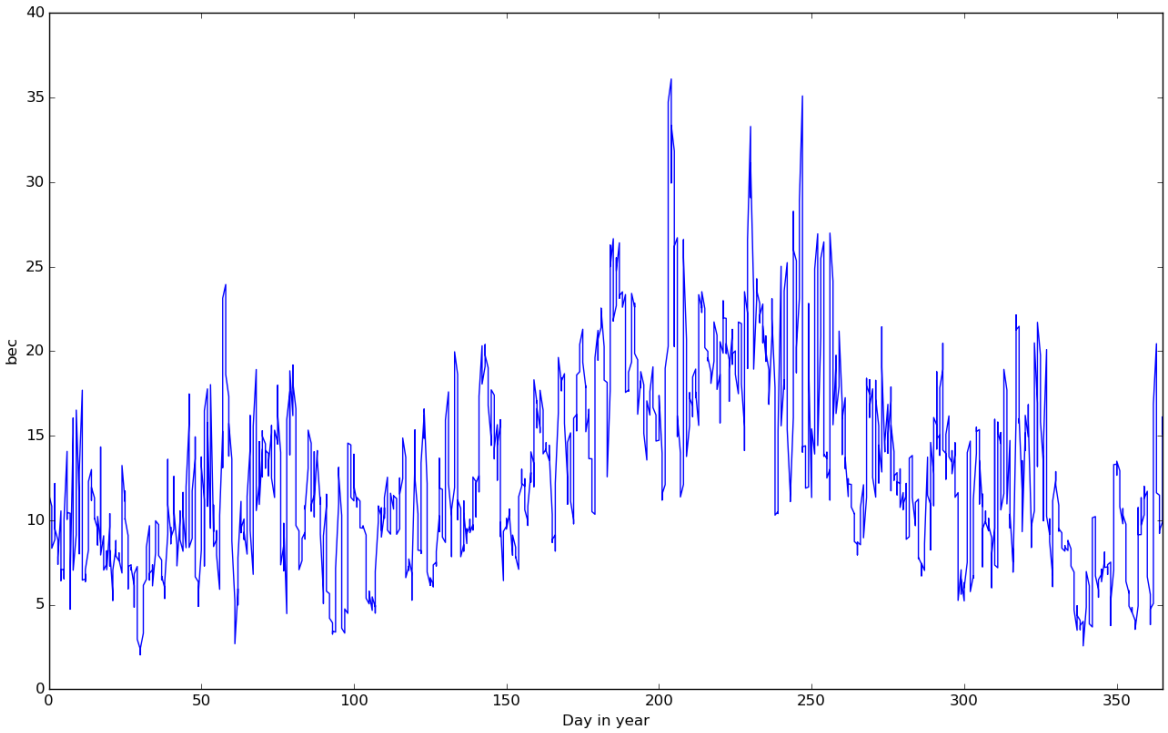


Figure 11: The bec values in 2012

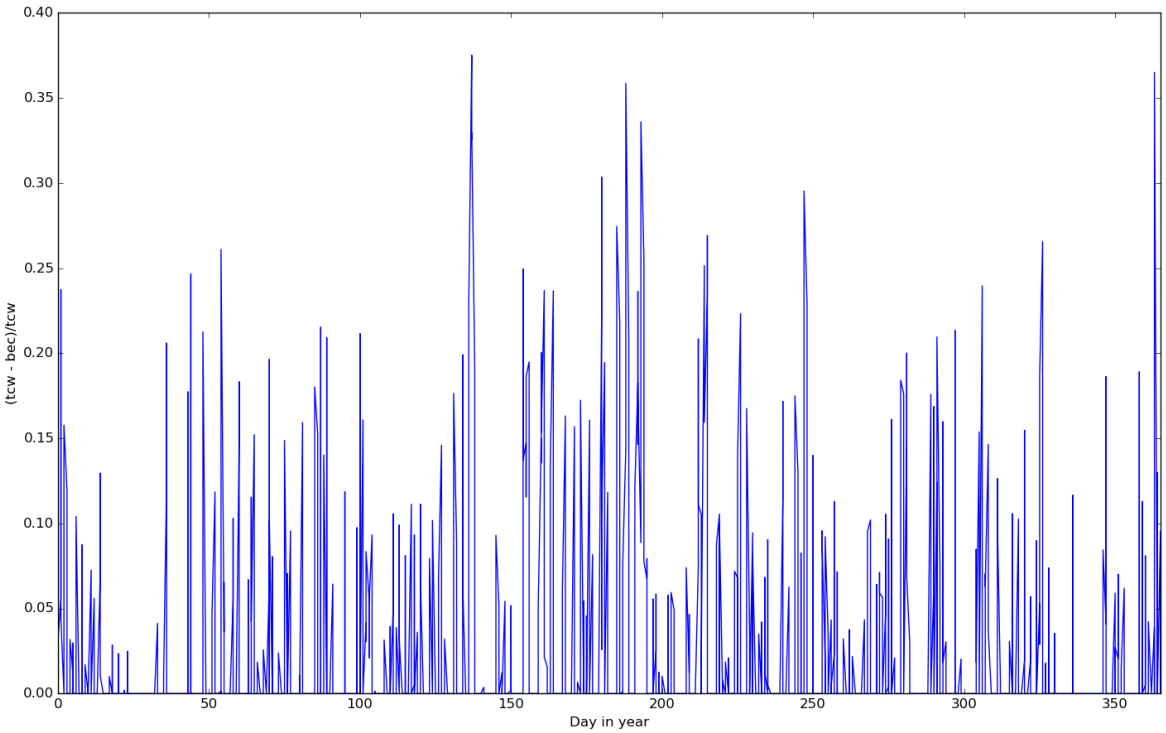


Figure 12: The relative difference between *tcw* and *bec* in 2012.

5.6 Time enters the picture

In this section I will describe the *tbec*, which is a persistence diagram computed for a given time interval. The general idea of the computation is to look at the complex in figure 13 (a $K(G)$ with an 'extra wall' added), and make a filtration on this using the *tcw*. In this filtration, paths from the equator to Bergen will correspond to a 1-cycle, so the hope was that computing the persistent 1-homology of this filtration would give a nice picture of how many paths there would be from the equator to Bergen at a given filtration step. Unfortunately, the cycles appearing when adding the extra wall are extremely few compared to the cycles only contained in the big cube, and not particularly persistent compared to them either. Comparing the persistent homology of the complex with and without the back wall, one can see that the total number of cycles number in the tens of thousands, while the difference of cycles number in the hundreds (and those cycles are not particularly persistent compared to the whole persistent homology.) It looks like it is impossible to see which cycles are interesting and which are not by only looking at the persistence. This means that some other criterion to choose only the interesting cycles is needed.

A natural solution to this would perhaps have been to look at the maps $H_1^i \to H_1^\infty \cong \mathbb{F}_2$ and see which classes are mapped to 0 and which are mapped to 1, as this would exactly tell the difference between cycles going the whole way round (and therefore corresponding to paths between the equator and Bergen) and those that don't. Unfortunately, this idea came to me rather late in the process, and I did not have time to implement it.

The solution I actually used is less natural because it depends on the choice of basis for the homology.

As already mentioned, the goal of this computation is to get a picture of the amount of atmospheric rivers reaching Bergen during a given time period. It is related to the *bec*, as it also looks for paths through high humidity from the tropics to Bergen. The result is a persistence diagram which I call the *tbec* (as it is in the same spirit of the *bec*, but with the time dimension added). The *tbec* is computed as follows:

- Choose time steps $t_{a-n} < t_a < t_{a+m}$. The goal is to find atmospheric rivers reaching Bergen during the time interval $[t_a, t_{a+m}]$, but the interval is expanded to allow for rivers originating earlier than t_a.
- Take the *tcw* values on the grid

  $$W \times \{t_{a-n}, \dots, t_a, \dots, t_{a+m}\},$$

  (where W is the same grid as in section 5.5.)
- Tweak this map by setting the value on the points Bergen$\times\{t_{a-n}, \dots, t_a, \dots, t_{a+m}\}$ to be slightly smaller than the minimum value, so that these points will appear at the very end of the filtration. This is effectively the same as removing these points[2]. The reason for this is that we are only interested in rivers reaching Bergen *after* t_a.
- Also set all the values at the equator to the maximum value, so that all of the equator will be included in the first filtration step.

[2] As long as removing these points don't change the homology of the final complex in the filtration.

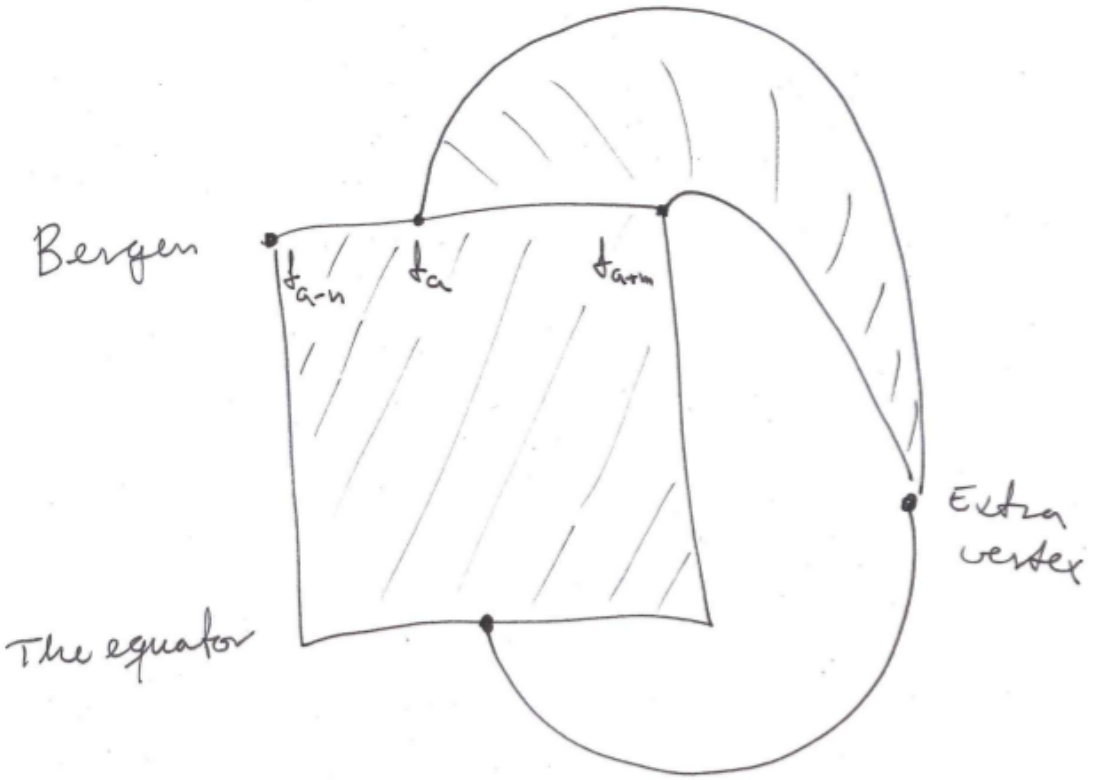


Figure 13: The complex constructed when doing the time-computation (the longitude is suppressed for better drawability.)

- Construct the top-down filtration induced by this grid map. This is a filtration of the complex $K(W \times \{t_{a-n}, \ldots, t_a, \ldots, t_{a+m}\})$, where the highest valued points are added first.
- Choose a point on the grid lying on the equator, add an extra vertex to the complex and connect this to the chosen point by a single line. Give both the extra vertex and line the same filtration value as the equator (so that they appear in the first filtration step.) Since the equator is contractible throughout the whole filtration, the homology does not depend on which point on the equator one chooses. (This extra line is added to make it easier to see which cycles use the back wall.)
- Also use lines to connect all points in Bergen$\times\{t_a, \ldots, t_{a+m}\}$ to the extra vertex, while adding 2-simplices between these lines to fill the holes between them. Give these cells simplices filtration values by the same principle as for constructing the top-down filtration (a cell appears in the filtration as soon as its whole boundary has appeared.) See Figure 13.
- Compute the persistent 1-homology of this filtered complex, keeping track of the basis for the cycles.

- Discard all the cycles in the basis that don't include the line from the equator to the extra vertex. (The basis for the cycles is the one found after column reducing the matrices, as described in section 3 This is the step depending on the chosen basis. (It is possible that it also depends on the point on the equator chosen a few steps above.)
- The persistence diagram *tbec* consists of the persistence intervals of the remaining homology classes.

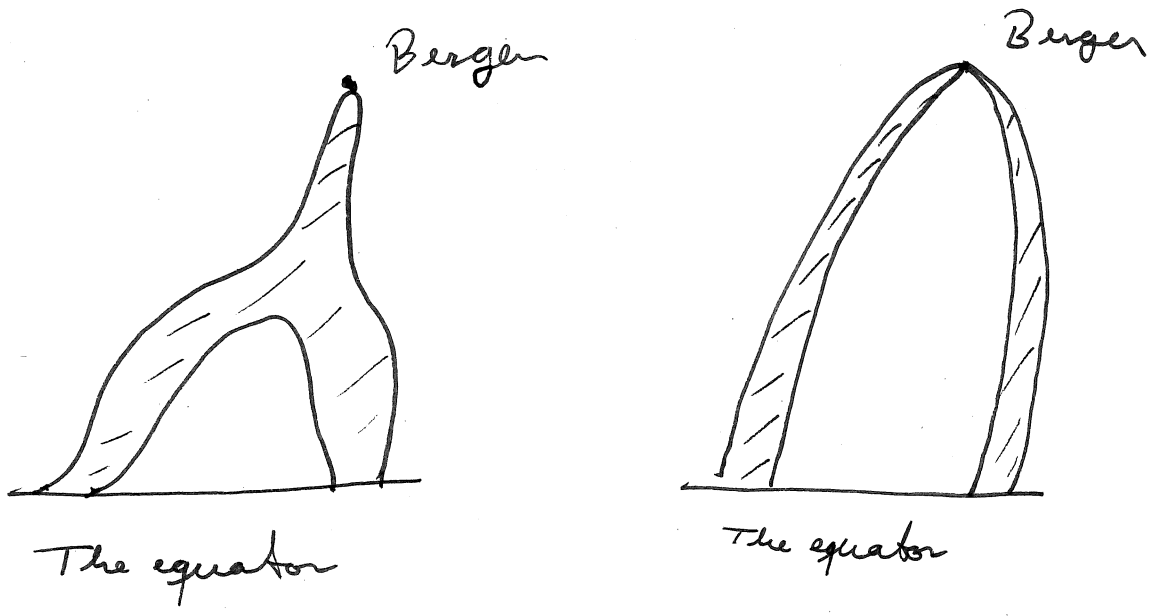


Figure 14: Two examples where it is not well defined how many homology classes we want to pick. The figure shows the filtered complex at the step where the cycles are born (the back wall is not drawn). As for now, my intuition (and testing on simple cases) suggest that the left-most example will just include one cycle and the rightmost will include two cycles.

The *tbec* has been computed for intervals $[t_a, t_{a+m}]$ corresponding to single months, and the intervals $[t_{a-n}, t_a]$ corresponding to about 10 days. The latter interval is chosen more or less arbitrarily, the only thing that has been checked regarding this is that there is a difference between this and $n = 0$. In section 6.3 you can see the *tbec* for all months in 2012 (except for January, which would need some slight alterations of the code to be computed, since the *tcw* values are stored yearly), and in section 6.4 you can see the diagrams for June in the years 2000-2011.

One way to extract information from the *tbec* is to observe that the more classes there are at high filtration values, the more (potential) atmospheric rivers are there in the chosen time interval.

One perhaps surprising observation is how much the number of persistence intervals in each diagram varies. The smallest have only about fifty intervals (see December 2012), while the largest have well over a thousand (for example June 2011). A possible explanation for this is that there is more turbulence in the summer months.

A path between Bergen and the equator will be detected as a class in H_1. The thought is that the amount of these paths (at least on high filtration levels) will give an indication of how many atmospheric rivers would hit Bergen during a given time interval. As mentioned, the persistent homology of the complex has quite a few more cycles than the ones detecting these paths, which is all the cycles that don't pass through the 'back wall' are discarded. What also is mentioned previously, is that this is dependent on the basis for the cycles, which makes things rather messy. For example, when two cells have the same filtration value, the algorithm chooses some implicit ordering on these, and the persistent homology is independent of this choice. The *tbec* diagram however, is not independent of this choice.

One example: There are two paths connecting the same component of Bergen×(time-interval) to the equator. Depending on choice, either one or both of the cycles these create will use the "back wall". In the *tbec* as it is computed now, this will be regarded as two rivers (I think).

There are probably more cases where it is not well defined what the result should be, but most of the cases that have been found are either not very persistent or else appears only at low filtration values. Nevertheless, the ambiguity of the *tbec* does make one feel that this is currently not a very good way to detect atmospheric rivers.

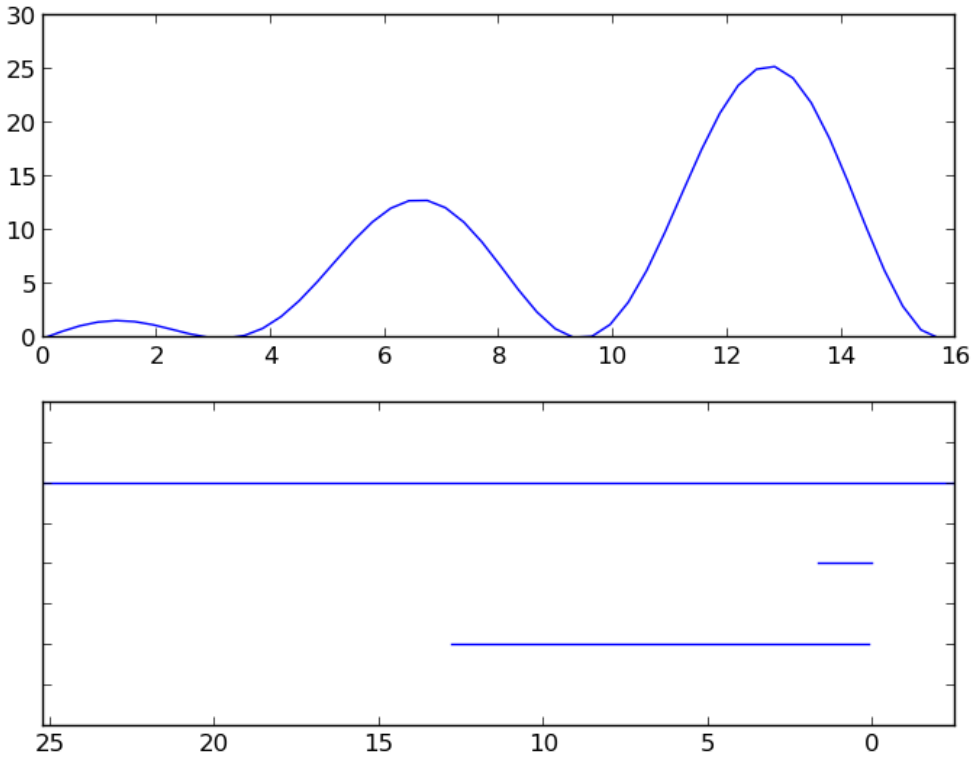


Figure 15: A function and the persistence diagram of its top-down filtration.

5.7 Comparing these two

Both the *bec* and the *tbec* are looking for paths between Bergen and the equator, so one would think that it would be possible to compare them somehow. For all the *bec* values, there should be a homology class in the *tbec* that detects the same path as the *bec* value. The *bec* can be regarded as a map on the grid $\{t_a, \dots, t_{a+m}\}$, so we can construct the top-down filtration $T(bec)$. As long as nothing too weird happens, the *bec* values in the same component of $T(bec)$ at a given filtration value should all correspond to the same homology class in the *tbec* at that filtration value (they all correspond to the same path as it persists over time). If we imagine that the function in Figure 15 is given by the *bec*, there is one path born at filtration level 25 lasting forever, and one path born at level 13 and one at level 2 both dying at level 0. So comparing the 0-homology of $T(bec)$ to the *tbec* should show some correspondence. This persistence diagram induced by the *bec* has much fewer persistence intervals than the *tbec*, which makes sense as the *bec* only catches the most dominant path at a given time step, while the *tbec* captures "paths through time" and may also capture several for one time step. In the figures in section 6.5 one can compare the most persistent homology classes of the *tbec* to the diagram induced by the *bec*. It would have been informative to compute the bottleneck distance (which is described in [1]) between these diagrams but, again, this was something that there would not be enough time for in this thesis. By inspection alone, there seems to be some correspondence between the two diagrams. Almost all of the intervals in the diagram induced by the *bec* seem to have a corresponding interval in the *tbec*. On the other hand there are quite a few intervals in the *tbec* that don't seem to have a pairing in the *bec*-diagram. If the *tbec* detects exactly what we want it to detect, these should correspond to "blobs" of humidity travelling from the subtropics to Bergen, or perhaps a path not detected by the *bec* because there was another more dominant at the same time step. The bottleneck distance would have provided a clear way to find these pairings, but as for now the reader must be content with looking at the pictures in section 6.5 and see if they agree with the author.

5.8 Dynamics

I mentioned a strategy using the dynamics of the atmosphere to detect atmospheric rivers, for which there would not be enough time to do. In this section I will describe the basic thoughts I had on this strategy.

From a point in W there is a 1-cell in $K(W)$ going north, and one going east. Assign the (absolute values of the) corresponding components of the integrated water transport (Definition 5.2) as values to these 1-cells. By generalizing the construction of the top-down filtration, extend this to an opposite cellular map on $K(W)$. Compute the persistent homology of the filtered complex, and see if there is some information to be extracted.

One could do the same as in the *bec* computation: Look after connectivity between Bergen and the tropics. Looking at the plots of the IVT in section 6.1 suggests that this would perhaps be a little restrictive, as there seem to be many cases of quite dominant atmospheric rivers there that would not be detected by this method. It seems to be more fruitful to find some way to see the extent and shape of the components of this filtered complex. This could perhaps be a way to find atmospheric rivers in general, and not just those reaching Bergen.

6 Results and discussion

6.1 Contour plots of the total water column when the bec is high.

In this section there are figures trying to illustrate how well the *bec* from section 5.5 captures atmospheric rivers. The upper part of the figures contain contour lines (or level curves) of the *tcw* (the coloured lines), and the vector field given by integrated water transport (the black arrows) at given time steps. The scales of the *tcw* and the water transport are not specified, as the discussion here is more qualitative than quantitative. The red contours bound the regions with highest humidity, and the blue contours bound the regions with lowest humidity. The size of the arrows say how large the water transport is.

The lower part of the figures contain a graph of the *tcw* and the *bec* around the given time step of the upper half. This time step is marked with a red cross. (In most figures this is in the middle of the graph, but in the one for day 0 it is to the left as the *bec* values to the left of this hasn't been computed.) Note that the scale on the graph can vary from figure to figure.

The time steps are chosen by finding the *bec* values that stand out the most compared to the general level in their vicinity. Looking at the figures in appendix B, the reader hopefully agrees that there are five that stand out the most: day 58 (February), day 204 (July), day 230 (August), day 247 (September), and day 363 (December). These five are shown first. Afterwards, some other time steps are included, where the *bec* still is high, but doesn't stand out as much.

Figure 16: The IVT and contours of the *tcw* at day 58. The *bec* is high:

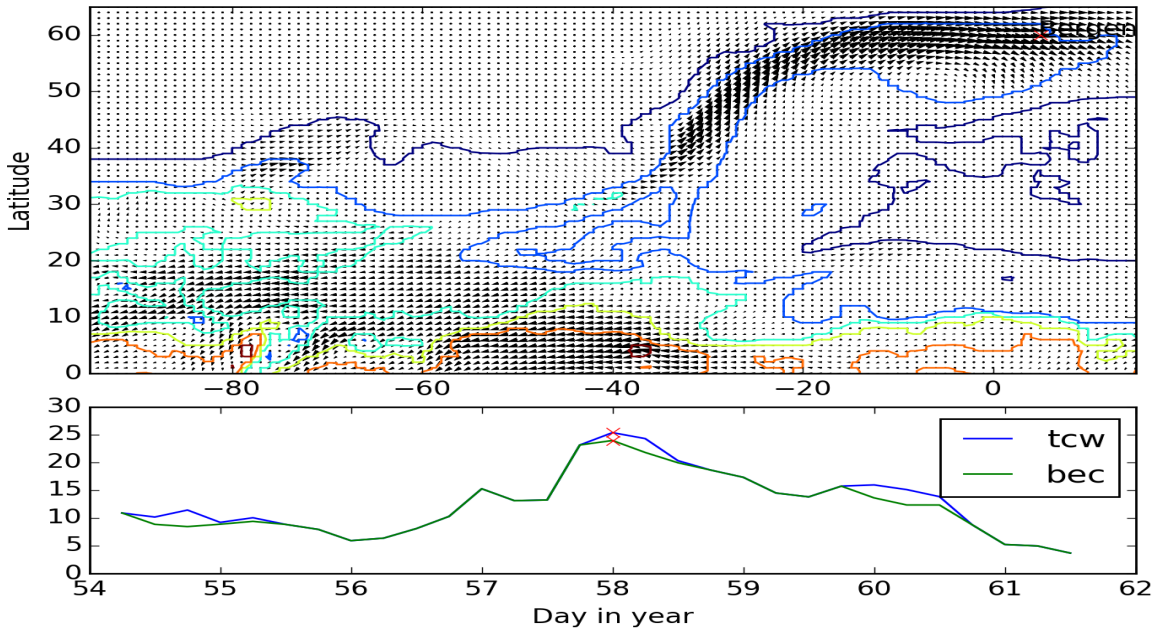


Figure 16 (day 58) seems to be a prime example of an atmospheric river hitting Bergen. It has a long, thin shape and along it is a considerable water transport which seems to hit Bergen straight on. This is the highest *bec* value in the winter/spring of 2012 (day 58 is in the end of February). It would probably be among the highest values for the whole year after normalizing for the seasonal variance.

Figure 17: The IVT and contours of the tcw at day 204. The bec is high:

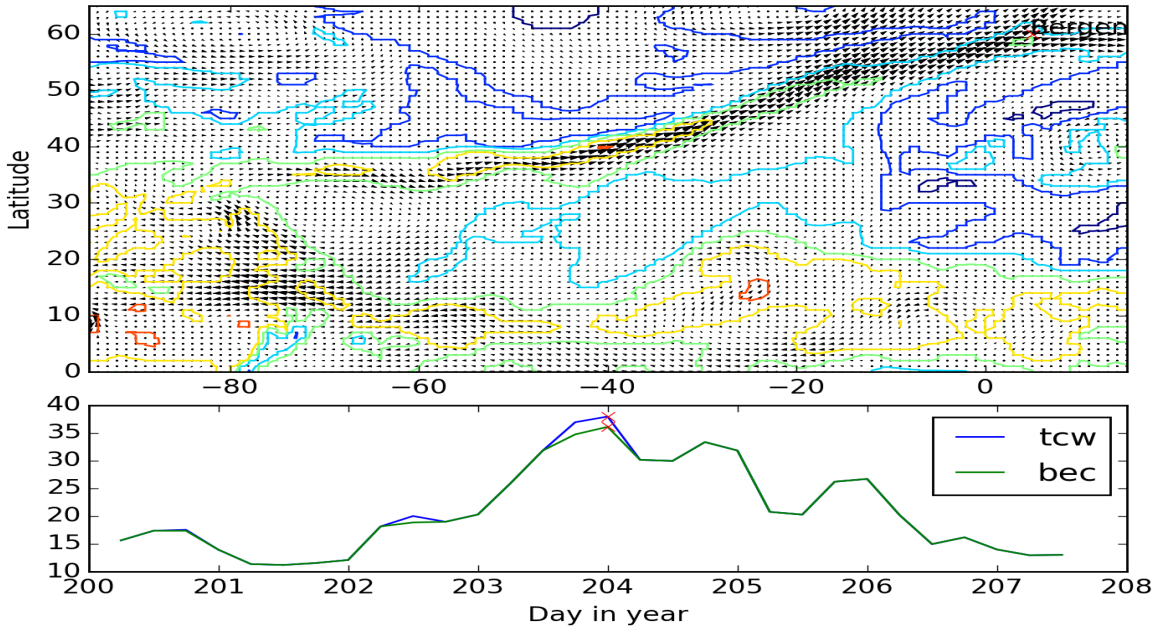


Figure 17 (day 204) also seems to depict a nice example of an atmospheric river. The contour of the tcw form a long, thin shape towards Bergen, and along this is quite high water transport. This is the highest bec value in the whole of 2012 (about 36.1). Figures 26 (day 204.75) and 27 (day 206) seem to show later instances of the same river, with less water transport for each iteration.

