CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Data conditioning and climate sensitivity analysis of a probablistic rainfall-runoff model
Authors	Nowbakht, Parvaneh
Publication date	2020-04-02
Original Citation	Nowbakht, P. 2020. Data conditioning and climate sensitivity analysis of a probablistic rainfall-runoff model. MRes Thesis, University College Cork.
Type of publication	Masters thesis (Research)
Rights	© 2020, Parvaneh Nowbakht. - https://creativecommons.org/licenses/by-nc-nd/4.0/
Download date	2026-11-03 00:38:19
Item downloaded from	https://hdl.handle.net/10468/9979

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Department of Civil Structural and

Environmental Engineering

University College Cork



Data Conditioning and Climate Sensitivity Analysis of a Probablistic Rainfall-Runoff Model

By

Parvaneh Nowbakht

A Thesis submitted for the Degree of Master of Engineering Science

April 2020

Acknowledgments

I would like to thank all of the people who have helped me with this project. This includes my supervisors Dr.Paul Leahy and Prof. Ger Kiely who have provided encouragement and guidance throughout the whole exercise. I would like to acknowledge the help that Dr. Lucía Hermida González offered at every stage of the project. Their criticism, encouragement and suggestions made possible the completion of this project. I also would like to express my thanks to the following people:

The helpful staff at the Office of Public Work, particularly Mr. Mark Hayes and Mr. Peter Newport

The helpful staff at Met Éireann , the Environmental Protection Agency and Environmental Institue .

Dr.Ciaran Lewis and Dr. Jeroen Pullens for their helpful advice on Geotop.

Dr Steven Cole, Group leader, Hydrological Forecasting Center for Ecology & Hydrology and helpful PDM support staff for their helpful advice on PDM.

This thesis has been prepared as part of the "*ClimAtt*" project at University College Cork and Environmental Change Institute Of Oxford university funded by the Irish Environmental Protection Agency (EPA) under Gran Award No. 2016-W-MS-23.

Abstract

The Munster Blackwater catchment, in the South West of Ireland, was regularly subject to flooding, prior to flood allevation works. The towns of Mallow and Fermoy within the catchment suffered many disturbances for their inhabitants with sometimes severe economic losses. A good knowledge of rainfall-runoff processes is important in order to understand the causes of flooding to be able to develop new infrastructure to manage flooding.

The first part of this project focuses on the rainfall and river flow data collection from different sources: the 15-minute time step precipitation data from the OPW, the 15-minute river level/river flow from the OPW and the EPA and the precipitation data from MÉRA (Met Éireann ReAnalysis-Climate ReAnalysis). MÉRA is a very high resolution climate reanalysis dataset which was used to calculate the monthly and annual rainfall in a specific year, for example for 2010 for selected locations (the nearest point to each rain gauge). Initial analysis of the measured OPW data shows significant numbers of missing values and outliers for the precipitation data. A method was developed to cluster the rain gauges with similar precipitation patterns based on the amount of precipitation of the nearest points to these rain gauges from MÉRA. Then a gap filling method was applied in each cluster to fill the missing values of each rain gauge with its cluster members. Other methods were also examined to obtain quality controlled data.

The second part of this project applies a conceptual hydrological model, PDM (Probability Distributed Model) developed by Moore (Moore, 2007) to the Munster Blackwater catchment. The model considers each point of a catchment as a single storage unit with a specific storage capacity (depth) that can be described by a Pareto distribution. PDM is suitable for a variety of catchments, and has minimal data and computational requirements. The input is 15-minute precipitation data from different rain gauges and 15-minute river level/river flow data from river stations along the river. The calibration was applied on three subcatchments of the Munster Blackwater catchment. The validation was applied for years between 2010 to 2017. The calibrations and validations indicate that the PDM model can explain most of the variability of observed flows in the different subcatchments over a period of years, especially when a high standard of data quality is available, for example in 2015. Then validation of the model for flood events was examined. Validation was applied for the highest flood event in each year during 2010 to 2017. The accuracy of the model runs are different for each subcatchment with the best accuracy of 93% in the Dromcummer subcatchment and the accuracies in Mallow Rail BR and Killavullen being 80 % and 78% respectively. The model estimates

the peak and low flow very well in Dromcummer. The computed flow is underestimated in Mallow and overestimated in Killavullen.

The third part of the project is to use the PDM model in a precipitation and river flow sensitivity analysis. This was achieved by increasing the precipitation amounts in the datasets by 10, 15, 20, 25 and 30% to examine how the peak flows and low flows respond. It was found that the peak flows increase by amounts similar to the precipitation increases. The low flows increase at a much lower rate than the precipitation increases. It is known that in a scenario of climate change for a warming world that the precipitation increases by a maximum of 7% per degree C increase in accordance with the Clausius-Clapeyron equation. However as a warming world also increases evaporation and will likely impact the soil moisture status, it is considered that flood flows might increase at a rate less than the precipitation increases. This can be examined by increasing the value of potential evaporation by 10, 15, 20, 25 and 30% .These conditions were not included in this and it is ecommended that further research be done in this area for Ireland.

Contents

Chapter 1 Introduction ... 8
1.1 Climate Change Attribution ... 9
1.2 Flooding ... 11
1.2.1 Effects of Human Activities on Flooding ... 12
1.3 Extreme Events in Ireland ... 13
1.4 Data Conditioning ... 14
1.5 Objectives ... 15
1.6 Layout of the thesis ... 16
Chapter 2 Literature Review and A Comparison of Hydrological Models ... 17
2.1 Review of Hydrological Models ... 18
2.2 Physically Based Models (White-Box Models) ... 19
2.2.1 GEOtop Model ... 19
2.2.2 tRIBS Model ... 25
2.3 Conceptual Models (Grey-Box Models) ... 27
2.3.1 Probability Distributed Model (PDM) ... 27
2.3.1.1 Probability Density Function and Cumulative Distribution Function ... 28
2.3.1.2 PDM and Pareto Distribution ... 29
2.3.2 TopModel ... 32
2.4 Metric Models (Black-box models) ... 35
2.4.1 K-Nearest Neighbours Model (KNN) ... 36
2.4.2 Artificial Neural Network (ANN) ... 37
2.5 Summary ... 39
Chapter 3 Data Sources And Descriptions ... 41
3.1 Site Description ... 42
3.1.1 Catchment Location ... 42
3.1.2 Topography and Digital Elevation Model (DEM) ... 43
3.2 Data Collection ... 46
3.2.1 Office of Public Works (OPW) Precipitation Data ... 46
3.2.2 Method of Analysis of Precipitation Data ... 47
3.2.3 MÉRA (Met Éireann Re-Analysis Climate Re-Analysis) ... 50
3.2.4 MÉRA Precipitation Data ... 51
3.2.5 River Level Data ... 56
Chapter 4 Analysis of Data and Methodologies Employed ... 58
4.1 Introduction ... 59
4.2 Kmeans Clustering ... 59

4.2.1 Elbow Method 60
4.3 Clustering of rain gauges 60
4.3.1 Clustering based on location 61
4.3.2 Clustering based on MÉRA precipitation 63
4.4 Gap filling 69
4.4.1 Rain Gauges Gap Filling With Cluster Members' data 69
4.4.2 Gap Filling of Rainfall Data Using Temporal Average of Rain Gauges' Nearest Record 71
4.4.3 Long Sequences of Precipitation and Outliers 71
4.4.4 River Level Data Gap Filling with Average of River Level's Nearest Record 77
Chapter 5 Probability Distributed Model (PDM) 78
5.1 Introduction 79
5.2 Data preparation for PDM Model 79
5.2.1 Rain Gauges Contributing Area 79
5.2.2 Thiessen Polygon 79
5.2.3 Inverse Distance Weighting (IDW) 82
5.2.4 Rating curves 82
5.3 Probability distributed soil moisture storage 85
5.5 Objective Functions 88
5.6 PDM model Updating Methods 89
5.6.1 State correction 90
5.6.2 Error prediction 91
5.7 Model Calibration 92
5.7.1 Input file 93
5.7.2 File -1 Data interface file 97
5.7.3 Alternative model configuration 97
5.8 PDM Model Inputs 98
Chapter 6 PDM Calibration and Validation 100
6.1 Subcatchments of the Munster Blackwater 101
6.2 Model Calibration 101
6.3 Model Validation 113
6.3.1 Model validation for a year 113
6.3.2 River Flow Comparison 120
6.3.3 Model Validation for an Event 134
6.3.4 Case 1: Event from $3th$ Feb 2014 to $3th$ Feb 2014 135
Chapter 7 Precipitation Streamflow Sensitivity Modelling 141
7.1 Precipitation Streamflow Sensitivity Modelling 142

7.1.1 Pattern 1: period containing the largest event (high flow) 146
7.1.2 Pattern 2: period containing the smallest event (low flow) 148
Chapter 8 Discussion and Conclusions 151
8.1 Discussion 152
8.1.1 Validation for Long period of time (annually) 152
8.1.2 Validation for short period of time (event) 153
8.1.3 Conclusion 154
8.1.4 Recommendations for Further Research 155
References: 156
Appendix A Annual and Event Validation- 2010 to 2017 162
A.1 Instructions On How To Run PDM Model 163
A.2 Example of annual runs of PDM for each subcatchment 166
A.2 PDM Event Hydrograph examples 185
Appendix B Python & R Code 207
B.1 Python Code MERA data using GRIBapi Package: 208
B.2 R Code for River level/ River flow data: 211
B.3 R Code for plot River level/ River flow data: 217
B.4 R Code rainfall data, Kmeans clustering, gap filling: 223

Chapter 1 Introduction

1.1 Climate Change Attribution

It is now well established that we live in a warming world. Global temperature has increased by approximately 1°C since the mid 1800s. With each 1°C warming, an additional 7% increase in precipitation occurs in accordance with the Clausius-Clapeyron equation (Li, Harrison, Bartlein, Izumi, & Prentice, 2013). So hydrologically we now have more precipitation, more evaporation and changes to river flow relative to the 1800s. In recent decades, there is more evidence of extreme climate phenomenon happening all over the world such as heat waves, floods, droughts, heavy rainfall and hurricanes. For example: Annual precipitation in Honghu region (located in Jianghan plain in the middle reaches of the Yangtze River in China) has the increase tendency in recent 50 years (1960 – 2011). The main reason is the increase of heavy precipitation event. The rainfall in from April to July count for 54% of total precipitation, and more than 90% of the heavy precipitation processes concentrated in this period, which leads to the increase of flood. The uneven distribution of precipitation time and decreased tendency of annual precipitation days lead to the increase of drought process in the region, especially the light drought (Liu, Liang, Zhou, & Liu, 2013). Using ocean and atmosphere observations demonstrate the link between increased upper ocean heat content due to global warming with the extreme rainfalls from hurricane Harvey in summer 2017 (Trenberth, Cheng, Jacobs, Zhang, & Fasullo, 2018). And the impact of The weather conditions (extremely high temperature conditions and low humidity, known as “heat wave”) of the city of Turin (Italy) on distribution system in 10 years from 2008 to 2017 was studied (Zhang, Mazza, & Bompard, 2019). The term “extreme event” can be explained as an intense or very severe rare phenomenon compared to what normally occurs over a period of time in a specific location. An event might be considered as an extreme event in a location but it might be a normal event somewhere else. Seasonality is another issue, as an unexpected event for a season might be considered as an extreme event while a similar event might be common in another season (Kiely et al., 2007). Climate is a description of atmospheric behaviour over long periods, typically years, decades, or longer (SciLine, 2017). For example, Ireland has a humid temperate climate. Climate change means a significant change in some features of climate (e.g. temperature, rainfall, wind, …). Climate change has happened several times over the long period of human history due to natural weather patterns or random climate variability and is now happening again. Most scientists believe in today’s world another important factor - an increase in CO_2 due to fossil burning - has a huge influence in climate change (Chen, Gao, Han, & Chen, 2009). They have found more evidence of impacts of human activities on climate change. Collecting different types of data from different sources with advanced technologies such as satellites and radar and using advanced computer processing power enable

scientists to analyse “big” climate data to study the many factors contributing to weather and which factors contribute to extreme weather events (Dabrowska et al., 2017; Fiore et al., 2016). Certain extreme weather events are unlikely to be happening without climate change. This is the science of extreme event attribution (SciLine, 2017). Two main factors, which can accelerate the process of climate change are:

- natural processes such as changes in the sun’s intensity, volcanic eruptions, and slow changes in the earth’s orbit around the sun
- Increases in greenhouse gases such as CO_2, N_2O and CH_4.
- human activity such as burning fossil fuel including coal, oil and gas producing carbon dioxide emissions, agriculture producing methane and nitrous oxide, land use change including deforestation, reforestation, urbanization and desertification (EPA, 2007).

The result of these human activities is increasing emissions of CO_2, which directly increases the greenhouse effect. The greenhouse effect is a mechanism that makes life possible on the earth. Roughly, 30% of the incoming energy to the earth by solar radiation is reflected back to space by the atmosphere and the earth, 50% is absorbed by the earth’s surface (the land and oceans) and 20% is absorbed by the atmosphere and clouds. The sun’s energy absorbed by the earth makes the earth warm (Figure 1.1).

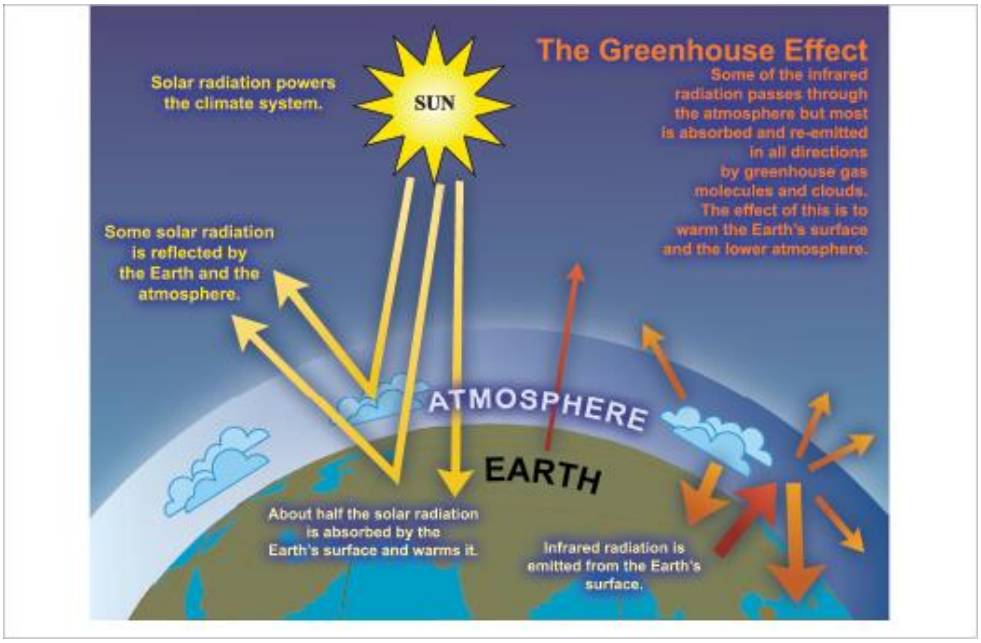


Figure 1.1 An idealised model of the natural greenhouse effect (Treut et al., 2007).

As Greenhouse gases (CO_2, CH_4, N_2O) have grown due to human activity over the past century, the absorbed energy by the earth and atmosphere has increased and the lost or reflected energy by the earth's surface is reduced, causing global warming (Treut et al., 2007). According to global temperature, analysis conducted by scientists at NASA's Goddard Institute for Space Studies (GISS), the average global temperature has increased by about 0.8°Celsius over the past century. Figure 1.2 shows the effect of a small increase on average of a variable with normal distribution, for example temperature. A small change on some climate attributions can produce big extreme events either in frequency or in value (Kiely et al., 2007).

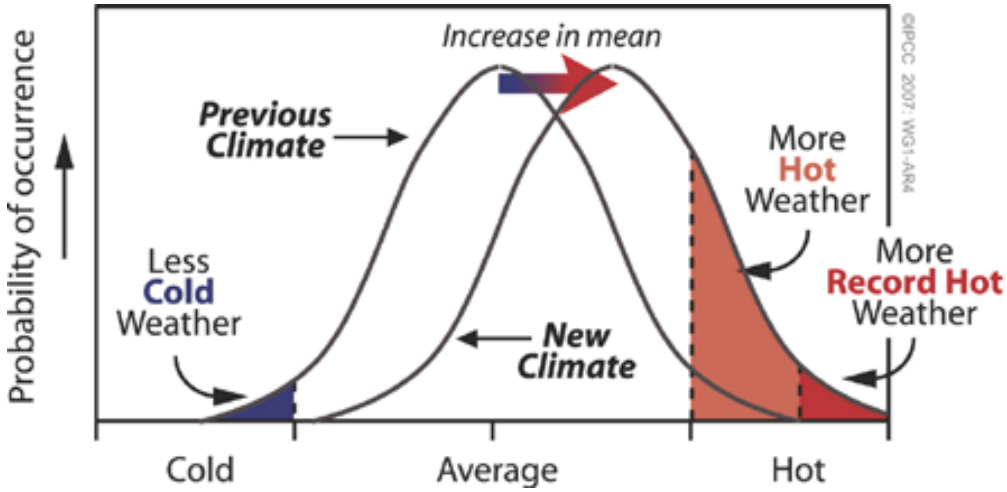


Figure 1.2 Schematic showing the effect on extreme temperatures when the mean temperature increases, for a normal distribution (Box TS.5, Intergovernmental Panel on Climate Change, Assessment Report 4: 53, 2007).

The detection and attribution of climate change studies use statistical methods such as:

- Optimal fingerprinting: a generalized multivariate regression
- Standard frequency approach
- Bayesian approach: Interest in a Bayesian approach is motivated by several factors. These include the ability to integrate information from multiple lines of evidence, and the ability to incorporate independent prior information into the analysis. There is increasing interest in the use of Bayesian methods of statistical inference in hydrology (Zwiers, 2004).

1.2 Flooding

Flooding is defined as water overflowing onto land that normally is dry. Flooding occurs mostly when rivers or streams overflow their banks and spread throughout relatively flat adjacent land, called a

floodplain (Corcoran, 2004). Floods can be the result of short duration intense rainfall or long periods of heavy precipitation, melting snow or dam breaks. Storm surge, hurricanes, waves and tides generate coastal flooding. Physical processes and the geographical locations (e.g low-lying) of the area where the flood occurs combine to generate floods. The geographic location is an important factor in determining the characteristics of a flood. For instance, heavy rains can generate floods in river valleys, but they do not generate floods on seacoasts. Storms may generate surges on low-lying shores of seas and lakes, but do not generate surges on shores with elevated and steep coasts. Floods can be classified according to the factors and conditions of their generation such as river floods, flash floods, urban floods and coastal floods. Floods depend on precipitation intensity, volume, timing, antecedent soil moisture conditions of their drainage basins (e.g., presence of snow and ice, soil characteristics, soil wetness, urbanisation, and existence of dikes, dams, or reservoirs) (Mandych, 2009).

Floods could be a natural disaster due to the power of moving water. As the energy is a product of mass by the square of speed, so the energy of moving water grows very fast. The increase in the amount of water causes an increase of pressure gradient and eventually increases the destructive effect of moving water. Floods can produce damage through the deposition of gravel and debris, suspended silt and potentially toxic microorganisms and dissolved chemicals. Although flooding has negative impact, it is part of natural processes and may have long-term positive impacts. Floods deposit nutrient-rich sediments, which support the animal and plant life and improves the land's fertility for agricultural use. Flood waters recharge the ground water. Soil deposited by floods prevents erosion and helps maintain the elevation of land masses above sea level (Doswell, 2003). The consequences of floods, both negative and positive, depends on their location, duration, depth and speed, as well as the vulnerability and value of the affected natural and constructed environments.

1.2.1 Effects of Human Activities on Flooding

Although floods can have disastrous consequences in both terms of life and economies, human activities in an area, including the construction of roads and buildings or even agriculture, may destroy the native vegetation and change the characteristic of the land and so, increase the flooding risk. Vegetation and soil store rainfall until saturated and then runoff is generated. In the absence of vegetation and changes in land use which reduce the interception and Infiltration to the soil, rainfall may start to runoff immediately and increase the flood risk (OPW, 2004).

1.3 Extreme Events in Ireland

Evidence of the effects of climate change on global temperatures and weather patterns has been clearly proven (Treut et al., 2007). On a regional scale, these effects have been detected in temperature and rainfall patterns in the UK and Ireland (Hoppe & Kiely, 1999) ;(Leahy & Kiely, 2010). Since 1975 in the western half of Ireland there has been an increase in the annual precipitation, which has an impact on flood frequency and flood magnitudes (Hoppe & Kiely, 1999).
Studies of the Irish Meteorological Service (*Met Éireann*) indicate the mean annual temperature increased by $1 - 1.6^{\circ}\ C$ by mid-century (1981-2000) over Ireland (Nolan, 2015), the largest increase seen in the east of Ireland. The highest temperature recorded is $32.5^{\circ}\ C$ at Boora Co.Offaly on 29^{th} June 1976 and the lowest temperature is $-18.8^{\circ}\ C$ at Lullymore Co.Kildare on 2^{nd} January 1979.
Ireland has engaged in global modelling by contributing to the scientific development of the EC-Earth global climate model. The EC-Earth and other global models are downscaled to provide information at a regional level for Ireland (Met Éireann, 2018a). Downscaling is a process by which global models, which have a typical resolution of 50km or more, are dynamically or statistically modelled onto a finer scale with a resolution as low as 4km, which allows a better resolution of coastlines and topography (Nolan, 2015). The downscaled simulations show, the mean annual precipitation decreases in Ireland over the full year, more during spring and summer and the rate of reduction is between 0 % to 20% (Nolan, 2015).

The impact of simulated global climate change on rainfall in Ireland was examined using the method of regional climate modelling for period (1981–2000) indicated there are significant increases of heavy rainfall events of about 20% during the winter and autumn (Nolan, 2015). The highest annual precipitation recorded is 3964.9 mm at Ballaghbeama Gap Co.Kerry in 1960, The highest monthly precipitation is 943.5mm at Gernapeka Co.Cork on December 2015 and the highest hourly precipitation is 52.2 mm at Clonroche Co.Wexford on 27^{th} June 1986. While the lowest annual rainfall is 356.6mm at Glasnevin, Co.Dublin in 1887 (Met Éireann, 2018c).

A study of impacts of climate changes in nine Irish catchments has been carried out (Steele-Dunne et al., 2008). Multiple catchments were selected to ensure varying topography, geology and climatology. Data from the European Centre Hamburg Model version 5 (ECHAM 5) GCM was used to force the Rossby Centre Atmosphere Model (RCA3) regional climate model to produce dynamically downscaled precipitation and temperature data. This data was used to force the HBV-Light conceptual rainfall-runoff model to simulate stream flow in the reference period (1961–2000) and in

the future (2021–2060). According to this study, temperature is expected to increase in all months in all catchments. The greatest increase is expected in August and the smallest increase occurs in June. All catchments showed an expected decrease in summer precipitation (greatest decrease in August) and an expected increase in winter precipitation (greatest increase in January). Comparison of the simulated flows for the period 2021-2060 with flows for the reference period 1961-1990 suggested that winter flows will increase and summer flows will decrease while river flows from October to April are expected to increase by up to 20% by 2060. Due to the combination of reduced summer precipitation, increased temperature and as a consequence increased evaporation, stream flow is expected to decrease by up to 60% from May to September. The study also points to an increase in the annual maximum daily mean flow indicating that the severity of large flood events will increase. A Similar study (Charlton, Moore, Sweeney, & Fealy, 2001) also produced the same prediction, reduction in flows for autumn months and higher flows in winter and spring, dry periods in the summer and autumn months in the majority of catchments are expected.

The difference between the study presented here and the above-mentioned studies is the long periods (yearly) and short period flood events were examined instead of changes in seasonal flows.

Absolute drought is defined by the world Meteorological organisation (WMO) as a period of 15 or more consecutive days, on none of which 0.2 mm or more of rain fell (Mac Carthaigh, 1996). Several absolute droughts were reported in Ireland as follows:

- The longest absolute drought from 3^{rd} April to 10^{th} (37 days) May 1938 in Limerick (Met Éireann, 2018c).
- The second largest absolute drought From 8^{th} Aug 1976 to 7^{th} Sep 1976 (31 days) in Claremorris Co.Mayo (Mac Carthaigh, 1996).
- The recent absolute drought from 22^{th} May 2018 to 12^{th} June 2018(22 days) in Belmullet (Met Éireann, 2018b).

Several extreme events happened in Ireland recently: for instance floods in November 2009 and the cold in January 2010, the Cork floods in 2009, the cold spell in Ireland in December 2009 and January 2010 and the drought of summer 2018.

1.4 Data Conditioning

Data is one of the most important factors in any research. Data might be collected from different sources with different formats. To make this data useful and applicable to feed the

application and transfer the data to information, first we need to be sure about the quality of the data. Data cleaning and conditioning are the process of detecting and correcting or removing of any incorrect and generally any bad values in the datasets. Analysis and visualization of the raw data is the first step to detect the problems in the datasets and then using statistical methods, machine learning and some knowledge of the research area can help to improve data quality.

The characteristic of data quality can be summarised as follow (Sattler, 2016):

- Accuracy: data were recorded correctly
- Completeness: All relevant data were recorded
- Uniqueness: no duplicate data exist
- Consistency: collect and store data by stable and steady mechanism
- Timeliness: keeping data up to date

1.5 Objectives

The objective of the project is to analyse the available data (rainfall and the Munster Blackwater river flows) and develop a procedure to obtain clean data. Then to simulate storm events (rather than long-term flows) with catchment data and sufficient meteorological rainfall data, using the PDM model (The Probability Distributed Model, Moore 2009) for hydrological simulation. These objectives were carried out as follows:

- The collection of the necessary input data (rainfall and river flows) to run the hydrological model. These data include observations of rainfall data from rain gauges and observed data of river level/river flow from river stations. These data also include the MÉRA data from Met Éireann ReAnalysis.
- Analyse all available data and develop the procedures for quality control of data
- The calibration of the parameters of the PDM model and their validation.
- The assessment of the potential use of the model for simulating river flow on the Munster Blackwater River and its subcatchments.

1.6 Layout of the thesis

Chapter 2 presents a comparison of several hydrological models in different categories and their characteristics, advantages and disadvantages to enable the choice of the most appropriate model for this study. Chapter 3 presents the description of the Munster Blackwater catchment and all data (rainfall and river flow) for analysis. Chapter 4 presents analysis of the data in more detail describes the method to classify the available rain gauges over the catchment and describes several gap filling methods to obtain clean and useful data, which includes using the MÉRA rainfall data. Chapter 5 describes the PDM hydrological model used in this research. Chapter 6 includes the calibration and validation of the PDM hydrological model. The model is calibrated by optimization of parameters for 2014 and validated for the years between 2010 and 2017. Chapter 7 includes results under modified climatic conditions for a rainfall-river flows sensitivity study. Two different patterns on growth of computed flows was distinguished by applying the PDM model on periods containing the largest events (high flows) and periods containing the smallest event (low flows). Chapter 8 includes a discussion of the events, conclusion and recommendation for future research.
References.
All analysis was done using R, Python and ArcGIS.

Chapter 2 Literature Review and A Comparison of Hydrological Models

2.1 Review of Hydrological Models

Over the past 50 years many different hydrological rainfall-runoff models, also called catchment hydrological models have been developed (Devia et al., 2015; Sitterson et al., 2017). One objective of catchment modelling is to gain an understanding of the hydrological processes operating in a catchment and how changes (e.g. climate or land use) in a catchment may affect the catchment hydrological response and thus the riverflow outputs.

Rainfall-runoff models are classified based on the extent or paucity of the models input data and parameters and the extent of physical principles applied in the model. Models can be fully distributed, semi-distributed or lumped.

In **distributed models** the parameters are expressed as a function of space and time and can contain extensive detail of the catchment topography, soils, land cover etc. The simulations of these models are distributed in space by dividing the entire catchment into small units, usually square cells or a Triangulated Irregular Network (TIN), so that the parameters, inputs and outputs can vary spatially. Distributed models are physically based models and demand a lot of parameters to solve the complex water and energy budgets or hydrological process and eventually to estimate the runoff in each cell or TIN.

Semi-distributed models divide the catchment into subcatchments units, each with unique uniform characteristics and a unique discharge point that can be either a node or another subcatchment. These models are not as complex as the fully distributed models and so require less input data and less parameterisations and are not as simple as the lumped models which require the least data (Pina et al., 2016).

In **lumped models**, the entire river catchment is considered as a single unit where spatial variability is not included in inputs and hence the outputs are generated without considering the spatial processes (Sitterson et al., 2017).

Rainfall-runoff models are classified into three categories according to (Wheater, Jakeman, & Beven, 1993):

- Physically based or mechanistic models (white-box models) based on the mathematical equations of the underlying physical processes and discretised physical equations of motions.

- Metric or empirical models (black-box models) which derive information from observational data (e.g. rainfall, river flow, river height) without considering the physical and hydrological processes of the catchment but using the statistical properties of the long time series of rainfall and flow.
- Conceptual or parametric models (grey-box models). These models can be considered as a combination of a physically-based model (white box model), and an empirical models (black box model), relying on a simplified representation of the physical system, which can be calibrated using historical data (Leahy, Kiely, & Corcoran, 2008).

The two most important inputs required for all models are rainfall and catchment data. Along with these inputs, catchment characteristics such as soil properties, vegetation cover, catchment topography, soil moisture content and, characteristics of ground water aquifers may also be considered. The best model is the one which is parsimonious, that generates results as close as possible to the reality using the least number of parameters and the least model complexity.

In this chapter, we describe some key characteristic of some hydrological models and compare distributed models, semi distributed models, lumped models and neural network type models or statistical models.

2.2 Physically Based Models (White-Box Models)

2.2.1 GEOtop Model

GEOtop is a physically based, fully distributed hydrological model that analyses the complete water and energy budgets in a catchment. GEOtop is an open source software available for Linux, Mac and Windows platforms (Endrizzi, Dall'Amico, Gruber, & Rigon, 2017).

GEOtop uses two different inputs: topographical and meteorological data. The model can incorporate complex topography. It considers the catchment as a Digital Elevation Model (DEM) which allows modelling the incident radiation on the topographical surface, both shortwave (including shadowing) and longwave (sky view factor), (Endrizzi et al. 2017).

GEOtop is a spatially distributed model so it divides the catchment into cells or pixels. Figure 2.1 illustrates how the Dripsey catchment (a 15 km^2 agricultural catchment in County Cork, Ireland) is split up into cells. The size of each cell coincides with the pixel size of the DEM (*dX,dY*). The model

also provides each cell in catchments with the topographical characteristics of the basin (elevation, slope, aspect, shadow, sky view factor), (Endrizzi et al. 2017; Lewis, 2011)

In GEOtop it is possible to have the soil profile with different layers, each layer may have a different depth. In other words, the depths of layers can be regular or irregular but uniform for all the catchment. The centre of each cell (on the *X-Y* axis) and the centre of the layer (on the *X-Z* axis) represents the cell information (see Figure 2.2, Figure 2.3).

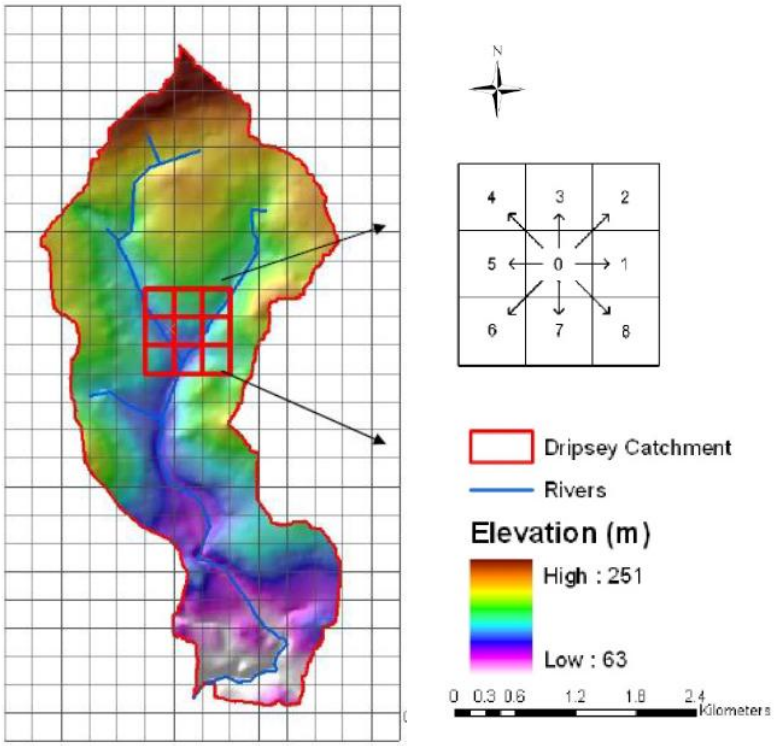


Figure 2.1: Dripsey catchment DEM and cell outline with D8 topology indicating surface flow direction (Lewis, 2011). In this 15 km^2catchment each cell had a size of 0.328 km * 0.328 km.