Figure 18: The IVT and contours of the tcw at day 230. The bec is high:

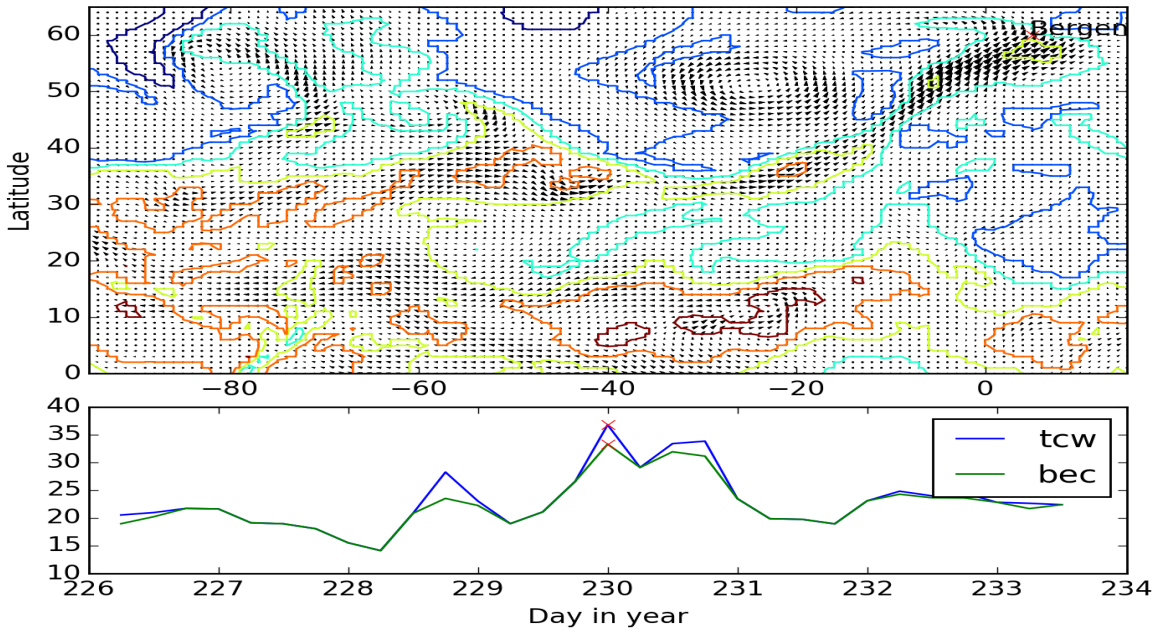


Figure 18 (day 230) is an example of something that looks like an atmospheric river, although here the water transport is slightly more concentrated at the tip of the river shape. The *bec* value here is also quite high (33.3).

Figure 19 (day 247) is another case where the shape of the contour is long and thin, but where all the water transport is concentrated at the tip. This looks like it is the same body of water that is at $50°N, -45°E$ in Figure 28(at least the size and direction seem to fit.) Also, in Figure 37 in section 6.2 one can see what happens right after this, where it seems like a region of high humidity has broken free and lies above Bergen. This is probably detected in the *tbec* (in September).

Figure 20 (day 363) also looks like a nice atmospheric river.

Figures 16 to 20 were the five figure where the *bec* values stood out the most. Figure 21 and following are figures where the *bec* value still is high, but does not stand out as much. These do not in the same degree have the characteristics of atmospheric rivers.

Figure 19: The IVT and contours of the *tcw* at day 247. The *bec* is high:

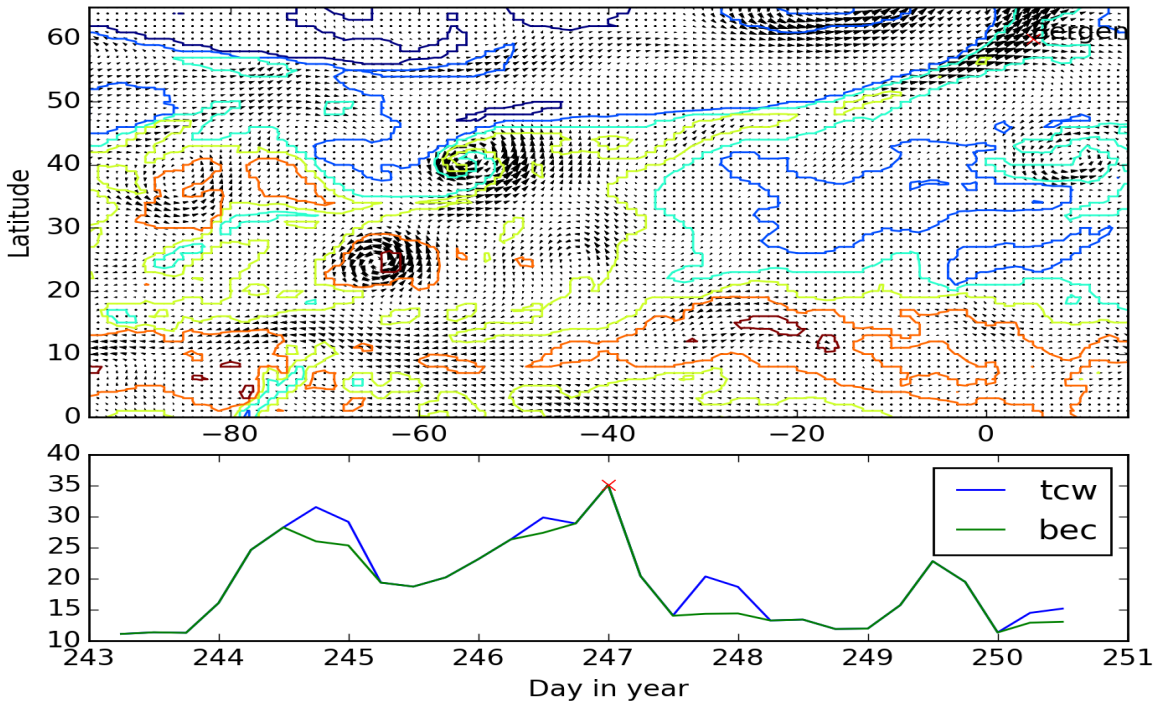


Figure 20: The IVT and contours of the tcw at day 363. The bec is high:

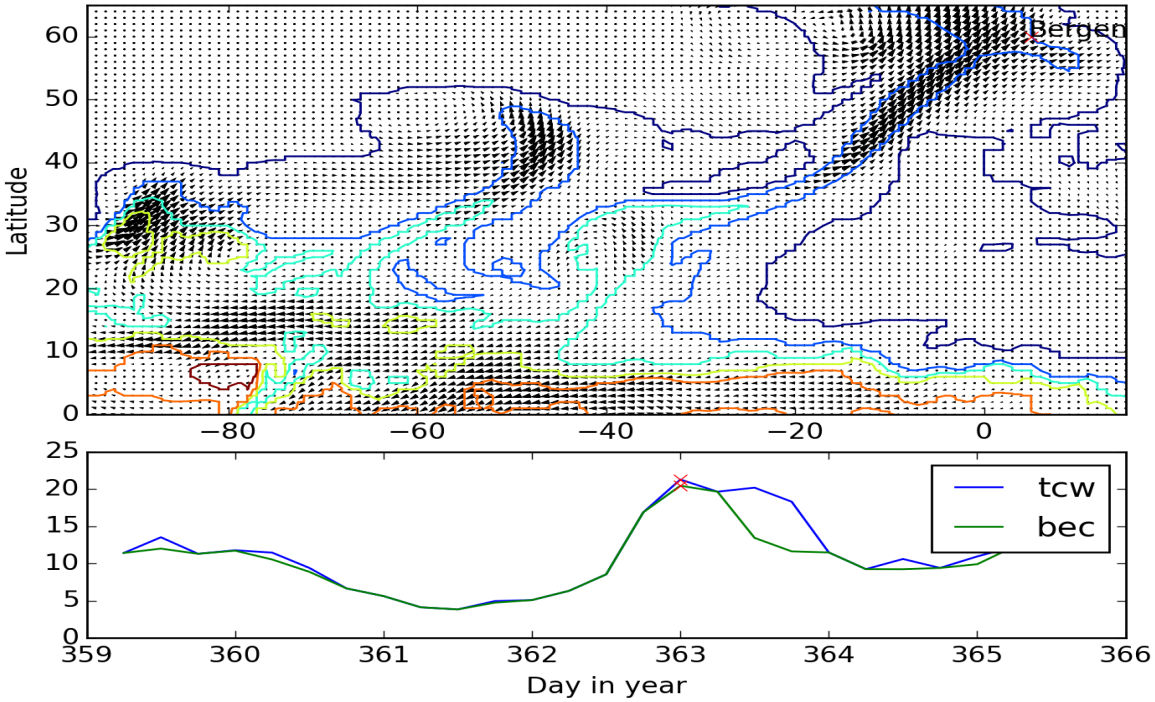


Figure 21: The IVT and contours of the tcw at day 0. The bec is high:

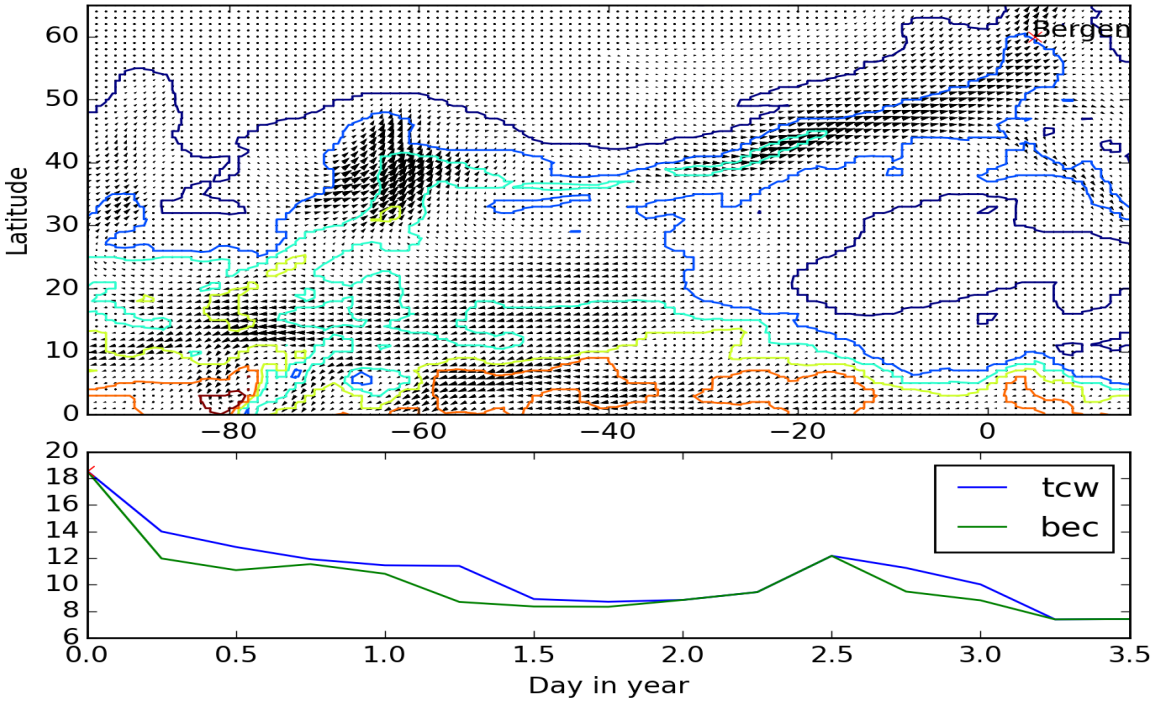
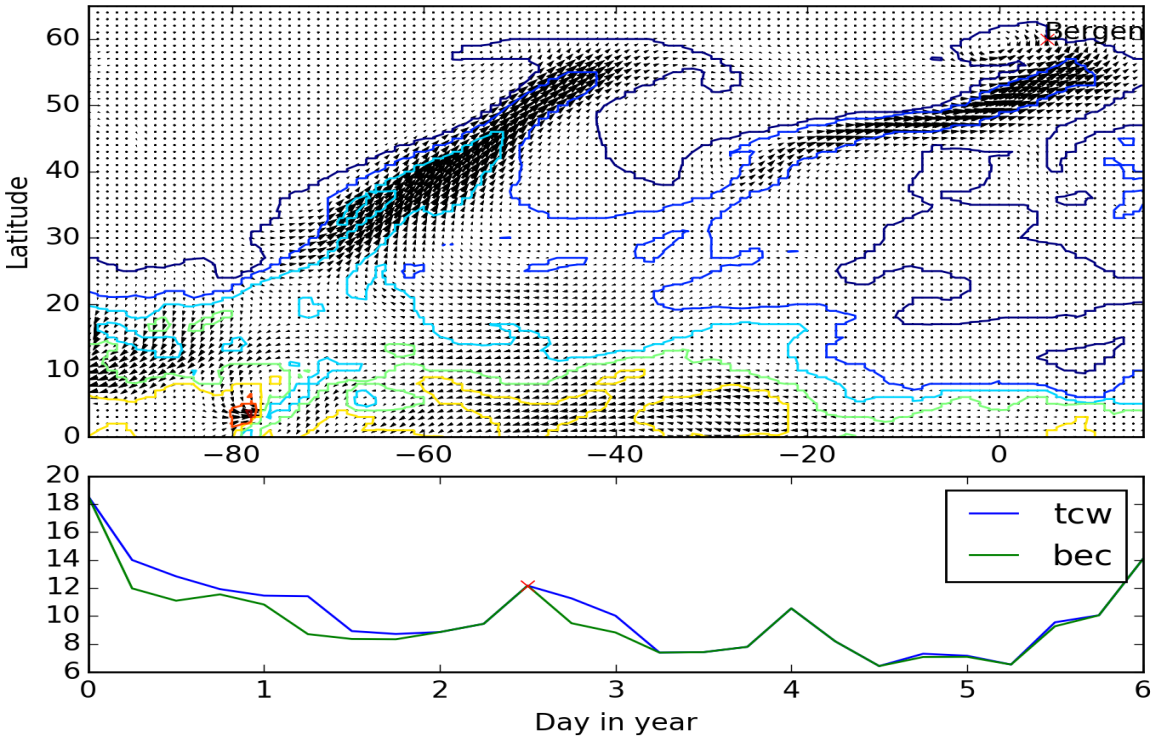


Figure 22: The IVT and contours of the tcw at day 2.5. The bec is high:



In Figure 21 (day 0) we see an example of a distinct long, thin 'river-shape'. There is also some inclination of water transport along this shape, although not as pronounced as in the ones above. This might be a remnant of a river starting in December 2011, as it here is detected in the very first time-step of the year.

In Figure 22 (day 2.5) we actually see two rivers, although only one is close to reaching Bergen at this point. There seems to be quite high water transport along the one reaching Bergen, but it doesn't seem to hit Bergen quite.

Day 0 and 2.5 have another feature that seems to repeat often: After a local maximum, the tcw value often decreases at a lower rate than the bec so that we get a small interval where the tcw is slightly larger. This strengthens the hypothesis that high bec coincides with high water transport, as it makes sense that the humidity stays high for a while after a period of high water transport, even after the transport has ended.

Looking at an animation of the tcw plots, one can see that the western river in Figure 22 (day 2.5) evolves to be the river reaching Bergen in Figure 23 (day 4). In this figure you can clearly see a river shape, and quite pronounced water transport along it. This also does not seem to quite hit Bergen. It should be mentioned that there *could* be time steps close to these were the water transport actually hit Bergen, but the bec is somewhat lower. This has not been checked.

Figure 23: The IVT and contours of the tcw at day 4. The bec is high:

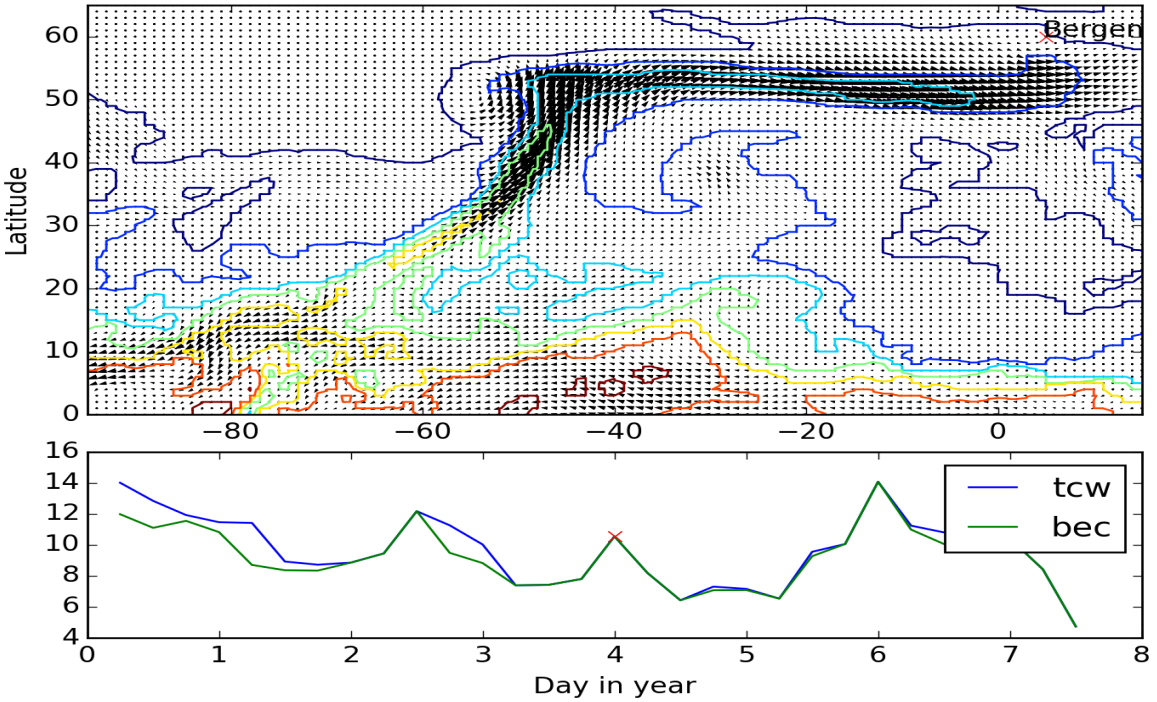


Figure 24: The IVT and contours of the tcw at day 6. The bec is high:

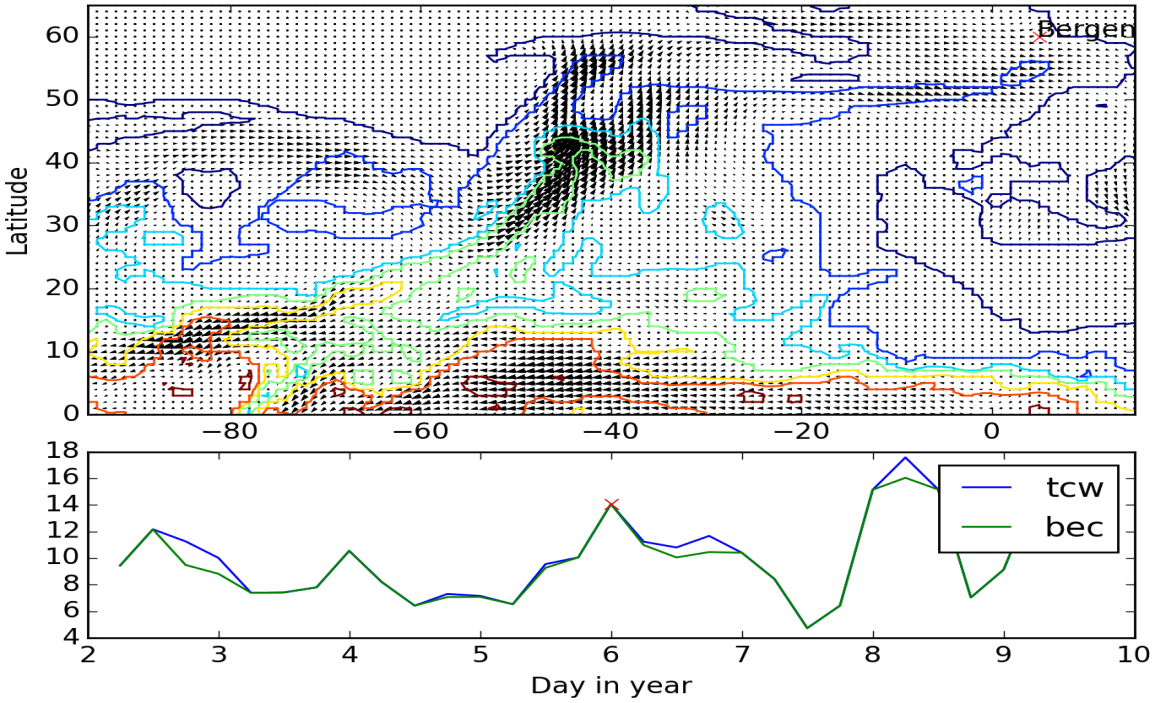
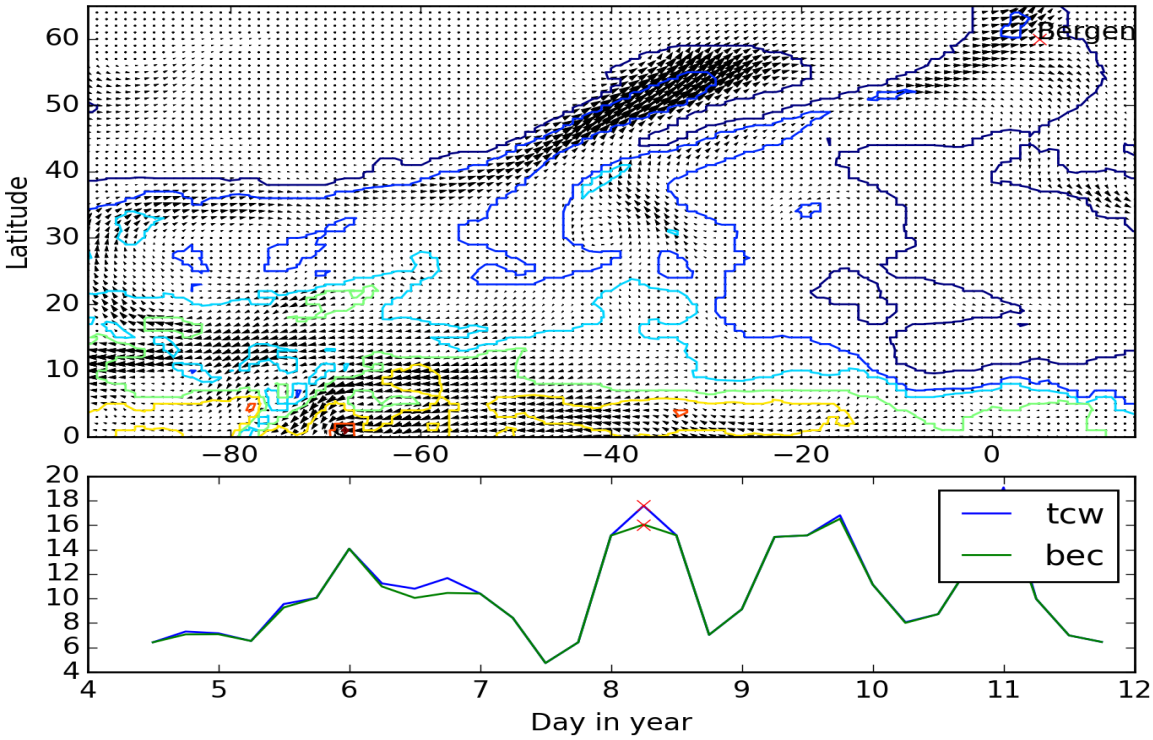


Figure 25: The IVT and contours of the tcw at day 8.25. The bec is high:



In Figures 24 (day 6) and 25 (day 8.25) there are river shapes, but it seems to be less water transport along them compared to the two above (days 2.5 and 4). This does not seem to reflect in the bec value, which is higher in the latter two, even though they have less pronounced rivers.

Figures 26 (day 204.75) and 27 (day 206) seem to show remnants of the river in Figure 17 (day 204).

Figure 28 (day 244.5) has the shape and some water transport along it so it looks like an atmospheric river. The water transport does not seem to hit Bergen. (That the tcw in Bergen continues to rise after this, while the bec falls probably means that the tip "breaks free" after this, creating a disconnected region of high humidity around Bergen.)

In Figure 29 (day 252) there does not seem to be an atmospheric river at all, despite the fact that the bec is quite high. The fact that there are several high bec values in the vicinity suggests that the high bec value is probably because of generally high humidity. With some goodwill one can see a river shape in the contours, but it curves quite a lot, and there is almost no water transport along it.