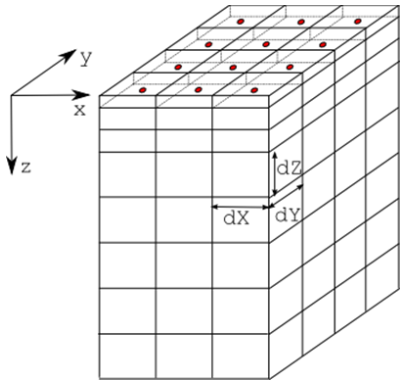


Figure 2.2: Three dimensional illustration of cells with different depth of layers. (Endrizzi *et al*. 2017)

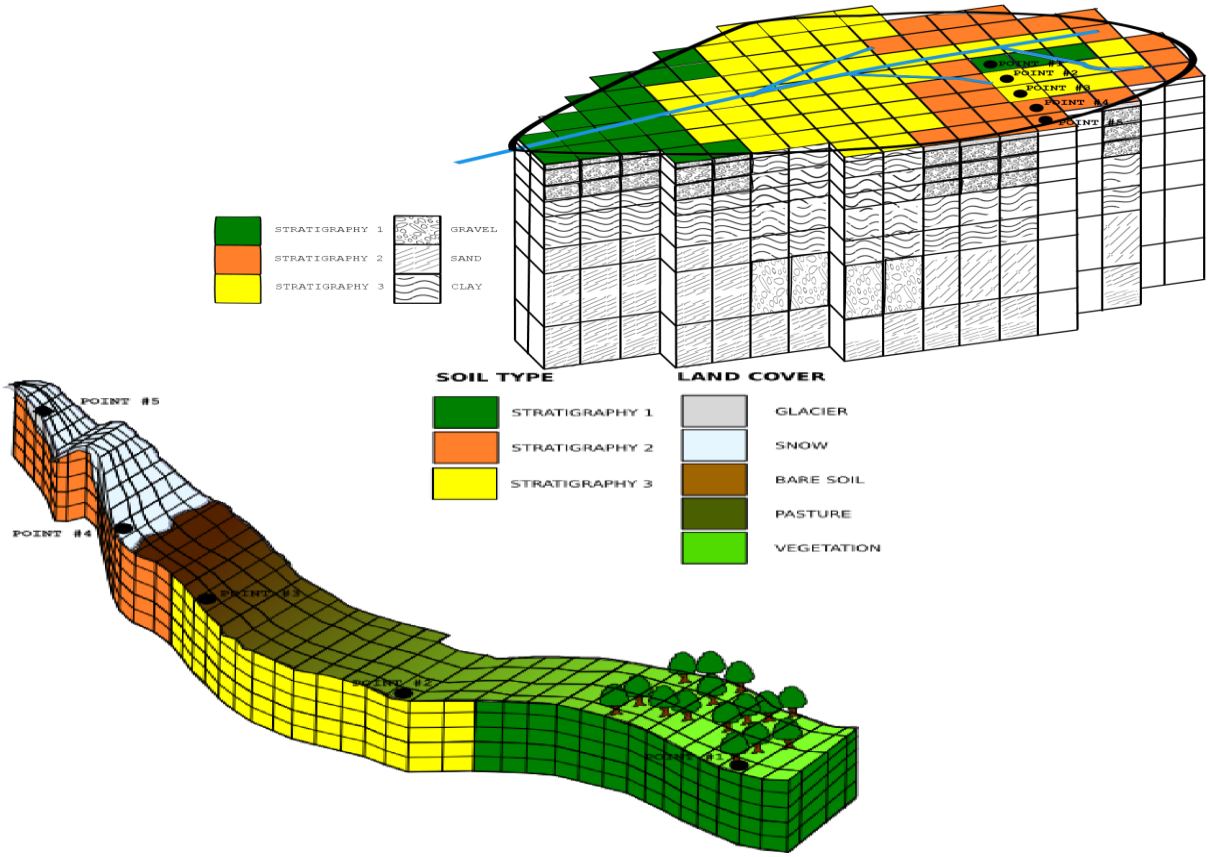


Figure 2.3: Land cover and Soil type map with multiple layers. (Endrizzi *et al*. 2017)

For every cell, the model solves both the energy and the water budget equations, divided into lateral and vertical flows. The flow of water in the unsaturated zone (below the surface but above the water table) is estimated by numerically solving Richard's equation, while the flow through the saturated zone (below the water table) is estimated from Darcy's law.

The general version of Richard's equation for three dimensional flow is:

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial x}\left[K(h)\frac{\partial h}{\partial x}\right] + \frac{\partial}{\partial y}\left[K(h)\frac{\partial h}{\partial y}\right] + \frac{\partial}{\partial z}\left[K(h)(\frac{\partial h}{\partial z} + 1)\right] - S_w \qquad \text{Eq 2.1}$$

where θ ($m^3 m^{-3}$ is the volumetric water content, *t (sec)* is the time, *h* (m) is the soil water matric head, *x*, *y*, and *z* are the spatial coordinates, S_w is a sink term which represents the volume of water removed per unit time from a unit volume of soil and $K(h)$ $(m.s^{-1})$ is the unsaturated hydraulic conductivity (Botros, Onsoy, Ginn, & Harter, 2012).

The Darcy's law equation is:

$$\frac{Q}{A} = K\,\frac{\Delta h}{\Delta l} \qquad \text{Eq 2.2}$$

where Q is the discharge in units of length cubed per unit time $(m^3.s^{-1})$, K is hydraulic conductivity in units of length per unit time $(m.s^{-1})$, $\frac{\Delta h}{\Delta l}$ = hydraulic gradient and A is the cross sectional area in units of length squared (m^2) (Brikowski, 2012) .

Surface flow might be generated when the rainfall intensity $(mm.hr^{-1})$ exceeds the infiltration capacity of the soil surface or when the top layer of soil becomes saturated (either from above or from below) and so any further precipitation on these soils results in surface flow. The surface flow is routed to the river channel using the D8 flow direction (see Figure 2.1). Imagine a cell with its 8 immediate neighbours, the flow direction from each cell to one of its adjacent or diagonal neighbours is calculated using the direction of steepest descent (Endrizzi et al. 2017; Lewis, 2011).

The distributed model GEOtop uses both topographical and meteorological data (Endrizzi et al .2017).

Topographical data includes:

- The land use i.e. vegetation, pasture, snow, glacier, forest, etc. This map is called land cover.
- The type of soil in terms of texture (gravel, sand, loam, clay: % of each) and bedrock depth, these data are used to estimate the hydraulic and thermal parameters of the soil. This map is usually called soil type.

Meteorological data may come from several stations. For each meteorological station, the following information is required:

- The meteorological station's unique number
- The coordinates (X, Y, Lat, Long)
- The elevation
- The Sky View Factor (SVF)*
- The standard time difference (of the time records with respect to Greenwich Meridian Time)
- The height of the wind speed and air temperature sensors.

*Sky View Factor (SVF) is defined as the ratio between radiation received by a planar surface and that from the entire hemispheric radiating environment (Svensson, 2004). In other words, the SVF is a measure of the openness of the sky to radiative transport in relation to a specific location. A value of 0 indicates complete obstruction and a value of 1 indicates complete open and unobstructed space, so SVF of locations with obstruction such as mountains, buildings and trees decrease proportionally, see Figure 2.4 (Zakšek et al, Oštir and Kokalj, 2011).

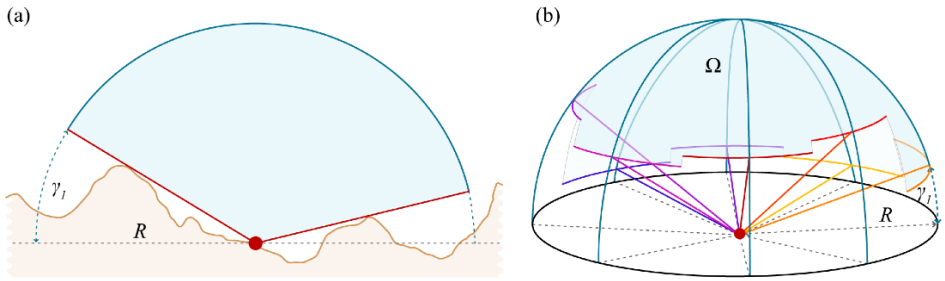


Figure 2.4 Two dimensional representation of SVF (a) Three dimensional representation of SVF (b) (Zakšek, Oštir, & Kokalj, 2011)

Table 2.1 shows the necessary, recommended and useful meteorological data for the model validation. To run GEOtop we need at least one weather station with hourly data, although more stations would be required as the catchment area increases.

Table 2.2 shows this same classification for distributed data.

We can summarise the main advantages and disadvantages of using GEOtop model as follows:

Advantages:

- Open source
- Provides a variety of outputs (time series and maps) based on user-defined requirements.
- Suitable for different types of catchments including mountainous catchments

Disadvantages:

- Complex installation
- Preparing the necessary required data is difficult and time consuming, as it needs a wide range of meteorological and topographical data. This may not always be available.
- Sensitive to some initial parameters. As GEOtop solves Richard's equation numerically, the model might show errors when the water balance does not converge.
- Requires another software or programming language for visualising the output e.g. ARCGIS or R or Matlab.
- The new model version is not compatible with the older version. The configuration files and input files in both versions are different, even the parameter names.

Table 2.1: Meteorological data classified as a function of their level of requirement for GEOtop model. (Endrizzi *et al*. 2017)

Necessary	Recommended	useful for validation
Air temperature (Ta) ℃	Longwave radiation (R ↓LW)	net radiation (Rn)
Relative humidity Rh or air water content or air vapor pressure or dew point. %	Net radiation (Rn)	soil heat flux (G)
Wind speed (v) $m.s^{-1}$	Diffuse shortwave radiation (R ↓SW)	sensible heat flux (H)
Shortwave radiation (R ↓SW)	Snow precipitation.	latent heat flux (ET)
Pressure (Pa) *pa*	Sky fraction covered by clouds.	
Precipitation (P) *mm*		

Table 2.2: Distributed data classified as a function of their level of requirement for GEOtop model. (Endrizzi *et al*. 2017)

Necessary	recommended	useful for validation
Digital Elevation Model (DEM)	Soil depth, how deep is the bedrock.	Brightness temperature.
Land use map of the basin. At least a classification with urban, forest, bare soil, agriculture, pasture.	Soil texture: fraction of rock, gravel, sand, loam, clay, organic matter, bulk density.	Surface Soil moisture.
	Surface roughness, including the height of obstacles on the surface (buildings, vegetation *)	
	Soil hydraulic conductivity $m.s^{-1}$	

* Information regarding vegetation is essentially of three types: vegetation density (fraction of vegetation cover and Leaf Area Index (LAI)- the ratio of total projected leaf area per unit ground area, and is widely used to characterize canopy light condition); seasonal variation of vegetation (Normalized Vegetation Deviation Index (NVDI)- a simple graphical indicator that can be used to analyse remote sensing measurements, typically, but not necessarily, from a space platform, and assess whether the target being observed contains live green vegetation or not); and vegetation height. These indices are often available with coarse resolution in global data sets from satellite products such as MODIS.
There are many standard MODIS data products such as:

- Level 1: raw radiance, calibrated radiances and geolocation fields
- Atmosphere products : cloud product, total precipitable water, atmospheric profile
- Land products : land surface temperature, land cover products, evapotranspiration
- Cryosphere products : snow cover, sea ice and ice surface temperature
- Ocean products: sea surface temperature, remote sensing reflectance, particulate organic carbon, etc.

2.2.2 tRIBS Model

The **T**IN based **R**eal-time **I**ntegrated **B**asin **S**imulator (tRIBS) is a physically based distributed model developed for real time, continuous hydrological forecasting at the Ralph M. Parsons Laboratory, Massachusetts Institute of Technology, (Vivoni & Ivanov, 2002). tRIBS is a model suitable for small to mid-size catchments (Wolock, 1995). tRIBS was applied in the Munster Blackwater catchment with total size of 1200 km^2 (Steinmann, 2005). The object-oriented programing language C++ was used to design and develop the model. Several modules were used to solve the different hydrological process equations. Table 2.3 lists these model processes. Figure 2.5 indicates the interaction between those different processes.

Table 2.3 Components of the tRIBS distribution hydrologic model (Vivoni & Ivanov, 2002)

Model Process	Description
Rainfall Interception	Canopy water balance model
Surface Energy Balance	Combination equation (lE), Gradient method (H) and Force restore equation (G)
Surface radiation Model	Short-wave and long-wave components accounting for terrain variability
Evapotranspiration	Bare soil evaporation, transpiration and evaporation from wet canopy
Infiltration	Kinematic approximation with capillarity effects; unsaturated, saturated and perched conditions; top and wetting infiltration fronts
Lateral Vadose Flow	Topography-driven lateral unsaturated and saturated vadose flow
Runoff Production	Infiltration-excess, saturation excess, perched return flow and groundwater exfiltration
Groundwater Flow	Two dimensional flow in multiple directions, dynamic water table
Overland Flow	Non-linear routing
Channel Flow	Kinematic wave hydraulic routing

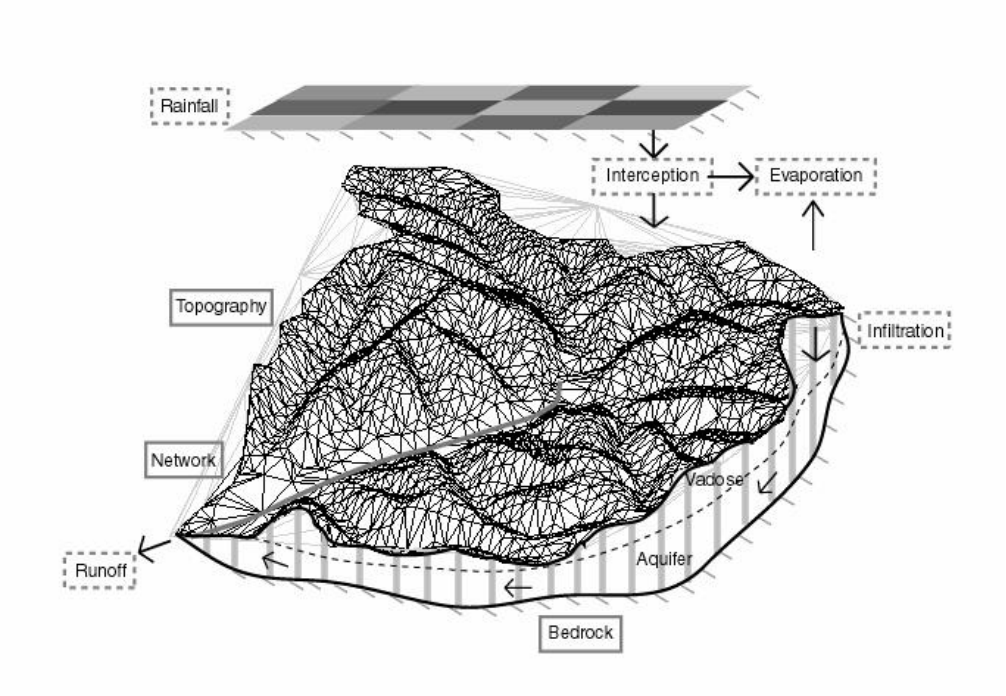


Figure 2.5 Schematic representation of the coupled hydrological processes in the tRIBS model. (Vivoni & Ivanov, 2002)

tRIBS uses different types of Geographical Information System (GIS) based input data such as the topography, soils and land use data to characterize the catchment, water table depth to initialise soil moisture, radar rainfall as precipitation input, and meteorological station data for modelling inter-storm periods. These inputs may have various types of data structures: grid data, TIN data, point data and text data. These data structures are described briefly below:

- Grid input: could include soil and land use, ground water table depth, depth to bedrock and radar rainfall grid.
- TIN input: topographical data are input to the model via different methods. One way is to generate a TIN mesh from ArcGIS Info and export it into a format that tRIBS can convert to a point file.
- Point data: useful to provide meteorological data from sets of weather stations or rain gauges data.
- Text data: various types of text files are used in the tRIBS model to specify its options, hydrologic parameters or control commands.

tRIBS creates different types of output. The main output is the outlet hydrograph, which is a riverflow time series at the outlet of the catchment at the time step defined in the main input file (Steinmann, 2005). According to (Steinmann, 2005) validation of the tRIBS model using calibrated parameters at

the Munster Blackwater for 2002 to 2005 has shown the timing and magnitude of the flood events were not accurate enough. The model could accurately predict the flood event only one out of five cases, the magnitude of the peak flows were underestimated between 17 % to 33 % for four cases. Inaccuracies of the model could be the results of unreliable calibration parameters and model inputs. We can summarise the main advantages and disadvantages of using tRIBS model as follows:

Advantages:

- Is created by a object-oriented C++ programming, take advantages and use various object oriented methods including inheritance, polymorphism and virtual functions.
- Provides a variety of outputs (time series and spatial outputs).
- Suitable for small to mid-size catchments.
- Provide source code to autorized developer to modify and update the code.
- Ability to use remote sensing data such as radar rainfall data.

Disadvantages:

- A research-based code does not provide as much error checking as could be possible, so the responsibility is placed on the user to provide the model with the appropriate inputs in the correct format.
- Preparing the necessary required data is difficult and time consuming, as it needs a wide range of meteorological and topographical data. This may not always be available.
- Requires another software or programming languages for visualising and analysing the output e.g. ARCGIS or Matlab.

2.3 Conceptual Models (Grey-Box Models)

2.3.1 Probability Distributed Model (PDM)

The Probability Distributed Model (PDM) is a conceptual rainfall-runoff model (Moore, 2007) suitable for a variety of catchments, and has minimal data and computational requirements – much less than the GEOtop model. PDM has a large number of model options; the “standard” model is suitable for most practical purposes. Rainfall and potential evaporation data are used as inputs to produce runoff at any point of the catchment, which is controlled by the absorption capacity (of the

canopy, surface and soil). PDM considers each point of a catchment as a single storage unit with a specific storage capacity (depth) that can be described by a Pareto distribution. The "standard" PDM consists of two linear surface storages and a non-linear groundwater storage (Figure 2.8).

2.3.1.1 Probability Density Function and Cumulative Distribution Function

In statistics, the distribution of a random variable can be discrete or continuous. For a discrete probability distribution, the probability that a random value *x* can take a specific value *p(x)* is *0 ≤ p(x) ≤ 1* and $\sum_i p(x_i)$ = 1.

For a continuous probability distribution, the probability is defined for an interval, not for single points so the probability of *p(X = x) = 0* and we can define a Probability Density Function (PDF) *f(x)* and the Cumulative Distribution Function(CDF) *F(x)* as follows:

$$f(x) = \lim_{\Delta \to 0} \frac{p(x < X \leq x + \Delta)}{\Delta} \qquad \text{Eq. 2.3}$$

$$p(x < X \leq x + \Delta) = F(x + \Delta) - F(x) \qquad \text{Eq. 2.4}$$

By using Eq. 2.3 and Eq. 2.4 we have:

$f(x) = \lim_{\Delta \to 0} \frac{F(x+\Delta) - F(x)}{\Delta}$ by using definition of derivative

$= \frac{dF(x)}{dx} = F'(x)$ if F(x) is differentiable at x

So the probability that *x* is in the interval *(a,b)* is: $F(b) - F(a) = \int_a^b f(x)dx$

Figure 2.6 Illustrates the PDF and CDF for the interval *(a,b)*.

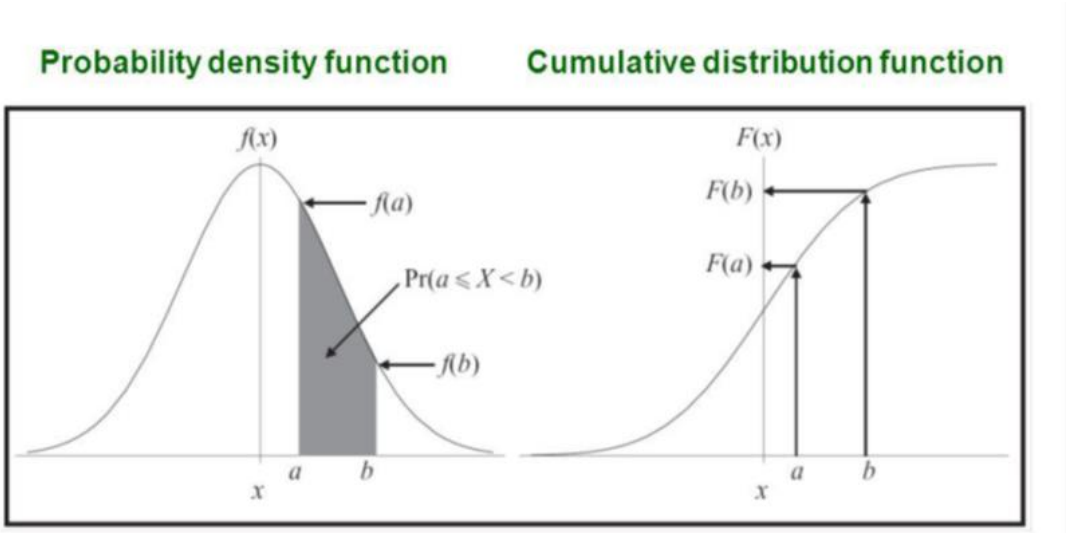


Figure 2.6 PDF & CDF for interval (a,b) (Probability density function, 2015)

2.3.1.2 PDM and Pareto Distribution

A Pareto distribution is a type of exponential distribution, usually skewed and heavy tailed which is used for social, scientific, geophysical, actuarial, and many other types of phenomena.

In PDM the Pareto distribution is used to describe the distribution of storage depth across the catchment which ranges from a minimum depth C_{min} to a maximum depth C_{max} with their CDF and PDF distribution of depth defined as:

$$F(C) = 1 - (1 - \frac{c}{C_{max}})^b \qquad C_{min} \leq C \leq C_{max} \qquad \text{Eq. 2.5}$$

$$f(c) = \frac{b}{C_{max}}(1 - \frac{c}{C_{max}})^{b-1} \qquad C_{min} \leq C \leq C_{max} \qquad \text{Eq. 2.6}$$

The parameter *b* controls the degree of spatial variability of storage depth over the catchment. Figure 2.7 illustrates the PDF for the Pareto distribution. When *0 < b < 1* there are more deep storage units (in the saturated zone), while with *b > 1* it is more likely there is more shallow storage (in the unsaturated zone) and when *b = 1* indicates a constant storage capacity.

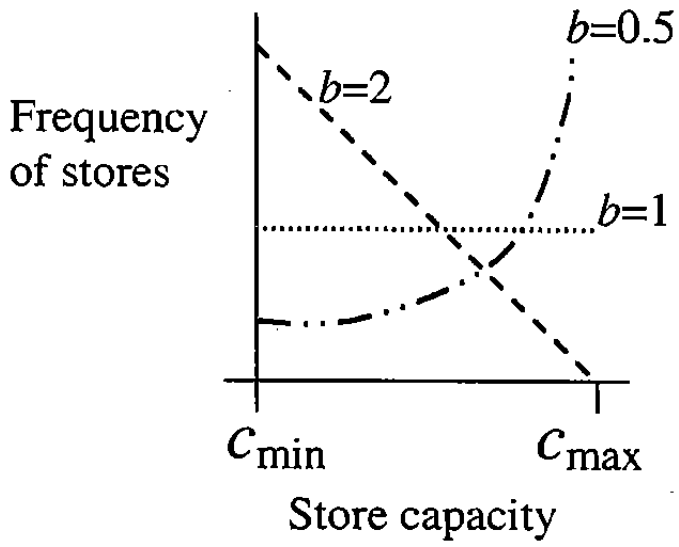


Figure 2.7 PDF for the Pareto distribution of different values of "*b*" (CEH, 2012).

PDM transfers rainfall to two parallel flows: direct runoff or fast flow (also known as surface runoff) and recharge or slow flow by recursive solution of the nonlinear storage equation known as the Horton-Izzard equation (Figure 2.8).

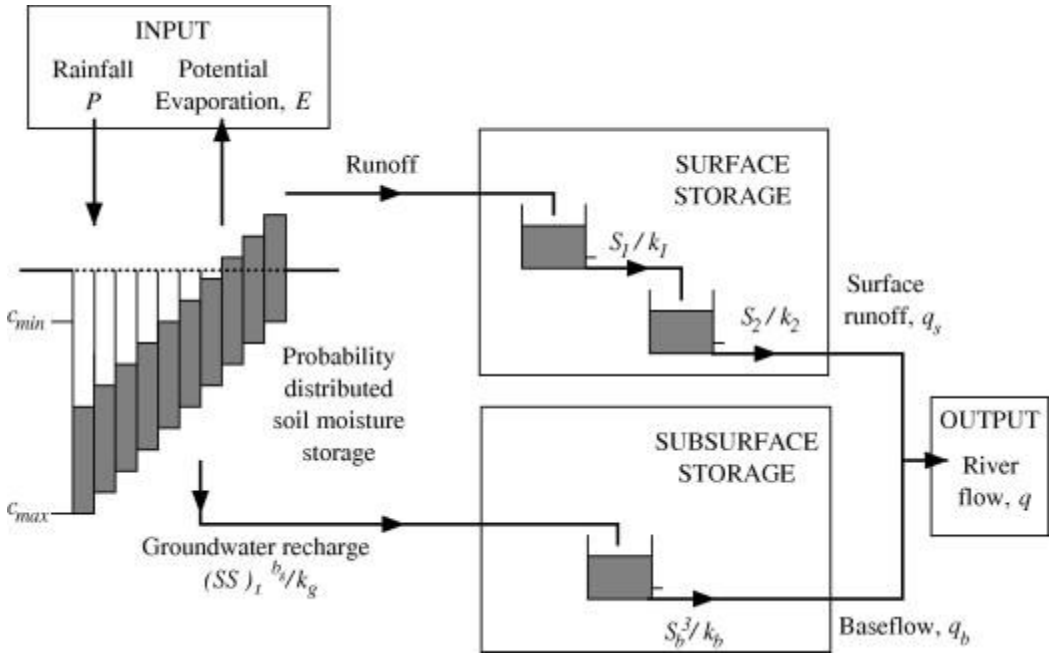


Figure 2.8 Structure of the PDM with a cascade of two linear storages for fast-runoff routing. (CEH, 2012)

The nonlinear storage model is specified by the Horton-Izzard equation:

$$\frac{dq}{dt} = a(u-q)q^b \qquad q> 0 , -\infty <b < 1 \qquad \text{Eq. 2.7}$$

where *q ≡ q(t)* is the rate of outflow and u ≡ u(t) is the rate of inflow to the store per unit area ($m.s^{-1}$), $a = mk^{\frac{1}{m}}$ and $b = \frac{m-1}{m}$, are two parameters related to those of the nonlinear storage equation $q = k\,S^m$. *S ≡ S(t)* is the volume of water held in the storage per unit area at time t. This ordinary differential equation can be solved exactly for any rational value of *m*. When *m = 1* (linear equation), *m = 2* (quadratic), exponential *b=1* and *m=3* (cubic) for a constant input u over the time interval *(t, t+T)*, (Moore & Bell, 2002).

The "standard" PDM surface storage consists of two linear storages with time constants k_1 and k_2 (in a linear storage, flow from the storage is equal to a constant multiplied by the water content of the storage). The moisture storage drains vertically to form the groundwater recharge as a nonlinear function of the storage content with time constant k_g. This recharge forms the input to the subsurface storage, which is usually taken to be a cubic storage with time constant k_b. Finally, the output from the subsurface storage (baseflow) is added to the surface runoff to form the total flow at the catchment outlet (Figure 2.8).

For the calibration, PDM generally uses a 15-minute time step over several years of historical record (using rainfall and potential evaporation as inputs and riverflow as output). However if these data are not available and we just have storm events in 15-minute time steps and a long period record of a daily time step, then it is good to start to run the model first with daily time steps to find the seasonal trends in the data and, then run the PDM with a combination of a continuous long period of daily time step and several 15-minute time step storm events (Moore, 2007).

After an initial manual model calibration, automatic optimization can be applied to minimise the objective function, which provides a quantitative measure of the difference between observed flows and computed or modelled flow. There are a number of objective functions in PDM to measure the model accuracy. The R^2efficiently and Root Mean Square Error (*RMSE)* are the most common objective functions used.

$$RMSE = \sqrt{(n^{-1}\sum {e_i}^2)} \qquad \text{where } e_i = Q_i - q_i \qquad \text{Eq. 2.8}$$

Q_i is the observed flow at time i, and q_i is the computed flow at time i.

$$R^2 = 1 - \frac{\sum {e_i}^2}{\sum (Q_i - \overline{Q})^2} \qquad \text{where } \overline{Q} \text{ is the mean of the observed flow} \qquad \text{Eq.2.9}$$

R^2 indicates what percentage of variability of observed flow can be accounted for the model.

The advantages and disadvantages of the PDM model can be summarize as below.

Advantages:

- Suitable for different types of catchment
- Large number of model options
- Minimal data and computational requirements
- Easy to install and run
- Automatically creates plots

Disadvantages:

- Initialization of some of the parameters based on trial and error
- Less outputs than GEOtop

2.3.2 TopModel

TopModel is a semi distributed conceptual rainfall runoff model which uses the topographic information of the catchment area as a DEM and soil transmissivity to generate the runoff at the catchment outlet based on the theory of hydrological similarity between points in a catchment based on a measure called the "topographic index" – *TI (Eq 2.11).* The simplicity of the model comes from the use of this index introduced by Beven (1986), which is used as an index of hydrological similarity (Tarboton, 2003). *TI* is the ratio of drained area per unit contour length to the local slope angle of the location, and represents the relation between the tendency of water to accumulate at any point in the catchment, and the tendency for gravitational forces to move that water downslope.

The model predicts the saturation excess and infiltration excess runoff (surface runoff and subsurface stormflow) (Lin, Zhang, & Chen, 2010). TopModel can be considered as a physically based model as its parameters can be physically or theoretically measured (e.g. hydraulic conductivity, porosity, soil depth, depth of water table). It can be used in single or multiple subcatchments. It can be applied to different types of catchment. For instance, it has been successfully used in humid temperate (Beven, 1993), drier Mediterranean regimes (Durand, Robson, & Neal, 1992), small humid tropical (1.36 km^2 in Ivory Coast (Quinn, Beven, & Chevallier, 1991)) and forested headwater catchment (Moličova, Grimaldi, Bonell, & Hubert, 1997).

TopModel uses four main assumptions:

1- Dynamics of the saturated zone can be approximated by a successive steady state representation.
2- The recharge rate (r [m/hr]) entering the water table is spatially homogeneous.
3- The effective hydraulic gradient of the saturated zone can be approximated by the local surface slope.

4- The effective down slope transmissivity T of a soil profile at a point is a function of the soil moisture deficit at that point. The saturated hydraulic conductivity decreases exponentially with depth.

By considering assumption (1) and (2), the down slope subsurface flow rate per unit contour width 'q' is: (Figure 2.9)

$$q = r\,a \qquad \text{Eq. 2.10}$$

where 'r' is the recharge rate ($m.hr^{-1}$) and 'a' is the drained area per unit contour length.

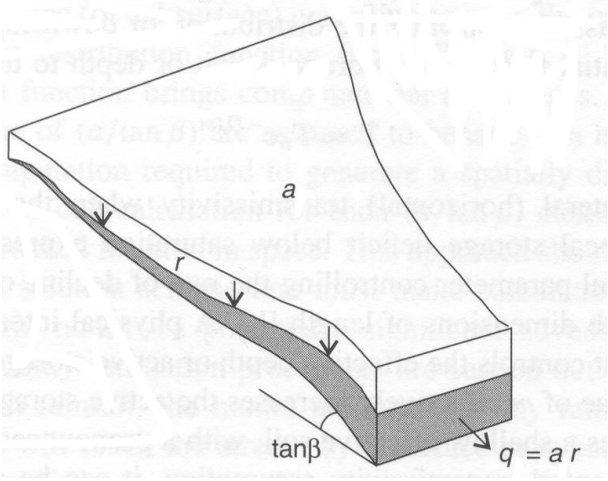


Figure 2.9 Definition of the upslope area draining through a point within a catchment (Tarboton, 2003).

The topographic wetness index λ is defined as:

$$\lambda = \ln\left(\frac{a}{tan\,\beta}\right) \qquad \text{Eq. 2.11}$$

Where 'a' is drained area per unit contour length and '$tan\,\beta$' is the slope of the ground surface at the location. Equation 2.11 indicates that areas with a high value of the topographic wetness index λ have large contributing area and low slopes, and are more likely to be saturated. These locations can be found along side streams. Areas with low values of λ have small contributing areas and high slopes, and so they are unlikely to be saturated. These latter areas are typically found at the top of hillslopes. Sometimes λ is referred to as a "similarity index" since it is used to group hydrologically similar areas in the catchment (Figure 2.10).

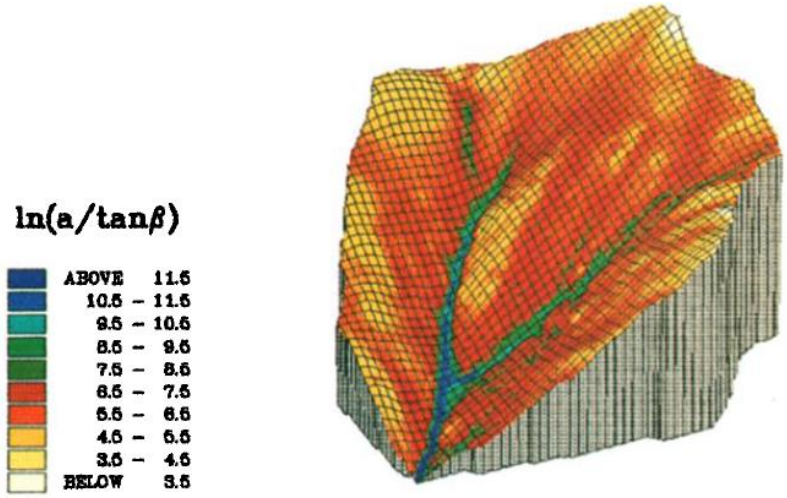


Figure 2.10 The topographic wetness index λ as a similarity index is used to group hydrologically similar areas. (Ambroise, Beven, & Freer, 1996)

TopModel computes the storage deficit or water table depth at any location. The storage deficit value 'D' is a function of the topographic index as:

$$D = -mln(r) + mln(T_0) - m\lambda \qquad \text{Eq. 2.12}$$

where $T_0 = \frac{k_0}{f}$, $m = \frac{\theta}{f}$ and $r = \frac{q}{a}$

'T_0' is transmissivity of the soil, 'k_0' is hydraulic conductivity at surface, 'θ' is an effective porosity and 'f' is a parameter that describes how hydraulic conductivity decreases with depth (Tarboton, 2003).

The model uses two different flow direction algorithms, D8 and Dinf. Dinf is a slower algorithm that generates more accurate results, The output is in the form of area maps or simulated hydrographs (Nourani, Roughani, & Gebremichael, 2011).

Previous studies have shown that DEM resolution has an impact on the topographic index. The mean of the $\ln\left(\frac{a}{tan\,\beta}\right)$ distribution increases as DEM resolution become coarser (Lin et al., 2010).

TopModel was applied in the study of the Dripsey River catchment located near the village of Donoughmore in rural Co.Cork, Ireland receiving an average of approximately 1400mm of rainfall per year (Scanlon, Kiely, & Xie, 2004). DEM data at a 10m resolution was used to calculate the distributed value of *TI* to find the effect of catchment topography on the phosphorus transport for three nested catchments. The results shows the mean, variance and skewness of *TI* distribution increase with increasing catchment size. Furthermore according to (Wolock, 1995), catchments with higher mean and variance of *TI* have higher percentage of surface runoff in the total streamflow. Therefore, it was expected for the catchment with the highest mean and variance in *TI* would have

the highest levels of phosphorus transport for a larger proportion of the catchment area, but the subcatchment, which has the highest mean, and variance in *TI* doesn't show the expected result, so the study suggested to examine another topographic descriptor length-slope factor called the LS-factor as an index of erosion to describe the phosphorus transport. The LS factor can be useful for identifying catchment areas that have the potential for erosion (Scanlon et al., 2004).
We can summarise the main advantages and disadvantages of TopModel as follows:

Advantages:

- Simple structure model and small number of parameters.
- Fast computational time.
- Suitable for single or multiple subcatchments and it can be applied to different types of catchments.

Disadvantages:

- No standard version available.
- Should met the four main assumptions.

2.4 Metric Models (Black-box models)

Metric models are mathematical and statistical methods that explain relationships between input data and output data or find patterns in data without considering the underlying physical process, catchment characteristics or physical parameters. Metric models require less input data than physical and conceptual models and so, this makes them generally fast to run and much faster to develop and useful for real-time forecasting and climate change attribution but as they are static, i.e. they cannot take change into account, e.g. changes in land use, so the model should frequently be updated or trained with new available data. Although black-box models do not require any understanding of physical and hydrological processes of the underlying system, some physical and hydrological knowledge would be useful to choose the most appropriate input data and output data to forecast (Ardabili, Mosavi, Dehghani, & Varkonyi-koczy, 2019; Napolitano, 2011).

Artificial Neural Networks (ANN) and K-Nearest Neighbours (KNN) are strongly nonlinear statistical models, which use machine learning techniques for classification.

A Classification algorithm is one type of black-box model which analyses and classifies historical and current data to predict future data trends. Firstly, the classification model trains the model, which means applying a learning algorithm to the training dataset to find the classification rules and then

applying these rules to a test data in order to assess the accuracy of the algorithm. The accuracy of the model is a percentage of samples or instances that are correctly classified (Gupta & Ghose, 2015).

2.4.1 K-Nearest Neighbours Model (KNN)

The KNN algorithm classifies the new object based on its similarity to the *K* nearest objects or its neighbours. Similarity between two objects defined by a mathematical function, usually the Euclidean Distance is used. Selecting the appropriate number of neighbours *K* is very important. If *K* is too small the effect of noise in the training dataset makes the KNN method not accurate. Selecting a large value for *K* reduces the effect of noise in the data but makes the boundaries between classes less distinct, and also make the algorithm very slow. There are many techniques for selecting a *K* value. Use the square root of the number of classified instances and N-fold cross validation for various values of *K* (see Figure2.11).

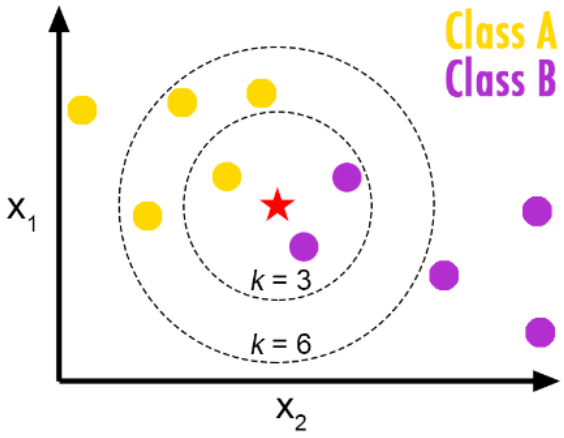


Figure 2.11 KNN Classification with different neighbours (Burton DeWilde, 2012)

KNN can be applied for hydrological modelling: for example for predicting rainfall or flood flows based on a long term rainfall dataset. The prediction of the next rainfall would be the average of the rainfall values of the *K* historical nearest neighbours (Gupta & Ghose, 2015). An improved KNN model with perturbation of the historic data has been applied to the Upper Thames River Basin (UTRB) in the Canadian province of Ontario to investigate the potential impact of several plausible climate change scenarios to extreme events, including both floods and drought (Sharif & Burn, 2006).

2.4.2 Artificial Neural Network (ANN)

ANNs are one of the main machine learning techniques that try to mimic the brain learning process, and were introduced by McCulloch and Pitts (1943). The ANN model gains its knowledge by detecting patterns and relationships in a dataset and learns through experience, not by programming. It also improves its knowledge by means of a backpropagation algorithm, a recursive algorithm which minimize the differences of output and computed output in two phases: a forward pass which, the processing of information occurs from the input layer to the output layer; and a backward pass, when the error from the output layer is propagated back to the input layer and the interconnections are modified (Govindaraju, 2000).

Since the early nineties, ANNs have been successfully used in hydrology-related areas such as rainfall-runoff modelling and forecasting, ground water modelling, water quality, water management policy and precipitation forecasting (Govindaraju, 2000).

ANNs have the following characteristics: (Govindaraju, 2000)

- Information processing occurs at many single elements called nodes, units, cells, or neurons.
- Signals pass between nodes through connection links.
- Each connection link has an associated weight that represents its connection strength.
- Each node typically applies a nonlinear transformation called an activation function to its net input in order to determine its output signal.

Each ANN has input and output layers and might have a number of layers between them called hidden layers. The number of hidden layers and the number of nodes in each hidden layer are usually obtained by trial-and-error. Inputs are multiplied by weights to increase or decrease emphasis on particular input, then the input function sums these weighted inputs, a mathematical function called an activation function is then applied to this value to generate the output which could be the input for the next layer (if there is more than one hidden layer) (see Figure 2.12 and Figure 2.13).

ANNs can be categorized based on the direction of information flow and processing. In a feedforward network, starting from a first input layer and ending at the final output layer. Information passes from the input to the output side. In a recurrent ANN, information flows through the nodes in both directions, from the input to the output side and vice versa.

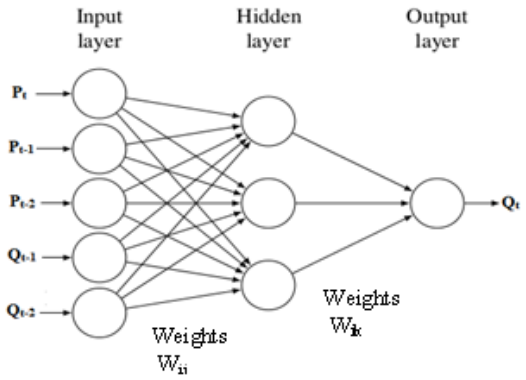


Figure 2.12: Neural Network nodes & layers (Govindaraju, 2000)

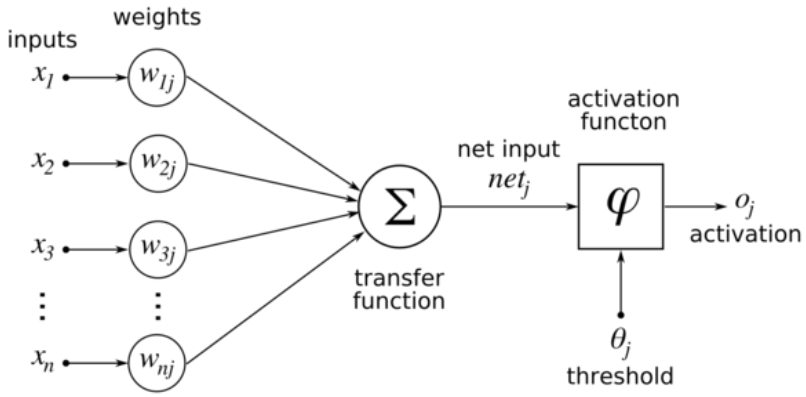


Figure 2.13: Neural Network process (Castrounis, 2016).

A study was carried out on the Munster Blackwater catchment in the southwest of Ireland in order to predict the river level as part of a project for flood forecasting (Leahy et al., 2008). The optimal ANN combined with simulated annealing and tabu search was used to minimise the value of the cost function and minimize the complexity of the ANN. To increase the accuracy of the ANN model we consider some physical parameters for selecting appropriate input (for example, upstream river levels are expected to be a good predictor of future river levels downstream in the same river system) (Govindaraju, 2000) to avoid the risk of having local minima. Too many inputs will result in a parameter space that is too large to be efficiently optimised during training. The properties of the ANN's structure also must be correctly chosen, the ratio of the number of training pairs to the number of neurons to be trained should be greater than 30. This study showed that the current value of river levels at the flood locations downstream and two upstream locations are better predictors of levels 5 hours ahead among other possible inputs such as mean sea level pressure (SLP) and the change in SLP over the previous 24 hours (Leahy et al., 2008).

The advantages and disadvantages of the Metric models can be summarize as below.

Advantages:

- Require less input data type than physical models.
- Generally fast to run and much faster to develop so useful for real-time forecasting.

Disadvantages:

- Should frequently be updated or trained with new available data.
- Require long term time series.

2.5 Summary

A large number of hydrological models with different characteristic and properties have been developed for different type of catchments, making it difficult to choose the most suitable model. For this reason several hydrological models from four main categories: lumped, semi-distributed, fully distributed and artificial neural networks have been reviewed above.

Generally lumped model parameters do not vary spatially within the catchment and the runoff is calculated just at the outlet. Parameters do not represent physical features of the hydrologic processes and the model is simple, easy to use, fast and needs minimal data. The PDM model is one such lumped model.

In a semi-distributed model, parameters are partially distributed over a number of small sub catchments. These models are more physically based than lumped models. They require more input data than lumped models but less than distributed models. TOPMODEL is one such semi-distributed model producing a hydrograph at the catchment and outlet.

In distributed models, parameters are fully distributed to square cells or to a triangulated irregular network (TIN). The model is physically based and requires large amounts of data. If there are accurate data available, the model generates results for each cell at any time but with high computational cost. GEOtop and tRIBS are fully physically distributed models (Xu, 2002).

The decision to use one model over the others is based on the answers required and the availability of data. If there are large datasets of topographical and meteorological data, the tendency is to use fully distributed physically based models. On the other hand, if there is limited or very little catchment distributed data, the trend is to use the lumped model approach. If there is a long meteorological datasets and limited catchment data, then one approach may be to use statistical or artificial neural network models.

In this research, the goal is to simulate the outflow hydrograph of storm events (rather than long-term flows) with limited catchment data but sufficient meteorological data, so the option is to look at either lumped or statistical models.

From the above review and experience, it was decided to use the PDM model in this research.

Chapter 3 Data Sources And Descriptions

3.1 Site Description

3.1.1 Catchment Location

The study area is the Munster Blackwater catchment, which is located in the south west of Ireland. It has a catchment area of approximately 3324km^2, which is 4% of the total area of Ireland (Corcoran, 2004). The Munster Blackwater catchment is a relatively broad and flat area which is surrounded by mountains: the ranges of Knockmealdown, Kilworth, Galtee, Ballyhoura and Mullaghareirk in the North and the Boggeragh range in the South (Figure 3.1). The source of Blackwater river is in the Mullaghareirk mountain near Ballydesmond in County Cork. The river flows from west to east. It flows in a southerly direction through Ballydesmond, then turns eastwards at Rathmore, after passing through the urban centre of Mallow and Femoy. It then enters County Waterford where it flows through Lismore and eventually turning south at Cappoquin and enters the sea at Youghal (Nicholson, 2012).

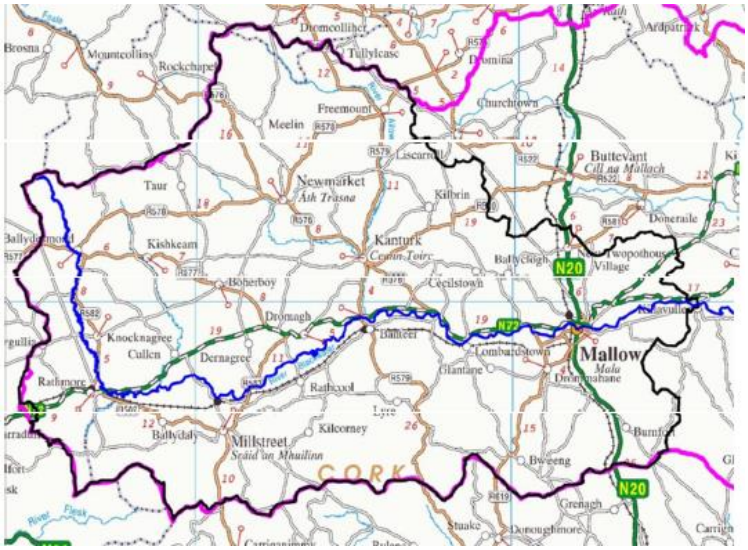


Figure 3.1 Mainstream of the Munster Blackwater from its source at Ballydesmond on the West to Killavullen (downstream of Mallow), (Nicholson, 2012).

The Blackwater catchment can be considered as four subcatchments. The four subcatchments considered are: Duarrigle, Dromcummer, Mallow Rail BR and Killavullen based on the availability and the quality of their data (Figure 3.2). Table 3.1 gives the area of each subcatchment calculated by ArcGIS.

Table 3.1 Area of the Blackwater catchment and its subcatchments

Station No	Subcatchment	Area(km^2)	Easting	Northing
18050	Duarrigle	247.93	124929.00	94373.00
18048	Dromcummer	862.50	139796.00	99320.00
18055	Mallow Rail BR(opw)	1168.65	155078.00	97842.00
18003	Killavullen	1248.32	164710.00	99738.00

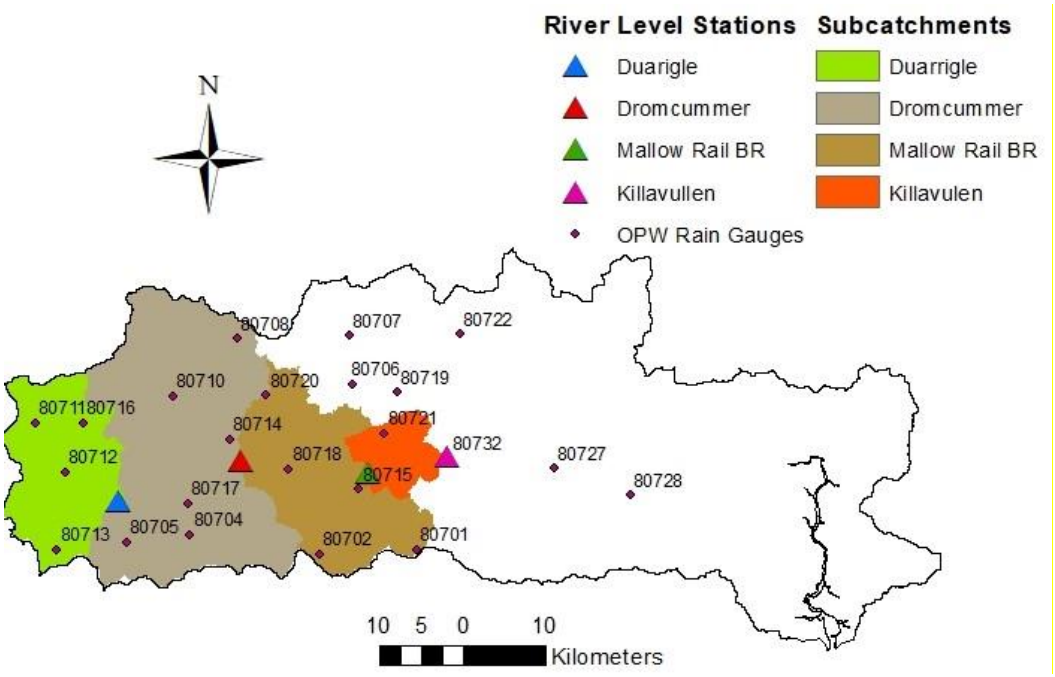


Figure 3.2 Munster Blackwater catchment and its four nested subcatchments.
See Table 3.2 for details of each numbered gauge

3.1.2 Topography and Digital Elevation Model (DEM)

The Open Topography Facility (Nandigam & Arrowsmith, 2018) provides online access to high-resolution topography data and related processing tools through cyberinfrastructure developed at the San Diego Supercomputer Center at University of California, San Diego (Nandigam & Arrowsmith, 2018)

The Digital Elevation Model (DEM) can be represented as a raster (a grid of squares or pixels). The DEM of the Munster Blackwater catchment was downloaded from Open Topography (Nandigam & Arrowsmith, 2018). It is a 30-m raster resolution DEM, which means that each pixel is approximately 30 meter wide. Table 3.2 shows the location and estimation of rain gauges elevation by using the DEM of the Blackwater catchment from Open Topography in ArcGIS10. Figure 3.3 and Figure 3.4

indicate the rain gauge 80713 (Duhallow Way – Reservoir) with an elevation of 293 m is the highest rain gauge located in the Duarrigle subcatchment and the rain gauge 80731 (Cappoquin) is the lowest rain gauge with 15 m height above sea level in the East of the Munster Blackwater catchment. Figure 3.5 illustrates the Blackwater catchment and drainage network delineation based on location of four river level stations and the DEM of Blackwater catchment using the Hydrology tools in ARCGIS. The subcatchments were then obtained from the derived drainage network (see Figure3.2).

Table 3.2 List of 32 OPW operated rain gauges on the Blackwater catchment

Station No	Station Name	Easting	Northing	Elevation(m)
80701	Bottle Hill - Pump House	161062.50	88465.60	214
80702	Bweeng - Pump House	149359.90	87868.70	218
80703	Lyre – Reservoir	140595.70	91935.70	277
80704	Kilcorney – Reservoir	133710.90	90320.80	216
80705	Millstreet – Reservoir	126049.30	89392.60	208
80706	Buttervant - Pump House	153378.70	108472.00	103
80707	Ballyhoura Way - Water intake works	152888.00	114398.40	87
80708	Freemount - Waste water treatment plant	139407.80	113998.00	138
80709	Ballinatona - Water treatment plant	129019.10	111884.80	204
80710	Newmarket – Reservoir	131637.10	107079.40	178
80711	Ballydesmond - Pump house	115026.40	103842.70	195
80712	Knocknagree - Old pump house	118576.60	97862.40	172
80713	Duhallow Way – Reservoir	117626.30	88508.80	293
80714	Kanturk - Waste water treatment plant	138451.30	101775.20	79
80715	Mallow - Pump house	154104.10	95790.90	67
80716	Kishkeam - Waste water treatment	120785.10	103812.30	199
80717	Rathcoole - Waste water treatment plant	133406.00	94119.80	94
80718	Pallas - Old pump house	145572.50	98161.10	108
80719	Doneraile – Pumphouse	158665.00	107508.00	72
80720	Kilbrin – Resevoir	142907.00	107149.00	190
80721	Two Pot House – Resevoir	157150.00	102581.00	122
80722	Ballygugroe – Landfill	166243.00	114549.00	220
80723	Kildorrey - Sewage Works	171727.00	110626.00	77
80724	Mitchelstown - Water Treatment Plant	180920.00	113349.00	78
80725	Castlecooke – Pumphouse	187734.00	104662.00	105
80726	Bartlemy – Pumphouse	181733.00	88500.00	133
80727	Fermoy – Pumphouse	177618.00	98283.00	41
80728	Coole – Pumphouse	186894.00	95063.00	77
80729	Tallow - Resovoir/Pumphouse	200822.00	92213.00	81
80730	Lismore - Resevoir/Pumphouse	206103.00	101402.00	176
80731	Cappoquin – Cappoquin	213336.00	97704.00	15
80732	Killavullen - Water Treatment Plant	164898.00	99446.00	43

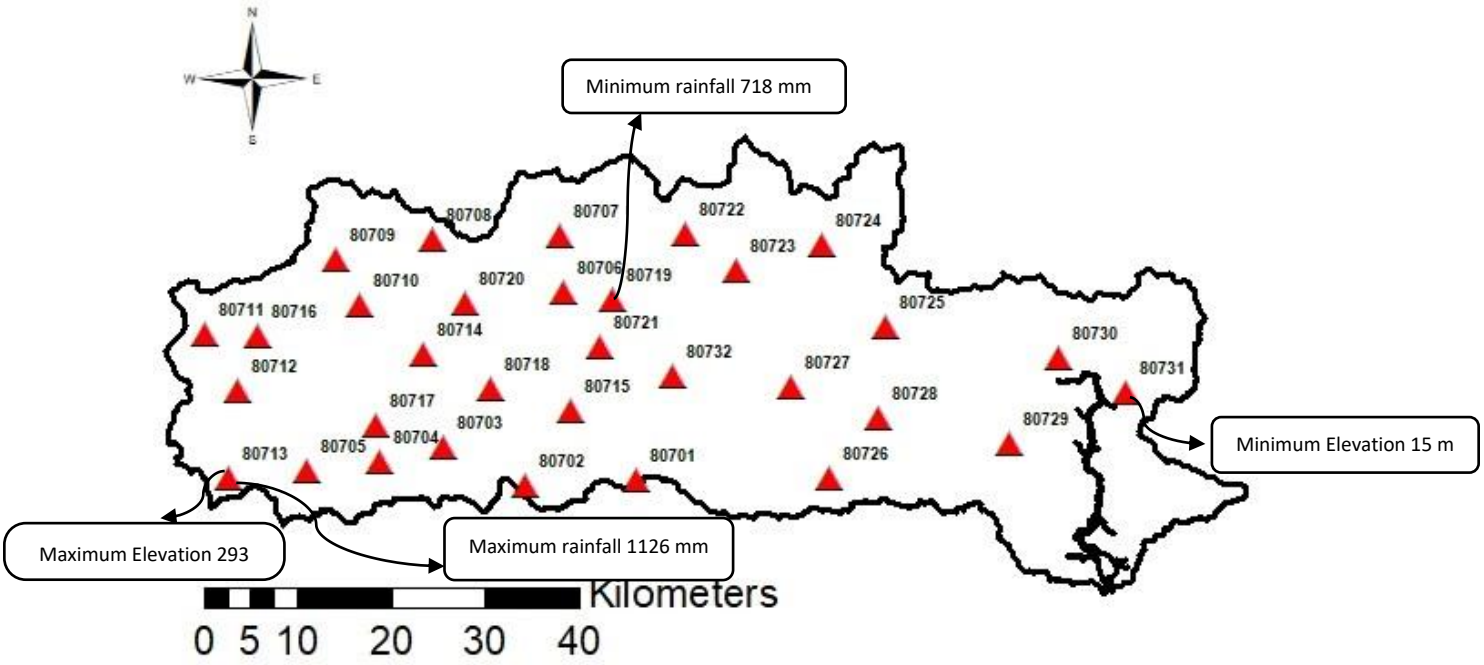


Figure 3.3 Location of 32 OPW rain gauges across the Munster Blackwater Catchment for 2010

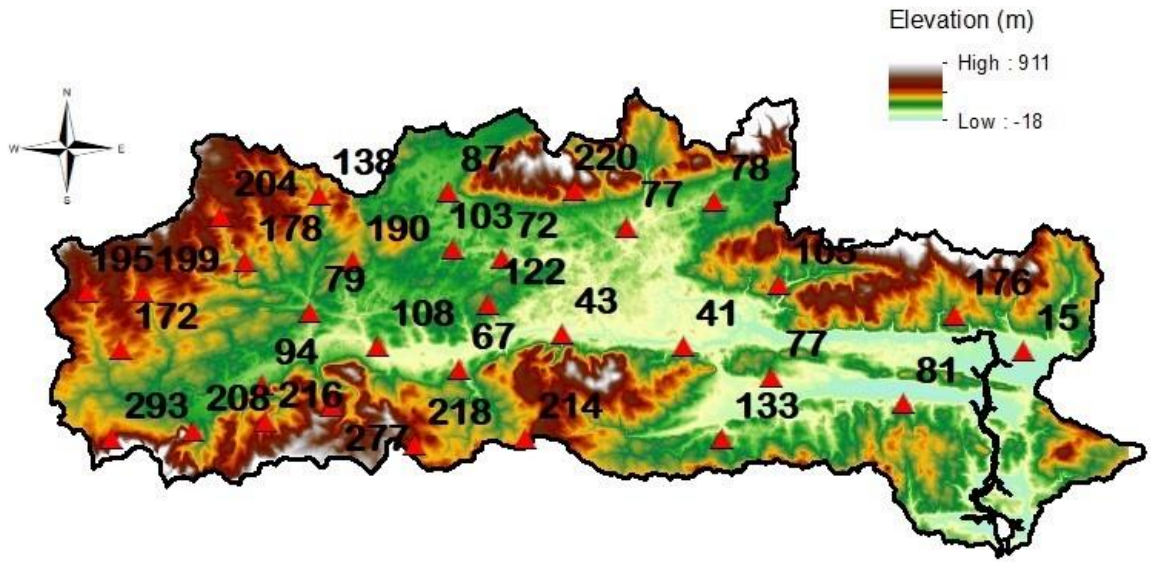


Figure 3.4 Munster Blackwater catchment topography and rain gauges elevation across the full Blackwater catchment.

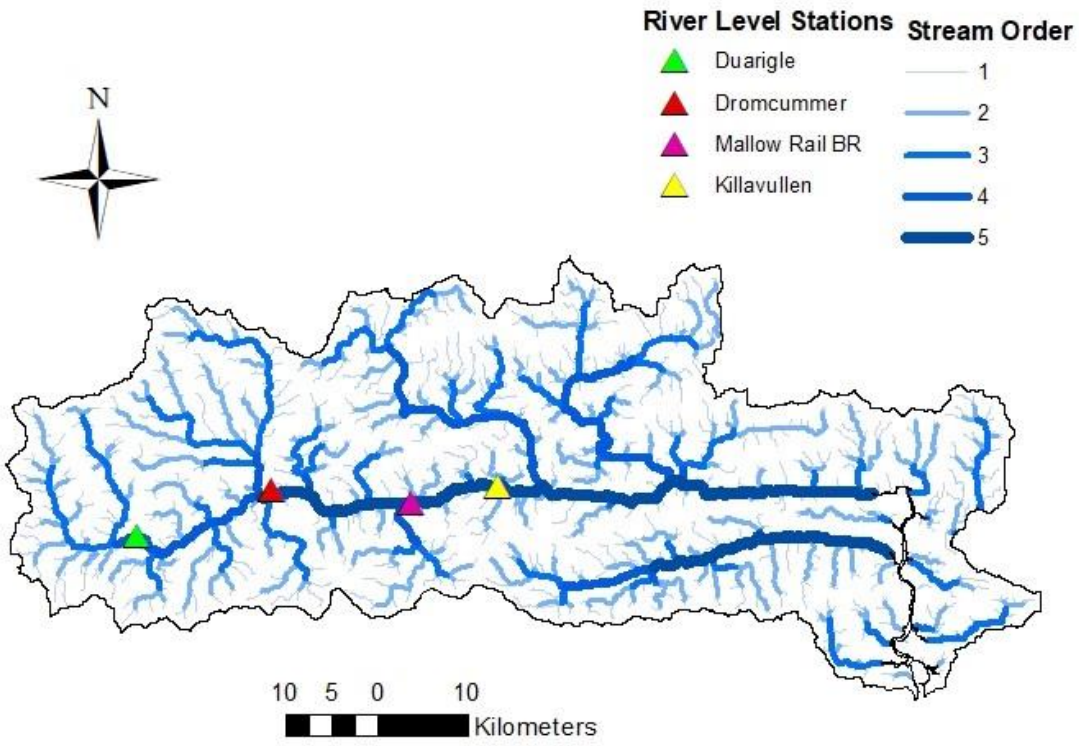


Figure 3.5 Munster Blackwater catchment topography and drainage network with four river level stations

3.2 Data Collection

The first step in this research is to understand the existing data, how it was collected and what is the relation between different parts of data and how we can obtain more data from different sources to gain maximum information from all data. Precipitation is an important input variable for hydrological models. The 15-minute time step precipitation and river flow are the two main data sources needed for running the PDM model. In this section, we first look at precipitation data from OPW and MÉRA and then river water levels from the EPA and OPW at different stations located in each subcatchment of the Munster Blackwater catchment to derive the river flow using rating curves for these locations.