Figure 26: The IVT and contours of the tcw at day 204.75. The bec is high:

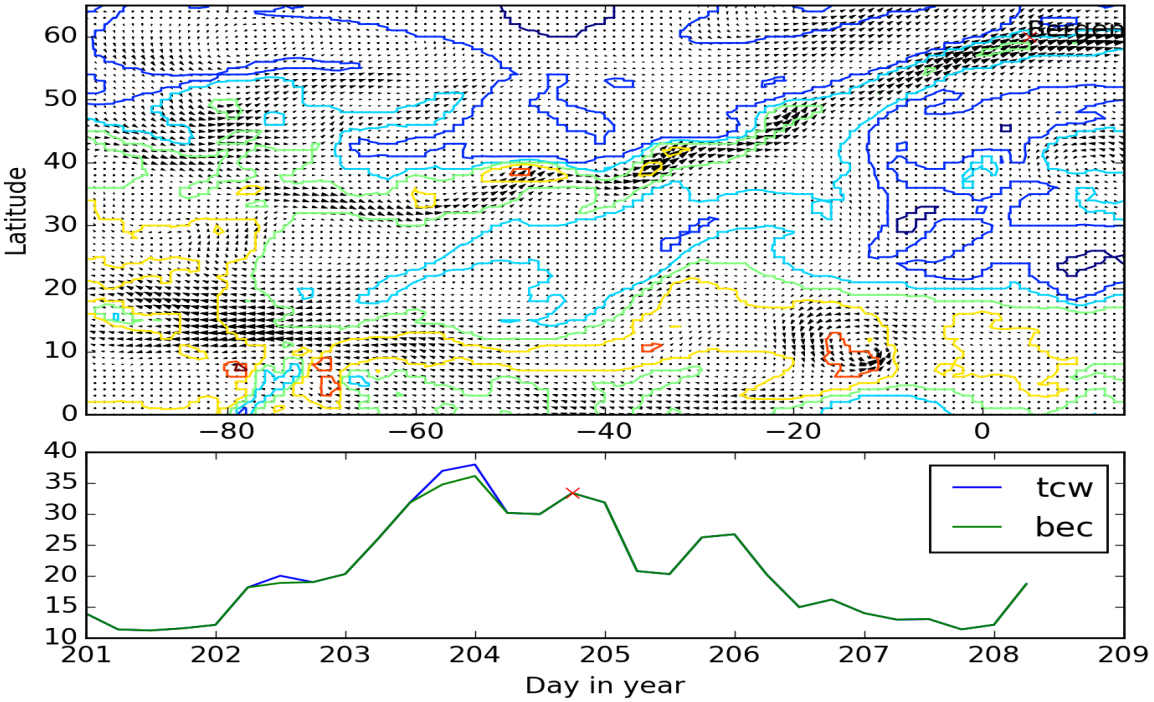


Figure 27: The IVT and contours of the tcw at day 206. The bec is high:

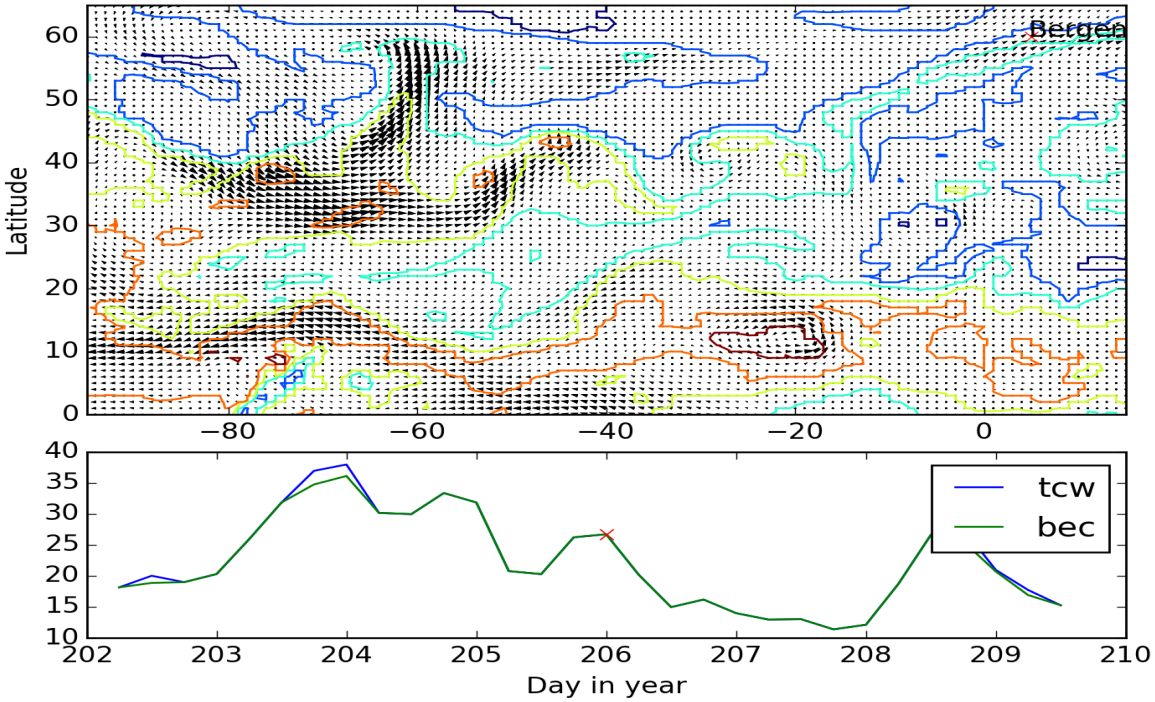


Figure 28: The IVT and contours of the tcw at day 244.5. The bec is high:

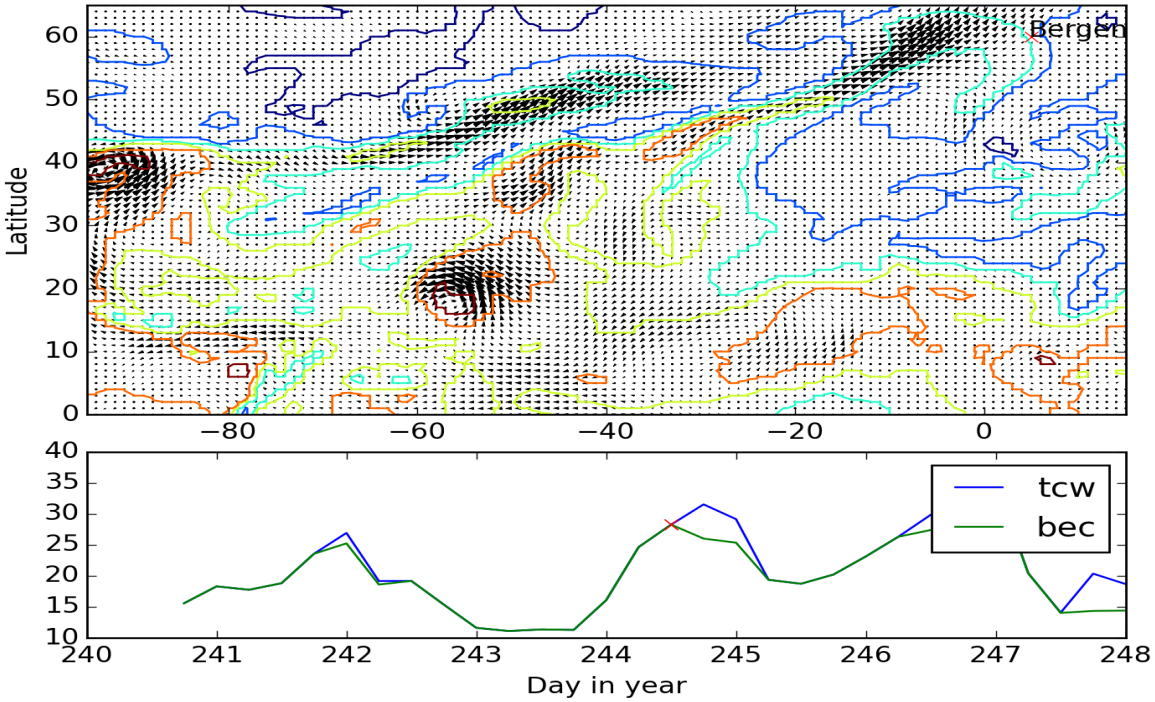


Figure 29: The IVT and contours of the tcw at day 252. The bec is high:

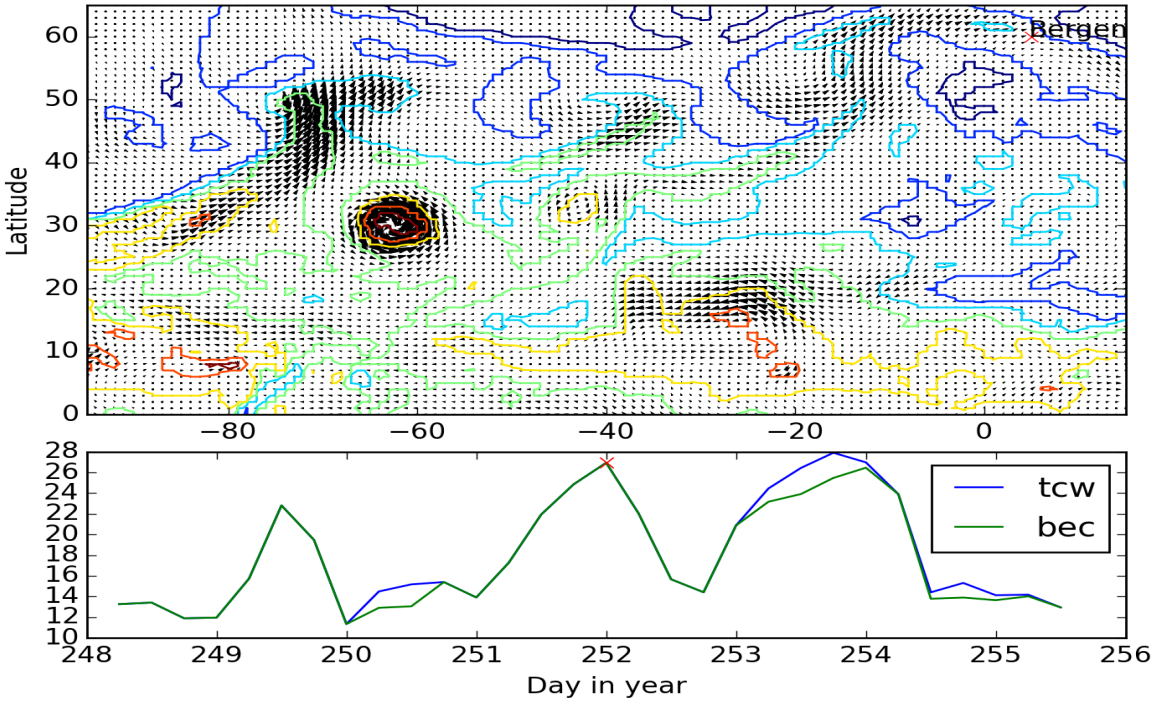
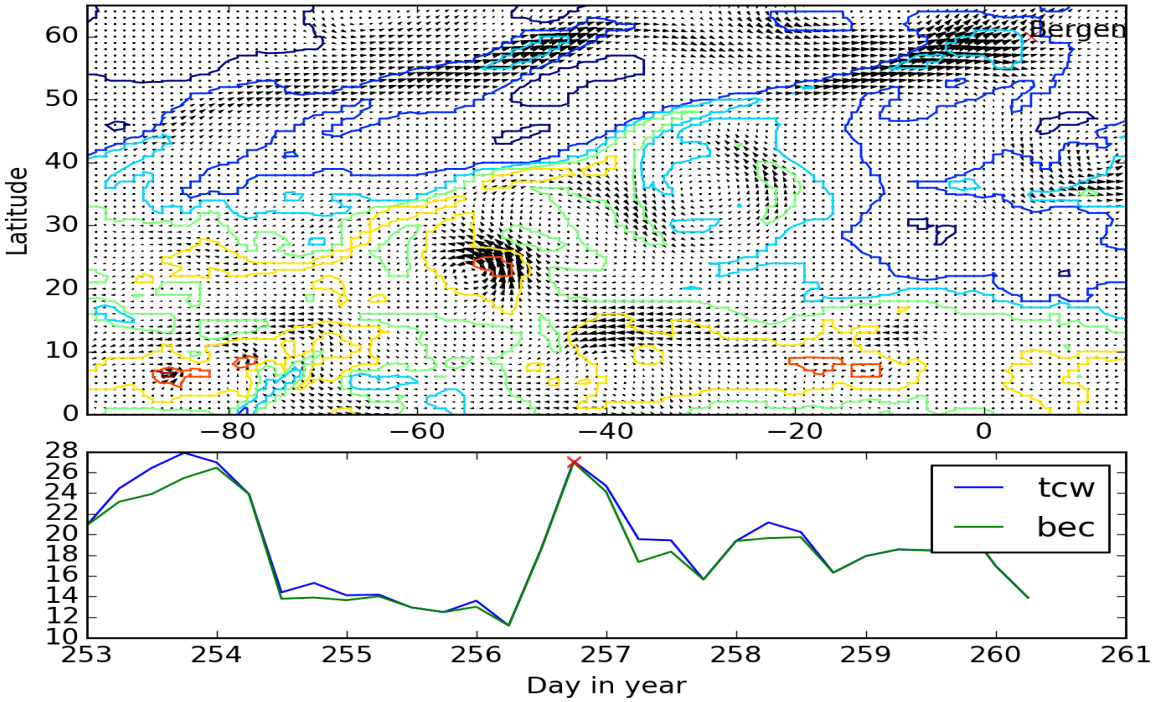


Figure 30: The IVT and contours of the *tcw* at day 256.75. The *bec* is high:



There is slightly more water transport in Figure 30 (day 256.75) than in the previous figure. This is also a case where it seems to be consentrated on the tip of the river shape. One thing that may be noticed here is that the water transport doesn't seem to lie along the drawn contour lines. This would perhaps be different with contour lines drawn at different levels, but if not this would be a curious case, as in all the other cases I've looked at, the transport goes along the contour lines. The *bec* is about the same for both of these time steps, even though the figures look very different. This looks like something that would be detected by the *tbec* (in September).

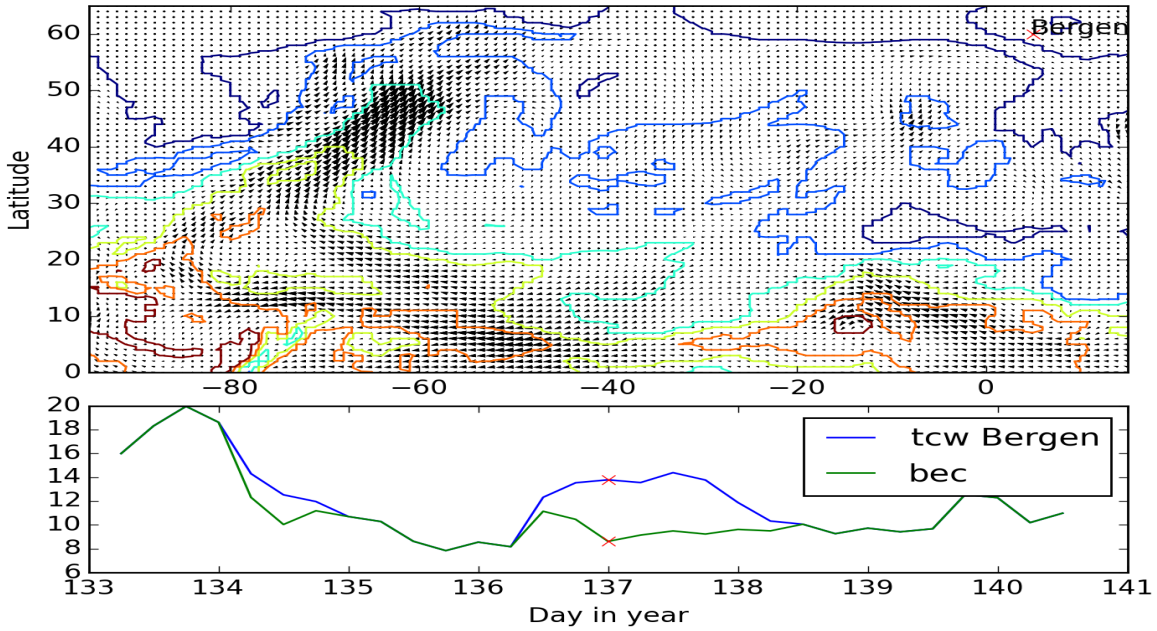
6.1.1 Conclusion

The *bec* seems to be quite good at catching the long, thin shape that characterizes atmospheric rivers. The five *bec* values that stand out the most in Figure 11 in section 5.5 (at least locally), day 58, 204, 230, 247, and 363, seem to have the most pronounced atmospheric rivers. Among these five, day 58 and 204 stand the most out, which looks like it reflects in the figures here, as these are the only two days where there definitely is a long, thing region of high water transport reaching Bergen. Looking at time steps where the *bec* values don't stand as much out, things get slightly more muddled.

6.2 Contour plots of the total water column when the relative difference is high.

This section has the same type of figures as section 6.1, only here the time steps are the ones where the relative difference between the *tcw* in Bergen and the *bec* is at the highest. (Since we take the relative difference the seasonal variance should not be an issue here.) We are trying to illustrate why the *bec* is a better way to detect atmospheric rivers than only looking at the *tcw*. (It would perhaps have been better to choose steps where the *tcw* is high independently of the *bec*, but as these two values are equal more often than not, this would not be independent on the *bec* anyway. Looking at the highest *tcw* values one sees that it is a bit mixed whether they occur at steps where the relative difference is high or not.)

Figure 31: The *IVT* and contours of the *tcw* at day 137. The relative difference is high:



In Figure 31 (day 137) there is nothing resembling an atmospheric river anywhere near Bergen. The only river shape is far away, and the water transport around Bergen is relatively low. The reason for that the *tcw* is so much higher than the *bec* is quite obvious: There is a region of high humidity around Bergen that doesn't connect to the equator (inside the dark blue contour). Note that this is an artefact of the choice of region to work over. If it had been extended slightly to the right the region around Bergen would have connected to the equator. This might be an argument in favour of choosing a region where Bergen is so far up in the corner (else the *bec* value would have been quite high here, with no atmospheric river in sight).

In Figure 32 (day 180.25) there looks like it is some turbulence around Bergen, but nothing resembling an atmospheric river. (Although there seems to be a rather major one farther west.) There is a quite large region around Bergen disconnected from the equator (bounded by a turquoise line). This would probably have been connected to the equator if our region was extended to the east.

Figure 32: The IVT and contours of the tcw at day 180.25. The relative difference is high:

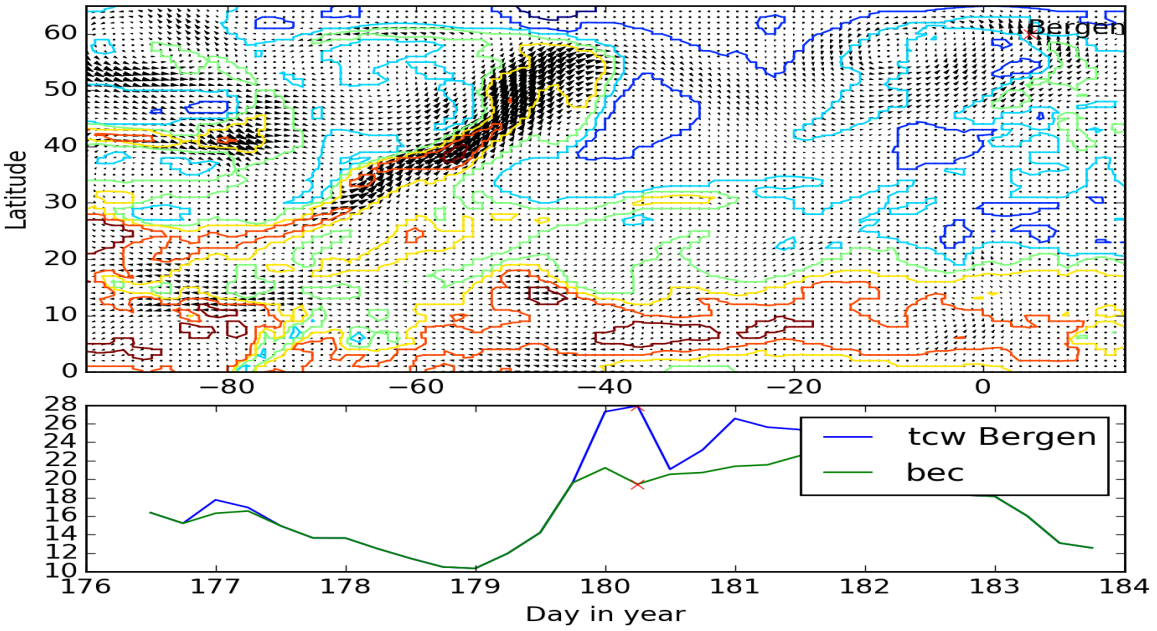
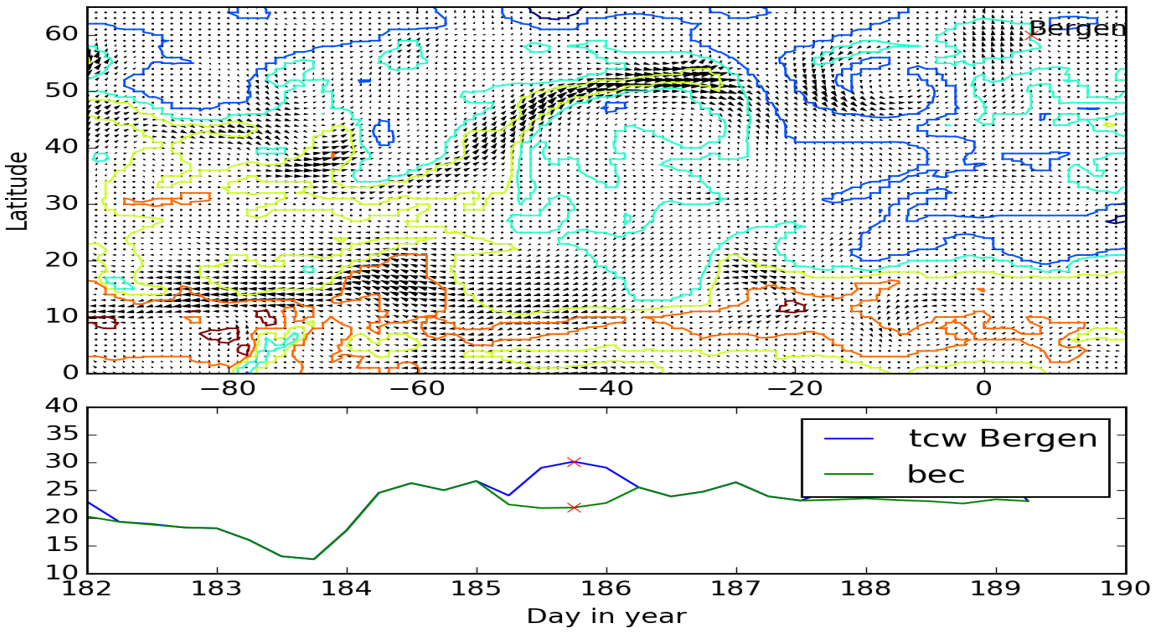


Figure 33: The IVT and contours of the tcw at day 185.75. The relative difference is high:



Figures 33 (day 185.75) and 34 (day 188.75) are close both in time and in features. Neither have any major water transport or a river shape in the vicinity of Bergen, but both have a small region of high humidity disconnected from the tropics around Bergen (with turquoise boundary). In fact you can see both regions in both the figures, and that they are moving eastward.

Figure 34: The IVT and contours of the tcw at day 188.75. The relative difference is high:

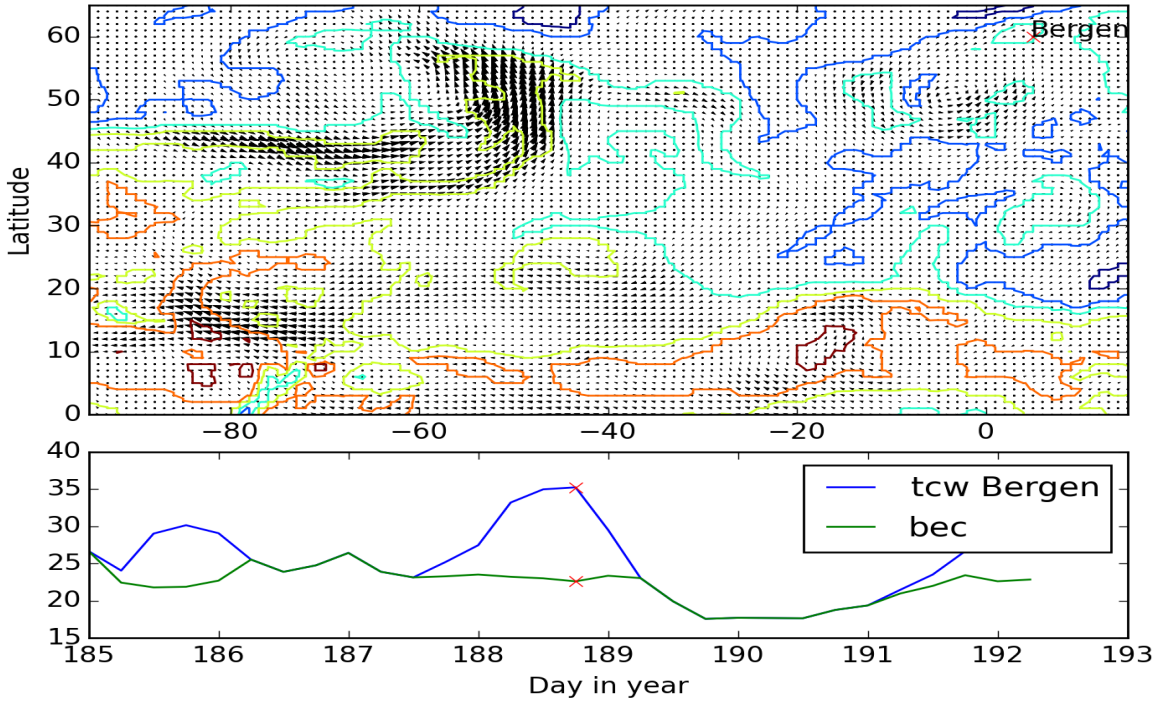


Figure 35: The IVT and contours of the tcw at day 193.75. The relative difference is high:

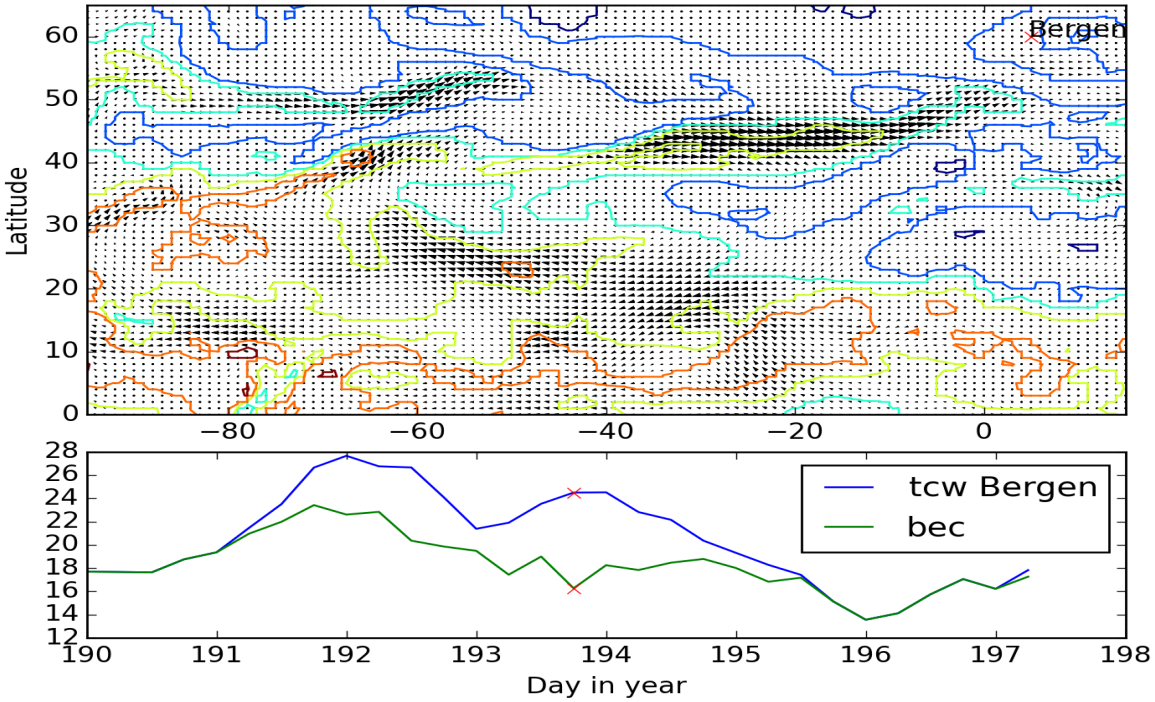
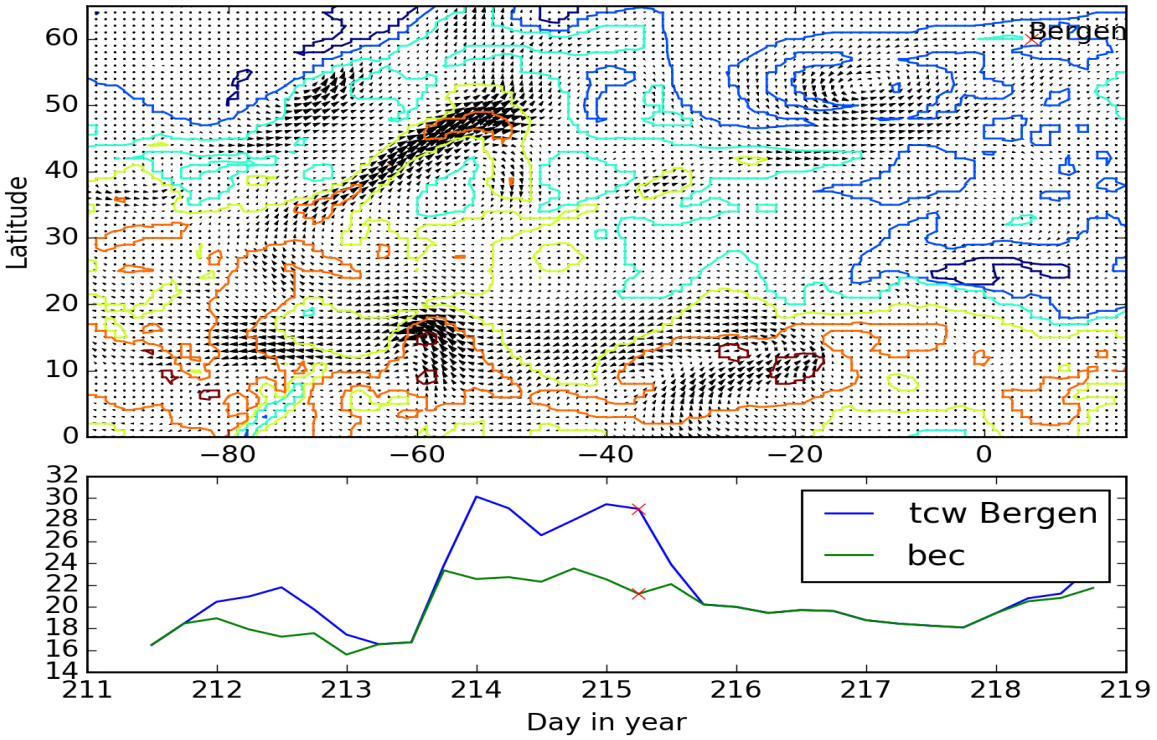


Figure 36: The IVT and contours of the tcw at day 215.25. The relative difference is high:



In Figure 35 (day 193.75) it seems to be an atmospheric river quite close to, but not reaching, Bergen. (It might be the same river that reaches Bergen in Figure 17 (day 204). Also here one would guess that the *bec* could have been higher if more points to the east were included.

Figure 36 (day 215.25) looks quite similar to Figures 33 (day 185.75) and 34 (day 188.75): There is nothing looking like an atmospheric river close to Bergen, and there is a disconnected region of high humidity around Bergen.