3.2.1 Office of Public Works (OPW) Precipitation Data

The key input data to be used in this project is 15-minute time step precipitation from 32 different rain gauges across the Blackwater catchment collected by OPW from 2007 to October 2017. As Figure 3.2 shows, the rain gauges are located throughout the catchment.

All rain gauges have gaps and outliers in their data (due to equipment failure or operator absence) which require to be quality controlled to create the most accurate estimation of rainfall.
The EPA Ballyguyroe Landfill Site Annual Environmental Report January 2013 – December 2013 (O' Brien, 2014) confirms this and emphasises that 'The OPW Hydrometric section, do not undertake any post processing of rainfall data. The data collected is only archived as it is received. Any rainfall data disseminated will be raw unchecked data and it is the responsibility of the end user to ensure its suitability.' According to this report, several issues can cause gauges to both over and under estimate rainfall events and the severity of this can change with wind direction and strength. A list of issues, which can affect OPW rain gauge sites are as follow:

- Incorrect distances to obstructions (e.g fencing, tanks and buildings)
- Incorrect mounting height (too low or too high, on building roofs)
- Over exposure (sites too exposed to elements, wind shear effects)
- Vegetation being closer than the desirable minimum distances
- Temporary malfunction of rain gauge unit
- Temporary malfunction of telemetry systems

This research examines a practical method to fill the data gaps and modify the outliers of each OPW rain gauge with its nearest neighbors. For each rain gauge, the nearest neighbor means the rain gauges with the most similar amount of precipitation. In chapter four, this method is described in more detail.

3.2.2 Method of Analysis of Precipitation Data

In order to prepare suitable data to run the hydrological model PDM, analysis of available data, in 15-minute time step precipitation from the 32 rain gauges operated by the OPW from 2010 to 2017 was carried out.

Four types of problems were encountered with these datasets:

1) Missing values: there are some periods where the date and times were recorded but, the precipitation was not recorded. The data entry is NULL (Table 3.3).
2) Unrecorded value: there are some periods where a date, time and precipitation were not recorded. We can consider this as a step change or jump in dates and times.
3) Outliers: there are some records where the date and time are correctly recorded but the amount of precipitation is not realistic. For example after some missing value, not recorded value or some period with zero precipitation there is a record with unrealistic extremely high precipitation. In this study, we assumed that precipitation of more than 15 mm in a 15-minute time step is not realistic (Table 3.4).

4) The long sequence of precipitation: there are some records, which were recorded correctly, even the values of precipitation are reasonable but there are long sequences of precipitation where the cumulation of these precipitation values are not realistic.

Table 3.3 Example of missing values of precipitation (NULL) & outliers in rain gauges data

Year	Month	Day	Hour	Minute	second	Rainfall
2012	11	15	22	0	0	0.0
2012	11	15	22	15	0	0.0
2012	11	15	22	30	0	*NA*
2012	11	15	22	45	0	*NA*
2012	11	15	23	0	0	*NA*
2012	11	15	23	15	0	*NA*
2012	11	15	23	30	0	*NA*
2012	11	15	23	45	0	*NA*
2012	11	16	0	0	0	*NA*
2012	11	16	0	15	0	145.0
2012	11	16	0	30	0	0.0

Table 3.4 provides a summary of these 32 rain gauges from 2010 to 2017. The last column (percentage) shows the percentage of all bad data values for each rain gauge. As previously noted unrecorded values are the most significant problem in each rain gauge. The rain gauges 80707 (Ballyhoura Way - Water intake works) with 6.8% of bad data has the best data quality and rain gauge 80731 (Cappoquin – Cappoquin) with 69.28% of bad data has the worst data quality. The three rain gauges (80731, 80730, 80729) have more than 60% bad data, data recorded until 2013. For the rest of this study the first 23 rain gauges will be used, which have less than 15% bad data and will try to clean these data by gap filling. Figure 3.6 shows the location of the rain gauges with more than 15% bad data quality and less than 15% bad data quality.

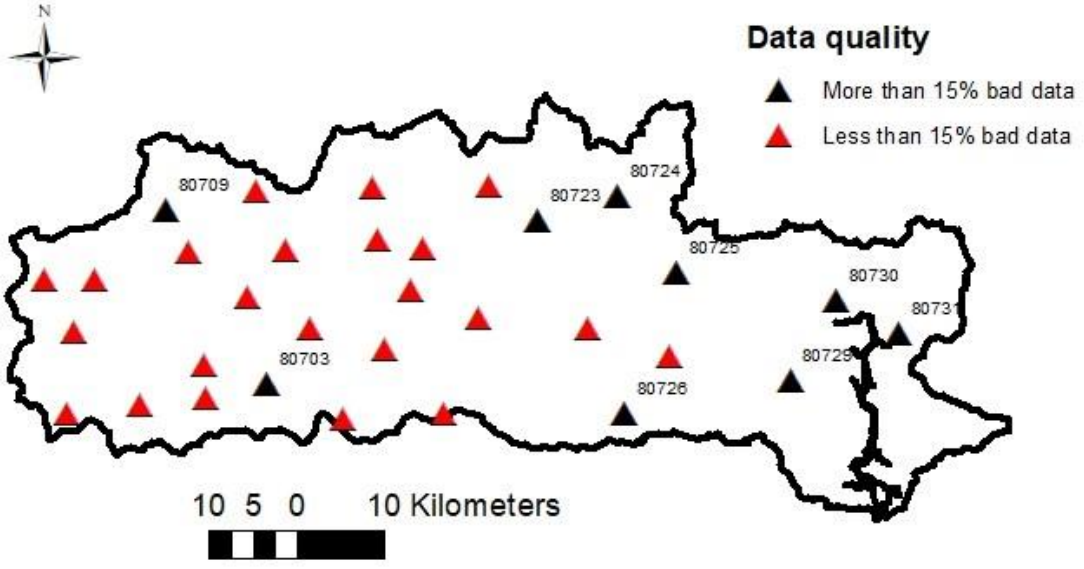


Figure 3.6 All OPW rain gauges and rain gauges with indication of data quality

Table 3.4 Summary of missing values and outliers of 32 rain gauges on the Blackwater catchment. Less than 15% bad data (Black) accepted, Greater than 15% bad data not accepted (Red)

Number of Missing and Outliers of Each Rain gauges				
Rain gauge No.	**Not Recorded**	**missing value**	**Rainfall > 15 mm**	**% of bad data**
80707	18914	117	59	6.97
80702	19273	196	0	7.11
80713	19518	151	345	7.31
80705	20744	62	9	7.6
80714	20944	223	87	7.76
80712	22833	289	25	8.45
80706	23386	232	0	8.62
80710	23593	334	0	8.74
80708	24310	215	8	8.96
80711	24170	542	251	9.12
80722	24761	89	172	9.14
80704	26403	190	3	9.71
80717	26735	101	145	9.85
80721	26495	268	213	9.85
80727	26963	113	3	9.89
80719	27232	256	17	10.04
80718	28153	135	1	10.33
80720	28359	934	2	10.7
80728	27039	3374	26	11.12
80701	24619	6389	1	11.32
80732	28996	2857	2	11.63
80716	31953	127	4	11.72
80715	32031	104	45	11.75
80726	51307	601	0	18.96
80723	56908	895	1	21.11
80703	67058	350	16	24.62
80724	68846	3749	1	26.51
80709	65408	20543	4	31.39
80725	88650	2367	16	33.24
80729	179265	88	2	65.5
80730	186513	438	8	68.27
80731	189360	334	3	69.27

3.2.3 MÉRA (Met Éireann Re-Analysis Climate Re-Analysis)

Climate reanalysis datasets are important sources of information for monitoring climate, validation and calibration of Numerical Weather Prediction (NWP) models. They are extensively used in diverse research areas such as food and agriculture, renewable energy, ecology, planning, economics and hydrology.

Reanalysis is a scientific method for developing a comprehensive record of how the weather and climate change over time, typically extending over several decades or longer, and covering the entire globe from the Earth's surface to well above the stratosphere. Reanalysis datasets are produced by data assimilation.

Data assimilation is a sequential time-stepping procedure: in each time step, it uses the past observational data from a wide variety of sources such as ships, satellites, ground stations, Radiosonde Observations (RAOBs) and radar to predict the current state. Then in a new time step, the observational data for the previous time step is compared with the predicted state and corrects its prediction. After correcting the previous state then all past data is used again to predict the new state and so on (Gleeson, Whelan, & Hanley, 2017) (Figure 3.7).

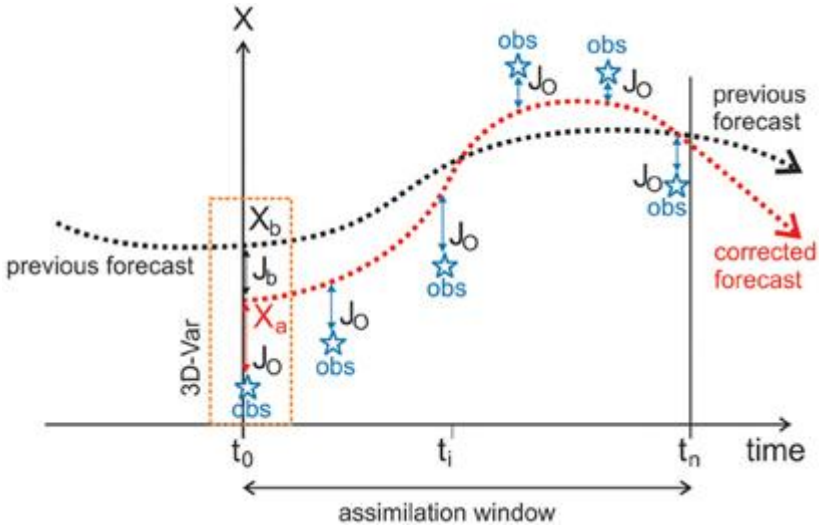


Figure 3.7 Correction of the model with the observations introduced by assimilation (Met, 2017).

Kalman filtering, smoothing, and variational analysis, Bayesian statistic and Markov chain Monte Carlo sampling are used in data assimilation (Wikle & Berliner, 2007).

Due to computational constraints, global reanalyses cannot be run at the very high resolutions required to resolve mesoscale (10 to 1000 *km*) processes. However, they are generally used to provide boundary conditions for regional reanalyses. The advantage of regional reanalyses is that

they can be run at high temporal and spatial resolution so that focus can be put on near surface variables, extremes and frequency distributions. Several regional reanalyses have already been produced. For example the EU European Reanalysis and Observations for Monitoring project (EURO4M: Klein Tank,2010) and the Irish reanalysis dataset, called MÉRA (Met Éireann ReAnalysis) .

The Irish Meteorological Services (Met Éireann) has carried out a 35-year very high resolution (2.5 km horizontal grid) regional climate reanalysis for Ireland using the ALADIN-HIRLAM numerical weather prediction system (www.hirlam.org). This reanalysis dataset spans over the period 1981 to 2015 and includes surface, near surface and atmospheric parameters. Conventional observations data (i.e. synoptic stations, ships, buoys, radiosonde ascents and aircraft) were assimilated. These observations are the same as those used by the European Centre for Medium Range Weather Forecasts, (ECMWF) ERA-Interim reanalysis. Locally available SYNOPs (surface synoptic observations) were used to fill gaps in the ERA-Interim SYNOP observation archive to ensure cycle continuity and successful data assimilation in each cycle. MÉRA uses a 3 hr forecast cycle with surface and upper-air data assimilation. Three-hour forecasts were produced for each cycle except the midnight cycle (00:00 UTC) when a 33 h forecast was produced. This provides a precipitation forecast each day that can be evaluated using locally available observations of daily accumulations of precipitation. One of the main advantages of very high-resolution reanalysis is to improve the precipitation forecasts. (Gleeson et al., 2017)

In this project, we use the MÉRA dataset to calculate the annual rainfall for specific locations on the Blackwater catchment using the GRIB format files from MÉRA.

3.2.4 MÉRA Precipitation Data

The Irish Meteorological Service, Met Éireann, has carried out a 35-year very high resolution (2.5 km horizontal grid) regional climate reanalysis over the period of 1981 to 2015, for Ireland called MÉRA. (Gleeson et al., 2017)

In this project, we use the precipitation from MÉRA to calculate the monthly and annual rainfall in each specific year. For example for 2010 for selected locations (nearest point to OPW rain gauges location) on the Blackwater catchment (Figure 3.8). MÉRA data is supplied in GRIB format. For access to these kind of files the GRIBapi package of Python in the Linux environment (Santoalla, 2016) is used to extract the hourly rainfall for each location and then calculate the monthly and annual rainfall. Monthly rainfall is used to classify the rain gauges based on their similarity of amount of precipitation. In Chapter 4 we will describe this in more detail.

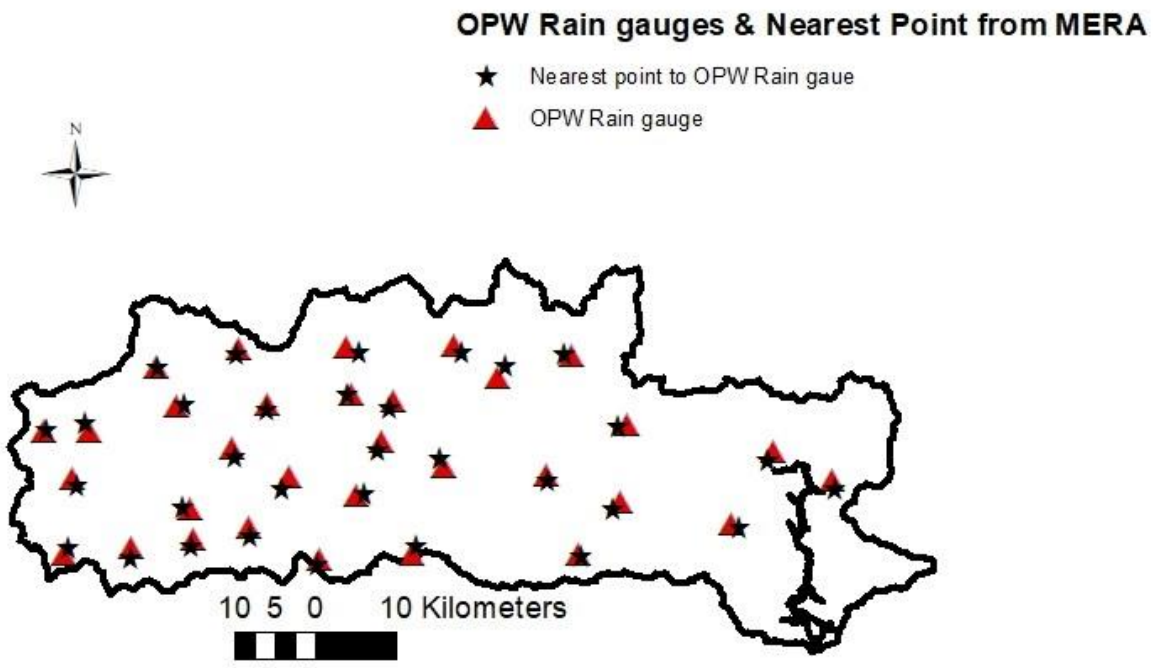


Figure 3.8 OPW Rain gauges locations and estimated nearest points to them from MÉRA

Figure 3.9 and Figure 3.10(a) and (b) illustrate the annual and monthly rainfall in mm for 2010 from MÉRA at the nearest locations to the 32 OPW rain gauges. As can be seen the rain gauge 80713 (Duhallow Way – Reservoir) with 1126.7 mm annual precipitation had the highest amount of annual precipitation while rain gauge 80719 (Doneraile - Pumphouse) with 718.5 mm annual precipitation had the minimum amount of annual precipitation. August with 23.04 mm mean precipitation is the month with minimum precipitation and has less variation of precipitation between rain gauges in the range of 13.06 mm to 32.64 mm. While January with 116.68 mm mean precipitation is the month of maximum precipitation in some rain gauges with more variation in the range of 71.98 mm to 116.41 mm across gauges. It is relevant to note that 2010 was a relatively dry year on the Munster Blackwater catchment.

The statistical summary of annual and monthly precipitation of all rain gauges is given in Table 3.5 and Table 3.6.

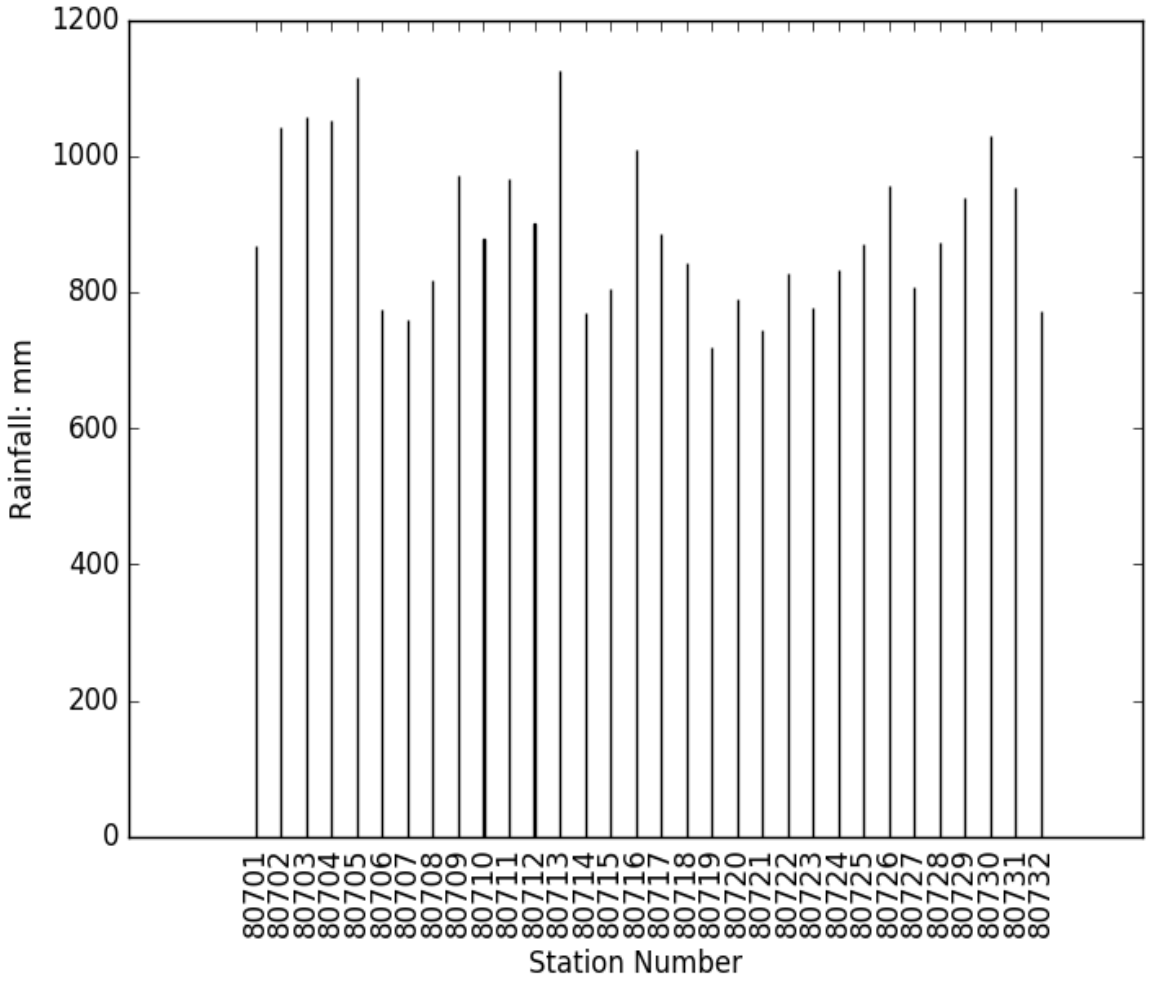


Figure 3.9 Annual rainfall for 2010 from MÉRA for 32 nearest points to 32 OPW rain gauges across the Munster Blackwater catchment

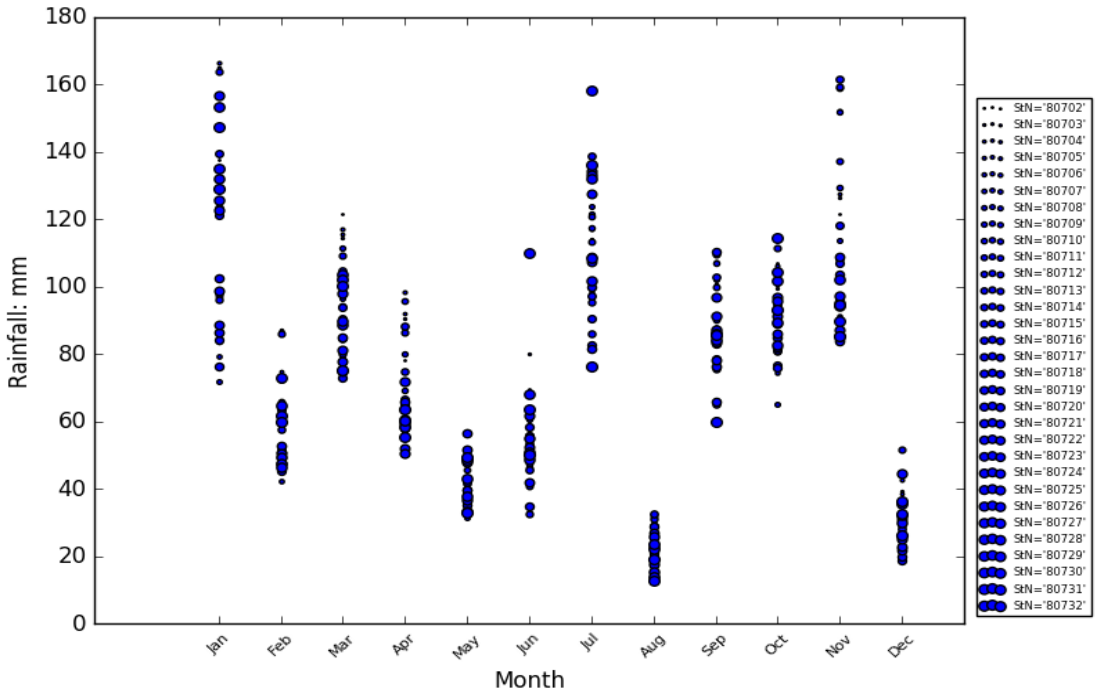


Figure 3.10 (a) Monthly rainfall for 2010 from MÉRA for 32 nearest points to 32 OPW rain gauges across the Blackwater catchment

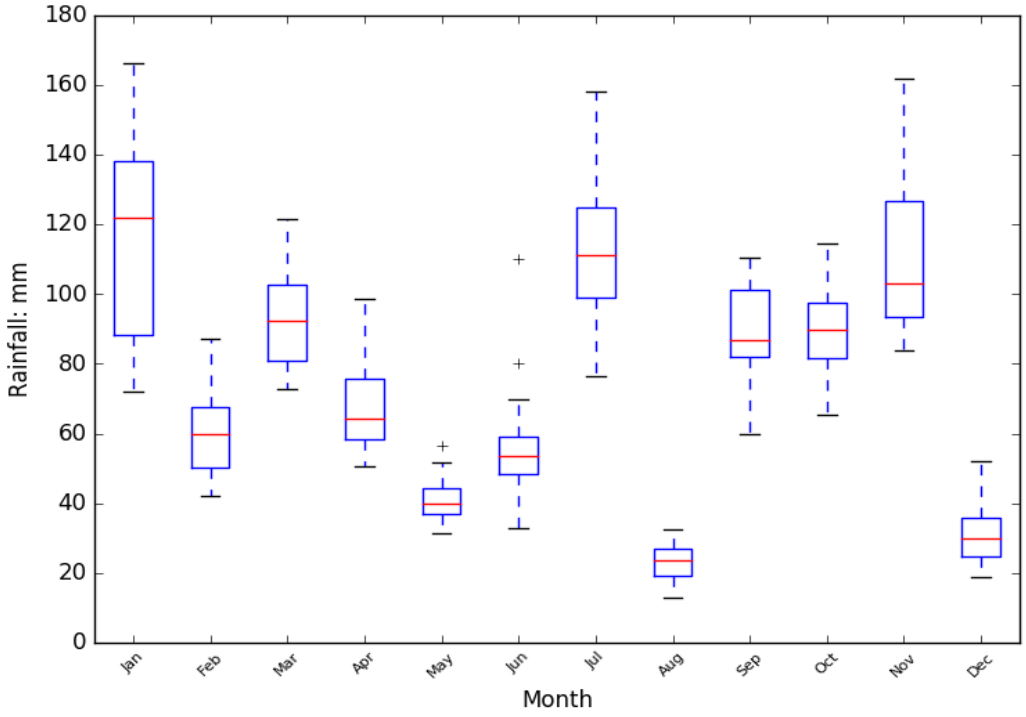


Figure 3.10 (b) Box-whisker plot of monthly rainfall for 2010 from MÉRA for 32 nearest points to 32 OPW rain gauges across the Blackwater catchment

Table 3.5 Statistical summary of annual precipitation for 2010 from MÉRA for 32 nearest points to 32 OPW rain gauges across the Blackwater catchment

MÉRA annual precipitation (mm) for 2010	
Min	718.6
Max	1126.8
Mean	892.3
Median	872.9
First quartile	801.9
Third quartile	968.5

The minimum annual precipitation was at (latitude 52.21236°N and Longitude -8.6138°W) corresponding to rain gauge 80719 which was 718.58 mm and the maximum annual precipitation was at (latitude 52.0499°N and Longitude -9.19639°W) corresponding to rain gauge 80713 which was 1126.77 mm.

Table 3.6 Statistical summary of monthly precipitation for 2010 from MÉRA for 32 nearest points to 32 OPW rain gauges across the Blackwater catchment

	MÉRA Monthly precipitation (mm) for 2010											
	Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
Min	71.97	42.24	72.92	50.73	31.39	32.73	76.33	13.06	59.95	65.35	83.96	18.95
Max	116.4	87.34	121.6	98.62	56.52	109.9	158.2	32.64	110.5	114.5	161.8	51.90
Mean	116.6	60.57	93.47	68.61	40.93	55.37	112.2	23.03	89.02	90.38	111.2	30.76
Median	121.9	59.7	92.36	64.35	39.76	53.65	111.1	23.45	86.75	89.61	103.0	30.02
First quartile	88.22	50.28	80.87	58.49	36.73	48.19	98.87	18.99	82.03	81.78	93.31	24.58
Third quartile	138.0	67.56	102.6	75.77	44.40	59.14	124.8	27.08	101.1	97.63	126.8	35.76

3.2.5 River Level Data

River level records are collected at a network of water level stations operated by local authorities in conjunction with the EPA and OPW throughout Ireland. In addition, the ESB, Teagasc and some private agencies collect river level data for a small number of rivers. Telemetric data loggers are used to digitally record the water level at 15-minute intervals using pressure transducers. These loggers then periodically transmit the recorded data back to the central database (typically every 6-8 hours) from where it is uploaded to epaHydroNet (epaHydroNet, 2016). All data loggers are set to GMT for the full year to avoid the time changes in Summer/Spring time (epaHydroNet, 2016).

15-minute time step river water level data were downloaded for four stations (each one located at the outlet of each of the four sub catchments of the Blackwater catchment). River level data for Duarrigle and Dromcummer stations are from the epaHydroNet website (epaHydroNet, 2016) and river level data for Mallow rail BR station and river flow data for Killavullen are from the OPW HYDRO website (OPW HYDRO website, 2013). These datasets need to be quality controlled before use. Table 3.7 shows the summary of unrecorded and missing values in these datasets. Figure 3.11 (a) and (b) shows the time series of river level and flow during 2010 to 2017 for Dromcummer station after data cleaning. It is notable that the percentage of bad data for the river level stations are significantly lower than those for the OPW precipitation data.

Table 3.7 Summary of unrecorded and missing value for four river level stations on the Blackwater catchment

Number of Gap and Missing value of river level or river flow in 4 stations					
Station Name	**Not Recorded**	**missing value**	**Outliers: level > 3 m or flow > 450**	**% bad data**	**Duration**
Duarrigle	861	11	103	0.35	2010- 2013
Dromcummer	6547	19	0	2.34	2010-2017
Mallow rail BR	8906	1	3289	4.35	2010-2017
Killavullen (flow)	2381	13	980	1.2	2010-2017

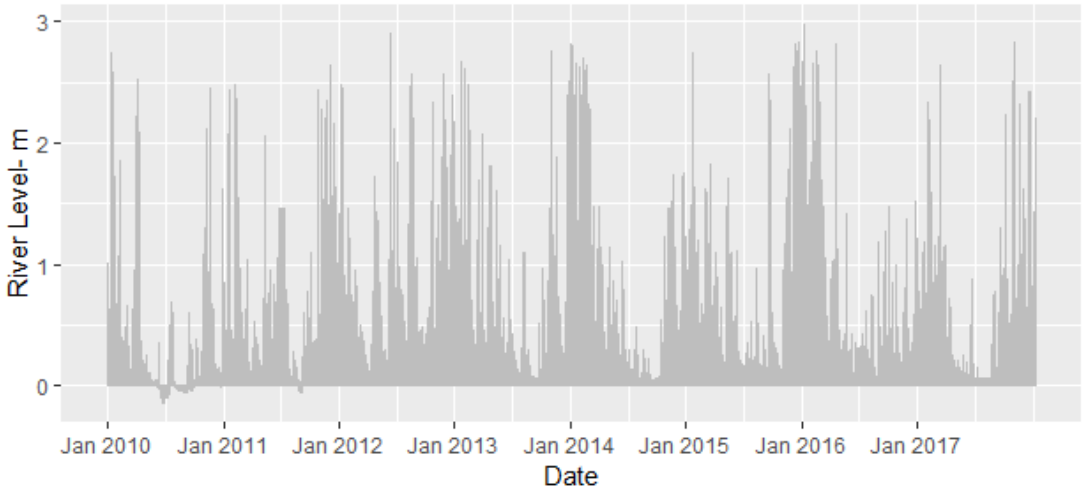


Figure 3.11 (a) Time series of 15-minute river level for Dromcummer during 2010 to 2017

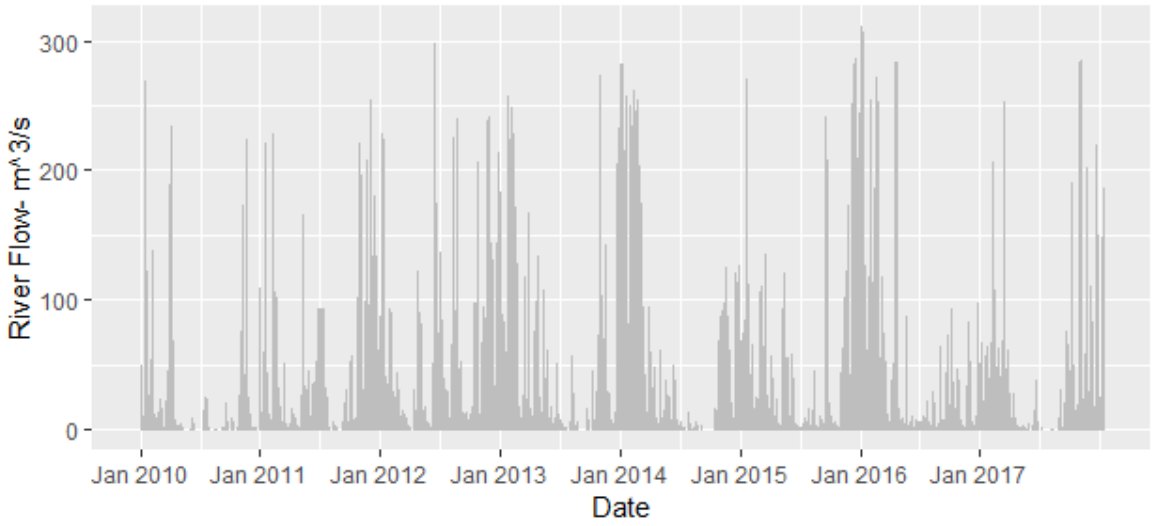


Figure 3.11 (b) Time series of 15-minute river flow for Dromcummer during 2010 to 2017

Chapter 4 Analysis of Data and Methodologies Employed

4.1 Introduction

Analysis of available data from different sources is one of the most important issues in hydrological modelling, the quality of available data can be poor and therefore there is a need for quality control and gap filling to obtain useful data as input for hydrological modelling

The aim of this chapter is to analyse the data, identify any missing values, outliers and any kind of error in precipitation data, and then develop a method to cluster the rain gauges based on similarity of their precipitation and eventually filling the gaps and correcting the errors in any individual rain gauge record by adjacent members of its cluster.

4.2 Kmeans Clustering

Kmeans clustering is an unsupervised Machine Learning technique which is suitable for finding the similarity of data when unlabeled data (data without defined category or group) is available.

Kmeans clustering partitions the dataset into *K* different groups or clusters of similar objects, where each member of each cluster is nearest to the centroid of that particular cluster. Usually the Euclidean Distance is used as a mathematical function to define the similarity between two objects in the dataset. The algorithms works as follows:

- Initialise the *K* value.
- Select k objects of the dataset randomly as a centroid of each cluster.
- For each object find the nearest centroid and assign the object to this cluster (Figure 4.1).

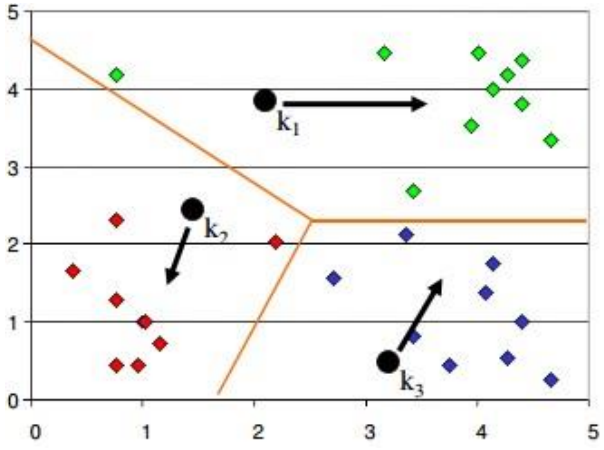


Figure 4.1 Kmeans clustering with 3 clusters ('K-means Clustering Tutorial-Machine Learning', n.d.)

Selecting the appropriate number of cluster *K* is important, but there is no theoretical method to find the optimal value of the number of clusters. The aim of finding the best value of *K* is to obtain the minimum variance in each cluster and maximum separation between clusters. There are several techniques for selecting a *K* value. A practical approach is to compare the outcomes of multiple runs with different *K* and choose the best one. The Elbow method (section 4.2.1) can be also applied to find the optimal value of *K.* In general, a large *K* decreases the error but increases the risk of overfitting (Asanka, 2017).

Another important issue in Kmeans clustering is to determine the features for clustering. This comes based on trial and error, also analysis of data to find out if there is any relation between features and clusters. Some knowledge about the problem also can help to choose the best feature (Trevino, 2016).

For example in this case it is assumed that there is a relation between the location of rain gauges and amount of precipitation. It is expected that precipitation is greater when the site elevation is higher. The Kmeans function in R was applied to cluster the OPW rain gauges. The code was developed in R using the output of the Kmeans function such as number of clusters, size, mean and sum of squares of each cluster and total sum of squares to summarise and visualize the outputs for further analyse, and to find the optimal value of the number of clusters (*K*) using the Elbow method (see Appendix B.4 for more information).

4.2.1 Elbow Method

The Elbow method is a method to find the optimal value of *K* in Kmeans clustering. The main idea of the Elbow method is to run the Kmeans clustering for different values of *K* and then choose one that has the minimum sum of squared error (SSE) differences between each member of the cluster with the mean of that cluster which, indicates the variation within the cluster. SSE is given by:

$$SSE = \sum_{i=1}^{n}(x_i - \bar{X})^2 \qquad \text{Eq 4.1}$$

where n is the number of individual objects in the dataset, x_i is each individual object and $\bar{X}$ is the mean of each cluster (Asanka, 2017).

4.3 Clustering of rain gauges

The OPW's rain gauges need to be clustered across the Blackwater catchment in order to fill the gaps of each rain gauge by its cluster members. Here are two approaches for clustering rain gauges. As the precipitation data from MÉRA were not available at the beginning of this project, clustering

based on rain gauges position was considered. After obtaining MÉRA precipitation data, the clustering based on rain gauges monthly and annual precipitation was examined and compared. Eventually clustering based on rain gauges monthly precipitation was used to cluster rain gauges.

4.3.1 Clustering based on location

It is assumed that there is a relation between the geographic position and the amount of precipitation, thereby suggest using geographical position to cluster the rain gauges. The assumption is that more rain falls on the West of the catchment than on the East, as it is known that the West of Ireland has more rain than on the East of the country. In other words, the rain gauges with similar geographical position and elevations are likely to have similar amounts of precipitation. Figure 4.2 shows the relationship between a group of three features (Easting, Northing and Elevation) with annual rainfall, calculated here from MÉRA data. There is moderate positive correlation between elevation and annual rainfall, moderate negative correlation between Northing and annual rainfall and weak correlation between Easting and annual rainfall as illustrated in Table 4.1.

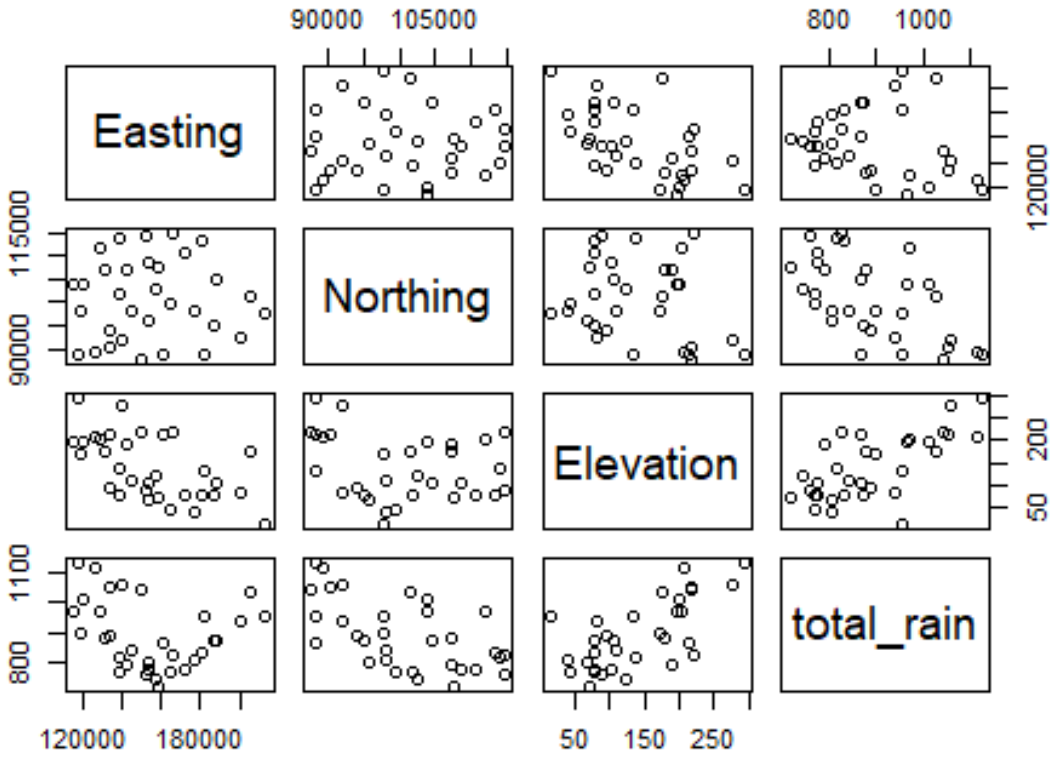


Figure 4.2 Scatterplots indicating correlation between the four features Easting, Northing, Elevation and annual rainfall for the MÉRA dataset for 2010

Table 4.1 Correlation between three features Easting, Northing, Elevation and annual rainfall of MÉRA dataset for 2010

	Easting	Northing	Elevation
Annual rainfall	- 0.18	- 0.61	0.65

The result of the Kmeans clustering with three features Easting, Northing and Elevation of rain gauges (see Table 3.2) when the value of k is six is illustrated in Table 4.2 and Figure 4.3

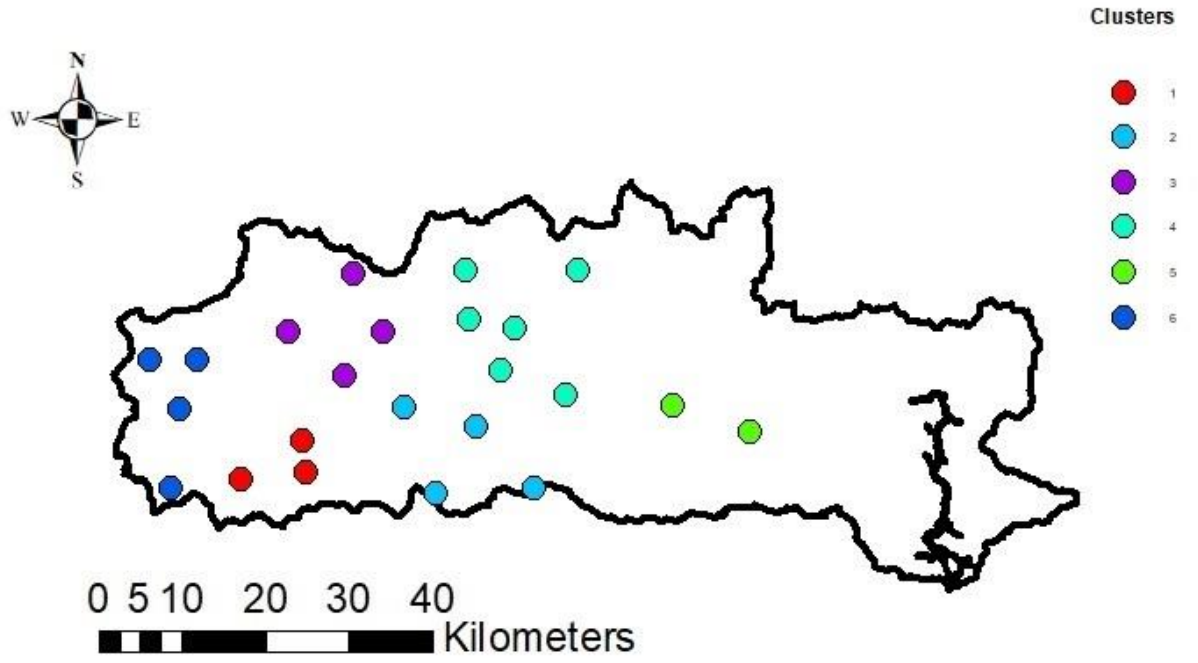


Figure 4.3 Members of each cluster when *K* = 6 with three features: Easting, Northing and Elevation

Table 4.2 Kmeans clustering results for K= 6 with three features: Easting, Northing and Elevation

Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**
80704	80701	80708	80706	80727	80711
80705	80702	80710	80707	80728	80712
80717	80715	80714	80719		80713
	80718	80720	80721		80716
			80722		
			80732		

4.3.2 Clustering based on MÉRA precipitation

The best way to classify the rain gauges is to find similarity of their precipitation during each time step. In general, a small time step means a large number of features for clustering which might cause overfitting and a large time step means a small number of features which might reduce the accuracy of clustering. It is suggested using the monthly rainfall during a specific year as 12 features for applying Kmeans clustering. Twelve features is a suitable number of features and represents the variation of rainfall during the year. Using annual rainfall is another candidate feature for clustering but just one feature cannot explain the characteristics of rain gauge precipitation and the similarity between gauges.

Hourly precipitation from MÉRA for 2010 is available, as MÉRA is high resolution (2.5 km) data so we can estimate the hourly precipitation for the nearest point of each rain gauge and then obtain monthly precipitation for these points (Figure 3.9 & Figure 3.10).

The next issue for applying Kmeans clustering is to determine the number of clusters *K*. The iterative code was developed in R to apply the Elbow method which applies the Kmeans clustering for different values of *K* and chooses the one with minimum sum of square error. The resulting plot of sum of square error for different *K* makes it easy to choose the best *K.* Figure 4.4 shows the plot of the Elbow method to obtain the optimal value of *K*, which is four.

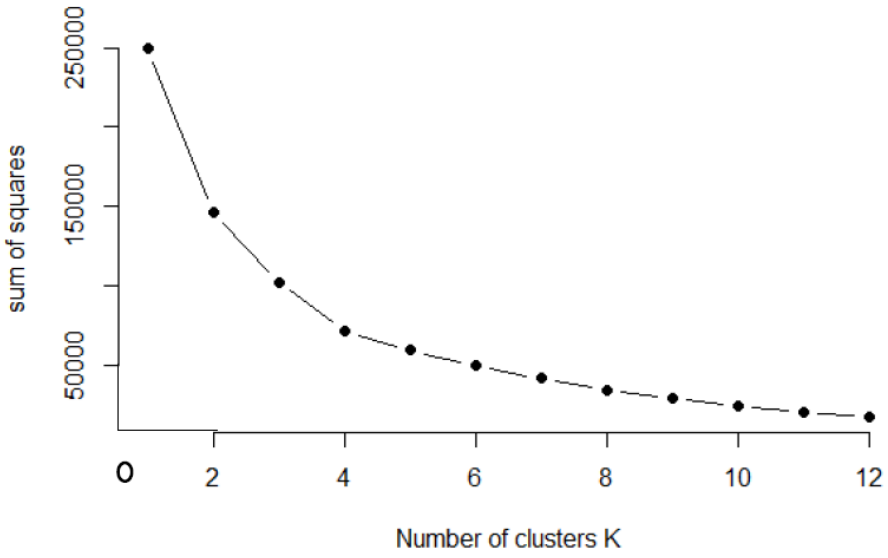


Figure 4.4 Elbow method to find the optimal value of *K.*

In this case the optimal value of *K* is not required, instead the number of clusters is required where there are more than two members in each cluster so that gap filling can be applied between members of each cluster to minimize the gaps. If the number of clusters is large then the number of members in each cluster is small. There might be less than two members in some clusters and therefore gap filling would not be possible using these clusters. If the number of clusters is small then the number of members of each cluster might be large so gap filling between each member of the cluster will reduce the gaps but might reduce the accuracy of data as well.

Therefore, it can be a trial and error process to find the best value of *K*. Table 4.3 shows the number of members in each cluster for different values of *K* between four and twelve.

Table 4.3 Number of clusters & number of members in each cluster for $4 \leq K \leq 12$

	Number of members For K								
Clusters	K = 12	K = 11	K = 10	K = 9	K = 8	K = 7	K = 6	K = 5	K = 4
Cluster 1	2	2	2	3	3	5	5	6	8
Cluster 2	2	3	2	3	3	2	4	3	4
Cluster 3	2	2	2	1	3	3	5	3	7
Cluster 4	2	2	2	4	3	3	3	7	4
Cluster 5	1	3	1	2	2	5	3	4	
Cluster 6	4	2	3	3	5	2	3		
Cluster 7	2	1	4	2	2	3			
Cluster 8	2	1	3	2	2				
Cluster 9	2	2	2	3					
Cluster 10	1	1	2						
Cluster 11	2	4							
Cluster 12	1								

In each cluster, at least three members are required to carry out successful gap filling. Evidently *K* = 6 is the best value for *K*. Table 4.4 and Figure 4.5 illustrate the result of Kmeans clustering for *K* = 6 based on monthly precipitation (12 features) and Table 4.5 and Figure 4.6 illustrate the result of Kmeans clustering for *K* = 6 based on annual precipitation (one feature).

Figure 4.7(a), Figure 4.7(b) and Figure 4.8 also show similarity of monthly and annual precipitation within each cluster. As described, we have used the *K*means clustering using monthly features (12 features) to cluster the rain gauges

Table 4.4 Kmeans clustering results for K= 6 based on monthly precipitation (12 features).

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
80701	80702	80706	80711	80715	80708
80717	80704	80707	80712	80721	80710
80718	80705	80714	80716	80732	80720
80727	80713	80719			
80728		80722			

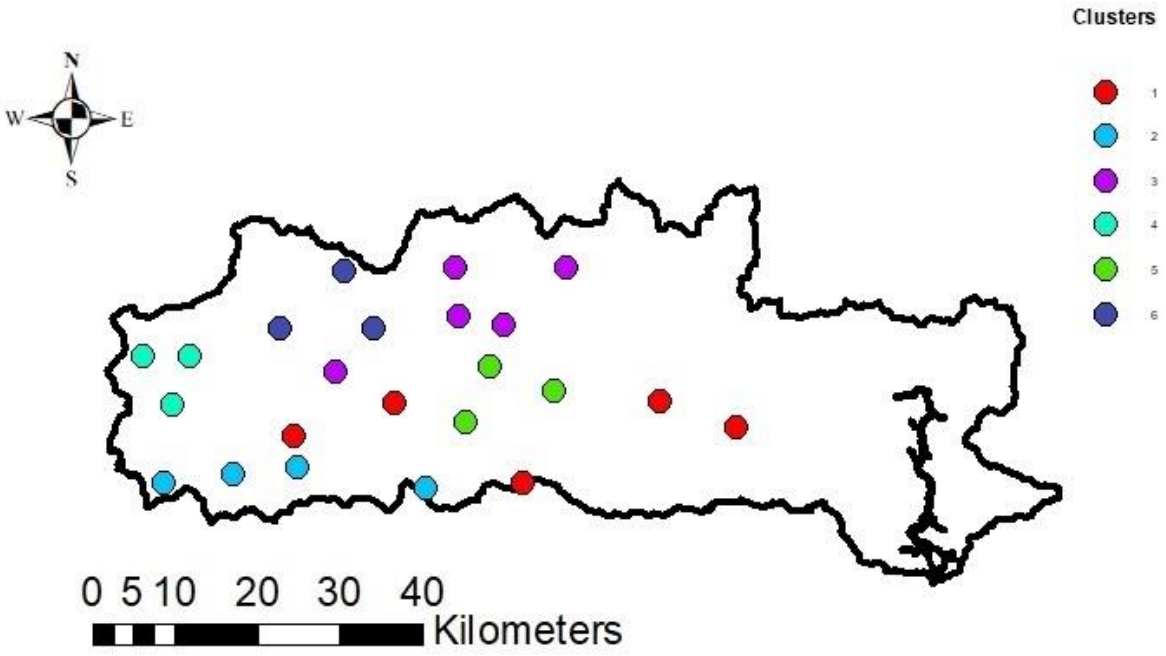


Figure 4.5 Members of six clusters based on the monthly precipitation (12 features).

Table 4.5 Kmeans clustering results for K = 6 based on annual precipitation (one feature)

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
80701	80706	80708	80702	80711	80705
80710	80707	80715	80704		80713
80712	80714	80718	80716		
80717	80719	80720			
80728	80721	80722			
	80732	80727			

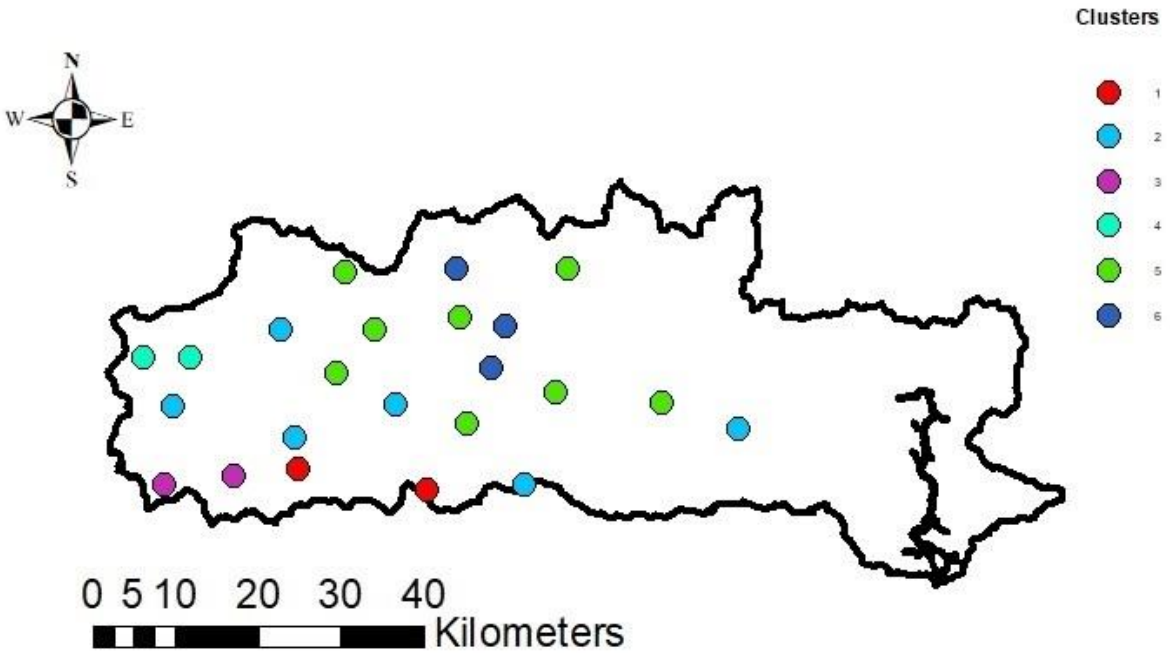


Figure 4.6 Members of six clusters based on the annual precipitation (one feature)

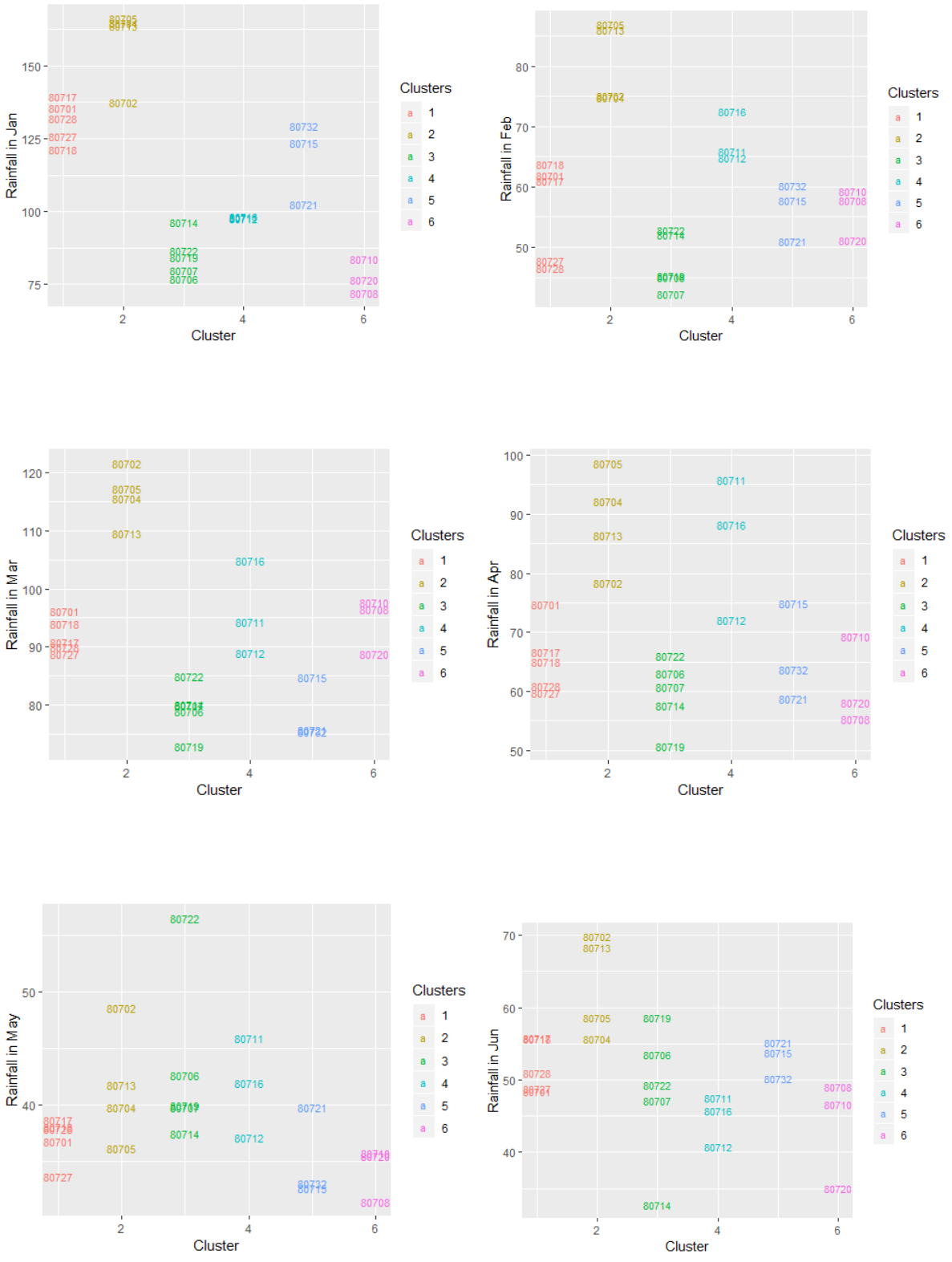


Figure 4.7(a) Similarity of monthly precipitation in each cluster January to June

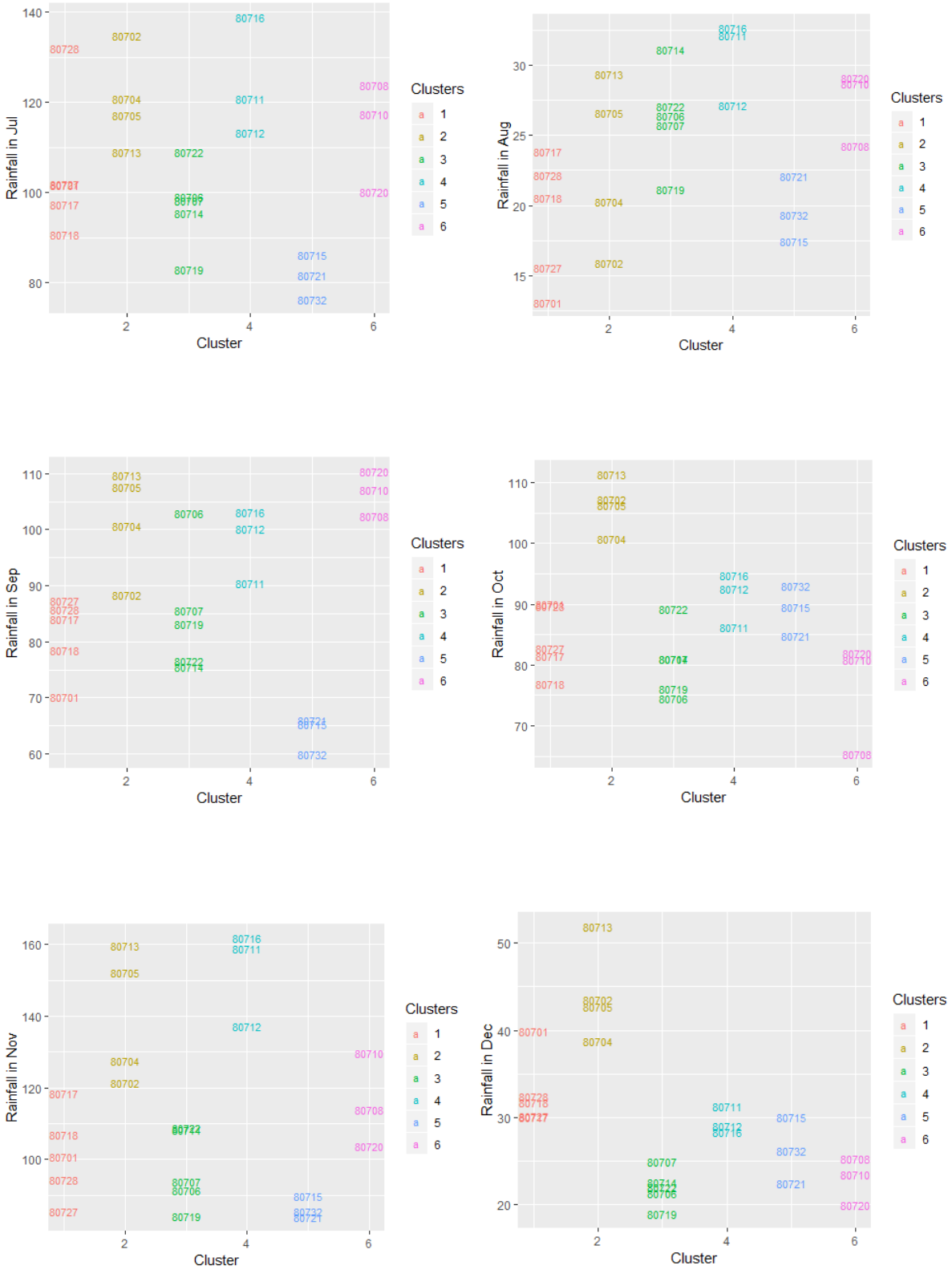


Figure 4.7(b) Similarity of monthly precipitation in each cluster July to December

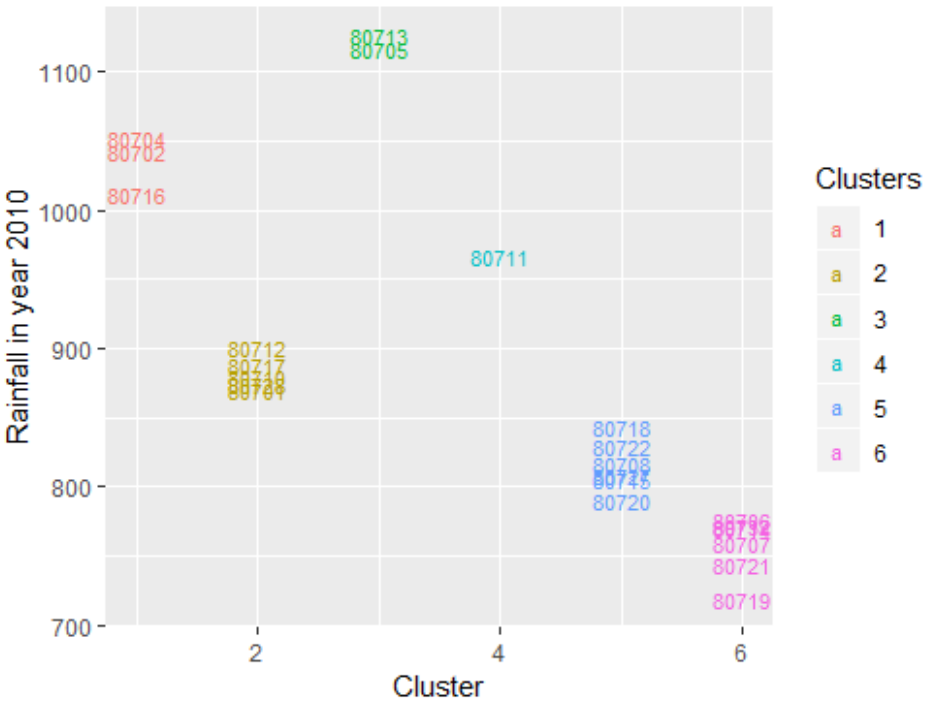


Figure 4.8 Similarity of annual precipitation in each cluster

4.4 Gap filling

4.4.1 Rain Gauge Data Gap Filling With Cluster Member Data

After identifying the missing values, outliers and errors in the 32 rain gauges with 15-minute precipitation data from OPW and then clustering 23 of them where they have less than 15 % bad data by Kmeans clustering with 12 features (monthly precipitation) and six clusters (K = 6), the method for filling the gaps and correction of outliers and errors in each cluster was developed as follows:

1. Filling the gaps (missing values and unrecorded values) and correct the outliers of the first member of the cluster with corresponding non-missing value and non-outliers of the next members of the cluster and vice versa
2. Repeat step 1 for next members of the cluster until the last members of cluster are reached
3. Repeat step 1 and 2 for the second, third and … until the one before last member of cluster is reached

Figure 4.9 illustrates this process in a cluster with four members. At the end each member of the cluster has minimal missing values and outliers, which we fill by another method.