Figure 37 (day 247.75) happens right after Figure 19 (day 247) in section 6.1. In the latter there was a river shape reaching Bergen, but with water transport only at the tip. Here it seems that the region with high water transport has broken off and is lying above Bergen. If this is an atmospheric river or not could perhaps be up to debate, but there is definitely a lack of a long thin shape of high water transport.

In Figure 38, the reason for the high relative difference (day 326) seems to be the tiny blue blob close to Bergen (it would probably be possible to extend this to a region including Bergen by choosing a slightly lower contour level). There is an atmospheric river here, but not reaching Bergen.

Figure 39 (day 363.75) is similar to Figure 37 (day 247.75), and the same comments are applicable here. This happens immediately after Figure 20 (day 363), where there is a noticable atmospheric river.

Figure 37: The IVT and contours of the tcw at day 247.75. The relative difference is high:

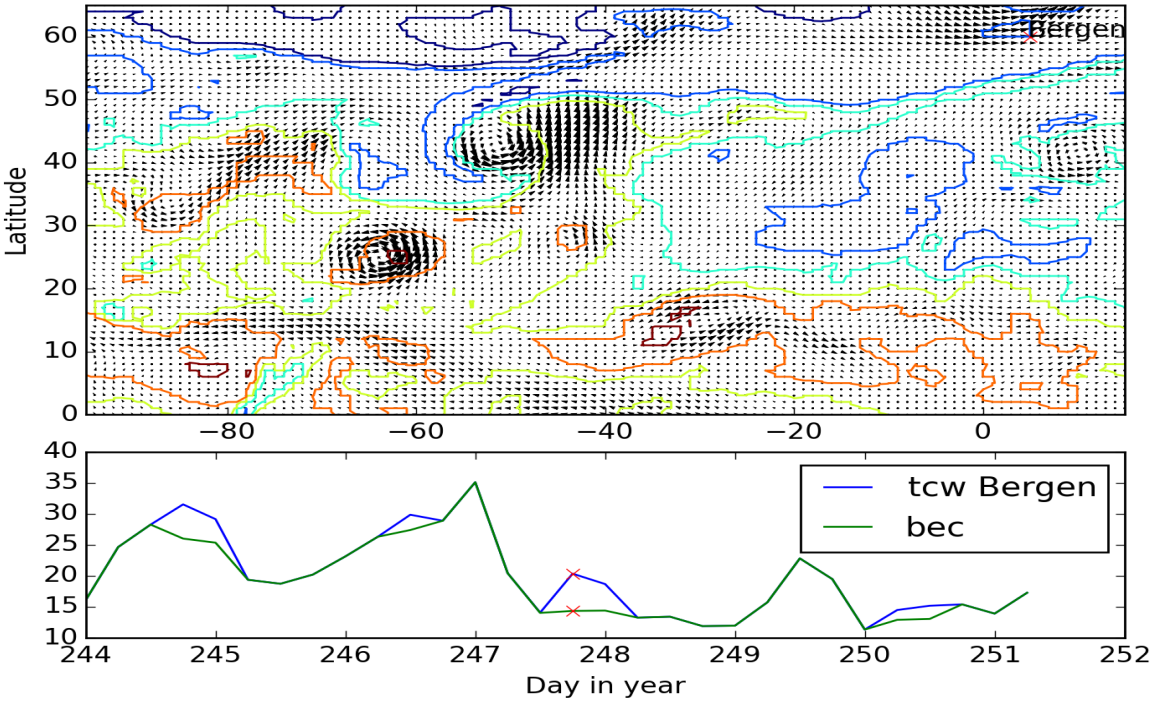


Figure 38: The IVT and contours of the tcw at day 326. The relative difference is high:

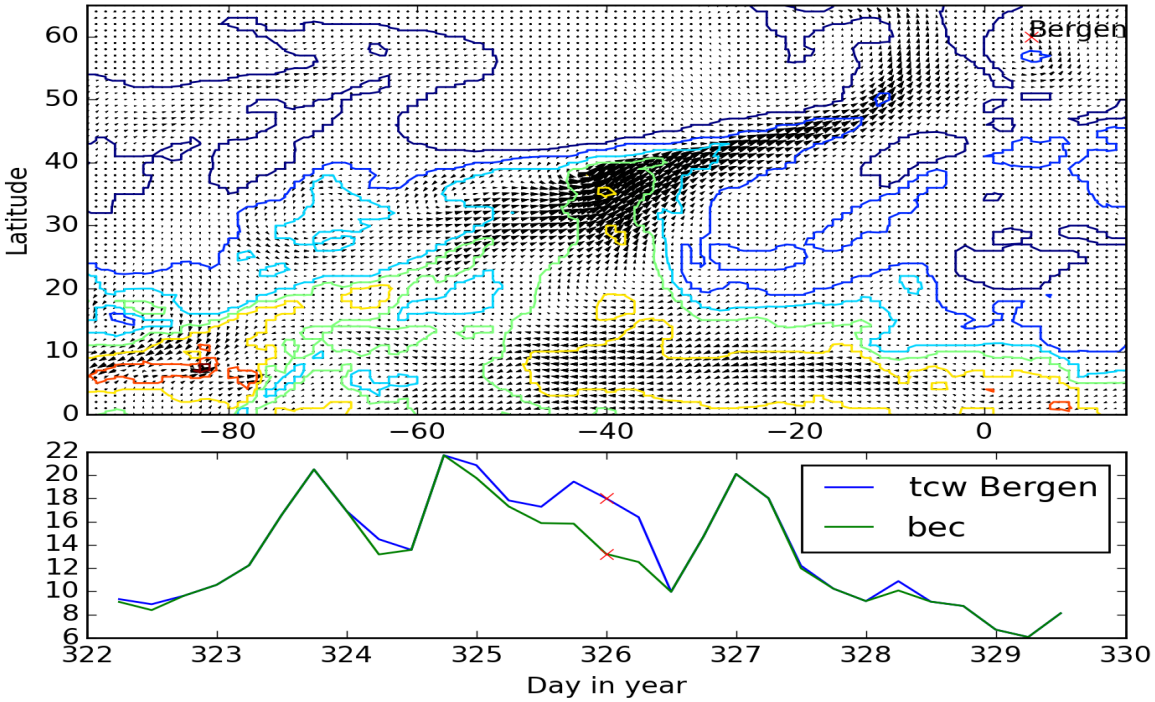
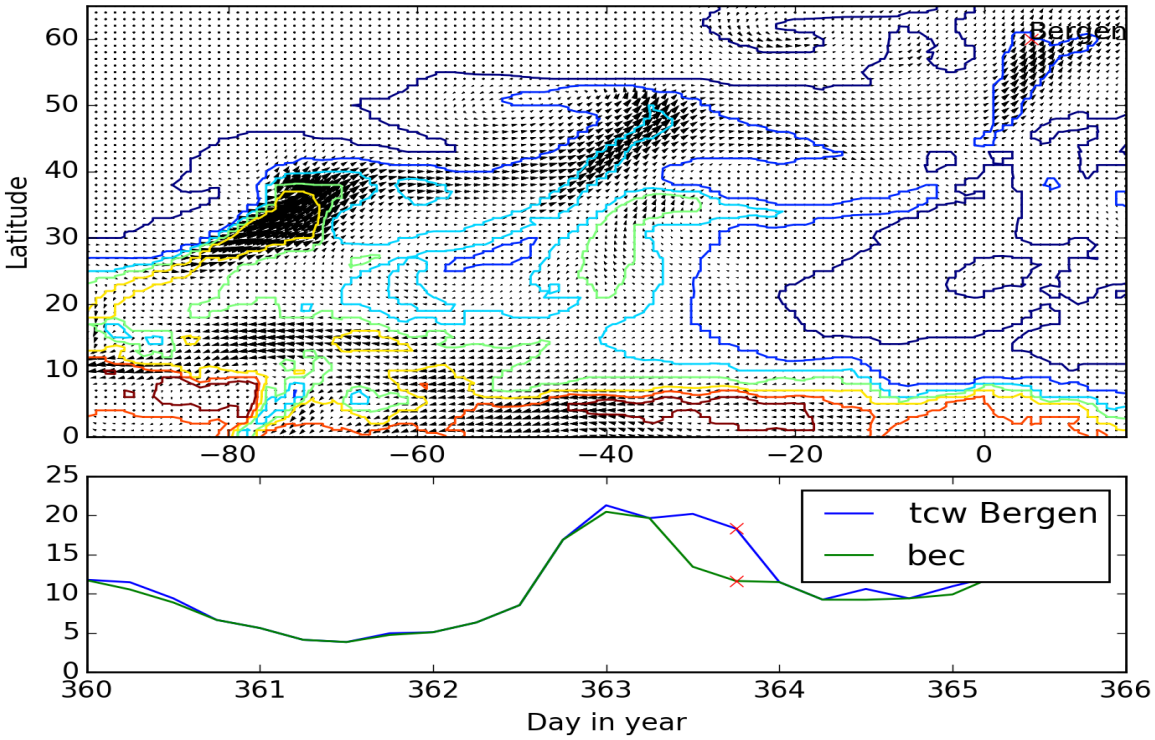


Figure 39: The IVT and contours of the tcw at day 363.75. The relative difference is high:



6.2.1 Conclusion

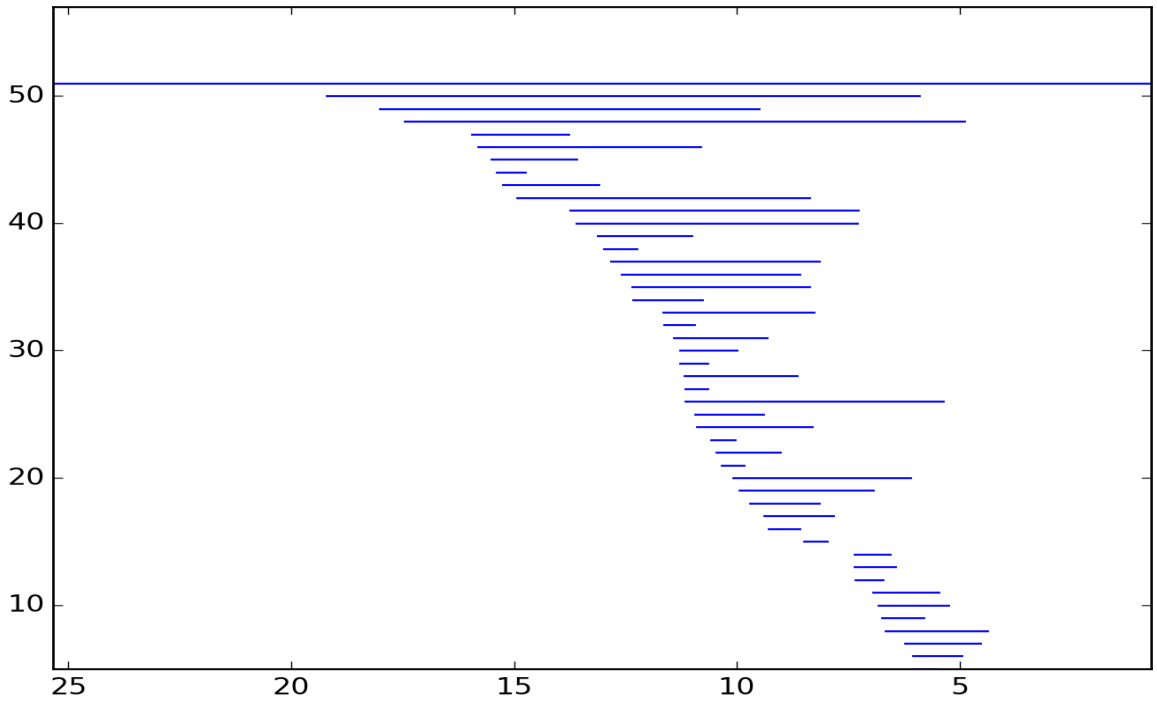
I feel that these figures confirm that the *bec* is better at detecting atmospheric rivers reaching Bergen than just the *tcw* value in Bergen. Also they justify laying the eastern border of our region so close to Bergen, as in all the cases where it looks like moving the border would increase the *bec* notably, there does not seem to be any atmospheric rivers reaching Bergen.

6.3 The tbec diagrams for February-December 2012.

In this section the bar codes of the *tbec* diagrams for the months in 2012 are shown. Only the most persistent third of the intervals are included, to filter out some noise. The interesting cycles here are the ones appearing at the highest filtration levels, as these correspond to paths through the highest humidity values. Note that the scale varies a bit from month to month, corresponding to the fact that the humidity varies with the seasons.

What I guess is that if there are cycles appearing much earlier than the cycles in general, this will correspond to a significant atmospheric river. On the other hand, if the births of the cycles are more evenly distributed, there are perhaps not so many rivers in the time period.

Figure 40: The *tbec* diagram of month 02 in 2012:



The most notable feature of the *tbec* diagram for February (Figure 40) is that there is one single cycle appearing much earlier than the rest. This coincides with the fact that one of the most noticable rivers detected by the *bec* appeared in February (see Figure 16 (day 58)). It seems that the most persistent intervals are concentrated among those that appear first in the filtration.

March (Figure 41) looks more top heavy at first glance, compared with February. This, however, is only because of the single cycle in February starting at a high filtration value stretching the scale. A more thorough look reveals that except that one cycle in February, they are quite similar. One difference is that in March there are two quite persistent intervals that appear quite late in the filtration.

The diagram for April (Figure 42) is quite similar to the one for March. Although the most persistent intervals seem to mostly be among those born early, there is one persistent one that is born quite late.

Figure 41: The *tbec* diagram of month 03 in 2012:

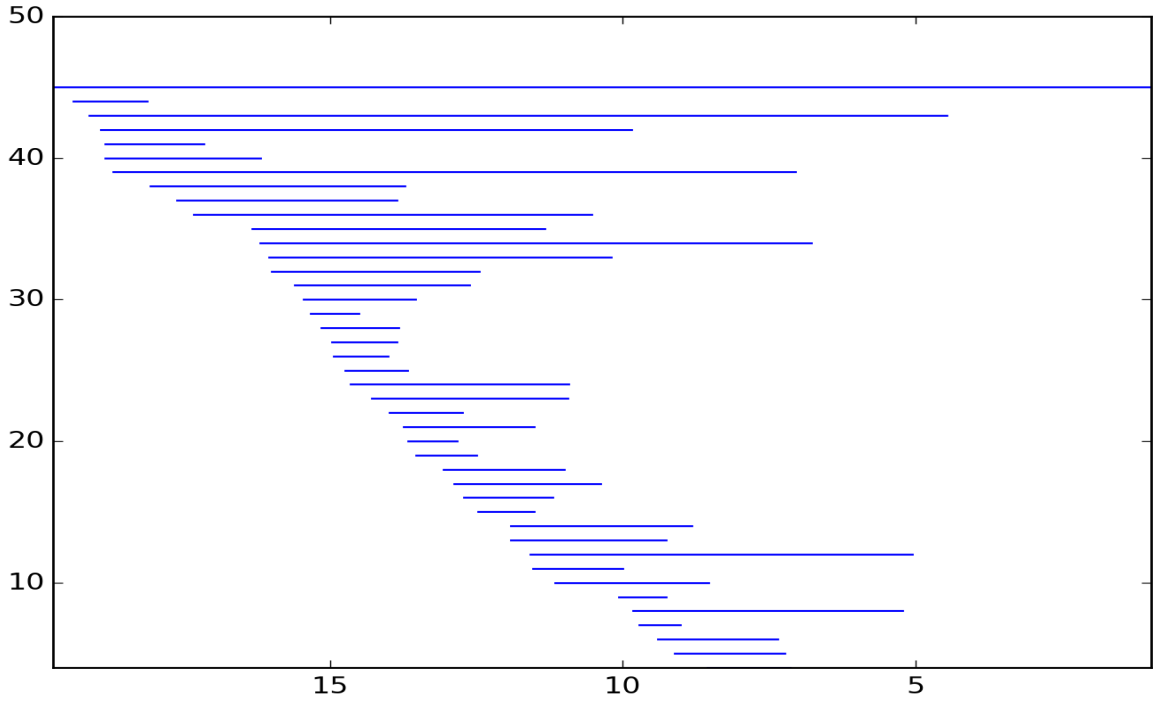


Figure 42: The *tbec* diagram of month 04 in 2012:

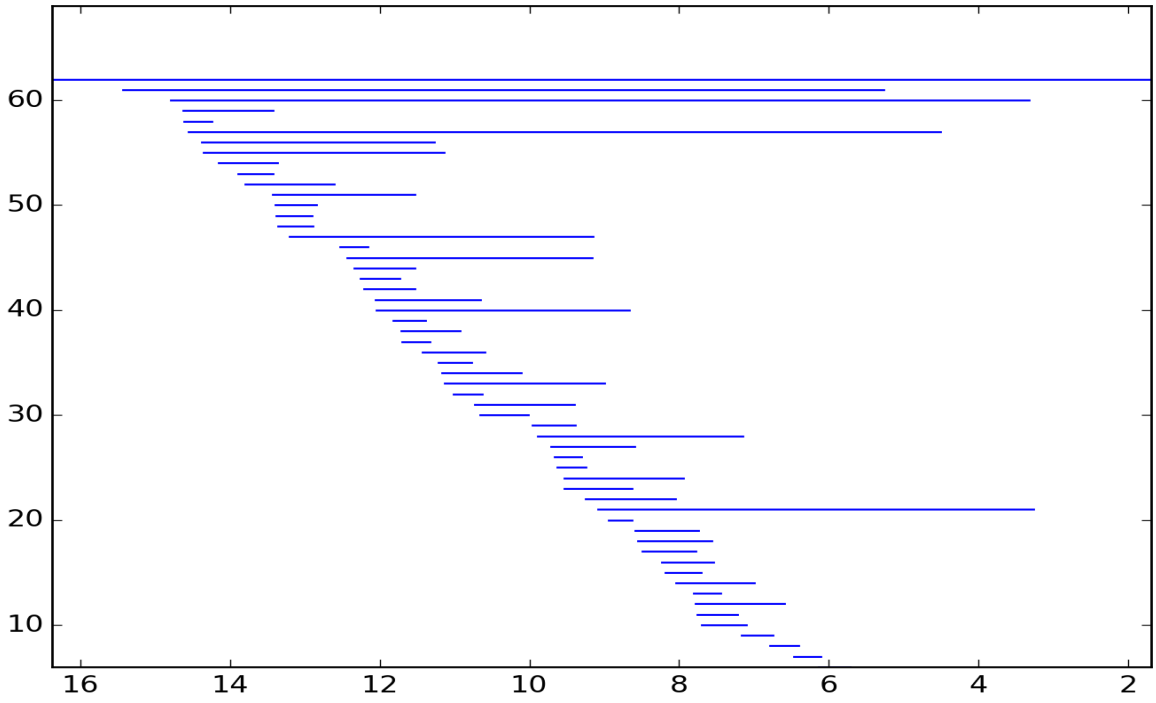
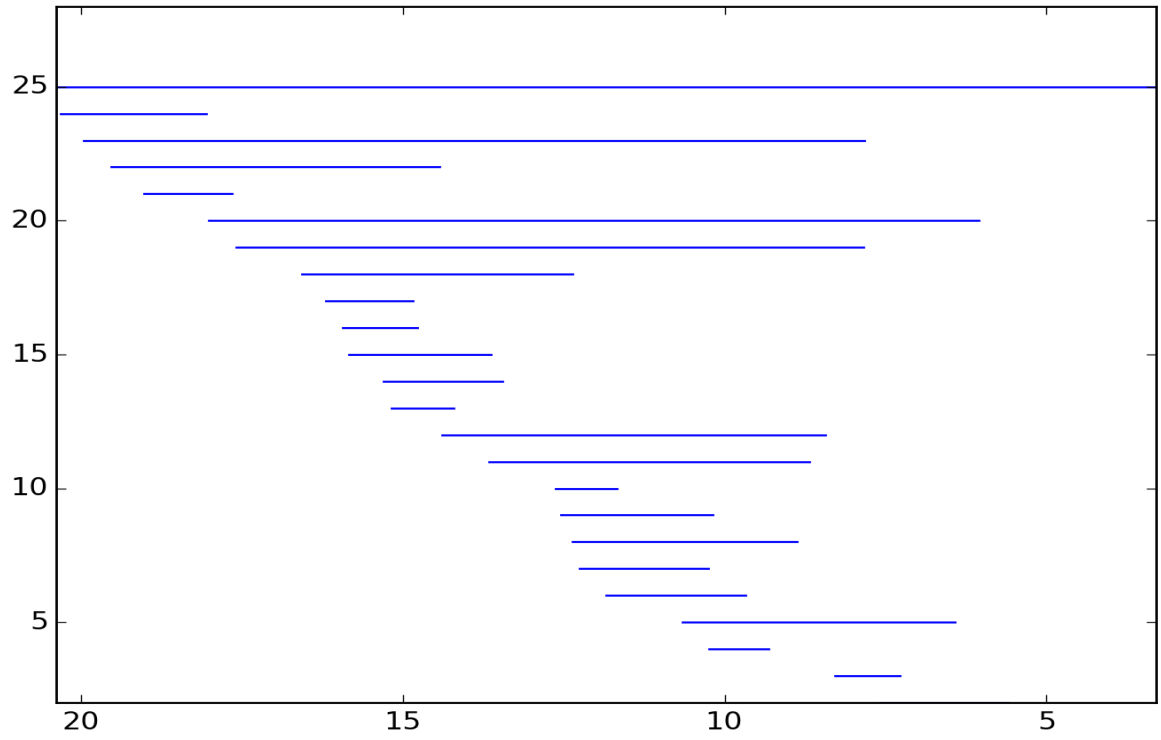


Figure 43: The *tbec* diagram of month 05 in 2012:



May (Figure 43) has very few persistence intervals compared to the previous ones. The births seem to be evenly distributed. The most persistent intervals seem to be born early.

June (Figure 44) has about the same amount of persistence intervals as February, March and April. Here the births don't seem as evenly distributed. Especially the three first seem to form a group of its own (Although one of these is not very persistent). The most persistent cycles are born quite early, although there are a few at the end that aren't negligible.

July (Figure 45) has extremely many persistence intervals compared to the other months. It has almost double the number of August, which comes second. It is difficult to say why. It could perhaps be because of more turbulence in the atmosphere, but this is pure speculation. There seems to be two cycles that are born considerably earlier than the others. This coincides with one of the most prominent atmospheric rivers detected by the *bec* (Figure 17), which is detected in July. The most persistent cycles seem to be more evenly distributed among the births than in the other months. This is a feature it shares with August 2012, and most of the Junes you can see in section 6.4, and so it looks like it is something typical of the summer.

Figure 44: The *tbec* diagram of month 06 in 2012:

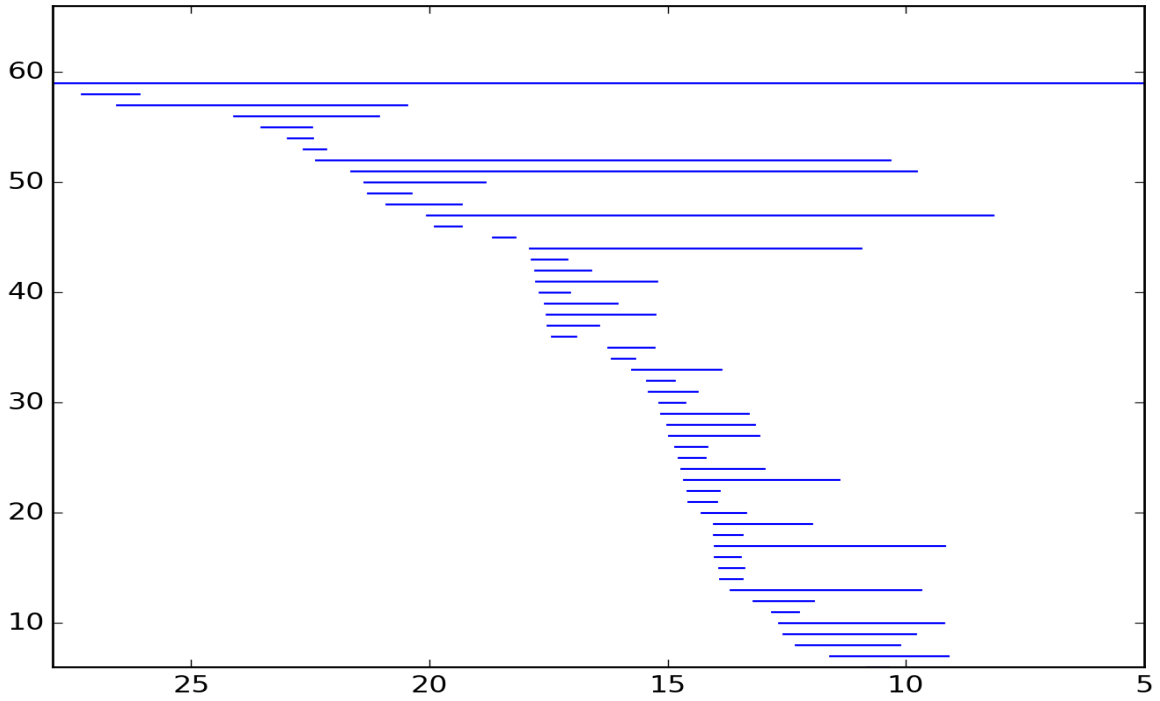


Figure 45: The *tbec* diagram of month 07 in 2012:

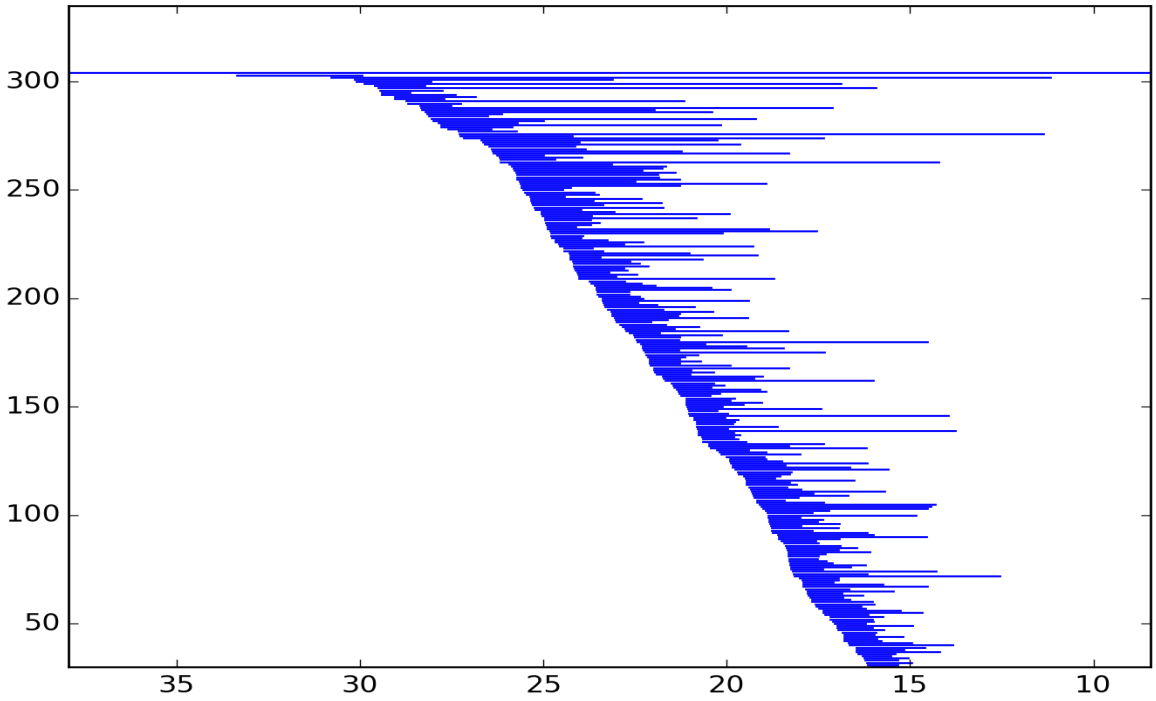
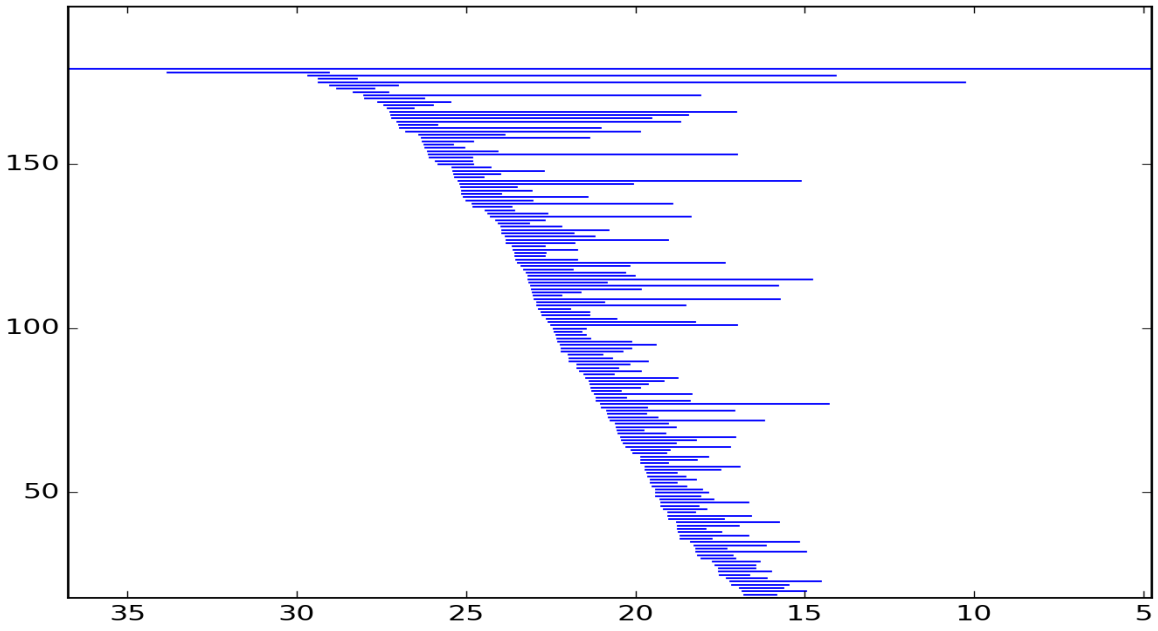
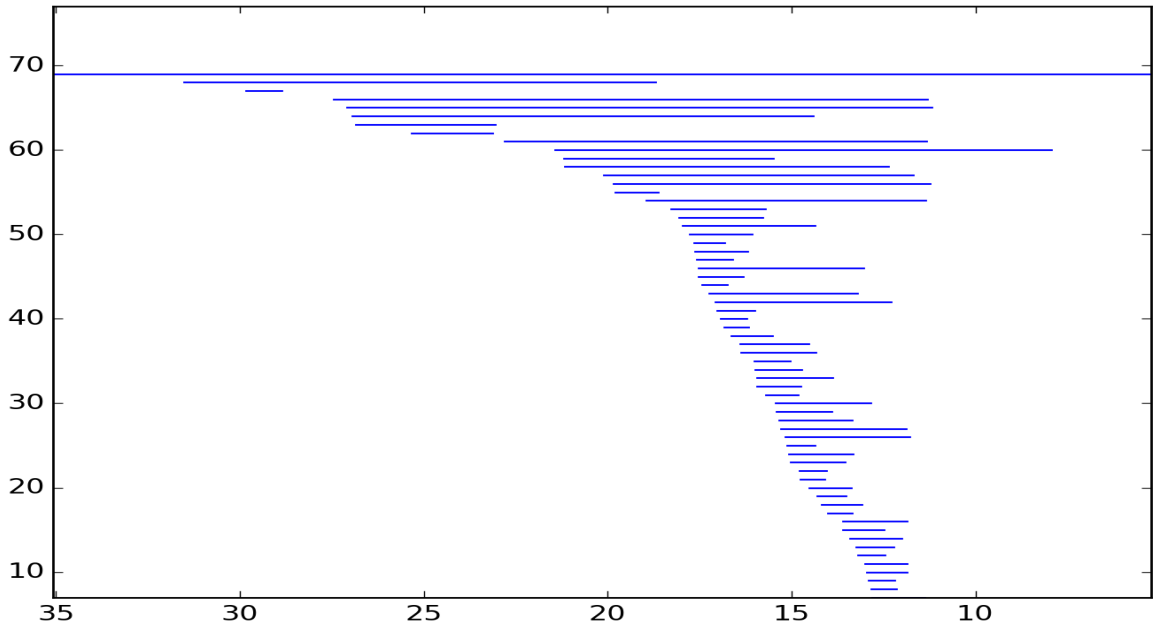


Figure 46: The *tbec* diagram of month 08 in 2012:



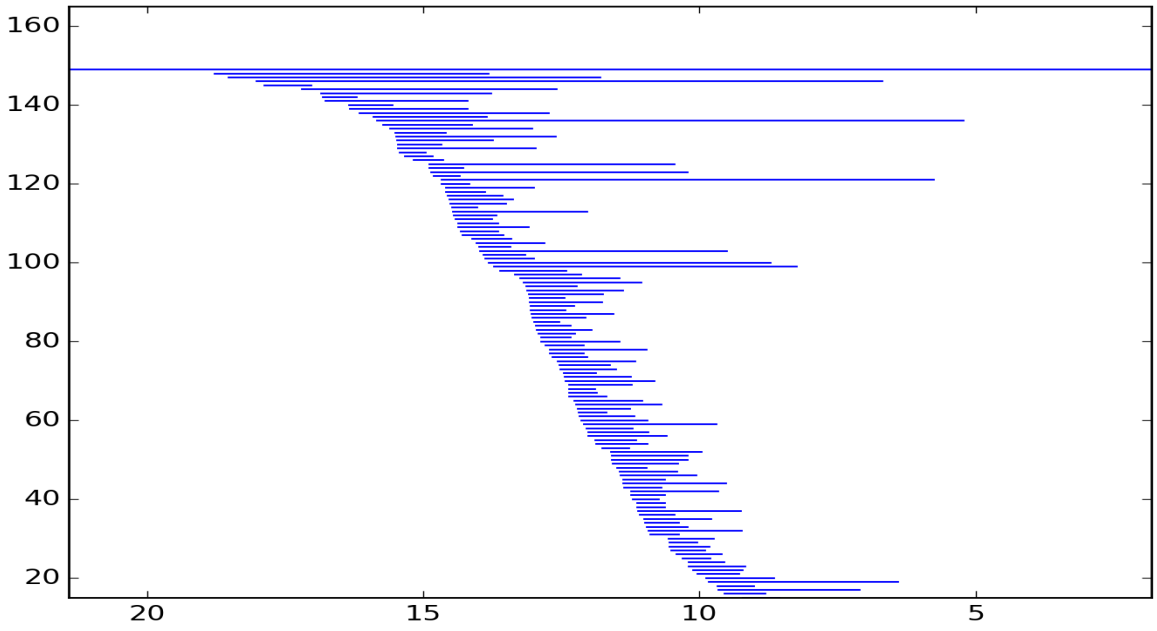
August (Figure 46) looks quite similar to July, but there are fewer persistence intervals. Also here there are two cycles born considerably earlier than the rest.

Figure 47: The *tbec* diagram of month 09 in 2012:



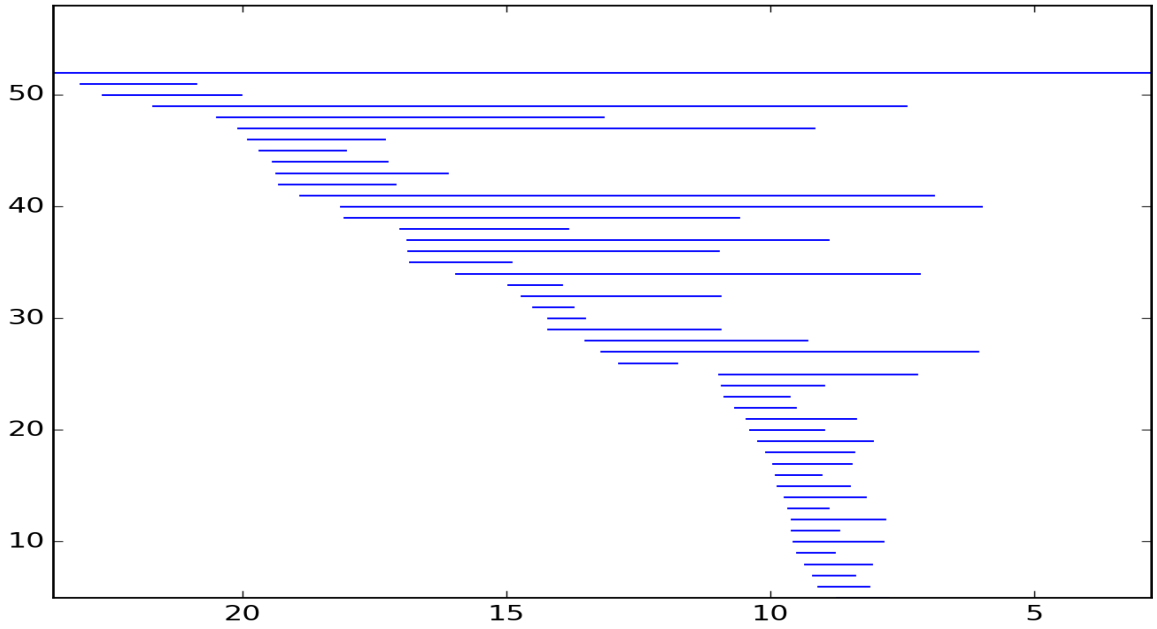
In September (Figure 47) the births seem to be split in three groups. The most persistent cycle appears early.

Figure 48: The *tbec* diagram of month 10 in 2012:



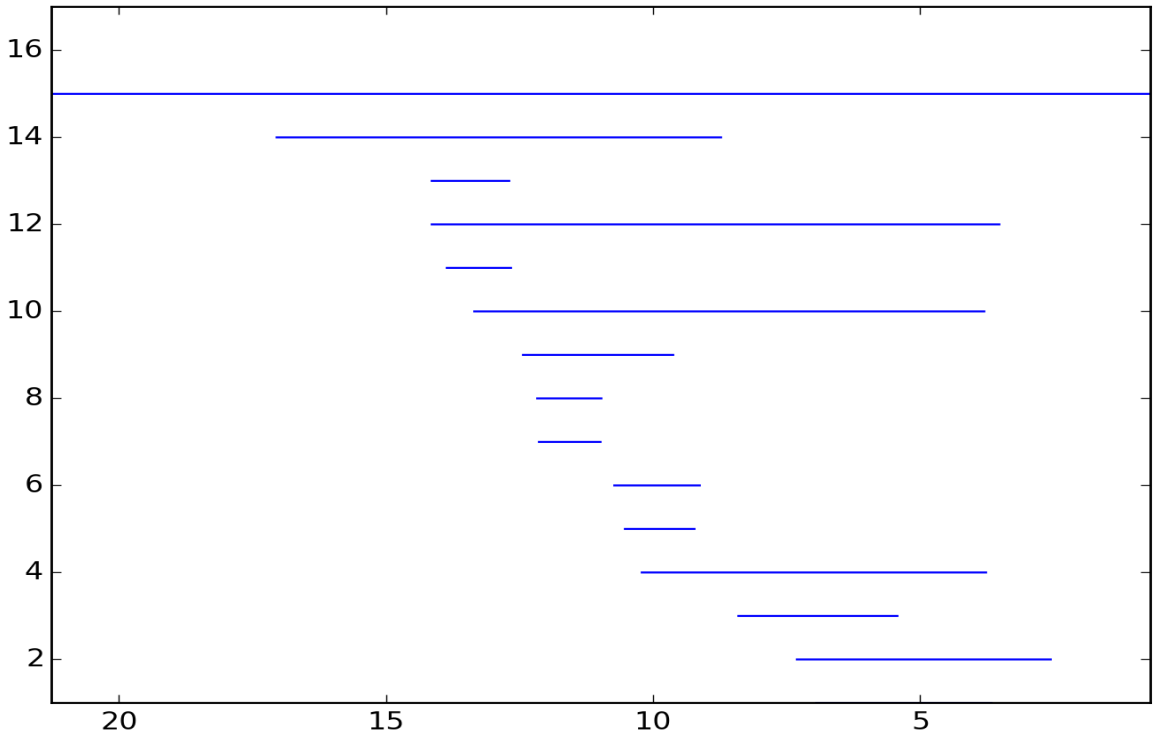
In October (Figure 48) there is one cycle appearing earlier than the others. The most persistent cycles appear early.

Figure 49: The *tbec* diagram of month 11 in 2012:



In November (Figure 49) the births are quite evenly distributed.

Figure 50: The *tbec* diagram of month 12 in 2012:

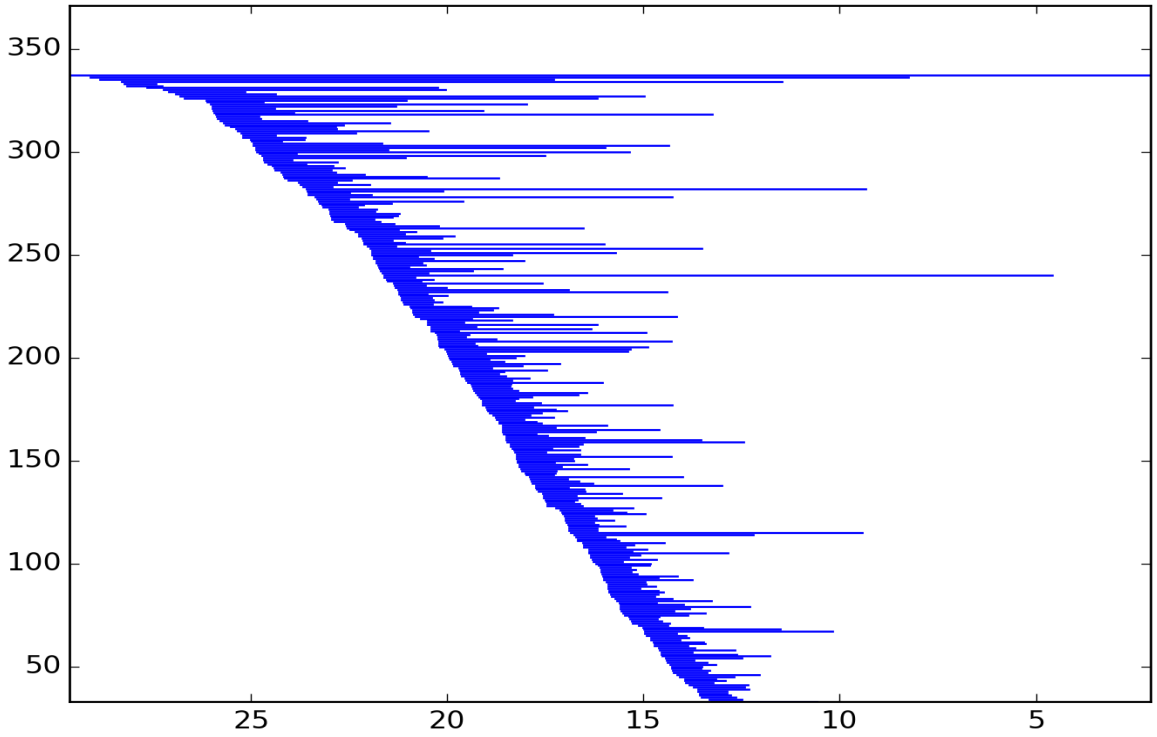


December (Figure 50) has very few persistence intervals. It is difficult to see any tendency when they are so few, but it looks like the first birth is notably earlier than the others.

6.4 The tbec diagrams for June 2000-2011

This section contains plots of the persistence diagrams for the month June in the years 2000 to 2011. Also here only the most persistent third is shown. The reason for including these figures is that it is interesting to compare the same period of time in different years, as here there is no effect from the seasonal variance[3]. You can notice this by the fact that the scale and number of cycles do not vary as much as in section 6.3.

Figure 51: The *tbec* diagram of month 06 in 2000:



June 2000 (Fig:51) has many intervals. The births look evenly distributed, and there are many intervals with high persistence that are born late in the filtration.

June 2001 (Fig:52) have a normal amount of intervals. There is one cycle born quite a bit earlier than the rest, and there are many intervals with high persistence that are born late in the filtration.

The diagram for June 2002 (Fig:53) look almost identical to the diagram for June 2001. The persistence intervals look like they are slightly shorter, though.

[3]Except that this variance may vary from year to year, of course.

Figure 52: The *tbec* diagram of month 06 in 2001:

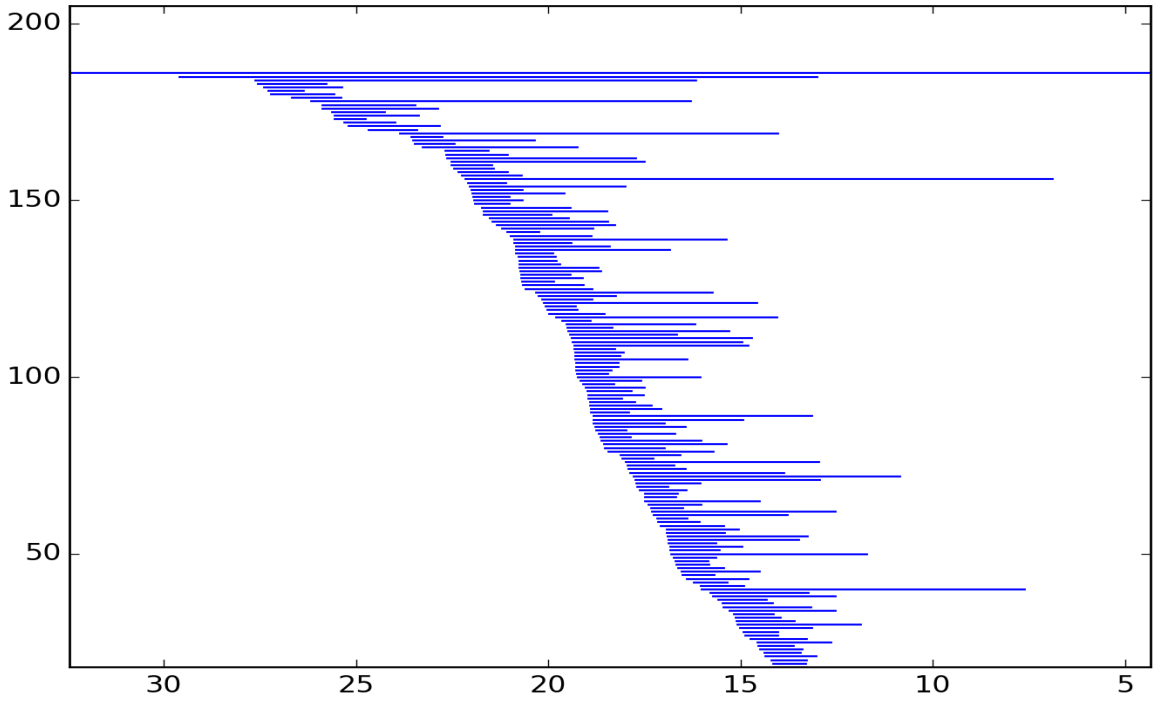


Figure 53: The *tbec* diagram of month 06 in 2002:

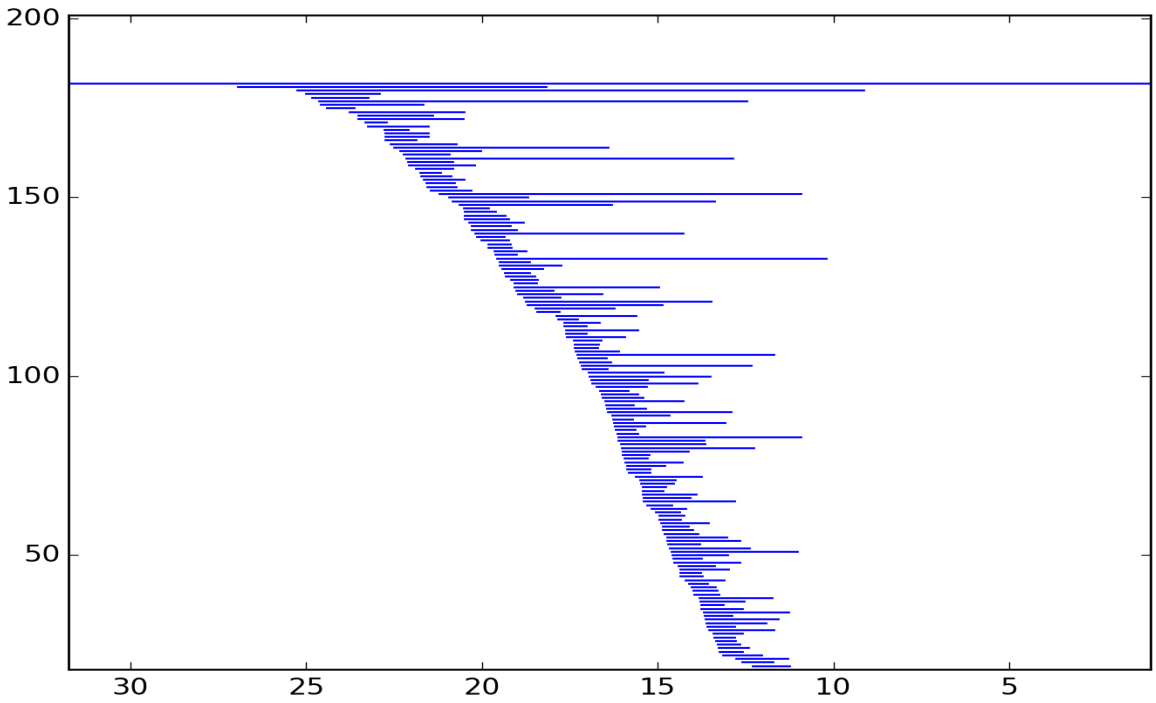
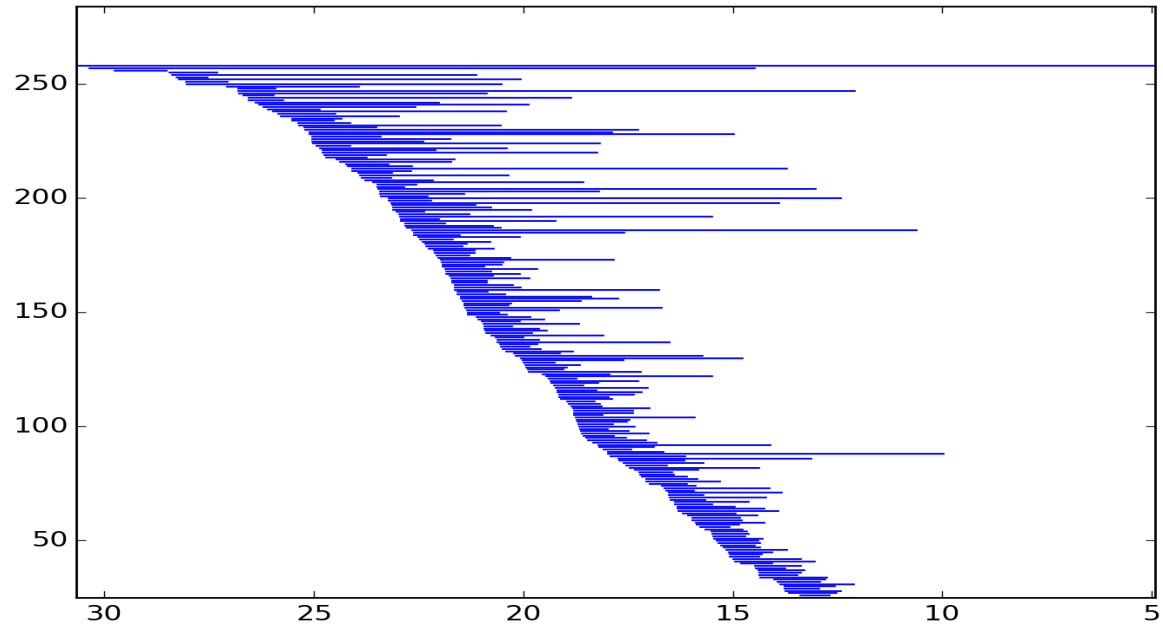
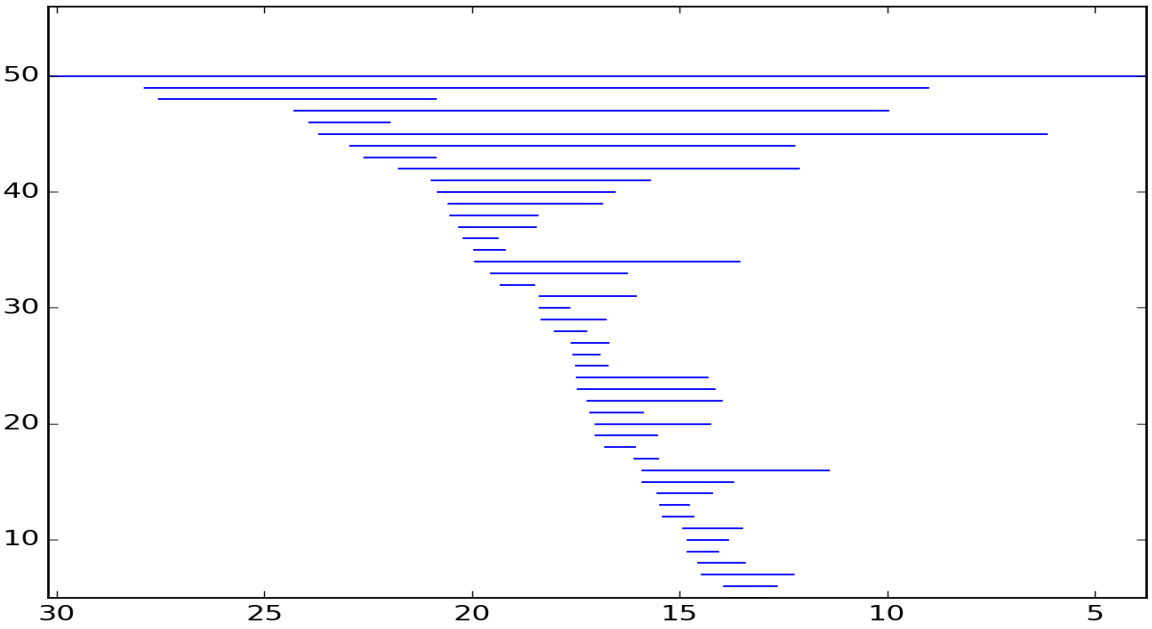


Figure 54: The *tbec* diagram of month 06 in 2003:



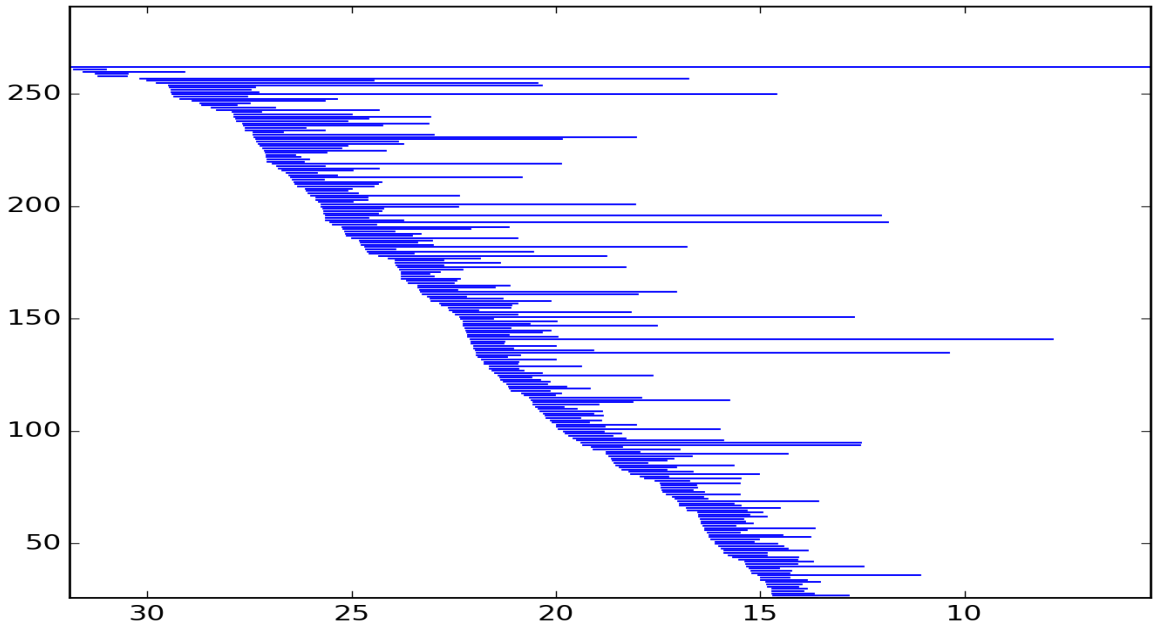
June 2003 (Fig:54) has a bit more intervals than normal. The births look evenly distributed, and the most persistent intervals seem to have mostly early birth values.