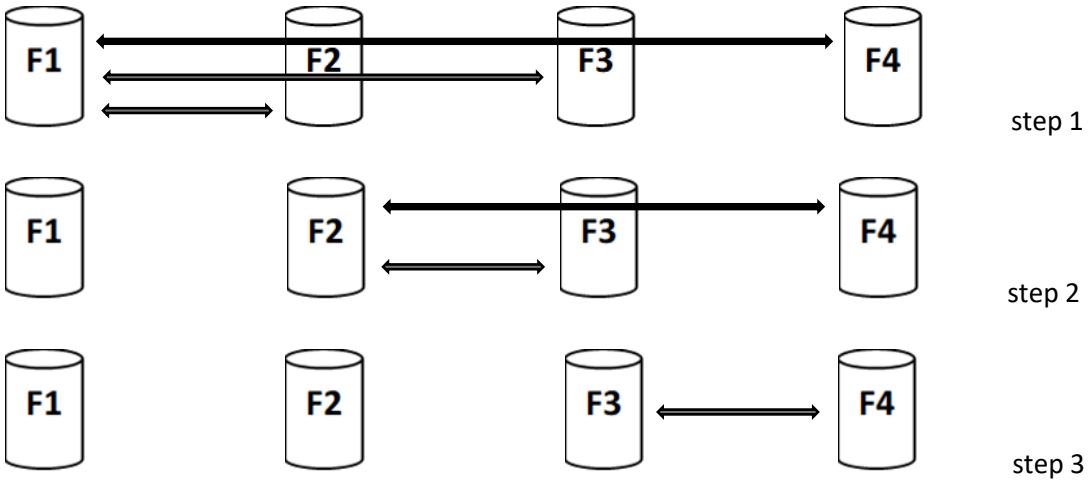


Figure 4.9 Gap filling process for a cluster with four members (F1,F2,F3,F4)

Table 4.6 shows the summary of each cluster after applying the gap filling method. As this method is an iterative method and fills the gaps of each member of cluster with at least two other members of cluster so it minimized the missing values and outliers and reduced bad data to 0.04 % to 0.09% for the period 2010 to 2017. The details of the bad data break down by year are shown in Table 4.7.

Table 4.6 Summary of missing values and outliers of each cluster after gap filling

Number of Missing and Outliers of Each Cluster of Rain Gauges after Gap Filling			
Cluster No.	**missing value**	**Rainfall > 15 mm**	**% bad data**
1	167	**0**	**0.06**
2	**139**	0	**0.05**
3	105	0	0.04
4	105	0	0.04
5	237	0	0.09
6	106	0	0.04

Table 4.7 Summary of number of missing values and outliers of each cluster for years 2010 to 2017

Year	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
2010	0	0	0	0	0	0
2011	0	0	0	0	0	0
2012	0	34	0	0	0	0
2013	70	8	8	8	8	8
2014	97	97	97	97	229	98
2015	0	0	0	0	0	0
2016	0	0	0	0	0	0
2017	0	0	0	0	0	0
Total	167	139	105	105	237	106

4.4.2 Gap Filling of Rainfall Data Using Temporal Average of Rain GaugeNearest Record

When there are no long sequences of missing values or outliers in the dataset the simple way to fill the gap or correct the outliers is to substitute the gap or outliers with the average of the two nearest non-missing and non-outlier records (before and after the gap (Table 4.8)). This method can be applied after the gap filling with the cluster members in order to minimize the amount of bad data in the datasets. A problem might occur when there is a long sequence of missing values or outliers in the dataset using this method, as the missing values or outliers practically will be filled with constant values. Further consideration might be needed in this case.

Table 4.8 Filling the gap with the two nearest record

Year	Month	Day	Hour	Minute	second	Rainfall-mm
2012	5	9	18	15	0	0.0
2012	5	9	18	30	0	0.2
2012	5	9	18	45	0	*0.2*
2012	5	9	19	0	0	***0.2***
2012	5	9	19	15	0	*NA*
2012	5	9	19	30	0	***0.4***
2012	5	9	19	45	0	*0.4*

(0.2 +0.4)/2 = 0.3

4.4.3 Long Sequences of Precipitation and Outliers

For many records there are periods where the values of precipitation were unrealistic where some might be large numbers (for example 145 mm in 15 minutes is considered an outlier) or some might be long sequences of reasonable values or long sequences of a constant value. Figure 4.10 shows the big jump in cumulative rainfall in a cluster before gap filling, which indicates the existence of outliers or a long sequence of precipitation.

The gap filling with the cluster members method was applied for correcting the outliers. Figure 4.11 shows the cumulative rainfall in a cluster after gap filling. There is a small jump in cumulative rainfall that indicates the gap filling method successfully corrected the outliers. The next step to optimizing data quality is to identify the long sequence of reasonable values in 15 minutes time steps. Determination of this value (number of sequences of reasonable values in 15 minutes time steps) is based on experience and some knowledge of the area. For example, more than four sequences of 15 mm of precipitation in consecutive 15 minutes time steps means more than 60 mm precipitation in an hour, which is not realistic.

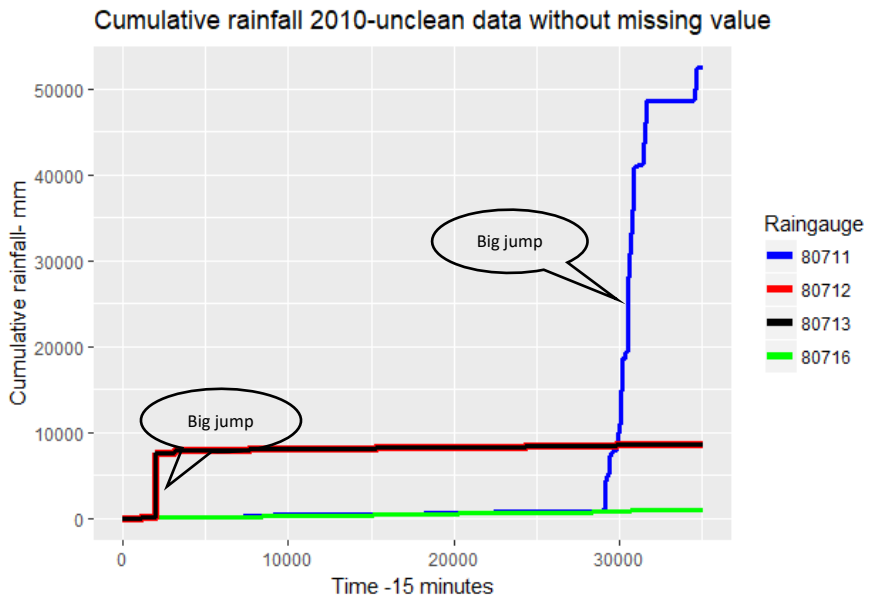


Figure 4.10 Cumulative rainfall for 2010 for individual rain gauges in a cluster before gap fillings. Cumulation of large outliers in a rain gauge causes the big jump.

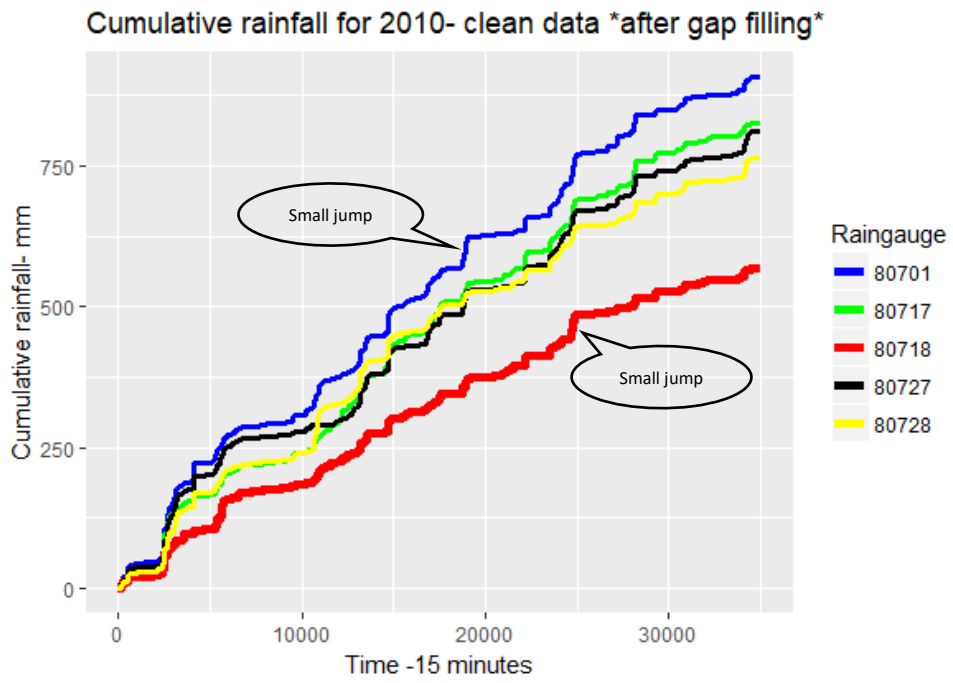


Figure 4.11 Cumulative rainfall for 2010 for individual rain gauges in a cluster after gap filling. Correcting large outliers reduces the cumulation of outliers.

In this study all rain gauge data were examined after gap filling. There are not more than four sequences of maximum 15 mm of precipitation in consecutive 15 minutes time steps which is realistic. So in this case further data cleaning is not required. If there were more than four sequences of 15 mm precipitation in 15 minutes, which indicates some error in the data, then applying the gap filling method for these periods is suggested.

A comparison of annual OPW precipitation data with annual MÉRA precipitation for 2010 is shown in Figure 4.12 for all 32 rain gauges to examine data quality of OPW precipitation data after these gap filling and outlier correction methods. This figure indicates that these methods were successful to obtain a high standard of data quality as the annual OPW data and MÉRA data are very close for those rain gauges which these methods were applied, except rain gauge 80711.

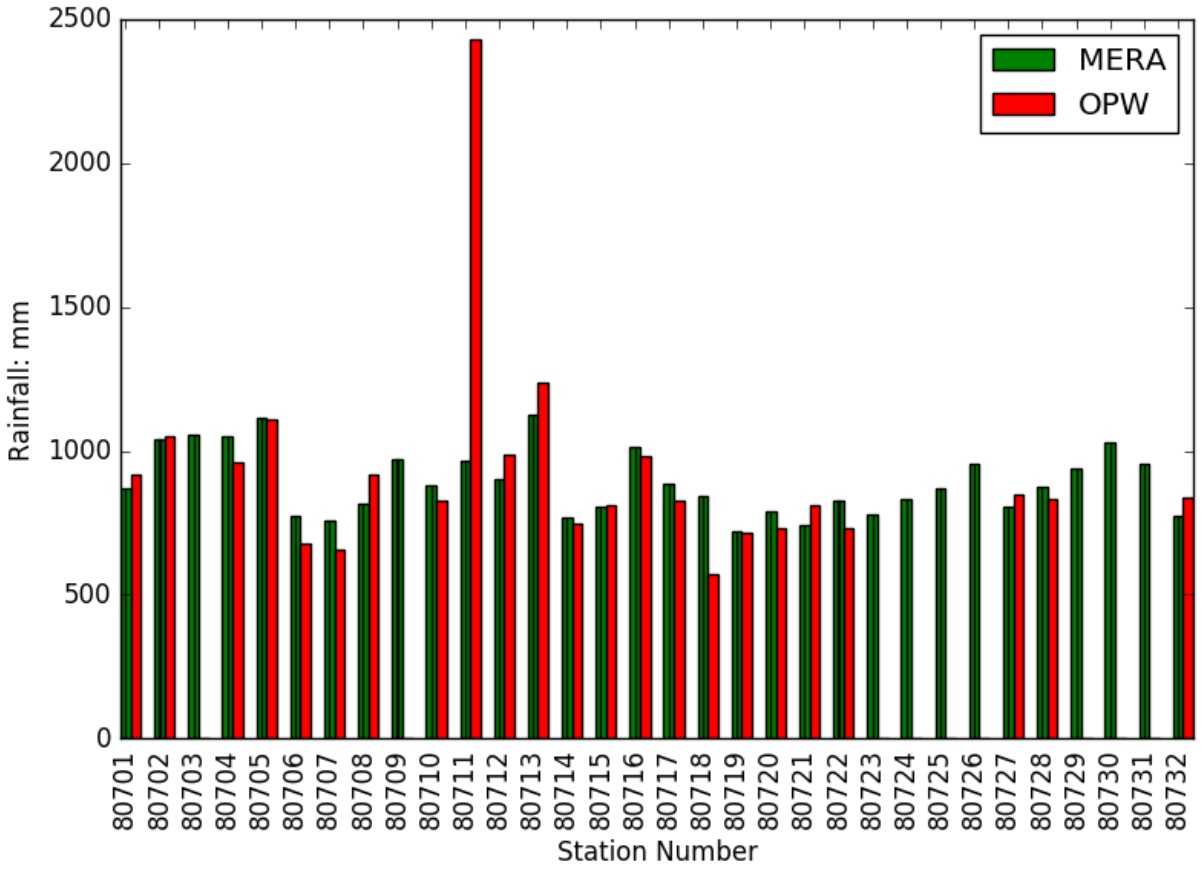


Figure 4.12 Comparison of annual OPW precipitation data after gap filling methods with MÉRA precipitation data for 2010.

Annual rainfall for rain gauge 80711 is very high so the comparison of monthly OPW precipitation and MÉRA precipitation was examined (Figure 4.13). The OPW Precipitation for November and December are very high and unrealistic for gauge 80711. Figure 4.14 shows the amount of rainfall recorded in November 2010 in OPW's rain gauge 80711 and Table 4.9 shows a part of these data. There are a lot of high amounts of rainfall but all less than 15 mm and not sequential, so our gap filling and outlier correction methods were not able to detect and correct them. These kind of outliers needs more consideration.

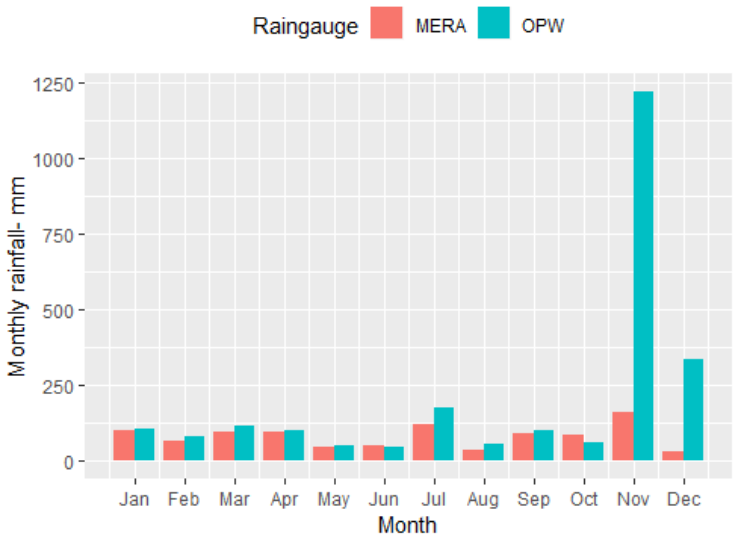


Figure 4.13 Comparison of monthly OPW precipitation data rain gauge 80711 after applying gap filling methods with MÉRA precipitation data for 2010.

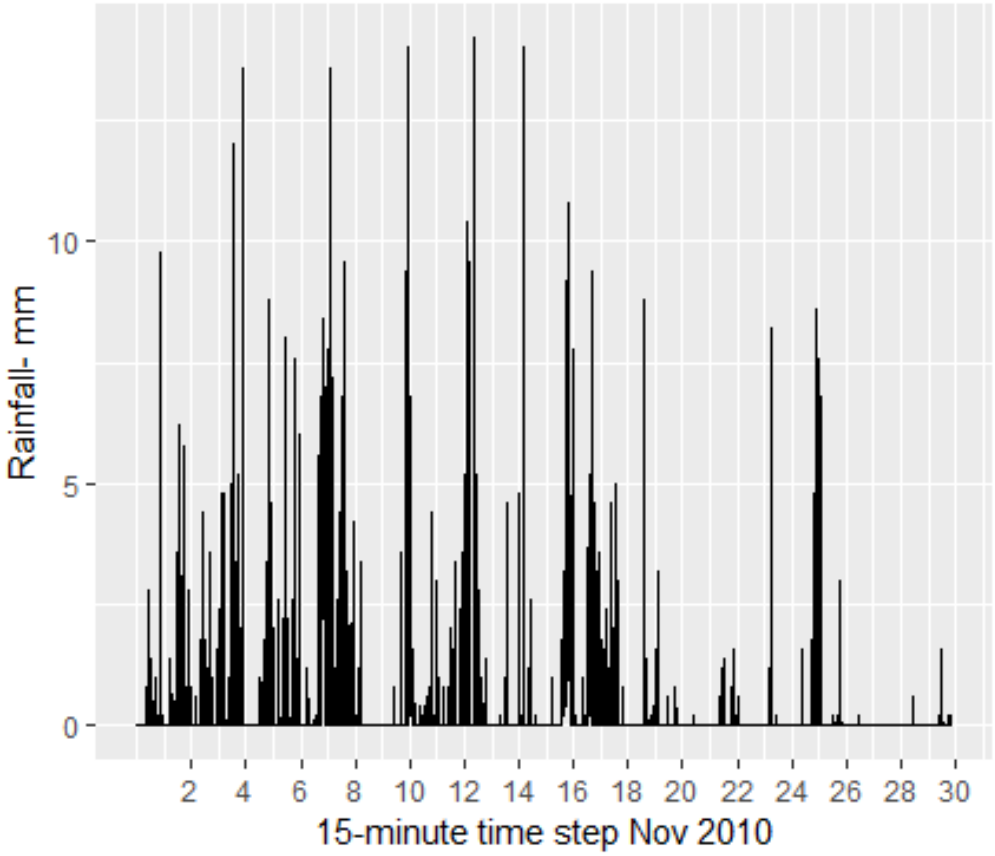


Figure 4.14 OPW precipitation data for rain gauge 80711 after applying gap filling methods, November 2010

Table 4.9 Part of OPW precipitation data (unrealistic) for rain gauge 80711 in November 2010

Year	Month	Day	Hour	Minute	second	Rainfall -mm
2010	11	7	17	30	0	3.2
2010	11	7	17	45	0	2.8
2010	11	7	18	0	0	*2.4*
2010	11	7	18	15	0	***2.0***
2010	11	7	18	30	0	*2.8*
2010	11	7	18	45	0	***2.2***
2010	11	7	19	0	0	*6.8*
2010	11	7	19	15	0	*6.4*
2010	11	7	19	30	0	*8.4*
2010	11	7	19	45	0	*6.4*
2010	11	7	20	0	0	*2.2*
2010	11	7	20	15	0	*5.4*
2010	11	7	20	30	0	*5.4*
2010	11	7	20	45	0	*3.8*

Figure 4.15 shows the monthly cumulative rainfall for OPW rain gauge 80702 (Bweeng - Pump House) for 2010 to 2016, indicating that the driest year was 2010 with annual rainfall around 1000 mm and the wettest year was 2014 with annual rainfall around 1600 mm.

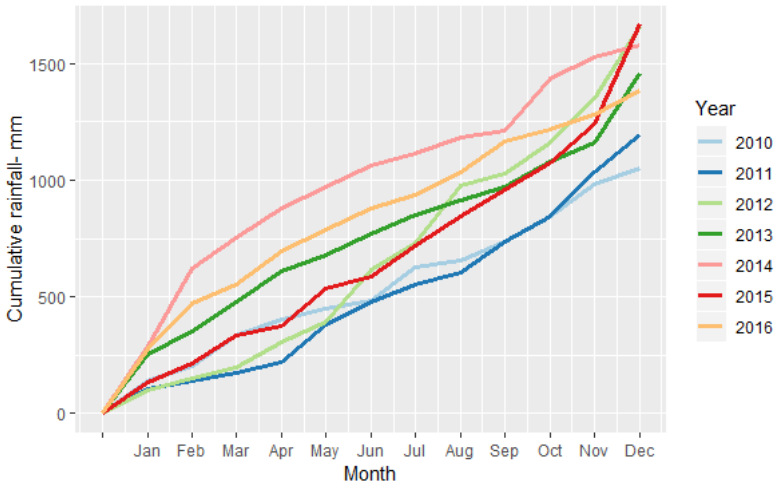


Figure 4.15 Monthly cumulative rainfall for OPW rain gauge 80702 (Bweeng - Pump House) for 2010 to 2016

This pattern is different in different rain gauges. The driest year for rain gauge 80722 (Ballygugroe – Landfill) was 2010 with annual rainfall around 700 mm and the wettest year was 2015 with annual rainfall 1600 mm while 2014 was the second wettest year with annual rainfall around 1000 mm (Figure 4.16).

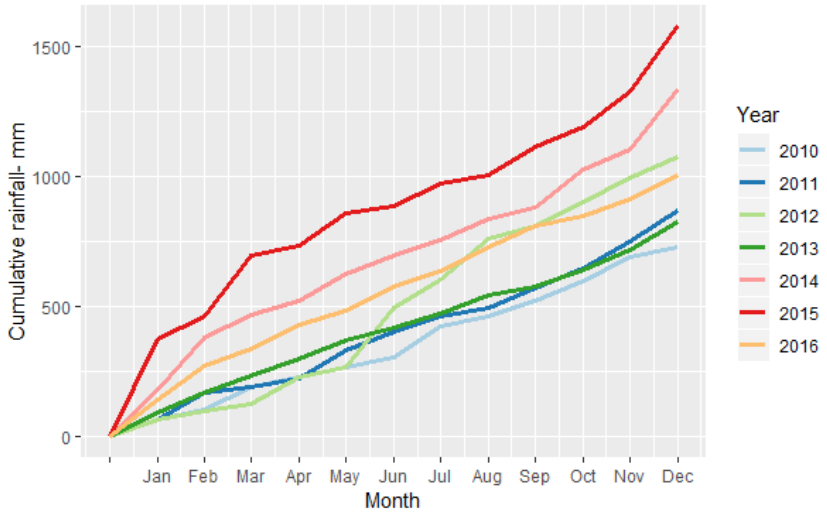


Figure 4.16 Monthly cumulative rainfall for OPW rain gauge 80722 (Ballygugroe – Landfill) for 2010 to 2016

This cumulative plot can be used to find any other outliers, which were not recognized by the outlier correction method for more investigation. Figure 4.17 shows a big jump in monthly cumulative rainfall for OPW rain gauge 80728 (Coole – Pumphouse) in September 2013, which indicates the existence of bad data (see Figure 3.3 for rain gauge locations).

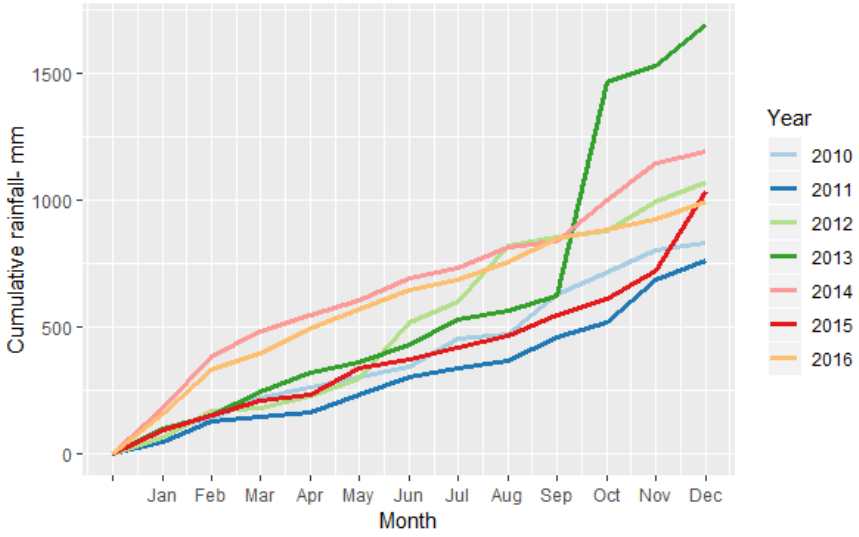


Figure 4.17 Monthly cumulative rainfall for OPW rain gauge 80728 (Coole – Pumphouse) for 2010 to 2016

4.4.4 River Level Data Gap Filling with Average of River Level's Nearest Record

Gap filling with the average of river level station nearest records as described in section 4.4.2 was applied to river level datasets to minimize the amount of bad data in these datasets. In Figure 6.12 (a), the time series of observed flows for 2011 and Figure 6.13 (c) the time series of observed flows for period of 29/12/2015 to 31/12/2015 indicate there were a long sequence of missing values or outliers in datasets which were filled by this method. The missing values or outliers practically were filled with the constant values. Further consideration might be needed in this case.
There is no bad data left on river level datasets following application of the procedure.
If there is not a large number of missing values and outliers in river level data, applying gap filling methods are not essential prior to applying the PDM model as PDM omits missing values.

Chapter 5 Probability Distributed Model (PDM)

5.1 Introduction

The Probability Distributed Model (PDM), is a conceptual rainfall-runoff hydrological model (Moore, 2007). The model inputs are rainfall and potential evaporation over the catchment and output is river flow at the catchment outlet. The main idea of this model is to describe the variation of soil moisture storage capacity of the catchment by a probability distribution. The standard model employs the Pareto distribution. Soil moisture varies over the catchment as it depends on the type of soil, geology, land use and topography. Instead of trying to model each of these factors, PDM introduces the distribution of storage depth over the catchment, which ranges from a minimum depth C_{min} to a maximum depth C_{max}.

5.2 Data preparation for PDM Model

5.2.1 Rain Gauges Contributing Area

The concept of contributing area of a rain gauge becomes an important issue when precipitation is available from multiple rain gauges across the catchment. The rain gauge contributing area defines an area of influence around a rain gauge. The Thiessen Polygon and Inverse Distance Weighted (IDW) methods in ArcGIS can be used to calculate the contributing area of each individual rain gauge.

5.2.2 Thiessen Polygon

The Thiessen polygon also known as Voronoi Diagram specifies the influence area (polygon) around each point of a given set of points (Yamada, 2017). Each point inside the polygon is closer to that specific point than any other points of the given set. Figure 5.1 illustrates the procedure to generate the Thiessen polygons associated with the set of points as follows:

- draw a line between each rain gauge locations (dash line)
- draw the perpendicular bisector of each dash line (solid line)

The solid lines construct the irregular polygons which divide the whole area into a number of smaller areas associated with each rain gauge of the given set of rain gauge locations. Therefore the area of the catchment (A) is the sum of the sub areas $A = \sum_1^n A_i$ where n is the number of rain gauges in the area and A_i represents the area of influence of each rain gauge .

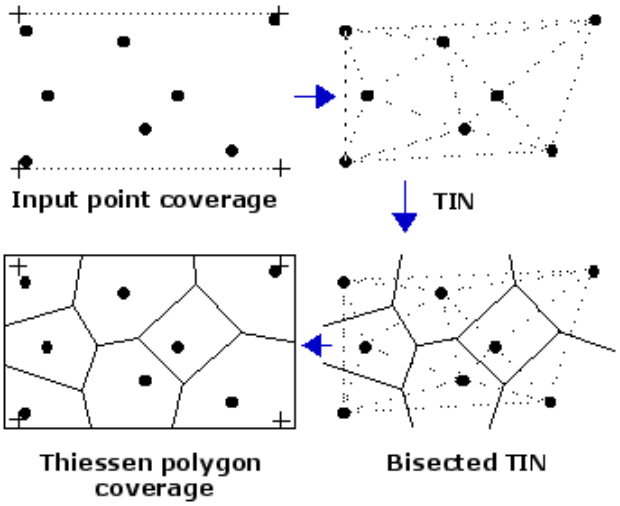


Figure 5.1 Partitioning area associated with a set of points by Thessien polygon method ('Creating Thiessen polygons', n.d.)

ArcGIS was used to generate the Thiessen polygons and to calculate the percentage of contributing area associated with each rain gauge across each of the four subcatchments (nested catchments). Figure 5.2(a) to (d) illustrates the Thiessen polygons for the four nested catchments within the Blackwater catchment (Duarrigle, Dromcummer, Mallow rail BR and Killavullen).

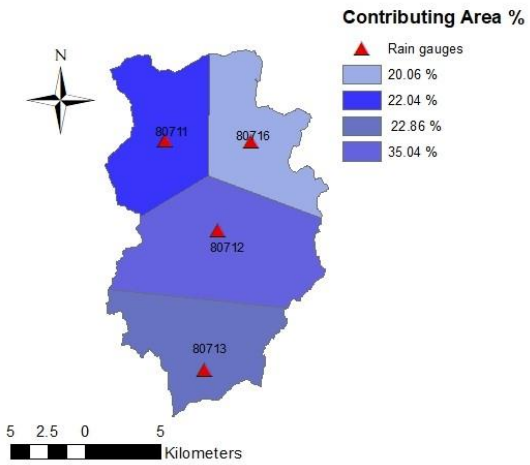


Figure 5.2(a) Thiessen polygon for Duarrigle subcatchment

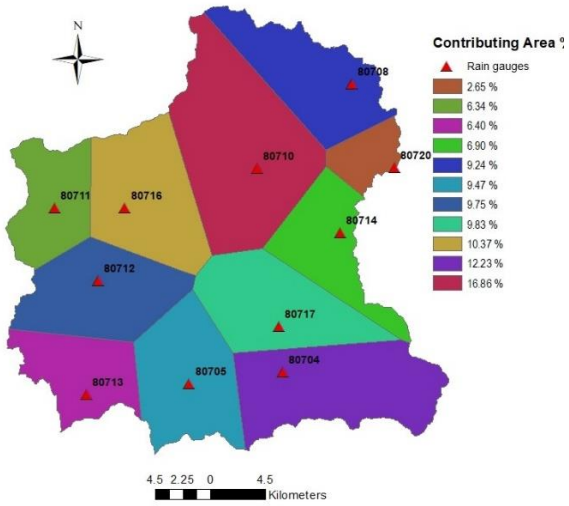


Figure 5.2 (b)Thiessen polygon for Dromcummer nested catchment

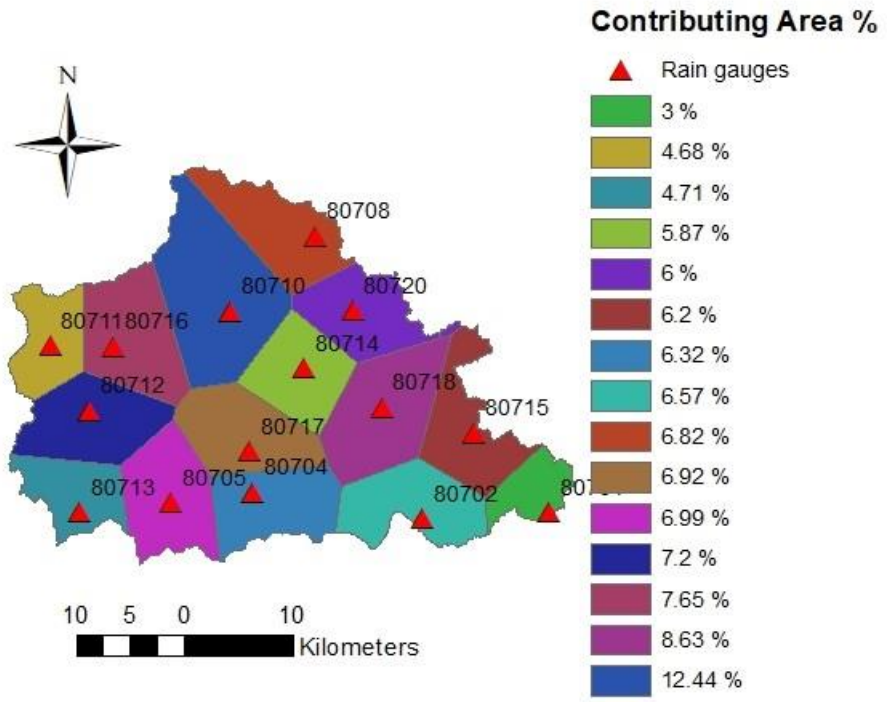


Figure 5.2(c)Thiessen polygon for Mallow Rail BR (Duarrigle & Dromcummer & Mallow) nested catchment

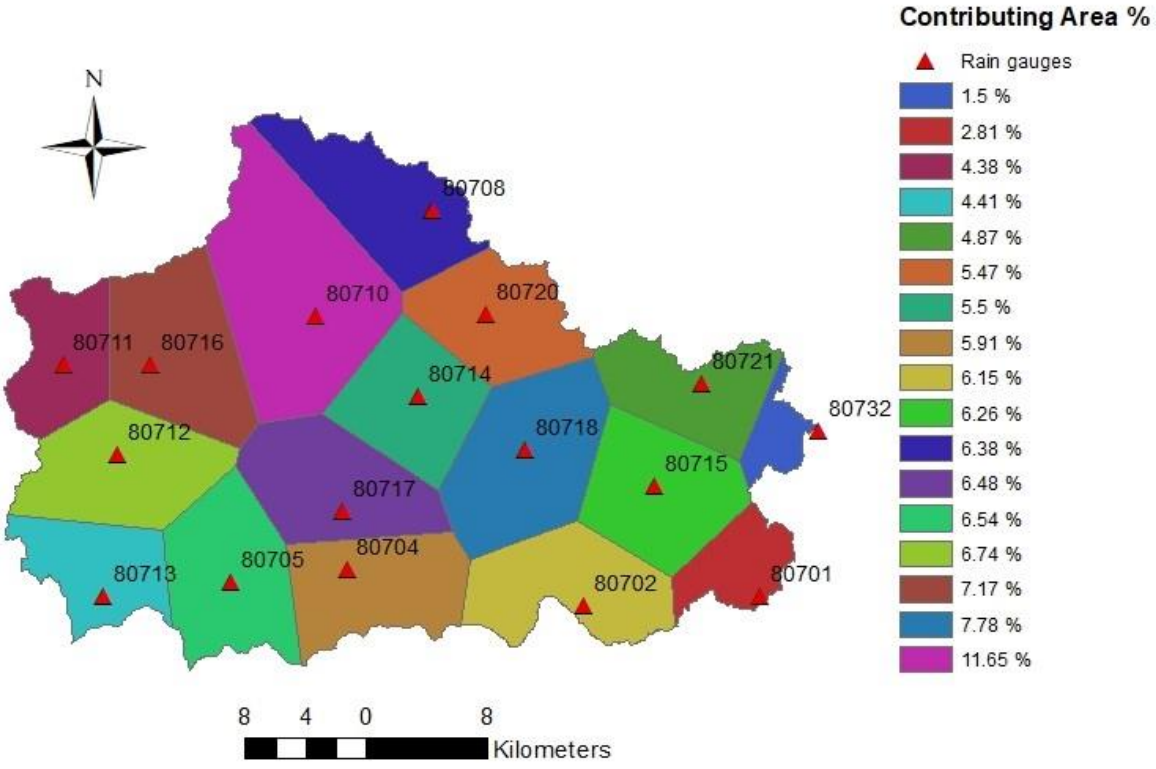


Figure 5.2(d) Thiessen polygon for Killavullen (Duarrigle & Dromcummer & Mallow & Killavullen) nested catchment

5.2.3 Inverse Distance Weighting (IDW)

The IDW is an interpolation method which assumes that points closer to each other have more correlation and similarity than those farther away. Measuring some variables (rainfall, temperature, soil moisture) across a spatial area is not possible. The IDW method is widely used to estimate the values at unmeasured locations in the area by averaging the value of some measured locations, where those locations closer to the unmeasured locations have more influence than those farther away. In other words weighting the average of measured locations by the inverse distance so the IDW method can be applied to divide the area into sub areas around the given set of points. The IDW method was applied on the subcatchment Duarrigle with three rain gauges. Figure 5.3 illustrates three-subdivision areas associated with each rain gauge.