Figure 55: The *tbec* diagram of month 06 in 2004:



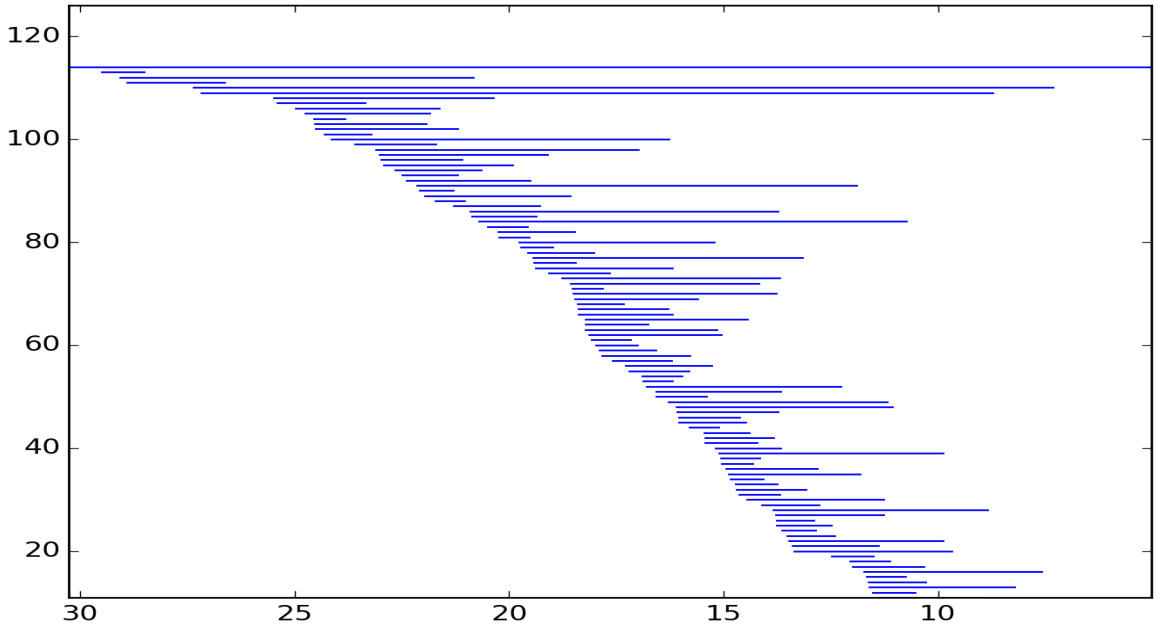
June 2004 (Fig:55) has very few intervals. The three first cycles are born a bit earlier than the rest, and the most persistent intervals seem to have mostly early birth values.

Figure 56: The *tbec* diagram of month 06 in 2005:



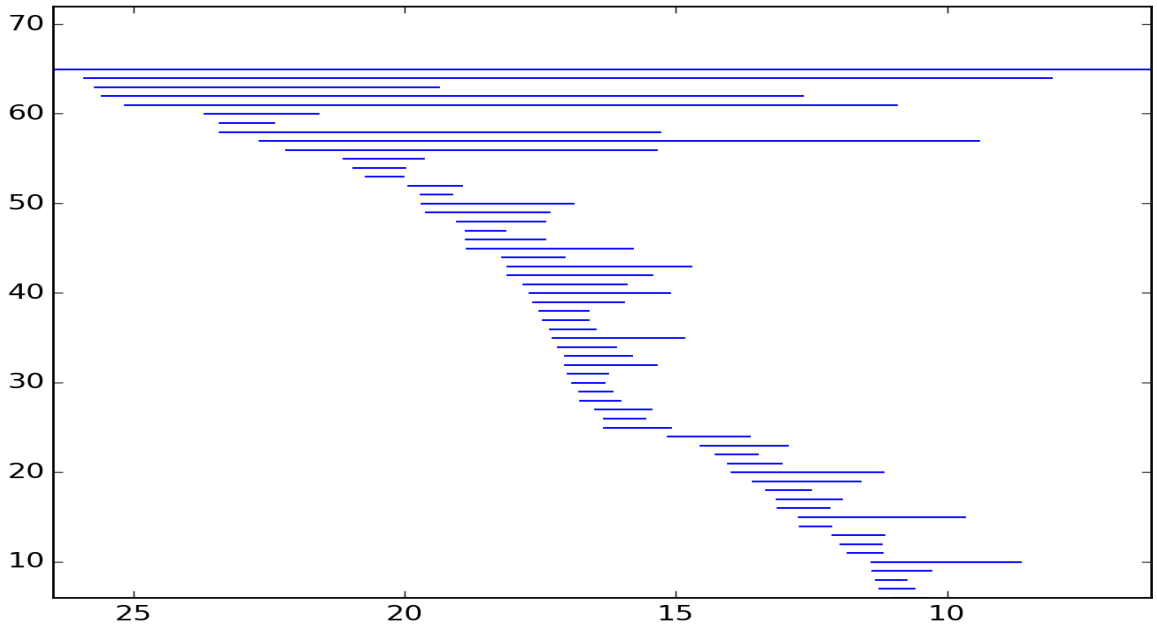
June 2005 (Fig:56) has a bit more intervals than normal. The births look evenly distributed, and there are quite a few persistent cycles that are born late.

Figure 57: The *tbec* diagram of month 06 in 2006:



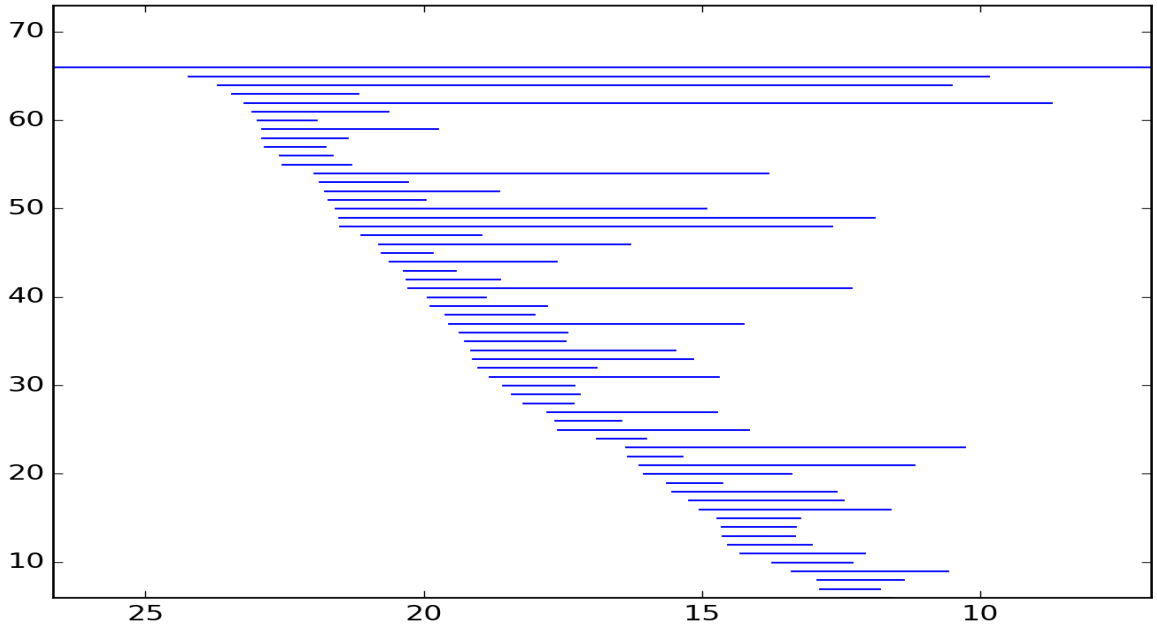
June 2006 (Fig:57) has a bit less intervals than normal. The births look evenly distributed, and the most persistent intervals seem to have mostly early birth values.

Figure 58: The *tbec* diagram of month 06 in 2007:



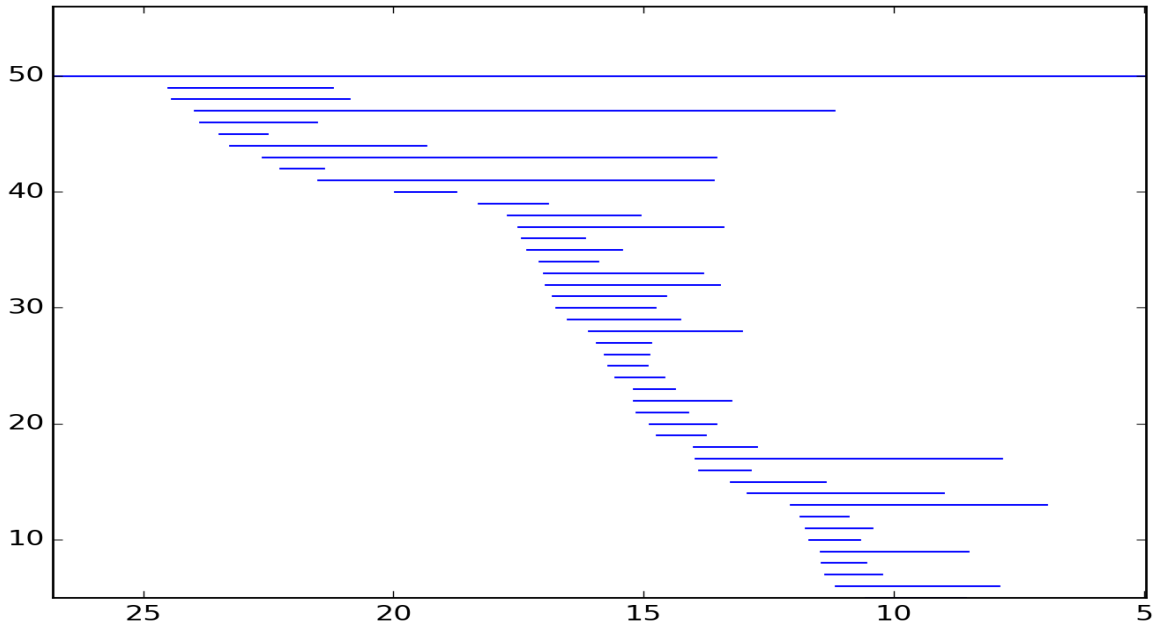
June 2007 (Fig:58) has few intervals. The births look evenly distributed, and the most persistent intervals seem to have mostly early birth values.

Figure 59: The *tbec* diagram of month 06 in 2008:



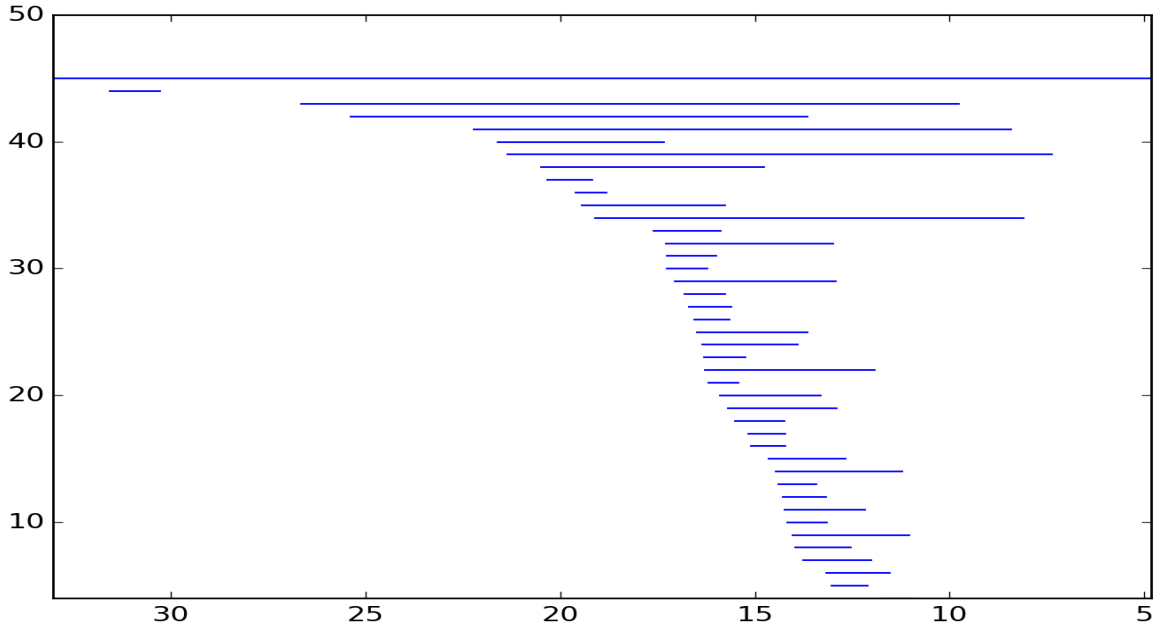
June 2008 (Fig:59) has few intervals. There is one cycle born earlier than the rest, and there are quite a few persistent intervals born late in the filtration.

Figure 60: The *tbec* diagram of month 06 in 2009:



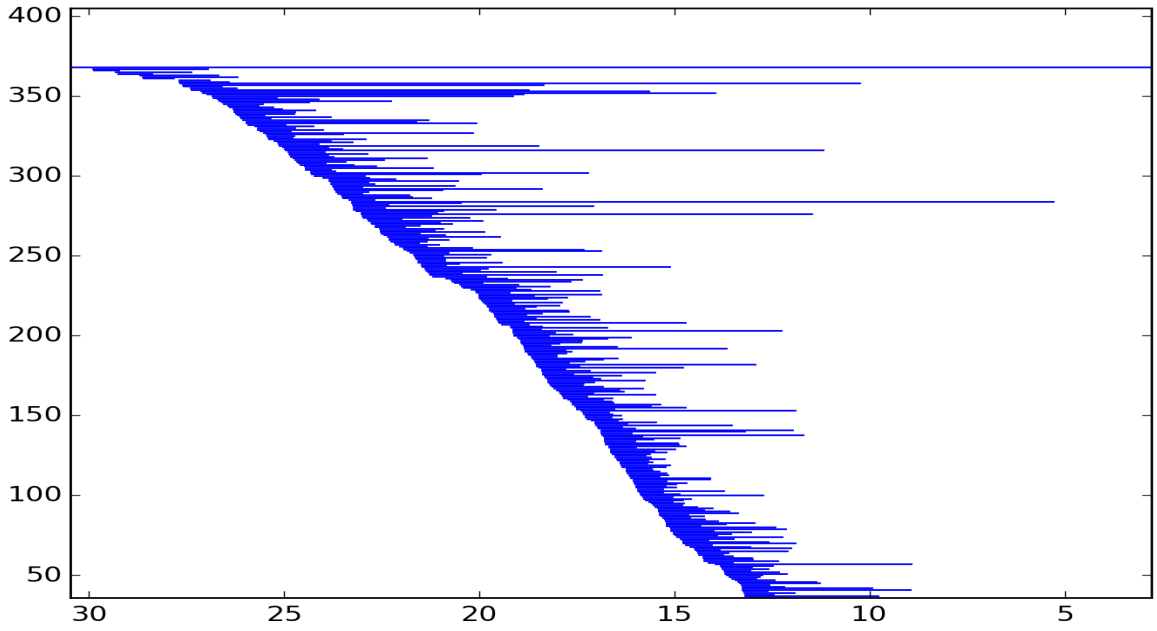
June 2009 (Fig:60) has very few intervals. There is one cycle born a little earlier than the rest, and there are a few persistent intervals born late in the filtration.

Figure 61: The *tbec* diagram of month 06 in 2010:



June 2010 (Fig:61) has very few intervals. There is one cycle born earlier than the rest (really two, but the second is not very persistent), and the most persistent intervals seem to be born early in the filtration.

Figure 62: The *tbec* diagram of month 06 in 2011:



June 2011 (Fig:62) has many intervals. The births seem evenly distributed. The most persistent intervals seem to be born early, but some persistent ones are born late in the filtration.

6.5 Diagrams comparing the tbec and the bec for February-December 2012.

In this section we compare the *tbec* diagrams to the *bec* as described in section 6.5. We make a persistence diagram induced by the *bec* by computing the persistent 0-homology of $T(bec)$. Looking at the pictures in appendix B, this amounts to letting a vertical line descend and at each height count how many different spikes the line goes through (Figure 15 in section 5.7 should provide a nice illustration for this.). Very persistent intervals corresponds to high spikes (which again hopefully correspond to an atmospheric river).

Ideally, the diagram induced by the *bec* should "almost" be a subset of the *tbec*, as the former detects paths only at single time steps and the latter should detect these and also paths going through time. So in the diagrams I try to find pairings between all the intervals from the *bec* and some of the intervals in the *tbec*. In the pairings, the one interval from the *bec* should be slightly smaller than the one in the *tbec* (again, because a path detected by the *bec* also should be detected by the *tbec*.

To make things slightly easier to see, the least persistent intervals are not included.

Unfortunately, the comparison here is not so very precise, as I have done no more than look at them and see if they look alike. A more rigid comparison could have been achieved using the bottleneck distance (which basically finds pairings of intervals from the two diagrams), but there was no time to implement this. I have tried to find pairings myself, by only looking at the pictures. When the pairings I guess are not immediate, they are indicated on the figure by a black line.

There are a few cases where there are intervals from the *bec* that I can't find a close pair for in the *tbec*. One possible explanation for this is one we do not like: It may be that the *tbec* fails to detect paths we want it to detect. Another (and perhaps more reassuring) explanation is that the path the *bec* detects could be detected by the *tbec* so much earlier that one cannot see on the diagrams that belong to each other.

Figure 63 compares the *bec* and *tbec* in February 2012. This is a very nice picture as all the red intervals (the ones from the *bec*) seem to have a corresponding blue interval more or less right next to it. The two cycles at the top probably both correspond to the atmospheric river in Figure 16 in section 6.1.

In Figure 64 (March) the picture is not as clear. There are more of my guesses of pairs that aren't immediate. Notably, there are a couple of red intervals that don't seem to have a pair (around cycle number 20 from the bottom). Also it seems to be quite a few blues that don't have pairings.

Figure 63: The *tbec* versus the persistence diagrams induced by the *bec* in month 02 in 2012:

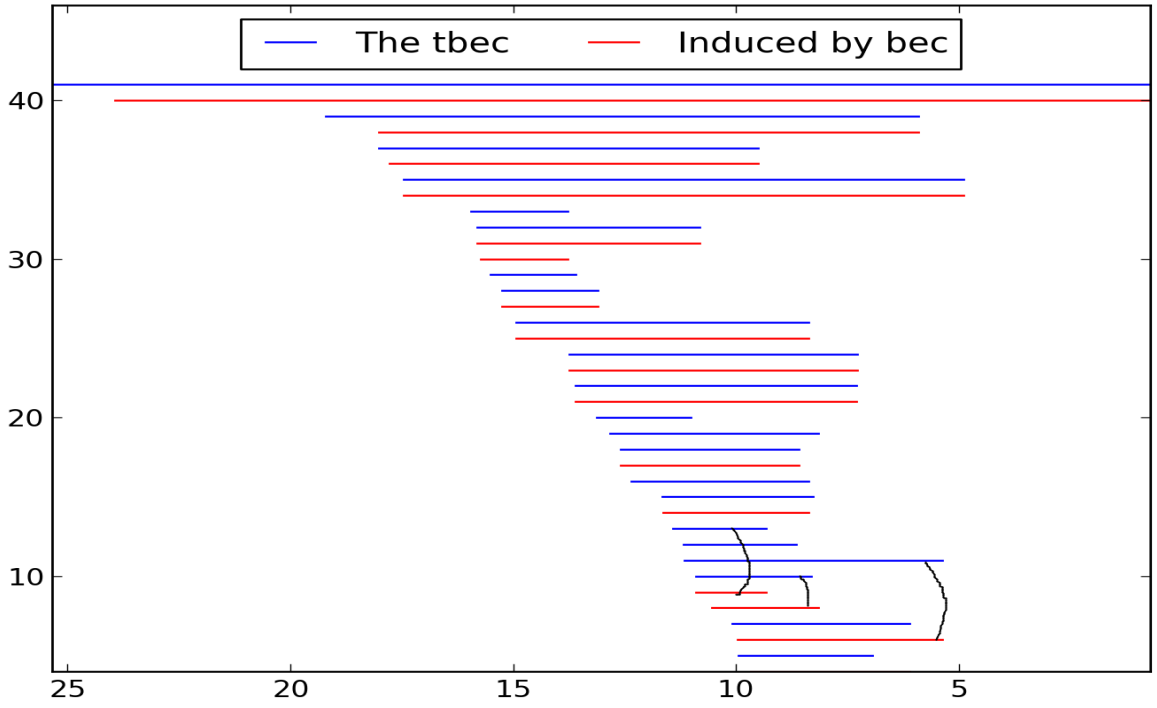


Figure 64: The *tbec* versus the persistence diagrams induced by the *bec* in month 03 in 2012:

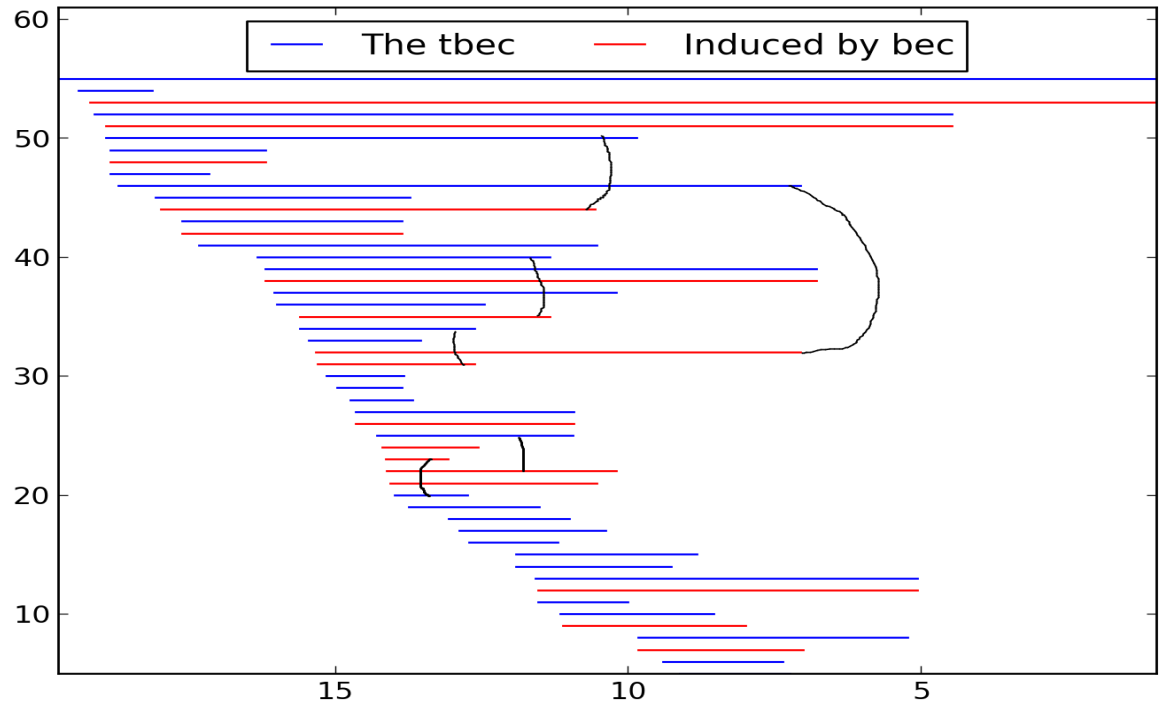
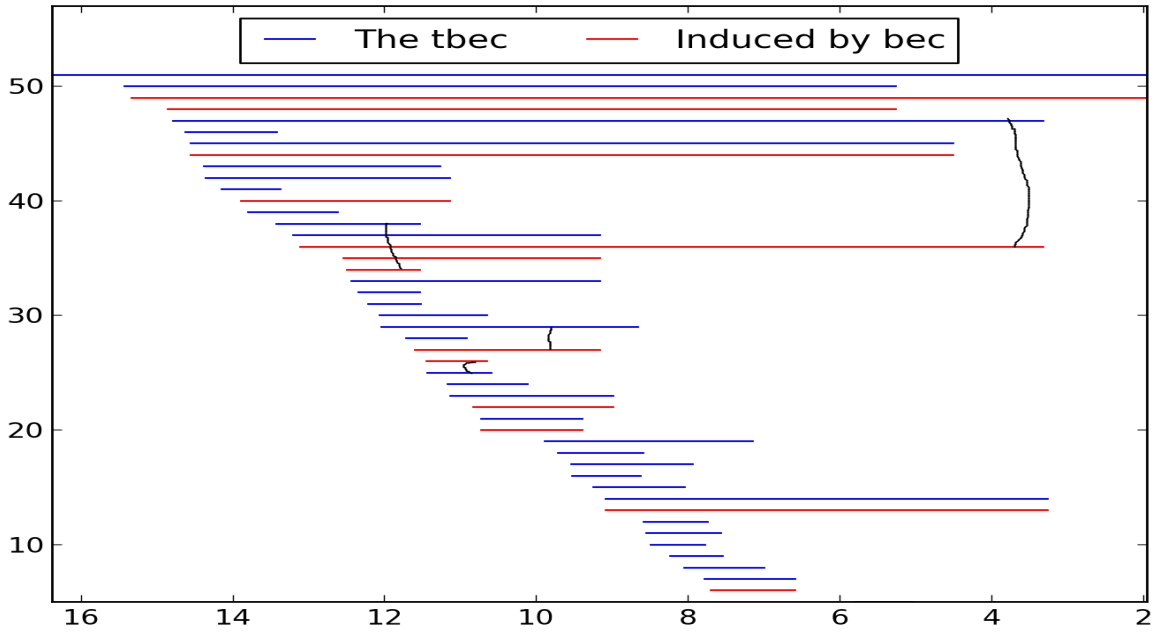
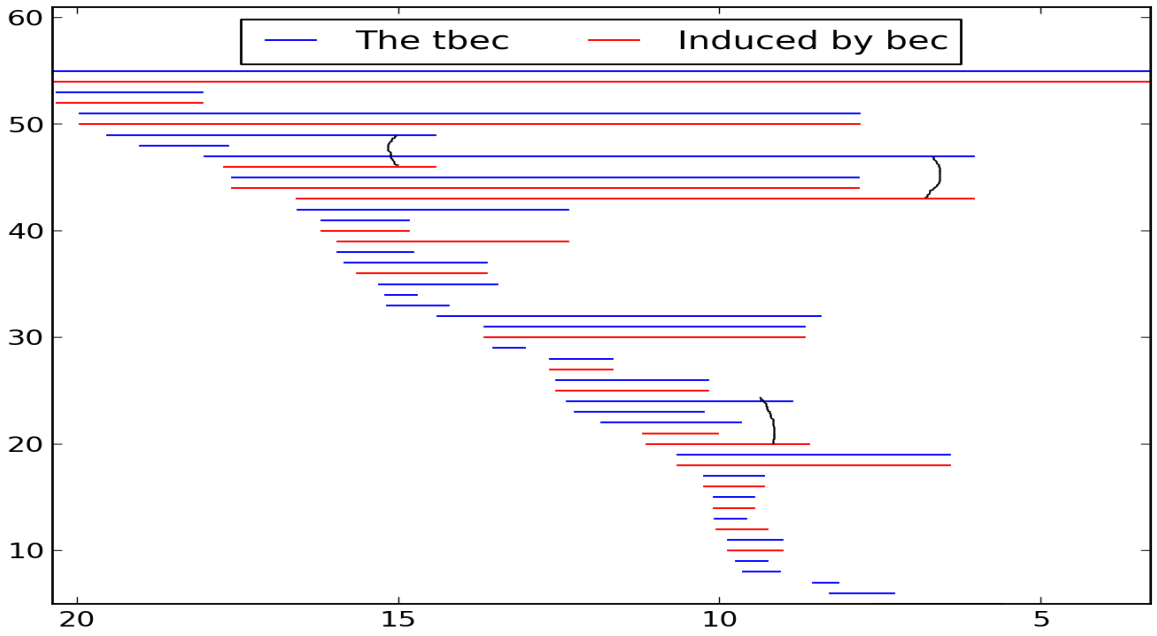


Figure 65: The *tbec* versus the persistence diagrams induced by the *bec* in month 04 in 2012:



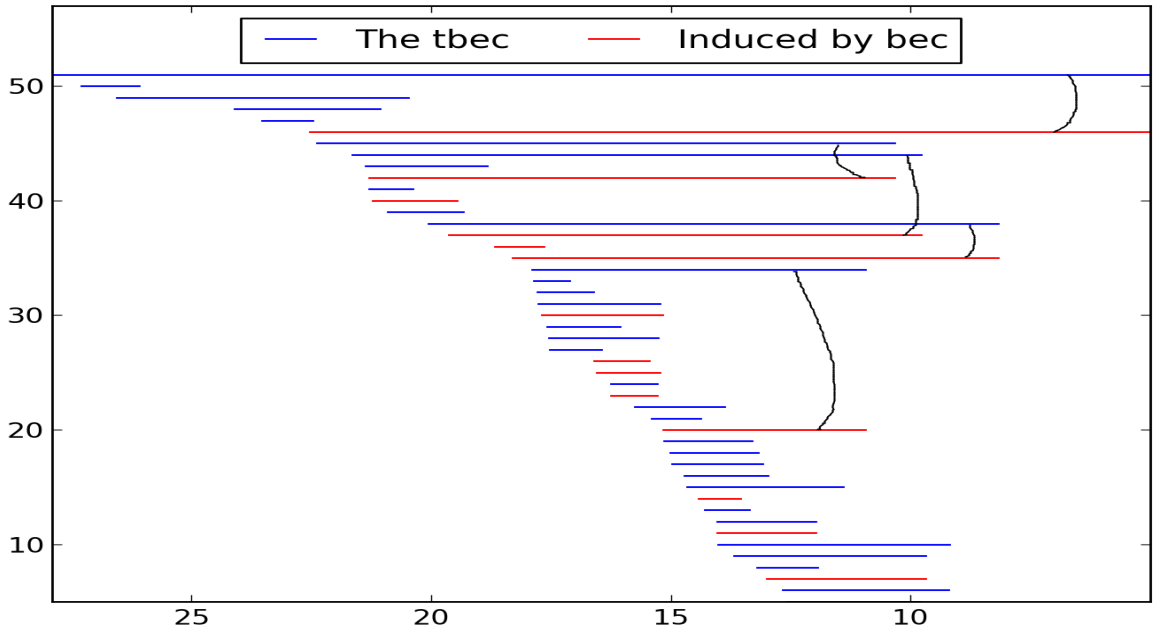
In Figure 65 (April) the reds mostly seems to have pairings with only slightly larger blues.

Figure 66: The *tbec* versus the persistence diagrams induced by the *bec* in month 05 in 2012:



Also in Figure 66 (May) the reds mostly seems to have pairings with only slightly larger blues, although there is a rather short one around the 20 mark doesn't seem to have a pair.

Figure 67: The *tbec* versus the persistence diagrams induced by the *bec* in month 06 in 2012:



One interesting thing in Figure 67 (June) is that the first intervals in the *tbec* start so much earlier than the ones coming from the *bec*. This probably means that there are some notable plumes (small, moving regions of high humidity). In the pairings I can find, the blues are often quite a bit larger than the reds (for example the pairing of the two that die at infinity). There is a short one (around number 25) that doesn't seem to have a pair.

Figure 68: The *tbec* versus the persistence diagrams induced by the *bec* in month 07 in 2012:

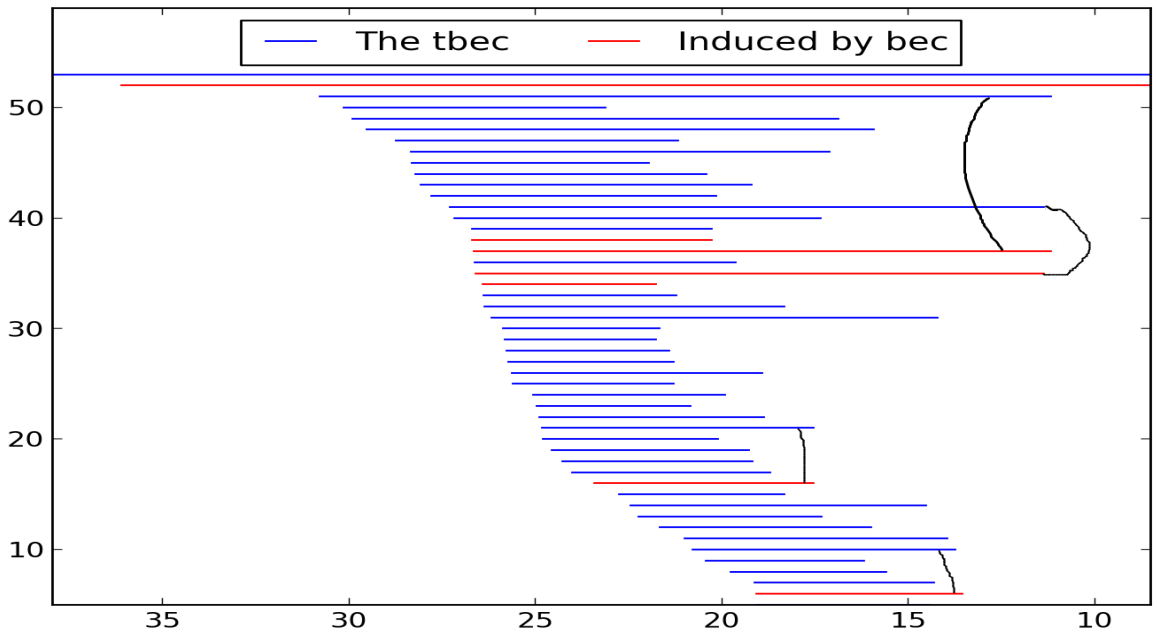
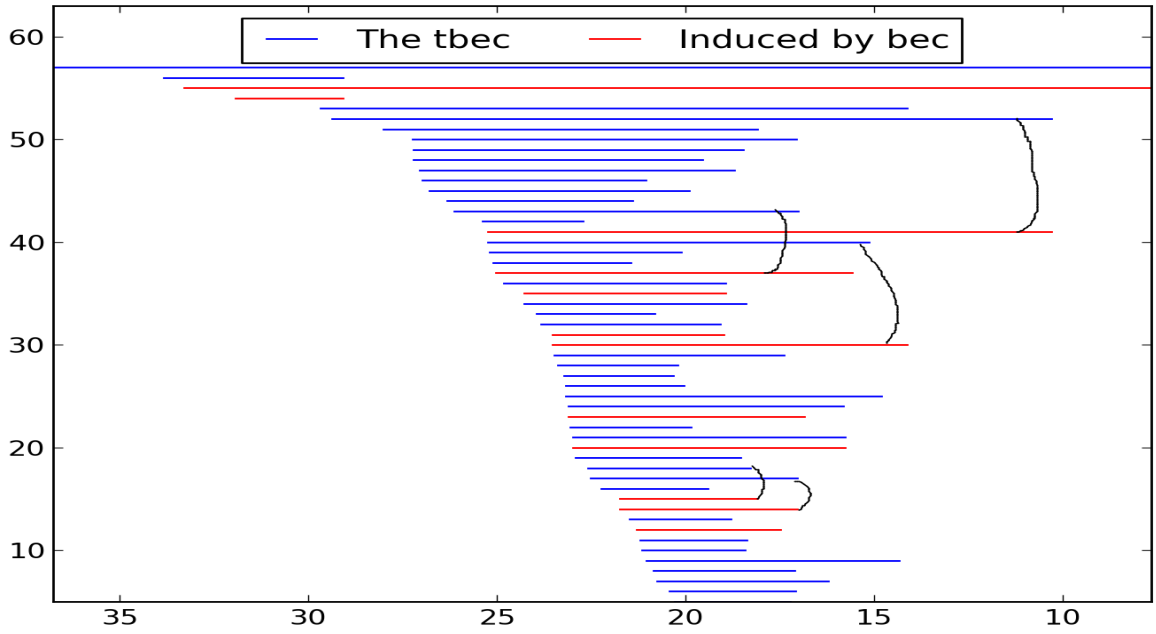


Figure 69: The *tbec* versus the persistence diagrams induced by the *bec* in month 08 in 2012:



In Figure 68 (July) most of the reds seem to have rather immediate pairs, although one (around the 35 mark) is paired with one that is quite a bit older. In Figure 69 (August) most of the reds seem to have pairings, although some are quite far away.

In Figure 70 (September) the two cycles dying at infinity seem to appear at the same filtration step. There is a rather short red cycle in the middle (around number 35) that doesn't seem to have an immediate pair, or maybe it should be paired with the much shorter blue right above it. There is perhaps a path here that the *tbec* should have detected, but didn't. If it is, it should not be a very big problem, as it is not at a very high filtration value, and it is not very persistent. The second red from the top looks like it should be paired with the blue right above, as they die at the same filtration step. This blue is born quite a bit earlier than the red. Looking at Figure 82 in appendix B it is possible to see that this red interval corresponds to the *bec* value at day 247 (the second highest spike that month). Figure 19 in section 6.1 explains (possibly) why the difference in birth values are so high: It shows that there is a disconnected region of high humidity over Bergen at this time step, and it is not difficult to imagine that this might have travelled all the way from the tropics, creating a path through considerably higher *tcw* values than the path restricted to the single time step. My guess is that something similar happens every time a red interval is paired with an older blue one.

Figure 70: The *tbec* versus the persistence diagrams induced by the *bec* in month 09 in 2012:

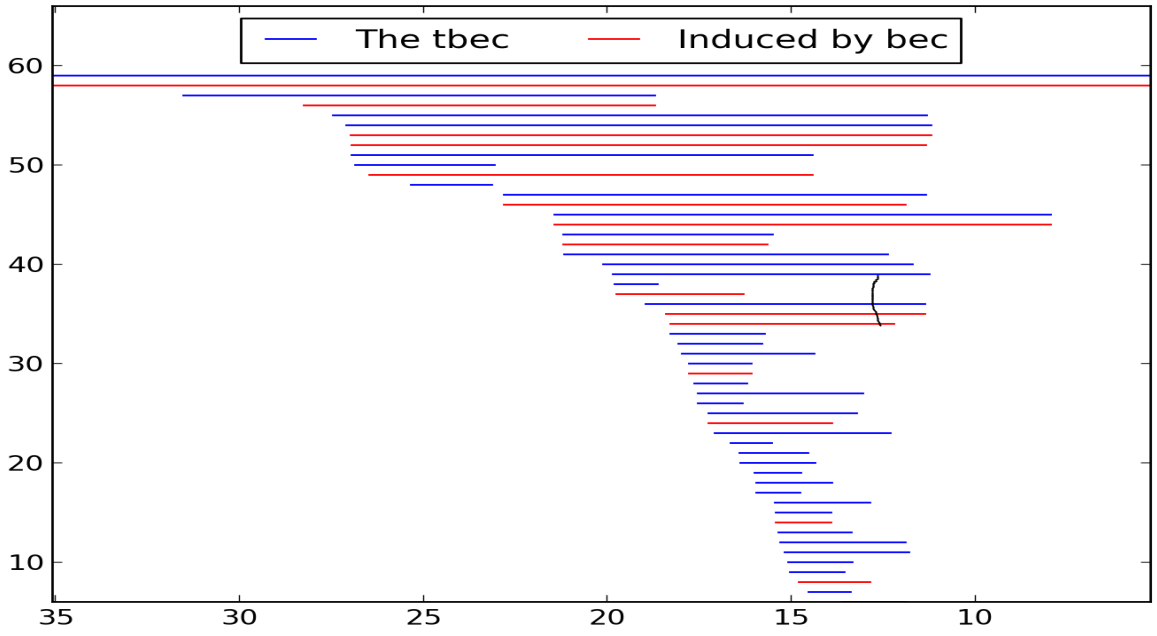
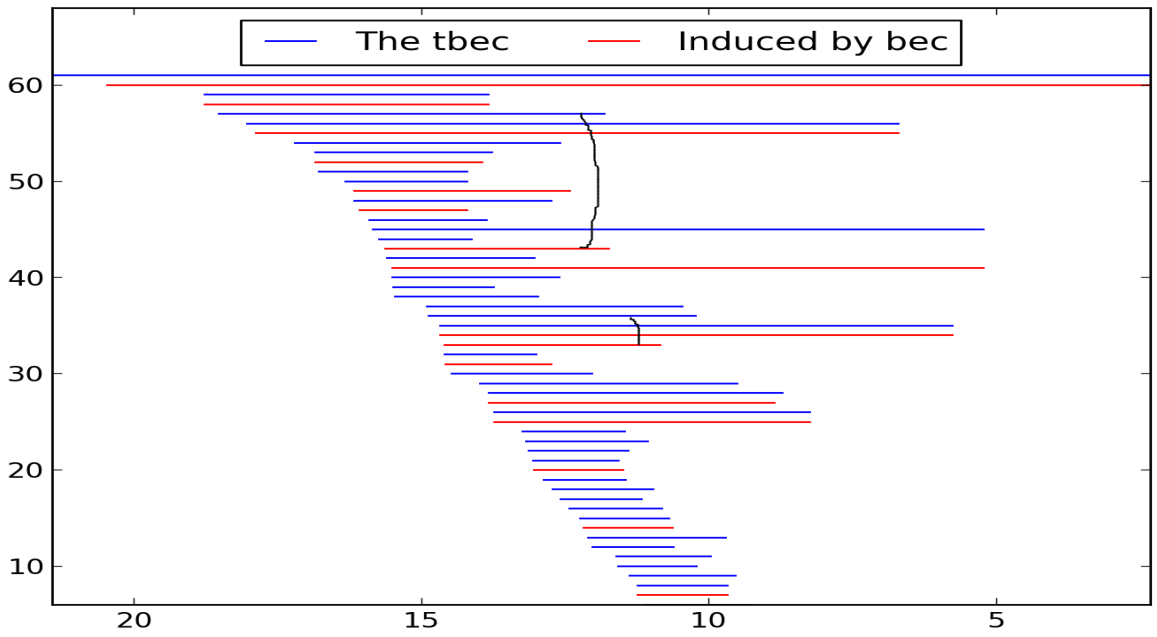
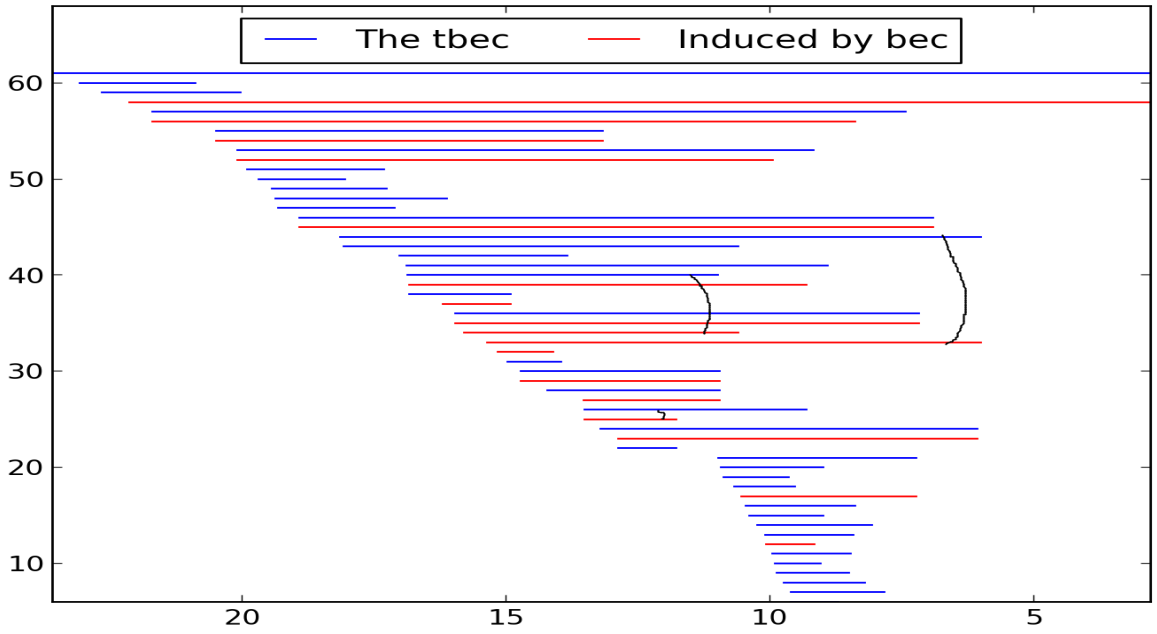


Figure 71: The *tbec* versus the persistence diagrams induced by the *bec* in month 10 in 2012:



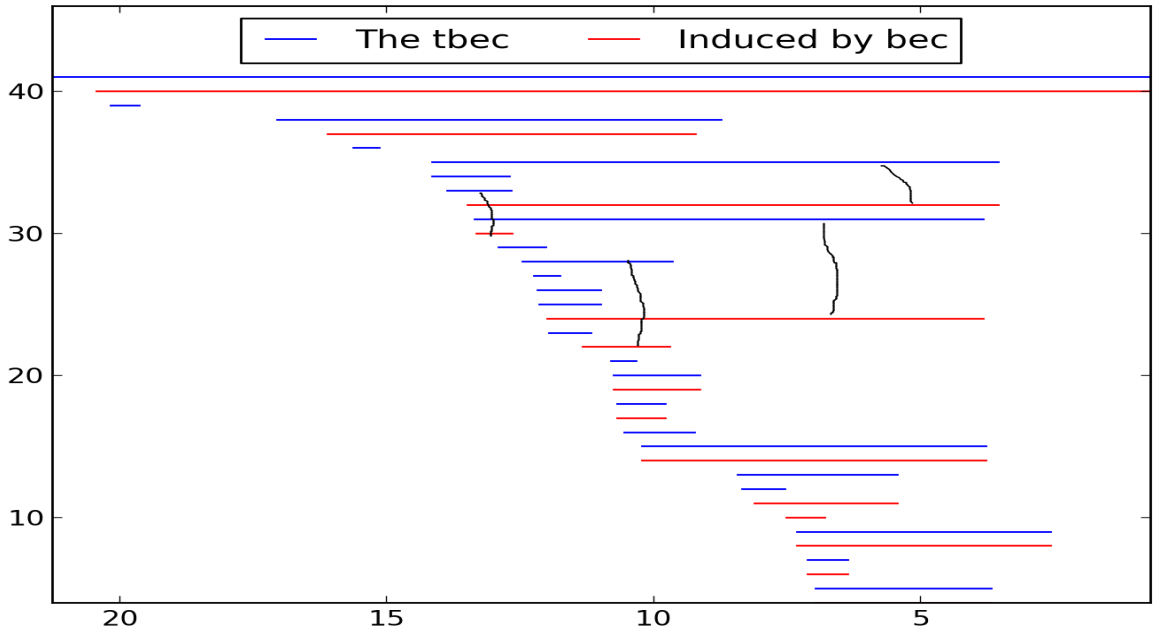
In Figure 71 there is one red (mid-length, right above the 40 mark) that is paired with a much older blue. Other than that, all the reds seem to have quite close pairings.

Figure 72: The *tbec* versus the persistence diagrams induced by the *bec* in month 11 in 2012:



In Figure 72 it looks like all the reds have nice pairings with the blues.

Figure 73: The *tbec* versus the persistence diagrams induced by the *bec* in month 12 in 2012:



In Figure 73 it looks like all the reds have nice pairings with the blues, except a short one (number 6 from the bottom).

6.5.1 Conclusion

After only having compared the diagrams quite superficially, it does look like the correspondence between the *bec* and the *tbec* is as we initially thought: Most of the intervals from the *bec* seem to have a pair in the *tbec* which is slightly larger. (And those that do not are usually not very persistent, nor at the highest filtration values.) If one wanted to do this more rigidly one should apply the bottleneck distance, and also try and see if the pairings actually correspond to pairs of cycles detecting the same paths. Another way would be to find at which time steps the paths the *tbec* detects hit Bergen, and compare with the *bec* around this time step.

7 What more can be done?

So, now we have discussed what has been done in this thesis: The *bec*, the *tbec*, and (somewhat in the background) the implementation of persistent homology. However, there is much more work that could be done on the problem of detecting atmospheric rivers using persistent homology.

For one thing, a more thorough analysis of what has been computed would be in order. One could do a similar analysis of the *bec* for more than only the year of 2012 and see if things look the same. It would also perhaps be fruitful to apply some way to eliminate the seasonal variance, so one could have a more general way of finding the *bec* values that stand the most out (the current one consists mostly of just looking at the pictures). It would also be interesting to compute the *bec* at times where there is known to be atmospheric rivers, and see how well it fares at detecting those.

For the *tbec*, one thing that could be done is to see if there is a better way of finding only the cycles we are interested in (namely those that are corresponding to paths between the equator and Bergen). Another thing that could be done is to investigate more thoroughly how bad the current way to do this really is. It has been tested on simple cases, and compared with the *bec*, and no fatal flaws has yet been unearthed. One could also try to get some overview of the climatological changes the last years, and see how these reflect in the *tbec* diagrams.

In addition to the comparison between the *bec* and *tbec* made in section 6.5, it is also possible to try and find a meaningful time step connected to the cycles in the *tbec* (e.g. the time step where the path it detects reaches Bergen) and compare this with the *bec* values at these time steps.

An can also of course try and carry out the dynamics strategy mentioned in section 5.8.

If the implementation made in this thesis should have any ambition of becoming a real asset for anyone other than the author, a lot of work remains. Perhaps most importantly, a good testing framework should be made to ferret out the bugs in the code that undoubtedly persist. In addition to this, the implementation of the bottleneck distance remains, which would eliminate some of the guesswork made in section 6.5. It would also be fun (but probably not necessary in this setting) to implement more field coefficients than only $\mathbb{F}_2$.

A Example using the implementation.

Here is a simple example of a calculation of persistent homology using the implementation of persistent homology made for this thesis.

```
from filteredcomplex import FilteredComplex
from cells import Simplex

#Zero dimensional simplices:
a = Simplex( 'a', filtrationValue = 0)
b = Simplex( 'b', filtrationValue = 0)
c = Simplex( 'c', filtrationValue = 1)
#One dimensional simplices:
bc = Simplex( ('b','c'), filtrationValue = 1)
ac = Simplex( ('a','c'), filtrationValue = 2)
ab = Simplex( ('a','b'), filtrationValue = 3)
#Two dimensonal simplex:
abc = Simplex(('a','b','c'), filtrationValue = 4)

#Construct complex and compute homology up to dimension maxDim.
simplices = [a,b,c,bc,ac,ab,abc]
f = FilteredComplex(simplices)
maxDim = 1
f.computePersistentHomology(maxDim)

#Print the homology.
for d in range(maxDim+1):
   print 'Homology of dimension %d: \n %s.\n' %(d, f.getDiagram(d) )
```

The output for this is:

```
Homology of dimension 0:
   [Persistencecycle <dimension: 0, birth: 0, death: 2>,
   Persistencecycle <dimension: 0, birth: 0, death: inf>].
Homology of dimension 1:
   [Persistencecycle <dimension: 1, birth: 3, death: 4>].
```

B Monthly plots of the bec and precipitation in Bergen.

In this section the plots of the *bec* and hourly precipitation in Bergen are shown one month at the time. It seems that all the spikes in the *bec* seem to coincide with a spike in precipitation, but not vice versa. This makes sense, as it should rain when an atmospheric river hits, but it could rain without any atmospheric rivers nearby. The reason the precipitation plots are "stowed away" in an appendix is that technically, precipitation is not a part of the definition of an atmospheric river, and so it should not play a part in the detection of them. The fact that it may rain without there being any atmospheric river nearby suggests that it also isn't a very good way to confirm the validity of the computation. Nevertheless, I do include the precipitation here, as it is nice to see that there is some connection (although the connection is not very surprising, giving the close connection between the *bec* and the humidity in Bergen).

Figure 74: The *bec* and the hourly precipitation in month 01 in 2012:

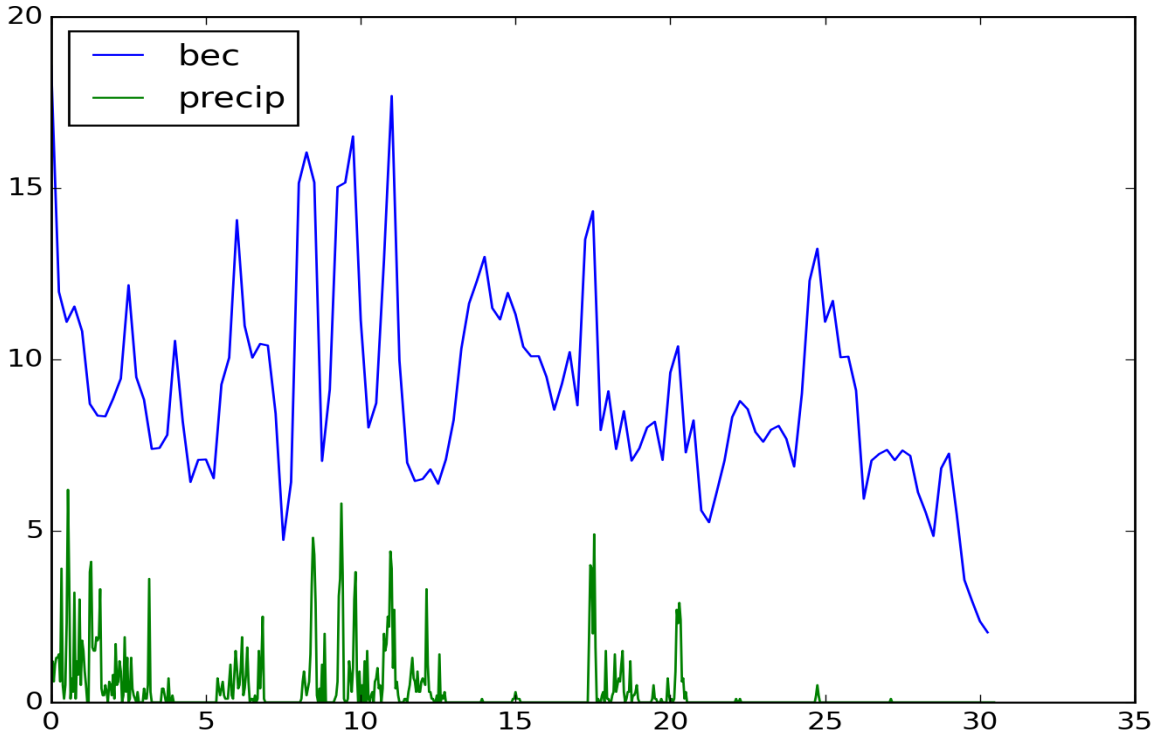


Figure 75: The *bec* and the hourly precipitation in month 02 in 2012:

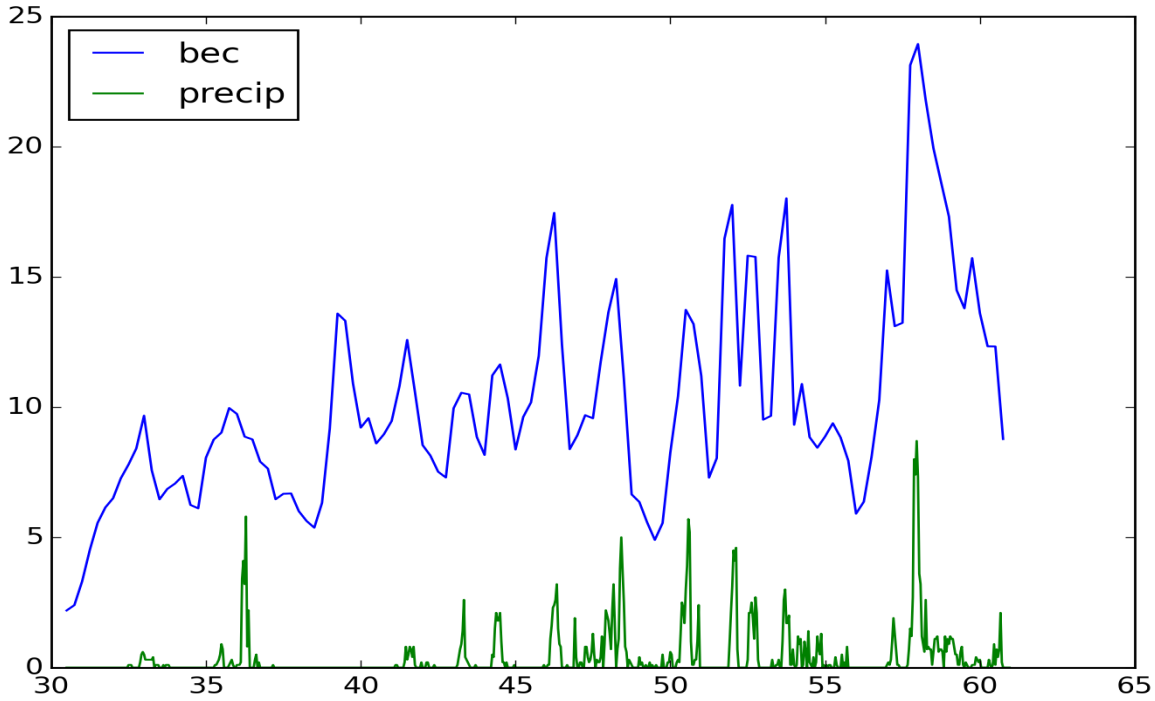


Figure 76: The *bec* and the hourly precipitation in month 03 in 2012:

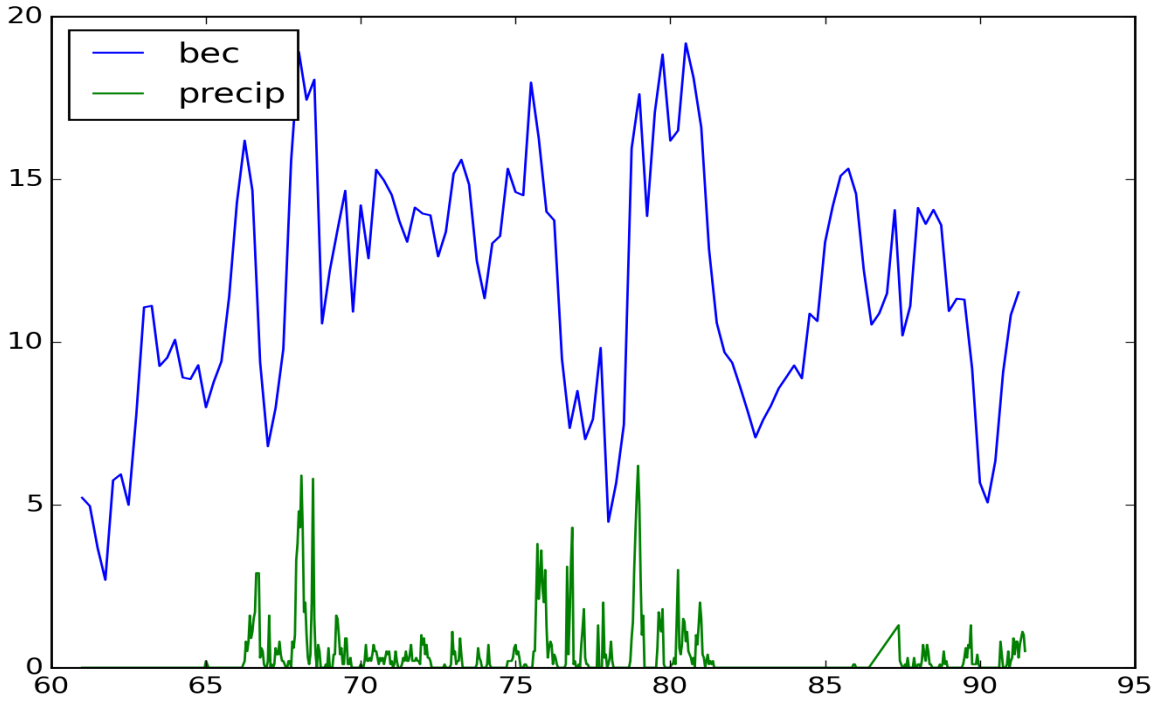


Figure 77: The *bec* and the hourly precipitation in month 04 in 2012:

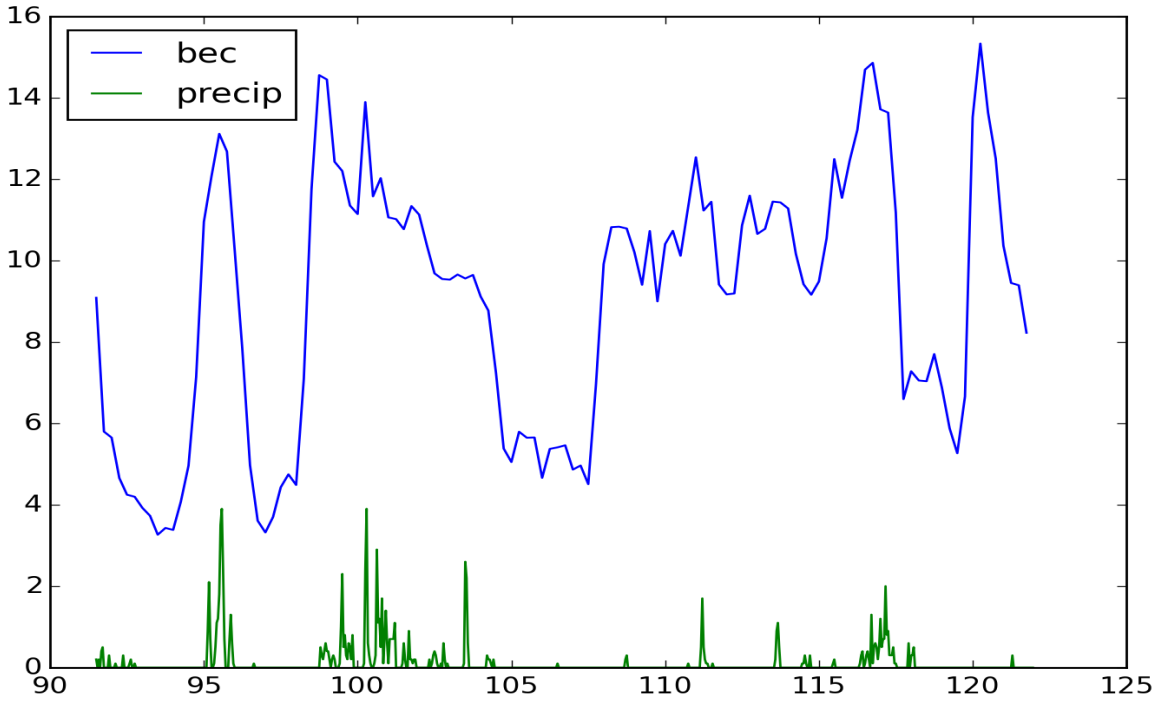


Figure 78: The *bec* and the hourly precipitation in month 05 in 2012:

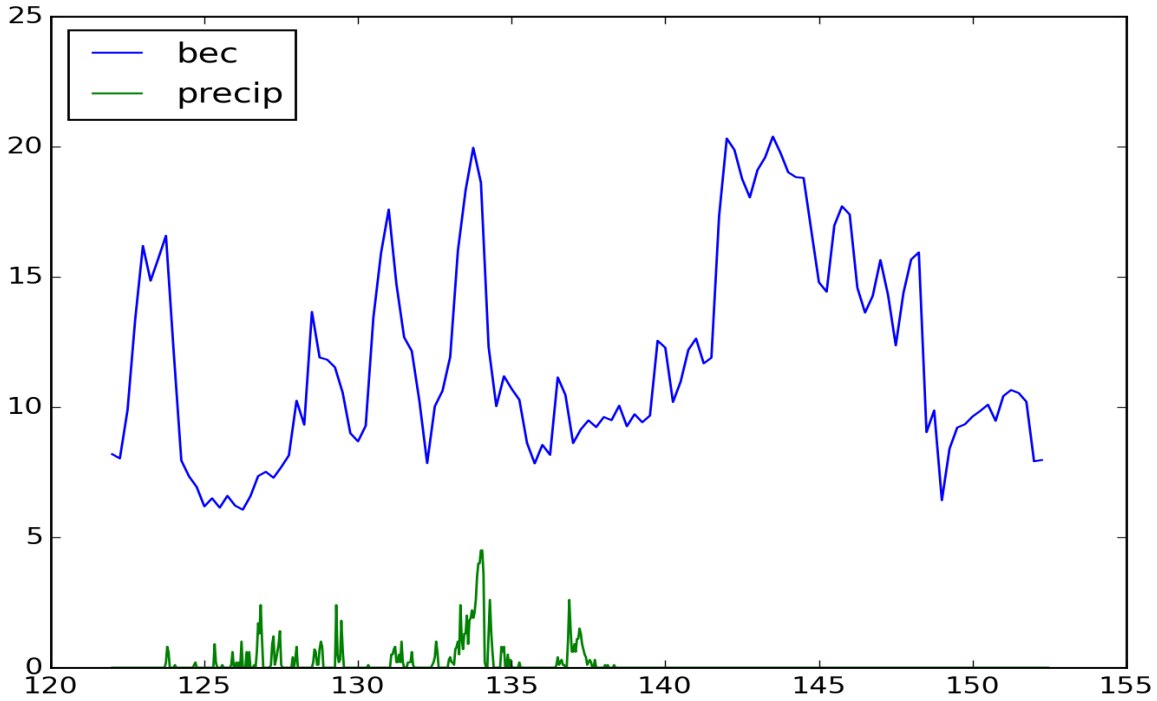


Figure 79: The *bec* and the hourly precipitation in month 06 in 2012:

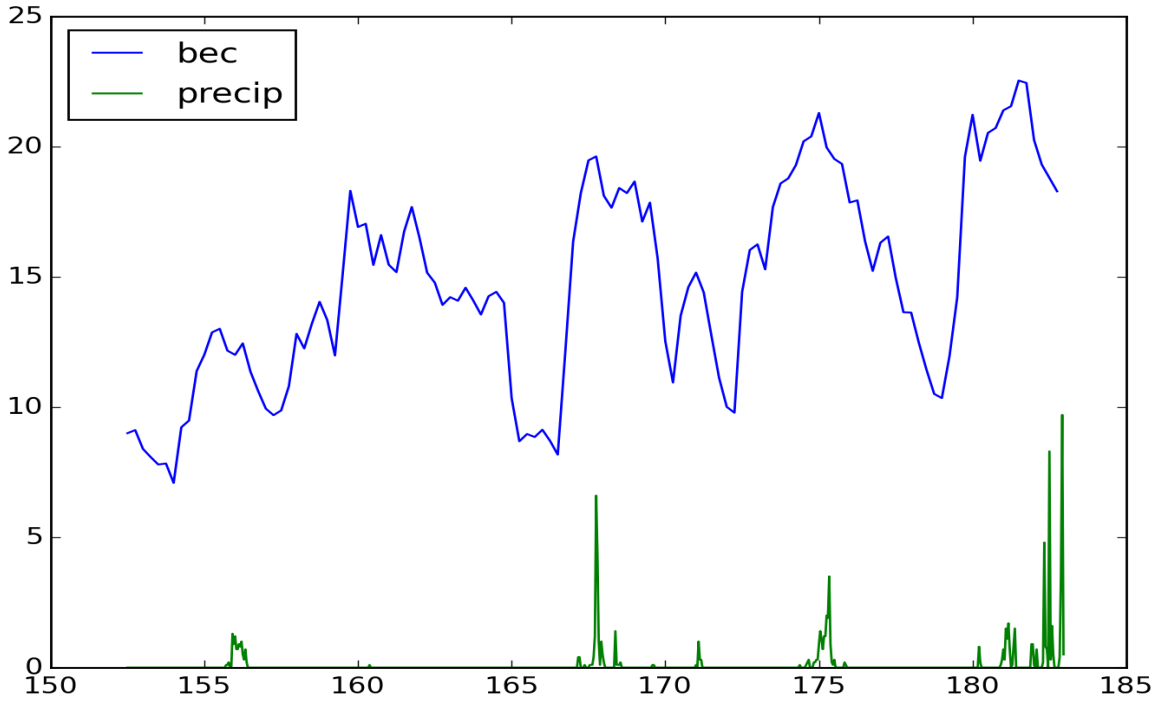


Figure 80: The *bec* and the hourly precipitation in month 07 in 2012:

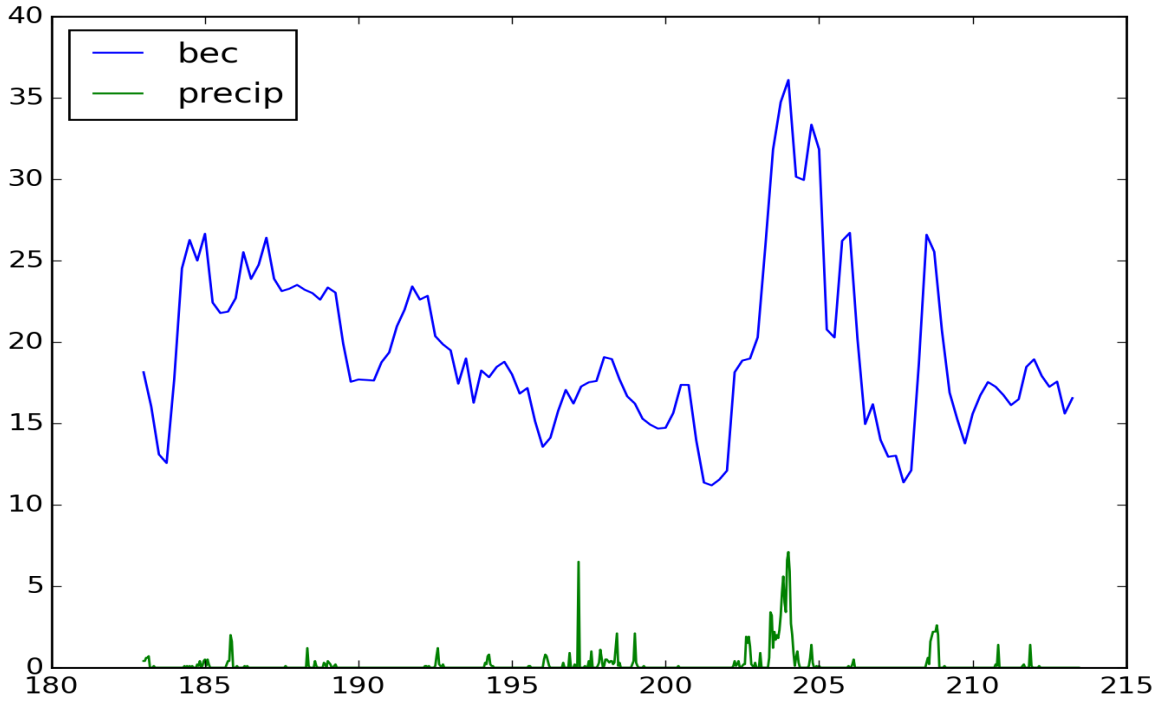


Figure 81: The *bec* and the hourly precipitation in month 08 in 2012:

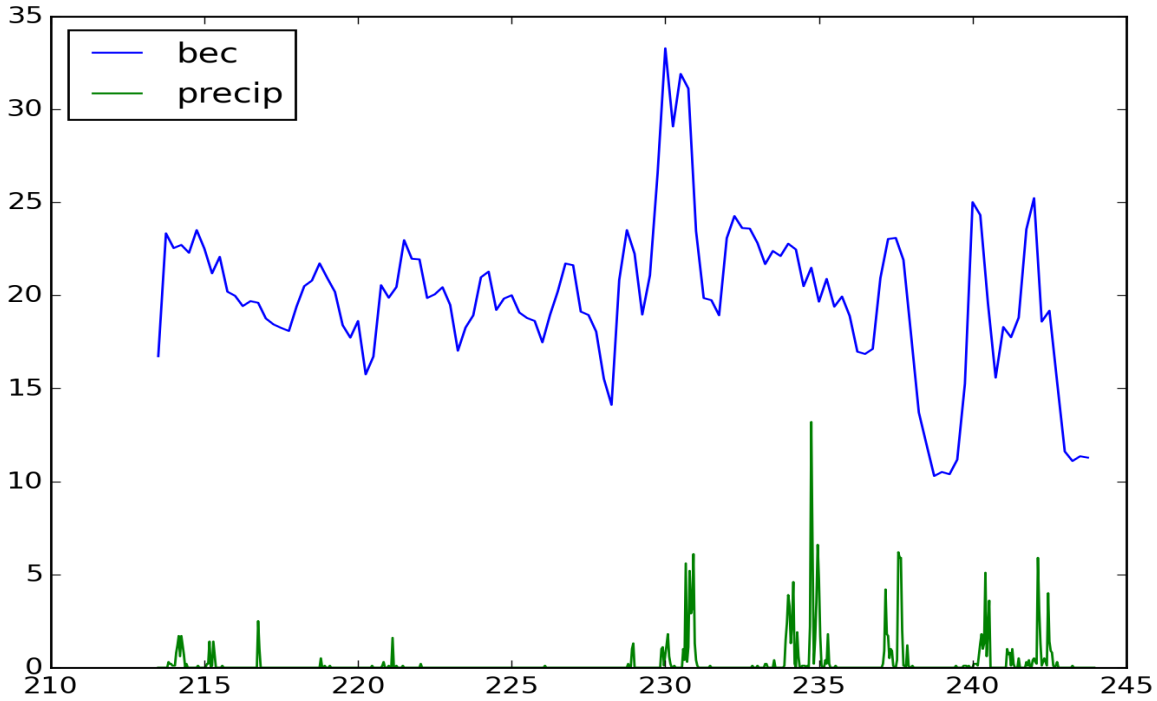


Figure 82: The *bec* and the hourly precipitation in month 09 in 2012:

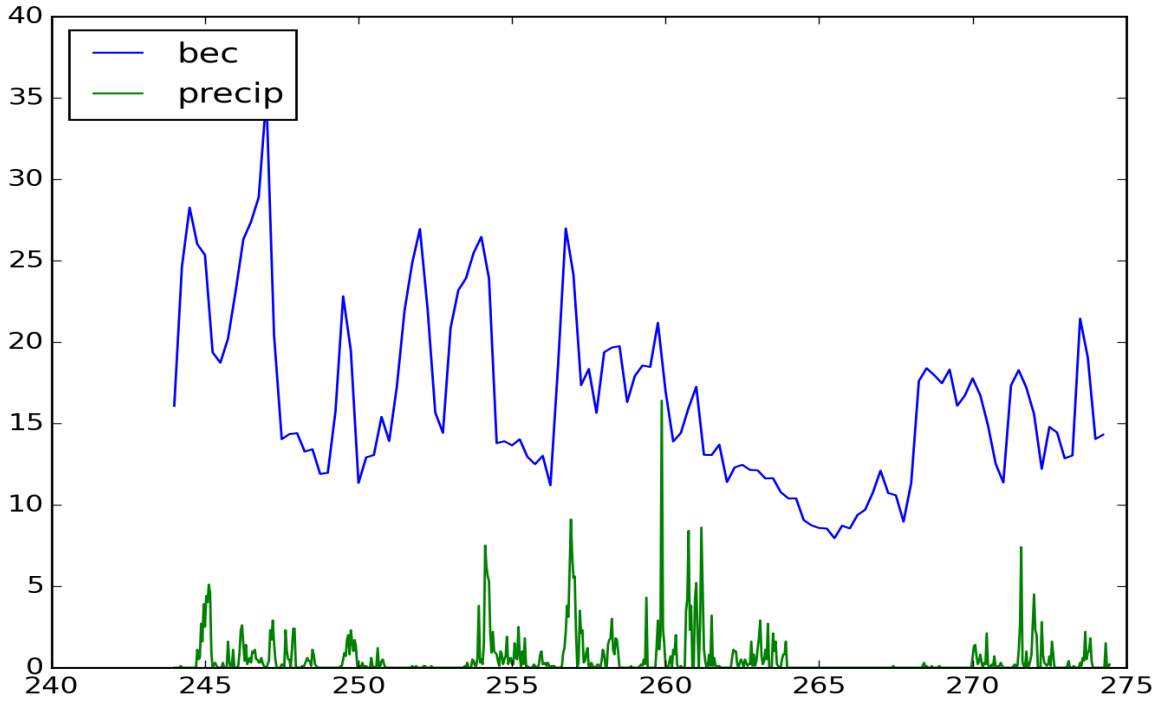


Figure 83: The *bec* and the hourly precipitation in month 10 in 2012:

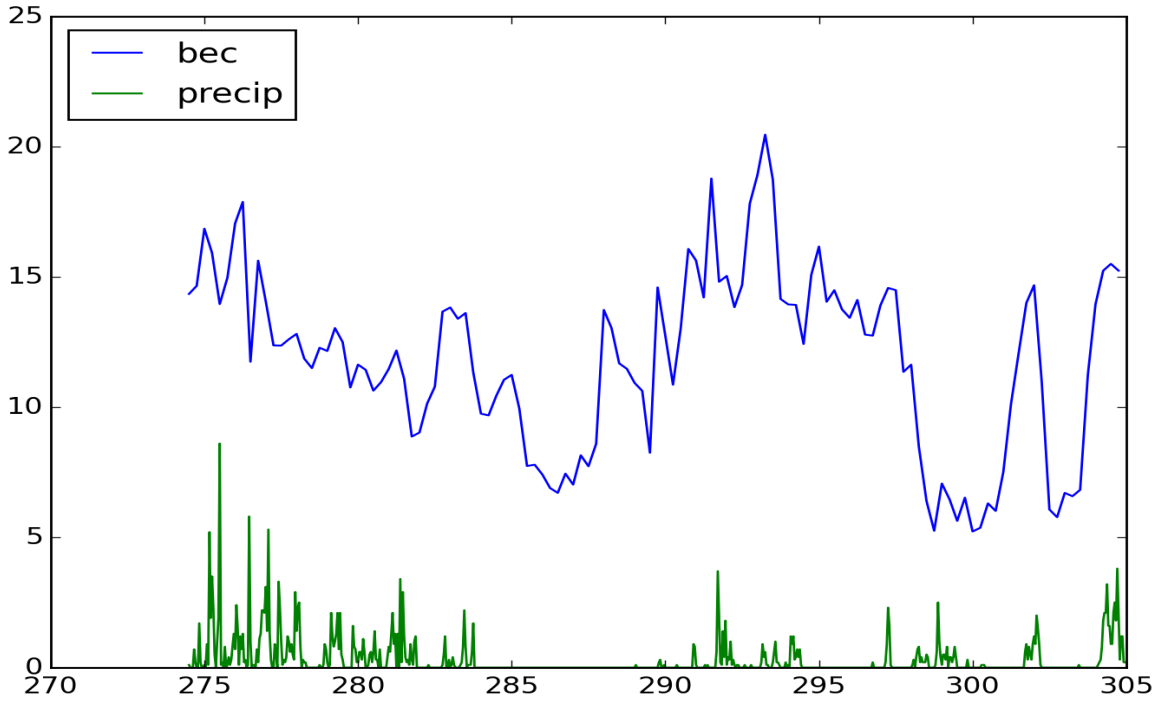


Figure 84: The *bec* and the hourly precipitation in month 11 in 2012:

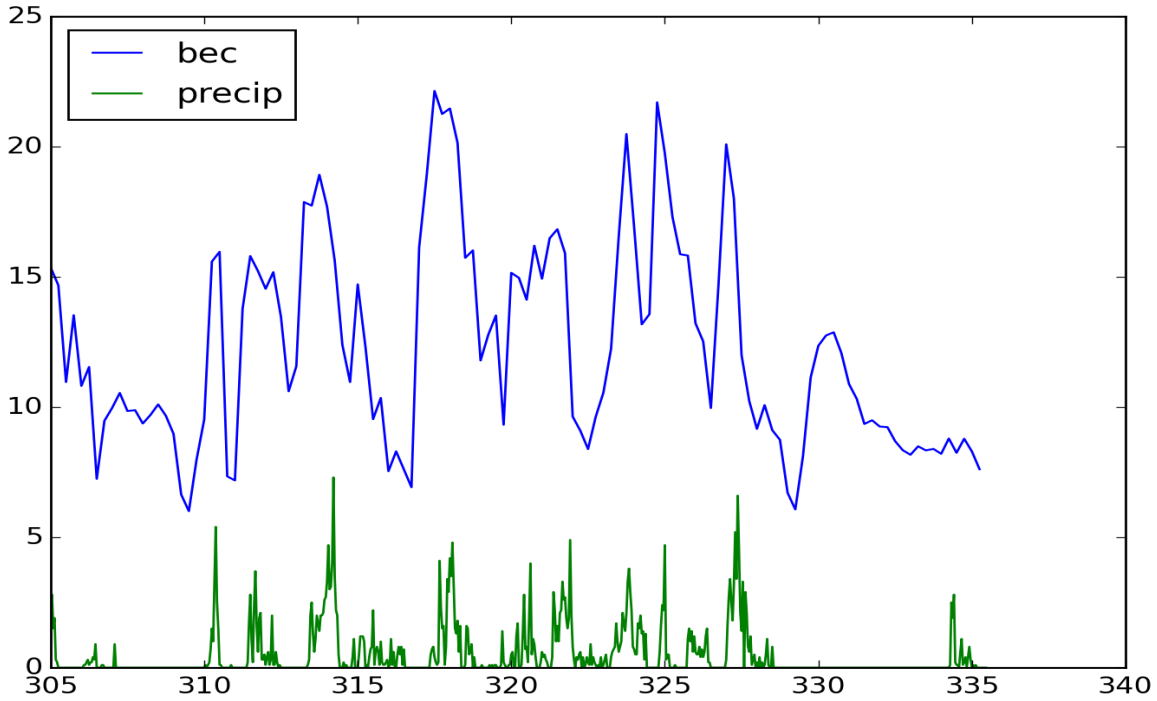
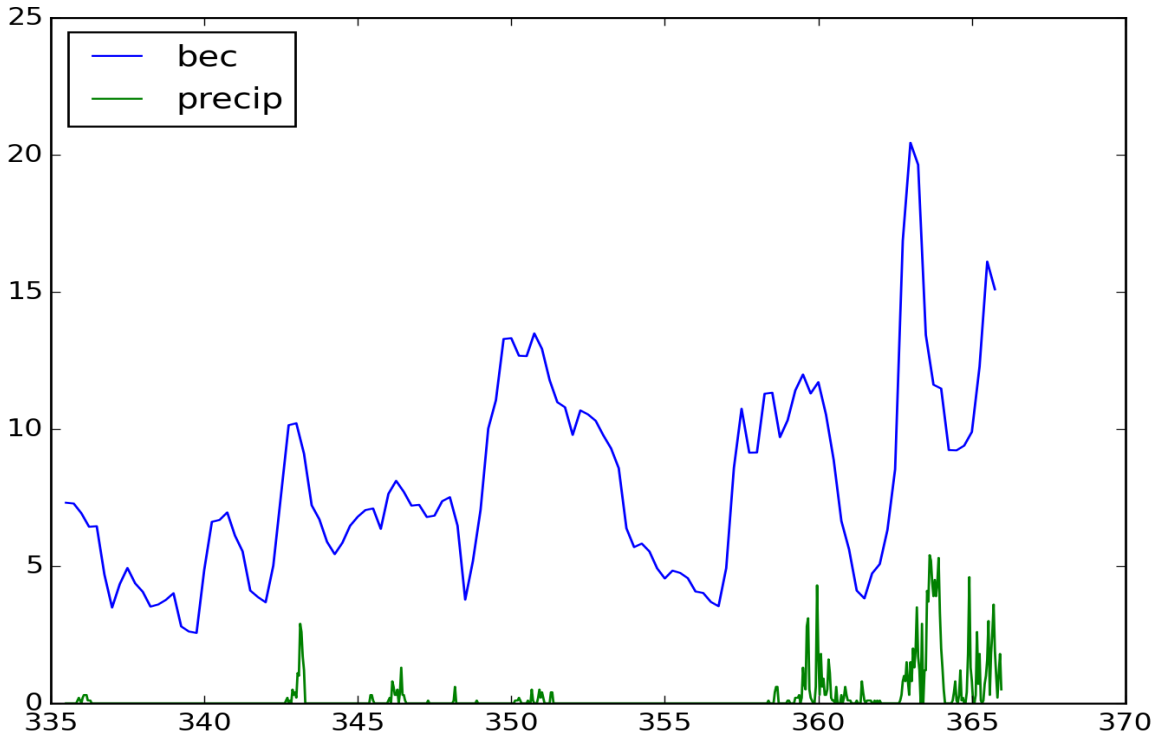


Figure 85: The *bec* and the hourly precipitation in month 12 in 2012:



References

[1] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1):103–120, 2007.

[2] D. P. Dee, S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Holm, L. Isaksen, P. Kallberg, M. Koehler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J. J. Morcrette, B. K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J. N. Thepaut, and F. Vitart. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY*, 137(656, A):553–597, APR 2011.

[3] Herbert Edelsbrunner and J. L. Harer. *Computational topology: An introduction*. American Mathematical Society, 2010.

[4] Allen Hatcher. *Algebraic topology*. Cambridge University Press, 2001.

[5] Dmitriy Morozov. Dionysus. Software available at `http://www.mrzv.org/software/dionysus/`, 2012.

[6] Jonathan J Rutz, W James Steenburgh, and F Martin Ralph. The inland penetration of atmospheric rivers over western north america: A lagrangian analysis. *Monthly Weather Review*, (2015), 2015.

[7] Andrew Tausz, Mikael Vejdemo-Johansson, and Henry Adams. Javaplex: A research software package for persistent (co)homology. Software available at `http://javaplex.github.io/`, 2015.

[8] A Zomorodian and G Carlsson. Computing persistent homology. *DISCRETE & COMPUTATIONAL GEOMETRY*, 33(2):249–274, FEB 2005.