Figure 5.3 Subdivion area of Duarrigle subcatchment generated by IDW

5.2.4 Rating curves

River flows can be calculated from river levels. The river levels are available in the epaHydroNet and OPW HYDRO website (http://waterlevel.ie/hydro-data/). The relationship between river level and river flow is given by the rating curve equation:

$$Q = a\,h^b \qquad \text{Eq 3.2.1}$$

where Q is the river flow in $m^3.s^{-1}$, a and b are dimensionless coefficients and h is the river level in meters. Each point along a river has a different flow/height (rating curve) relationship as the cross-sections of a river are usually not constant over the river length. The rating curve is less accurate for high flows than low flows as measuring the high flow is more difficult and sometimes is not safe or possible to do.

Coefficients *a* and *b* can be derived for a specific location by measuring river level *(h)* and river flow *(Q)* plotting the *h* against Q for different heights and flows. The non-linear regression model was built in R, based on the relationship between measured river level and river flow. For the period up to the year 2006, the estimation of a and b are 22.09 and 1.61 for the Duarrigle. Figure 5.4 compares the actual rating curve and the estimate of the curve by estimating of these coefficients.

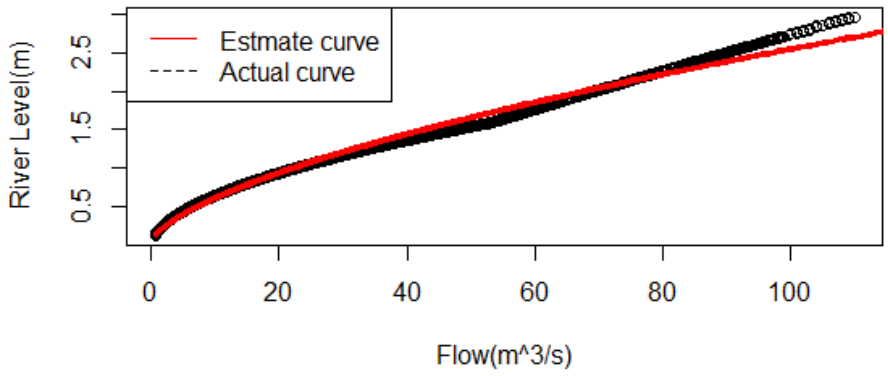


Figure 5.4 Acual Rating curve and it's estimation in Duarrigle subcatchment for 2006

$Q = a\,h^b$ where $a \approx 22.09$ & $b \approx 1.61$

These coefficients are different for different places along the river. For example for the Duarrigle subcatchment *a* and *b* were calculated by the EPA as follow.

There are three curves:

- For water level between 0 to 0.2 m *a = 8.857* and *b = 1.25151*
- For water level between 0.201m to 1.591 m *a = 22.6327* and *b = 1.83451*
- For water level above 1.592 m *a = 30.9276* and *b = 1.16186*

As coefficients *a* and *b* vary from one place to another along the same river it is imperative to consider different equations for the three different stations in the Blackwater catchment. Table 5.1 and Figure 5.5 show these coefficients for these three stations in the Blackwater catchment (for simplicity in Table 5.1 just one set of values of coefficients were presented).The river flow data is available on OPW HYDRO website for Killavullen station.

Table 5.1 Rating curve equation coefficients (a and b) for 3 stations on the Blackwater catchment (Steinmann, 2005)

Station	a	b
Duarrigle	22.1187	1.6256
Dromcummer	48.4194	1.7037
Mallow Rail Bridge	27.2661	1.967

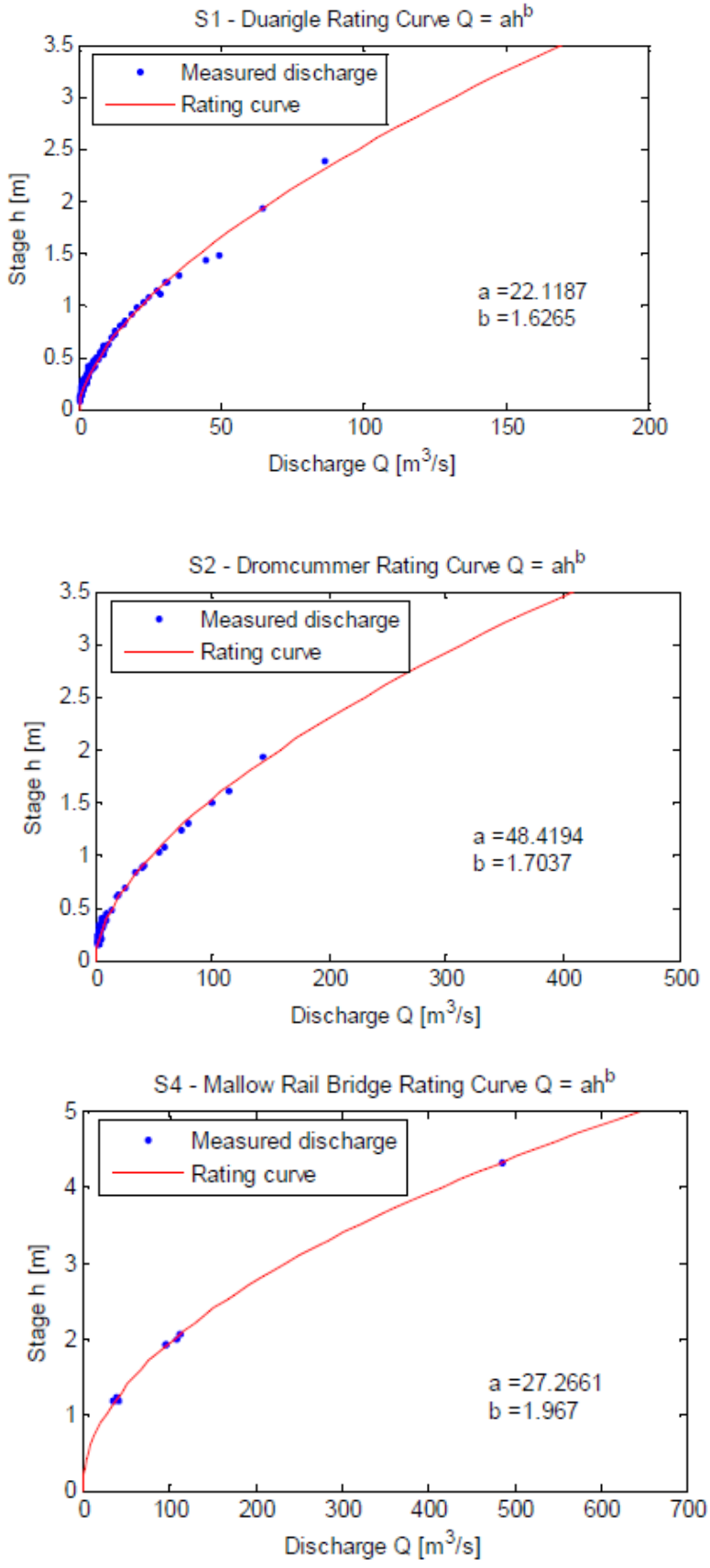


Figure 5.5 Rating curves for three stations along the Blackwater river for data collected up to 2005 (Steinmann, 2005).

5.3 Probability distributed soil moisture storage

Each point of a catchment is considered as a single storage or tank with specific storage capacity (depth) of c' (Figure 5.6) representing the absorption capacity of the soil at that point (CEH, 2012; Moore, 2007).

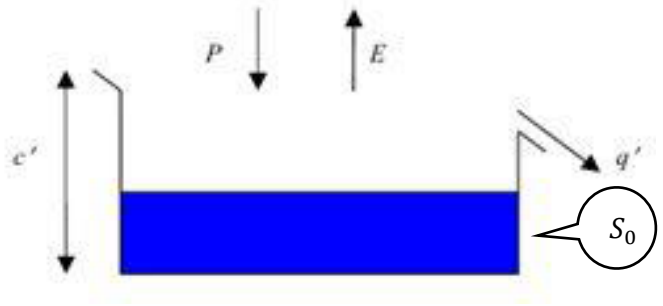


Figure 5.6 Point representation of runoff production by a single store (CEH, 2012).

The direct runoff production (q') is represented by precipitation (P) and losses of evaporation (E) at any point as follow:

$$q' = \begin{cases} P - E - (c' - S_0) & P > c' + E \\ 0 & P \leq c' + E \end{cases} \quad \text{Eq 5.1}$$

where S_0 is the initial depth of water in storage. The storage capacity (c) at any point can be considered as a random variate with probability density function $f(c)$ and Cumulative Distribution Function $F(c)$ as described in section 2.3.1, $F(c) = \int f(c)dc$, so the proportion of the basin containing stores of capacity less than or equal to C^* is :

$$P(c \leq C^*) = F(C^*) = \int_0^{C^*} f(c)dc \quad \text{Eq 5.2}$$

where $C^*(t)$ is the critical capacity storage below which all the stores are saturated at time t and start to generate runoff. To understand the storage capacity distribution across the basin, imagine that all stores with all different depths (storage capacity) are rearranged in order of depth and with their open tops arranged at the same height as shown in Figure 5.7.

Assume all stores are initially empty (dry) so rainfall at rate P for a unit duration will start to fill the stores. During the interval of raining the stores with depth less than the P will start generating direct runoff and gradually stores with larger depth and finally stores of depth P will produce direct runoff. The direct runoff production also can be represented as:

$$q(t) = \pi(t)F(C^*(t)) \quad \text{Eq 5.3}$$

where $\pi(t)$ is net rainfall rate at time(t), and $\pi(t) \ = P(t) - E(t)$, if we assume the precipitation $P(t)$and evaporation $E(t)$ occur at constant rate at time (t). The critical capacity will increase over the interval $(t, t + \Delta t)$ according to:

$$C^*(\tau) = C^*(t) + \pi(\tau - t) \qquad \text{Eq 5.4}$$

The volume of direct runoff per unit area during the interval $(t, t + \Delta t)$ will be:

$$V(t + \Delta t) = \int_t^{t+\Delta t} q(\tau) d\tau = \int_{C^*(t)}^{C^*(t+\Delta t)} F(c) dc \qquad \text{Eq 5.5}$$

and the water in storage over the basin as a whole, $S(t)$ will be :

$$S(t) = \int_0^{C^*(t)} \big(1 - F(c)\big) dc \qquad \text{Eq5.6}$$

So S_{max}, the total available storage is given by:

$$S_{max} = \int_0^{\infty} \big(1 - F(c)\big) dc = \overline{c} \qquad \text{Eq 5.7}$$

where $\overline{c}$ is the mean storage capacity over the basin.

For a given value of storage $S(t)$, $C^*(t)$ can be calculated and then the volume of direct runoff $V(t + \Delta t)$ calculated by Eq 5.4 and Eq 5.5.

Further loss can be introduced to the model as a recharge to groundwater. Assume the rate of drainage over the interval $(t, t + \Delta t)$ is $d(t)$ then the net rainfall at time (t) is defined as:

$$\pi(t) = P(t) - E(t) - d(t) \qquad \text{Eq 5.8}$$

There are several types of distribution that can be used to represent the soil moisture storage capacity variation over the catchment such as Pareto, Rectangular, Exponential, Lognormal and Triangular distributions. The Pareto distribution was suggested for the PDM standard model, which is widely used in practice, and we will use it in this study. The probability density function and the cumulative distribution function for the Pareto distribution were given in Eq 2.6 and Eq 2.5 and the following relations can be obtained from Eq 5.4 to Eq 5.8:

$$S_{max} = \frac{c_{max}}{b + 1}$$

$$S(t) = S_{max} \left[1 - \left(1 - \frac{C^*(t)}{c_{max}} \right)^{b+1} \right]$$

$$C^*(t) = c_{max} \left[1 - \left(1 - \frac{S(t)}{S_{max}} \right)^{\frac{1}{b+1}} \right]$$

$$V(t + \Delta t) = \pi \Delta t - S_{max} \left[\left(1 - \frac{C^*(t)}{c_{max}} \right)^{b+1} - \left(1 - \frac{C^*(t + \Delta t)}{c_{max}} \right)^{b+1} \right]$$

Figure 5.7, Figure 5.8 and Figure 2.6 show that the shape of this distribution varies with a parameter b. When $0 < b < 1$ there is more deep storage than shallow storage, when $b > 1$ there is more shallow storage.

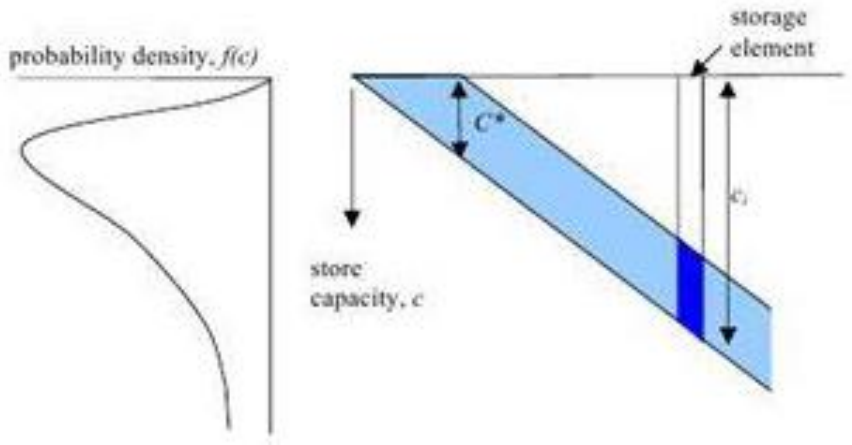


Figure 5.7 Storage elements of different depth and their associated probability density function (CEH, 2012).

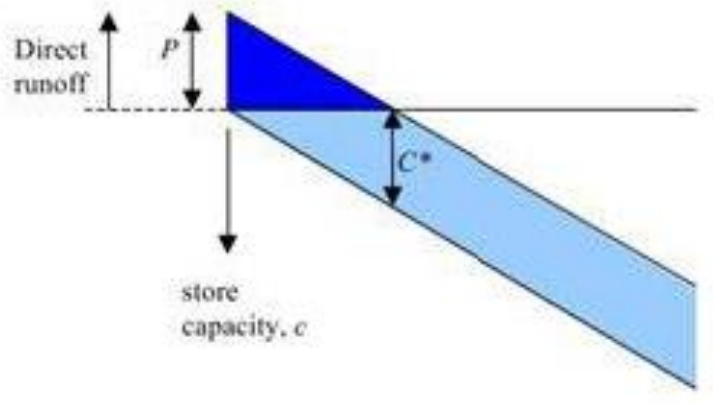


Figure 5.8 Direct runoff production from population of stores (CEH, 2012)

5.4 Surface and subsurface storage

Runoff generated from the saturated probability distributed stores contributes to the surface storage, representing routing of water via fast pathways to the basin outlet. This is usually represented in the PDM by a cascade of two linear reservoirs recast as an equivalent transfer function model (O'Connor, 1982). The groundwater storage, representing routing of water to the

basin outlet via slow pathways, is usually taken to be of cubic form, with outflow proportional to the cube of the amount of water in the store. The outflow from surface and groundwater storages, together with any fixed flow representing, say, compensation releases from reservoirs or constant abstractions, forms the model output.

5.5 Objective Functions

The PDM model provides two ways of model calibration which is obtained by comparing the observed and modelled flows. One way is manual adjustment to reduce the visual differences between observed and modelled flows and another way is automatic optimisation to minimise an objective function which is defined as a function to measure the difference between the observed and computed flows by the model. The most common functions used to measure the model performance are RMSE and R^2 Efficiency. If Q_t is the observed flow and q_t is the computed flow at time *t*

- the Root Mean Square Error (RMSE) is defined as:

  RMSE = $\sqrt{(n^{-1}\sum e_t^2)}$ where the model error is $e_t = Q_t - q_t$
- R^2 statistic defined as:

  $R^2 = 1 - \frac{\sum e_t^2}{\sum(Q_t - \overline{Q})^2}$ where $\overline{Q}$ is the mean of n observed flows. The R^2 indicates the proportion of variability in observed flows which the model can account for. A value of one indicates a perfect model while a value of zero indicates the model is only as good as using the mean of observed flows as a predictor. R^2 can be negative which indicates the model is even worse than using the mean of observed flows as a predictor.

Table 5.2 shows the most useful objective functions used for model calibration. The variable **objtyp** specifies the type of objective function to be used. Objective function type 21 is generally recommended.

Table 5.2 The most useful objective functions available within the Model Calibration Shell (CEH, 2012)

Objtyp (this is entry in input file)	**Description**	**When to use this type of objective function**
21 (or 221 for daily flows)	Root mean square error(RMSE) $\sqrt{n^{-1}\sum e_i^2}$ where $e_i = Q_i - q_i$	**Recommended** Has the advantage of giving reasonable weight to errors in high flows compared to errors in lower flows .provides an error measure in the units of river flow (or level).
22 (or 212 for daily flows)	RMSE of log(base *e) error* $\sqrt{n^{-1}\sum e_i^2}$ where $e_i = \ln(Q_i) - \ln(q_i) = \ln(Q_i/q_i)$	**Good for a special use.** Use this option if you want to pay attention to lower flows. The use of logs makes it more sensitive to errors in low flows.
24 (or 214 fo daily flows)	RMSE of error in square roots $\sqrt{n^{-1}\sum e_i^2}$ where $e_i = \sqrt{Q_i} - \sqrt{q_i}$	Use this option as an intermediate between ordinary and logarithmic errors.
31 (or 231 for daily flows)	Mean absolute error $n^{-1}\sum \lvert e_i \rvert$ where $e_i = Q_i - q_i$	**Good for a special use.** Compared to those based on squared errors, places less emphasis on reducing the largest errors. Can remove distortion in calibration caused by a few large errors. Provides an error measure in the units of river flow (or level).

5.6 PDM model Updating Methods

The PDM model is said to operate in simulation mode if observed flows are used for initialisation the model and transforms the rainfall and potential evaporation input to river flow. In order to improve model performance, current and past measured values of flow are used to continuously update the value of modelled flow. Two updating methods are used in PDM: state correction and error prediction. State correction is suitable for smaller catchments while error prediction may be better for larger catchments. For some cases state correction may be preferred as it needs only two extra parameters and automatic optimisation can calculate them.

5.6.1 State correction

The term "state" is used to describe a variable of a model which mediates between inputs to the model and the model output (Sozollosi-Nagy, 1976). In the PDM model, the main input is rainfall and output is basin flow. So state variables are the water content of the probability distributed soil storage, the surface and groundwater stores ($S_1, S_2\ and\ S_3$) and the flow out of the surface storage and flow out of the groundwater storage ($q_s\ and\ q_b$). There is always some error between the actual (observed) value and the predicted (modelled) value due to for example, errors in measuring the observed value and uncertainty of the model. In rainfall runoff models such as the PDM model, the error occurs between the model predictions flows q and the observed value of runoff Q as follows:

$$\varepsilon = Q - q = Q - (q_s + q_b) \qquad \text{Eq 5.9}$$

The procedure of state correction can be applied to minimize this error. The Kalman filter algorithm is a formal approach for state correction (Jazwinski, 1972) which provides an optimal adjustment scheme for incorporating observations, through a set of linear operations for linear dynamic models and an extended form of Kalman filter for nonlinear dynamic models such as the PDM. By apportioning the error, ε between the surface and groundwater stores in proportion to their contribution to total flow as follows:

$$q_b^* = q_b + \alpha\, g_b\, \varepsilon \qquad \text{Eq 5.10}$$

$$q_s^* = q_s + (1 - \alpha)\, g_s\, \varepsilon \qquad \text{Eq 5.11}$$

where the superscript * indicates the value after adjustment, the "gain" coefficients, g_b and g_s and α is:

$$\alpha = \frac{q_b}{q_s + q_b} \qquad \text{Eq 5.12}$$

If g_b and g_s are equal to one, then the $q_b^* + q_s^*$ is equal to the observed value flow. Automatic optimisation can be applied to get other values (not equal to one) of model parameters g_b and g_s to find the best fit between state- adjusted predictions and observed flows. So α can be defined as:

$$\alpha = \frac{q_b}{\beta_1 q_s + \beta_2 q_b} \qquad \text{Eq 5.13}$$

where β_1 and β_2 are parameters to weight the apportionment to the flow components.The recommended values for β_1 and β_2 are 10 and 1.1 to apportion more of the error adjustment to the

surface store. An adjustment to the probability distributed soil moisture may also be made, either of the proportional form:

$$S_1^* = S_1 + \alpha\, g_g\, \varepsilon \qquad \text{Eq 5.14}$$

or, the direct form of gain with α equal to unity.

5.6.2 Error prediction

Error prediction is an alternative method in PDM for forecast updating in real-time. Let $q_{t+\ell}$ be a predicted value of the observed flow $Q_{t+\ell}$ at time $t+\ell$ obtained by running PDM in simulation mode, which transform rainfall into flow. The error $\eta_{t+\ell}$ associated with this simulation mode forecast is defined as:

$$Q_{t+\ell} = q_{t+\ell} + \eta_{t+\ell} \qquad \text{Eq 5.15}$$

Let $\eta_{t+\ell|t}$ (the suffix notation $t+\ell|t$ should be read as a forecast at time $t+\ell$ given information up to time t) denote a prediction of the simulation mode error $\eta_{t+\ell}$, ℓ steps ahead from a forecast origin at time t using an error predictor. Then $q_{t+\ell|t}$ a real-time forecast made ℓ steps ahead from a forecast origin at time t is given by:

$$q_{t+\ell|t} = q_{t+\ell} + \eta_{t+\ell|t} \qquad \text{Eq 5.16}$$

and the real-time forecast error is given by:

$$a_{t+\ell|t} = Q_{t+\ell} - q_{t+\ell|t} \qquad \text{Eq 5.17}$$

which is smaller than the simulation-mode forecast error (Eq 5.15 and Eq 5.16).
An autoregressive (AR) model and autoregressive-moving average (ARMA) model can be used for error prediction. In an AR model, the future value of a variable is assumed to be a linear combination of *p* past observations and a random error with a constant term so the simulated mode error at time *t* can be mathematically expressed as:

$$\eta_t = -\phi_1\eta_{t-1} - \phi_2\eta_{t-2} - \cdots - \phi_z\eta_{t-p} + a_t \qquad \text{Eq 5.18}$$

where $a_t$is the residual error (uncorrelated) and $\{\phi_i : i = 1,,2,\ldots,p\}$ are model parameters and the number of parameters *p* is known as the order of the model (Agrawal, 2013). A moving average (MA) model uses the past errors as the explanatory variables so the MA model can be expressed as:

$$\eta_t = \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} + a_t \qquad \text{Eq 5.19}$$

where a_t is the residual error (uncorrelated) and $\{\theta_i : i = 1,,2, \ldots., q\}$ are model parameters, and the number of parameters q is the order of the model. An ARMA model is a combination of AR and MA models, which can represented as:

$$\eta_t = -\phi_1 \eta_{t-1} - \phi_2 \eta_{t-2} - \cdots - \phi_p \eta_{t-p} + \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} + a_t \quad \text{Eq 5.20}$$

The ARMA model may be used to give the following error predictor:

$$\eta_{t+\ell|t} = -\phi_1 \eta_{t+\ell-1|t} - \phi_2 \eta_{t+\ell-2|t} - \cdots - \phi_p \eta_{t+\ell-p|t} \qquad \text{Eq 5.21}$$

$$+ \theta_1 a_{t+\ell-1|t} - \theta_2 a_{t+\ell-2|t} - \cdots - \theta_q a_{t+\ell-q|t} \qquad \ell = 1,2, \ldots$$

where

$$a_{t+\ell-i|t} = \begin{cases} 0 & \ell - i > 0 \\ a_{t+\ell-i} & \ell - i \leq 0 \end{cases} \qquad \text{Eq 5.22}$$

and $a_{t+\ell-i}$ is the one-step ahead predictor error:

$$a_{t+\ell-i} = a_{t+\ell-i|t+\ell-i-1} = \eta_{t+\ell-1} - \eta_{t+\ell-i|t+\ell-i-1}$$

$$= Q_{t+\ell-i} - q_{t+\ell-i|t+\ell-i-1} \qquad \text{Eq 5.23}$$

and $= \eta_{t+\ell-i|t} = \eta_{t+\ell-i} = Q_{t+\ell-i} - q_{t+\ell-i}$ for $\ell - i \leq 0$ Eq 5.24

Eq 5.21 can be used recursively to calculate the error prediction $\eta_{t+1|t}$, $\eta_{t+2|t}$, … , $\eta_{t+\ell|t}$ from available values of a_t , a_{t-1} , … and η_t, η_{t-1} , ….

The PDM model simulation-mode value is $q_{t+\ell}$. The error prediction $\eta_{t+\ell|t}$ is obtained by Eq 5.21. Then the real-time forecast $q_{t+\ell|t}$ is calculated by Eq 5.16.

All of the above can be studied in more detail in the PDM manual and in the Moore's papers.

5.7 Model Calibration

The PDM model is normally run within the Model Calibration Shell environment and calibrated using interactive menus of Tscal-monitor (Tscal for Time-series Calibration). We need to provide two files to run and calibrate the model.

- Model input file (file extension is .inp e.g test.inp)

- File-1 Data Interface file (file extension is .fli e.g test.fli)

and Model data (e.g River flow data, rain gauge data,etc). Figure 5.9 illustrates the PDM model configuration and data requirements. For simplicity the model input files and File-1 Data Interface files are referred to by *.inp and *.fli respectively.

5.7.1 Input file

The input file contains model configuration information and event dates. The parameters and structure options that required to run the standard PDM model are summarised in Table 5.3, with suggested values for these parameters. The model parameters including the model specification variables and details of the external hydrometric data sources are defined in an input file (*.inp). The input file can be divided into 8 parts as follows:

1. Description, analysis and output required.
2. Model parameter list.
3. Contour plot generation.
4. Extended groundwater model.
5. Data source selection and model structure.
6. Event selection.
7. Error setting.
8. Forecast settings.

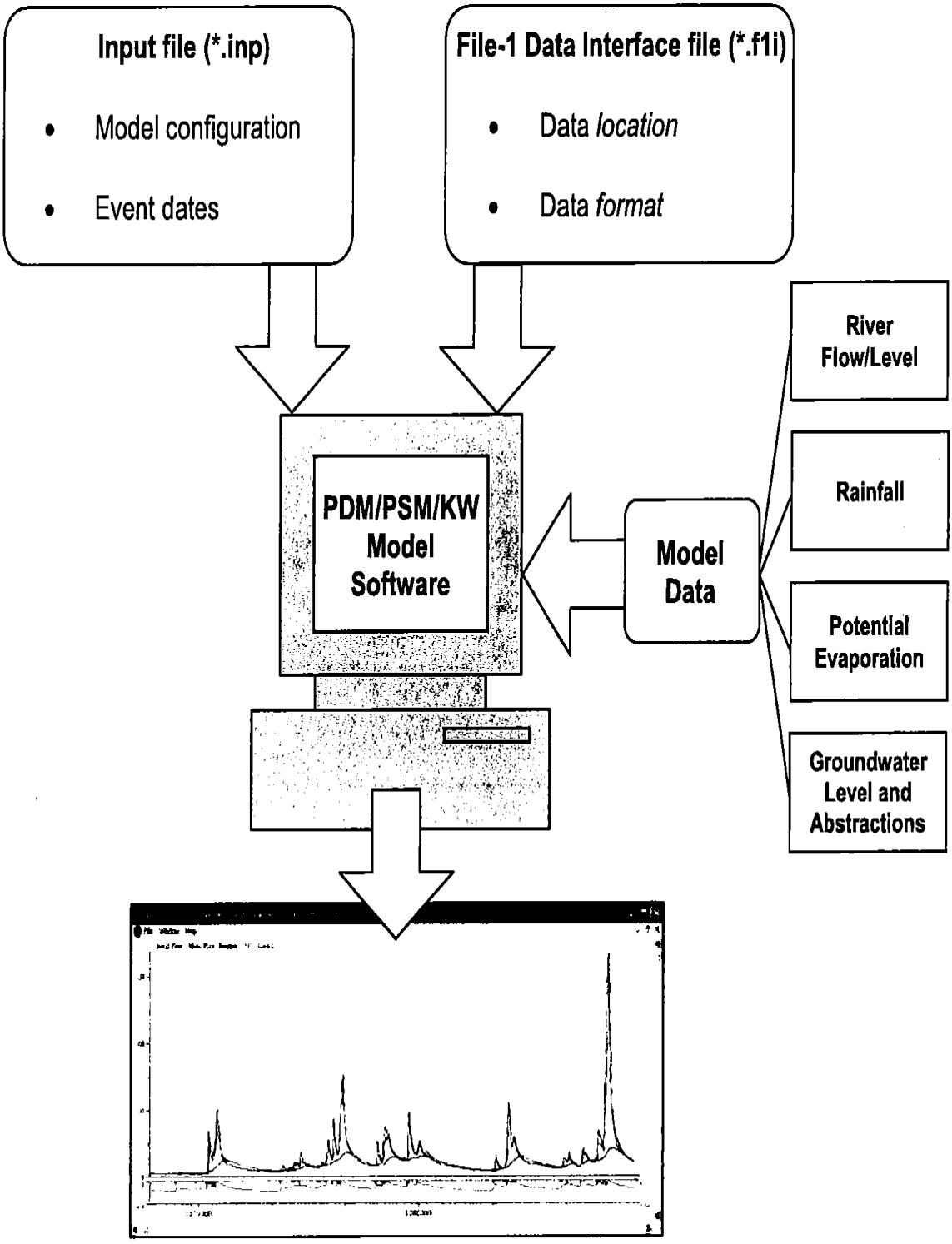


Figure 5.9 PDM model configuration and data requirement (CEH, 2012)

Table 5.3 Standards PDM model parameters (CEH, 2012)

Param-eter	Referred to input file as	Unit	Description	Function	Suggested values
f	Rainfac	none	Rainfall factor	Controls runoff volume	1
τ_d	Tdly	h	Time delay	Shifts hydrograph vertically along flow axis	0
Probability distributed store					
C_{min}	Cmax	mm	Minimum store capacity	Affect time of onset of runoff and rate of wetting up, but also feeds back to evaporation and recharge rates	0
C_{max}	Cmin	mm	Maximum store capacity		75
b	B	none	Exponent of Pareto distribution controlling spatial variability of store capacity		
Evaporation function					
b_e	Be	none	Exponent in actual evaporation function	Affect variation between seasons or years	2.5
Recharge function 1: standard					
k_g	Kg,sigg	hmm^{b_g-1}	Groundwater recharge time constant	Controls rate of aquifer recharge	10^5
b_g	Bg	none	Exponent of recharge function	Increase to magnify sensitivity of recharge rate to soil dryness	1.5
S_t	St	mm	Soil tension storage capacity	Increase to prevent complete drainage of soil moisture store and maintain evaporation	0
2: Demand-based					
α		none	Groundwater deficit ratio threshold		
β		none	Exponent in groundwater demand factor function		
q_{sat}		mmh^{-1}	Maximum rate of recharge		
3: Spliting					
α		none	Runoff factor controlling the split of rainfall to surface and groundwater storage routing when no soil recharge is allowed		
Surface routing					
k_1	K1,sigs	h	Time constants of cascade of two linear reservoris	Controls peakiness of hydrograph	1-20
k_2	K2,sigs	h			1-20
Groundwater storage routing					
k_b	Kb,sigs	hmm^{m-1} (m ≠3)	Baseflow time constant	Controls length of recession	10-200 (m = 2)
m		none	Exponent of baseflow nonlinear storage		
q_c	qconst	m^3s^{-1}	Constant flow representing returns/abstractions	Shifts hydrograph vertically along flow axis	0

The description of the model parameter list is shown in Table 5.4. Figure 6.1 shows the input file for running PDM for the Duarrigle subcatchment.

Table 5.4 Input file Part2, Model parameter list

Parameter	Description
indx	Switch to allow parameter value to vary(1) or not(0)
xs	Model parameter value: initial value for parameter.
smpstp	Initial step size for simplex optimisation.
smptol	Convergence tolerance for simplex optimisation.
xs name	Name for parameter(in single quotes)
/	Slash terminating list of model parameters.

Adjusting the values of these parameters in the input file can be obtained by trial and error. To adjust a parameter, set indx (first column in input file) to 1. The following guidance can help to choose these parameters:

- *'Objtyp'* indicates the type of objective function for automatic optimisation (Table 5.2)
- If the PDM with the interactive mode is used '*maxfun'* is set to 9999, otherwise to number of function evaluations during calibration (200 is recommended)
- Variable analyse indicates the unit of measurement for analysing the time-series and generating plots, usually set to two stands for units of $m^3 s^{-1}$.
- Table 5.5 indicates the setting of parameters for standard PDM model structure (configuration). Alternative model configurations are explained in section 5.7.3.
- To invoke the multiple rain gauges instead of name of a rain gauge under the title of rainfall, use the key word "MULTIPLE" then specify the number of rain gauges and their names and also rain gauge weight contributing area as described in section 5.2.2 and 5.2.3.

Table 5.5 Setting parameters for standard PDM model configuration (CEH, 2012)

Parameter in input file	setting	Meaning
Loss model(indl)	1	Pareto distribution
Runoff model(indr)	0	Normal PDM
Surface model(inds)	22	Cascade of 2 linear reservoirs
Baseflow model(indb)	3	Cubic storage

5.7.2 File -1 Data interface file

The location and format of the required data such as rain gauge data are specified in *.fli. A variety of data formats can be handled in *.fli. In this study 'text' file was used. Figure 5.10 shows the format of these data.

Year	Mo	Day	Hr	Mi	Se	Value
2010,	11,	5,	19,	0,	0,	0
2010,	11,	5,	19,	15,	0,	0.2
2010,	11,	5,	19,	30,	0,	0.4
2010,	11,	5,	19,	45,	0,	0
2010,	11,	5,	20,	0,	0,	0.2
2010,	11,	5,	20,	15,	0,	0.2
2010,	11,	5,	20,	30,	0,	0.2
2010,	11,	5,	20,	45,	0,	1
2010,	11,	5,	21,	0,	0,	0.2
2010,	11,	5,	21,	15,	0,	0

Figure 5.10 Example of rain gauge data and river data:

Format (year, month, day, hour, minute, second and value). Value could be rainfall, river flow or river level

5.7.3 Alternative model configurations

The standard PDM model structure with a Pareto distribution of storage depth, a cubic groundwater storage and a cascade of two linear reservoirs for surface storage is usually the appropriate model for many catchments (Table 5.5). There are some alternative model configurations which may be appropriate for specific catchments.

1. Alternative distribution function (*indl*)

   Despite the option of alternative distributions, the Pareto distribution is normally used while others are unlikely to be used such as the triangular, power, lognormal and exponential distributions.

2. Alternative runoff models (*indr*)

   Table 5.6 shows the different way of partitioning of rainfall between soil storage, fast and slow response paths. If *indr = 1* the proportion of runoff entering the surface store is α and

the proportion entering the base flow store is $(1 - \alpha)$. When the runoff splitting option is used (*indr = 1 or 2)* then two additional parameters need to be used. Alpha (α): proportion of runoff entering surface flow store, the recommended value is 0.5. Theta (θ): Initial value of moisture storage as proportion of maximum storage with recommended value of 0.7 to 1. If *indr = 2,* the proportion entering the base flow store is unrestricted and defined by an additional parameter β and not $(1 - \alpha)$.

3. Surface-flow models (*inds*)
   Non-identical reservoirs by setting the time constant k_2 to a non-zero positive value can provide great flexibility in the shape of the hydrograph. Figure 6.5 and Figure 6.6 (a) and (b) show hydrographs when k_2 is non-zero positive value.

4. Base flow models (*indb*)
   For small upland catchments in cases where the recession in not so sustained quadratic (*indb* = 2) or exponential (*indb* = 4) storage functions can be useful.

Table 5.6 Alternative runoff models (indr) (CEH, 2012)

Indr (runoff)	**Description**	**Use**
0	Normal PDM,Direct runoff goes to surface flow store, groundwater recharge goes to the baseflow store	Preferred choice for most applications
1,2	No groundwater recharge. Simple split of direct runoff between fast and slow paths.	May be useful for largely impermeable catchments with little baseflow
3,4	Direct runoff to fast path with demand moderated recharge to slow (aquifer path)	Aquifer-dominated catchments.

5.8 PDM Model Inputs

PDM is generally applied at a 15-minute time-step. The daily time step can be useful when 15-minute time step data are only available for the short periods of storm events (CEH, 2012). To run the PDM model generally we need three main inputs:

1) rainfall data
2) river flow data
3) potential evaporation data

Rainfall data from at least one rain gauge is required. In this study the rainfall data from 23 rain gauges across the Blackwater catchment were used. Analysis and converting these data to a format which is suitable for the PDM Model was described in chapter 4.

River flow data are required to compare the simulation output with the observation to calibrate/validate the model. These data are derived from river heights measured by automatic stations and transformed into flow data using rating curves. The standard way to run the PDM is using river flow data. When only river level data are available then there are two ways in PDM to convert the river levels to river flows. If the rating curve is known (preferred option) then it can be included in the configuration file (PDM calls this configuration file as an input file). If the rating curve is unknown then the river level can be fitted to a rating curve by automatic optimization, in PDM start with sensible initial parameters and then apply automatic optimization.

There is an alternative way to convert the river level to river flow using a rating curve before running the PDM. This method was used to obtain river flow in advance.

Required potential evaporation data can be obtained from external sources such as climatological observation station making daily observation or can be internally generated by a yearly sine curve with average value over the year. In this study, the sin curve PE profile over the year with averaging 1.4 is used. Figure 5.11 illustrates monthly mean potential evaporation from 2016 to 2019 at Cork Airport station, which was obtianed from Met Éireann website (monthly data section). Yearly PE is 516.3 *mm* and daily average is 1.4 (516.3/ 365 = 1.4).

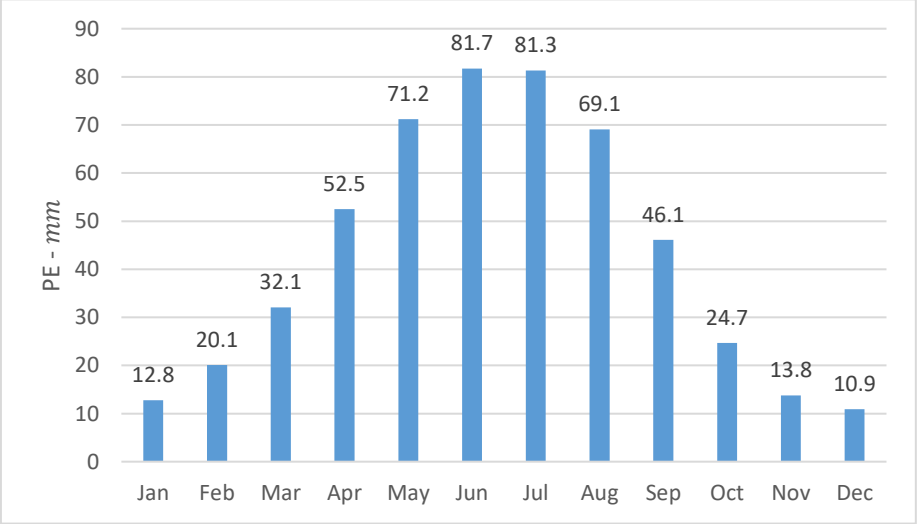


Figure 5.11 Monthly meanPotential Evaporation 2016 to 2019 at Cork Airport

Chapter 6 PDM Calibration and Validation

6.1 Subcatchments of the Munster Blackwater

We examine four of the Munster Blackwater subcatchments. These are subcatchments to Duarrigle (area 247km^2), to Dromcummer (area 862 km^2), to Mallow Rail BR (area 1168 km^2) and to Killavullen (area 1248 km^2). Table 6.1 shows available quality controlled river data in each subcatchment. The river flow data are available for Killavullen. The river flow data were defined by applying rating curves to water level recorded at Duarrigle, Dromcummer and Mallow Rail BR. The river level data are available for the period 2010 to 2013 in Duarrigle, while river flow/river level data for the other three subcatchments are available for the period 2010 to 2018. Rainfall data for all subcatchments are available for the period 2010 to 2017. Therefore, the PDM hydrological model is run for the last three subcathments of Blackwater catchment (Dromcummer, Mallow Rail BR and Killavullen) over the period 2010 to 2017. The Duarrigle subcatchment was omitted for analysis due to data unavailability for the period 2013 to 2017.

Table 6.1 Available quality controlled river level/river flow data in each subcatchment.

	15-Minute Rainfall	River level/river flow
Duarrigle	01/01/2010 -22/10/2017	01/01/2010 -26/12/2013
Dromcummer	01/01/2010 -22/10/2017	01/01/2010 -27/09/2018
Mallow Rail BR	01/01/2010 -22/10/2017	01/01/2010 -31/08/2018
Killavullen	01/01/2010 -22/10/2017	01/01/2010 -31/08/2018

6.2 Model Calibration

The PDM model was first applied to the Dromcummer subcatchment of area 862.50 km^2 using a 15-minute time step of river flow data and precipitation from 11 raingauges over the period 2010 to 2017. The year 2014 was chosen for detailed analysis.

To represent the PDM functionality the PDM was run for Dromcummer for 2014, as this year has the best data quality for both precipitation and river level/river flow. This was done as a calibration, step by step to show both manual and automatic optimisation. Adding or adjusting different parameters to the model can help to improve the model simulation (see Appendix A.1 for how to prepare to run

the PDM model in more detail). Figure 6.1 illustrates the input file (*.inp) with recommended initial parameters. Figure 6.2 shows the modelled river flow generated by running PDM for this input file. The modelled flow show peaks that are too high and low flows that are too low. The measure of performance indicates the quality of model performance, $R^2 = 0.7047$, which is good but still not sufficient. To improve the model performance, some facts about the area and trial and error of parameters adjustment was used. Some of the parameters were manually adjusted and then automatic optimisation was applied. PDM provides an automatic optimisation method to improve the model accuracy through an iterative procedure from the initial parameters. The performance of the optimisation depends on the choice of initial parameters. If the initial parameters are poorly adjusted then optimisation might not converge. The combination of manual adjustment of parameters with automatic optimisation is the optimum way to adjust the parameters to improve the performance of the model.

The parameter adjustment after a mixture of manual and automatic optimisation is shown in Figure 6.3, the resulting hydrograph is shown in Figure 6.4 and Figure 6.4(b) is shown a zoom in a particular region of the hydrograh. $R^2 = 0.9576$ indicates sufficient improvement in the model from an initial R^2 of 0.7047. Table 6.2 illustrates the comparison between these two model runs.

This process was repeated for Mallow Rail BR with an area of 1168.65 km^2 and using 15 rain gauges. The model was also run for Killavullen with an area of 1248.32 km^2 and using 17 rain gauges over the year 2014 as a calibration. Figure 6.5 and Figure 6.7 show parameters adjusted after this process and the resulting hydrographs are shown in Figure 6.6 and Figure 6.8.

```
indx        xs            smpstp          smptol       xs name
   1       1.0000         5.00000E-02     1.00000E-02  'rainfac'
   1      3.33508E-03     4.00000E-03     8.00000E-04  'cmin'
   1       43.445        -0.43445         8.00000E-02  'cmax'
   1      0.16433         1.00000E-02     2.00000E-03  'b'
   1       88.740        -0.88740         1.00000E-02  'be'
   1       9.0275         9.02752E-02     1.80550E-02  'k1'
   1       5.1081        -5.10808E-02     1.02162E-02  'k2'
   1       45.210         0.45210         8.00000E-02  'kb'
   1       2166.5         -21.665          3.7997      'kg'
   1       23.083         0.23083         2.00000E-03  'St'
   1       2.2647        -2.26472E-02     2.00000E-03  'bg'
   1      0.74125         7.41246E-03     1.48249E-03  'qconst'
   1      0.54890        -5.48899E-03     1.09780E-03  'tdly'
 /
```

Figure 6.3 Part of the PDM input file for Dromcummer subcatchment after manual adjustment and automatic optimisation. See Table 5.4 for parameter descriptions.

```
PDM: BlackWater at Dromcumnr (Munster)
 objtyp  maxfun  konvge    objtol
  11     1000       4     8.00000
analys    qmin      qmax       pkmin    pkrge
   2      0.00     1000.0        8.00    48
 prtfnf  prtopt   common file name of plot files
   0       3     'USER_EVPLOT'
indx     xs    smpstp  smptol    xs name (<= 8 chars)
 1      1.0     0.01     0.05    'rainfac'
 1      0       0.2      0.02    'cmin'
 1      75     1.20      0.8     'cmax'
 1     2.       0.1      0.05    'b'
 1     2.5      0.3      0.01    'be'
 1    8.        0.1      0.02    'k1'
 1    3.        0.04     0.09    'k2'
 1    150       0.4      0.08    'kb'
 1 100000.00   306.0    10.00   'kg'
 1    0        0.23      0.002    'St'
 1    1.5      0.013     0.002   'bg'
 1    0.0      0.01      0.002   'qconst'
 1    0.0      0.01      0.002   'tdly'
/
PROBABILITY DISTRIBUTED LOSS MODEL
 Flow    Rainfall   Climate-station   Daily Rain
'Drom_flow_10_17'  'MULTIPLE'  'sin curve    1.40000'  'NULL'
loss  runoff  surface  base  tdly   i_sdp  nsdr  sqmodel  oap
  1     0     22      3    -13      0      0    'NONE'  'FLOW'
Multiple gauges (number)
 11
'BMET_80711'  'BMET_80712'  'BMET_80716'  'BMET_80713'  'BMET_80705'  'BMET_80704'  'BMET_80717'
'BMET_80714'  'BMET_80710'  'BMET_80720'  'BMET_80708'
 Number rain zones and gauge weights
  1
 0.063 0.096 0.102 0.064 0.095 0.121 0.097 0.078 0.166 0.026
 0.092
Catchment area (km**2)
862.5
Event Begin       Event End       Noi     End 15min Rain
'00:00  01 Jan 2014'  '23:45 29 12 2014'    /
/
error parameters:  0(SIM) /,  or  1(UPD) upd miter gaintyp,  or  2(ARMA) p q
0 /
forecast mode:  0 = no forecast,  1 = fixed origin,  2 = fixed lead-time
0 /
forecast times ( 1 read-statement per event)
/
prtflg
-1
```

Figure 6.1 PDM input file for Dromcummer subcatchment with recommended initial values. See Table 5.3 and PDM User Manual for definition of model parameters

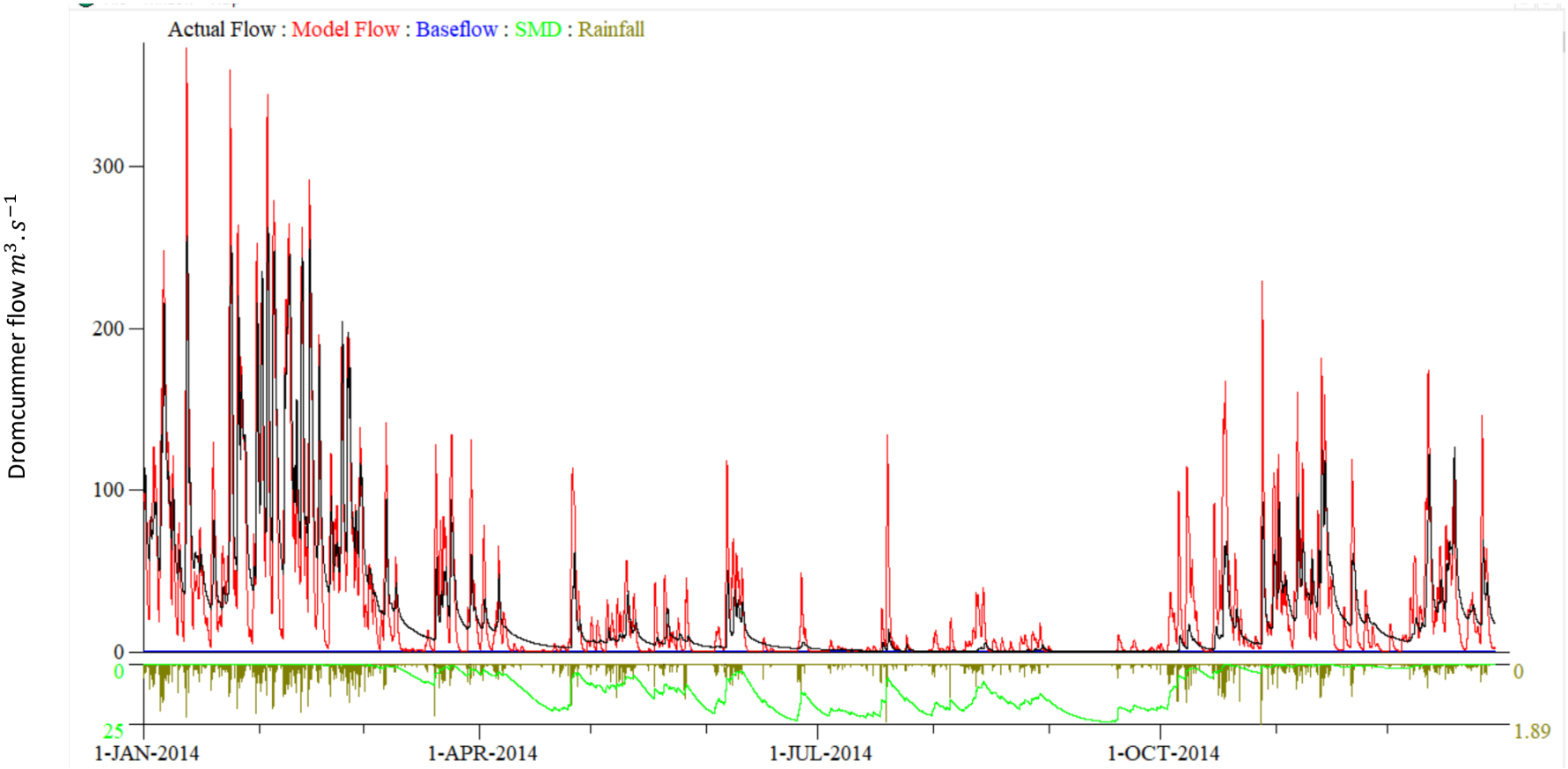


Figure 6.2 15-Minute precipitation & PDM simulation for 2014 at Dromcummer - observed versus modelled riverflow with recommended initial value (no manual adjustment or optimisation), $R^2 = 0.7047$

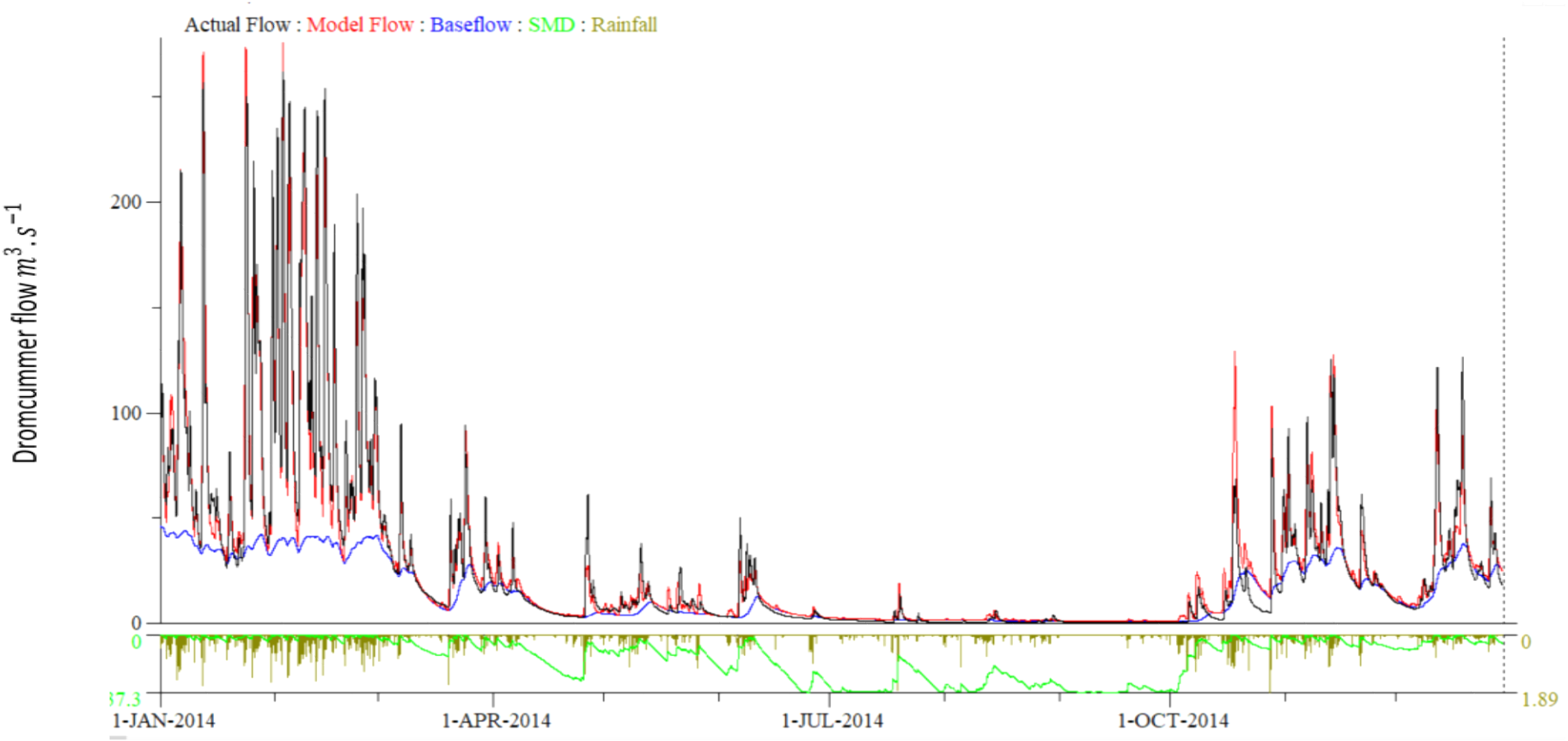


Figure 6.4 (a) 15-Minute precipitation & PDM simulation for 2014 at Dromcummer - observed versus modelled riverflow after manual parameter adjustment and automatic optimisation , $R^2 = 0.9576$.

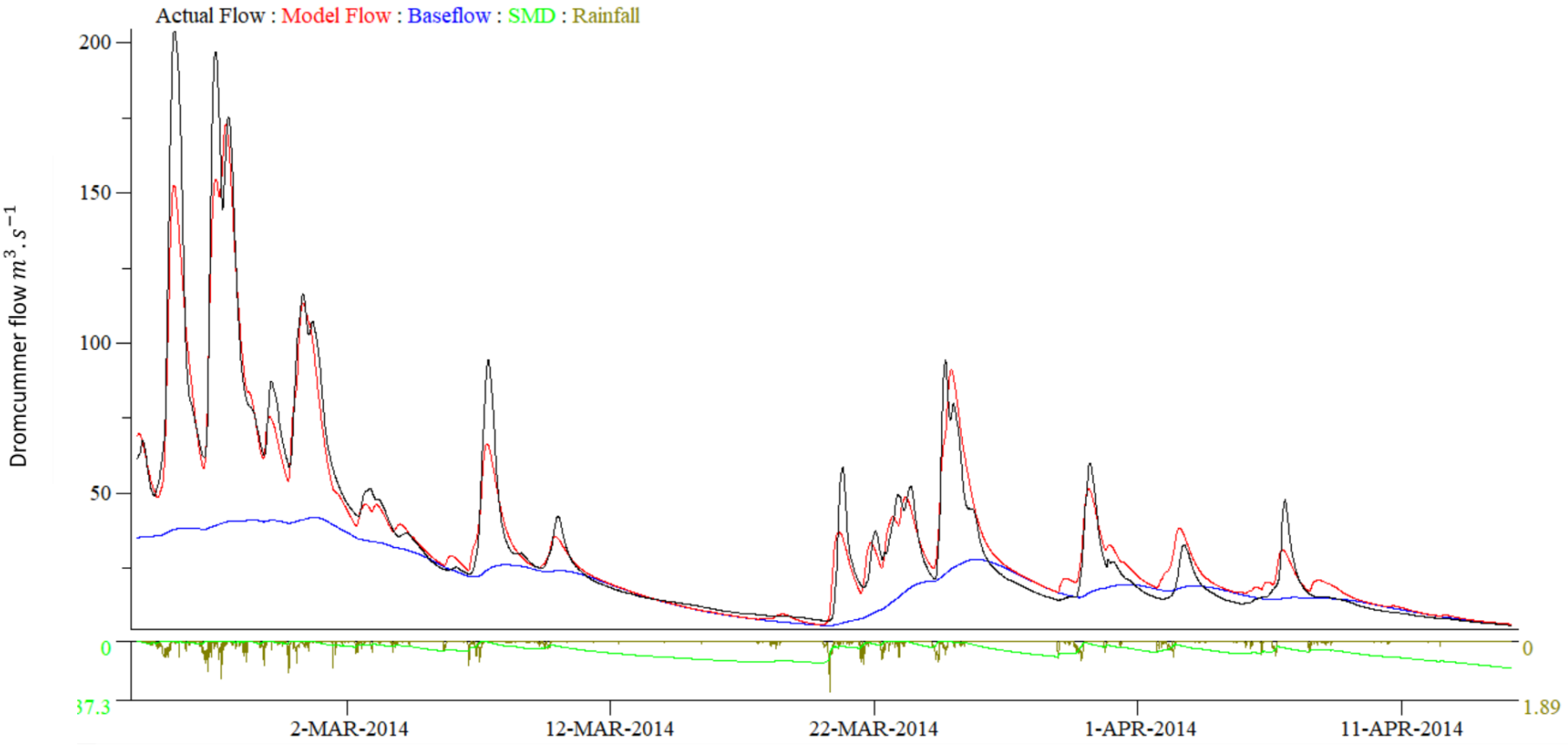


Figure 6.4 (b) 15-Minute precipitation & PDM simulation from March to April 2014 at Dromcummer - observed versus modelled riverflow after manual parameter adjustment and automatic optimisation $R^2 = 0.9576$. It is now noticeable that the modelled peaks are much closer to the observed peaks than in Figure 6.2 (the PDM run without manual adjustment and without automatic optimisation). The modelled low flows are also much closer to the observed flows. The baseflow is also much better.

Table 6.2 Summary of PDM Model outputs for Dromcummer for the full year of 2014 before and after manual and automatic optimisation. Note the significant improvement of the optimisation.

	R^2	Rainfall (mm)	Computed baseflow (mm)	Observed flow (mm)	Computed flow (mm)	Potential evaporation (mm)	Actual evaporation (mm)	Net rainfall (mm)
Default initialisation	0.704	1399.35	8.52	988.37	1044.94	511.00	361.87	1037
Manual and optimization	0.957	1399.35	513.52	988.37	1001.87	511.00	439.71	959.64

```
PDM: BlackWater at Mallow Rail BR(Munster)
 objtyp  maxfun  konvge    objtol
  11     1000        4     8.00000
analys    qmin       qmax         pkmin    pkrge
   2      0.00      1000.0          8.00    48
 prtfnf  prtopt   common file name of plot files
   0       3     'USER_EVPLOT'
indx  xs      smpstp       smptol       xs name
1  1.0000    5.00000E-02   1.00000E-02  'rainfac'
1  1.2281   -1.89606E-02   1.11309E-03  'cmin'
1  59.209    0.59209       8.00000E-02  'cmax'
1  0.19672  -1.96723E-03   3.64016E-04  'b'
1  386.79    3.8679        1.00000E-02  'be'
1  9.1712   -9.17123E-02   9.37700E-03  'k1'
1  5.9750    6.28698E-02   1.25740E-02  'k2'
1  86.496   -0.86496       7.19672E-02  'kb'
1  150.28    1.5028        0.29627      'kg'
1  39.481   -0.39481       2.00000E-03  'St'
1  1.5664    1.56636E-02   2.00000E-03  'bg'
1  7.9126   -7.91259E-02   1.65727E-03  'qconst'
1  3.5011    3.50114E-02   2.00000E-03  'tdly'
/
PROBABILITY DISTRIBUTED LOSS MODEL
 Flow    Rainfall   Climate-station   Daily Rain
'Mall_flow_10_17' 'MULTIPLE' 'sin curve 1.40000'  'NULL'
loss  runoff  surface  base  tdly    i_sdp  nsdr  sqmodel  oap
   1      0     22      3    -13      0      0    'NONE'  'FLOW'
Multiple gauges (number)
  15
'BMET_80711' 'BMET_80712'  'BMET_80716'  'BMET_80713'  'BMET_80705'
'BMET_80704' 'BMET_80717'  'BMET_80714'  'BMET_80710'  'BMET_80720'
'BMET_80708' 'BMET_80718'  'BMET_80702'  'BMET_80701'  'BMET_80715'
 Number rain zones and gauge weights
   1
0.0468  0.072  0.0765  0.0471  0.0699  0.0632  0.0692  0.0587  0.1244
0.06   0.0682  0.0863  0.0657  0.03    0.062
 Catchment area (km**2)
1168.65
Event Begin       Event End       Noi     End 15min Rain
'00:00  01 Jan 2014'  '23:45 29 Dec 2014'    /
/
```

Figure 6.5 PDM input file for Mallow Rail BR subcatchment after manual and automatic optimisation

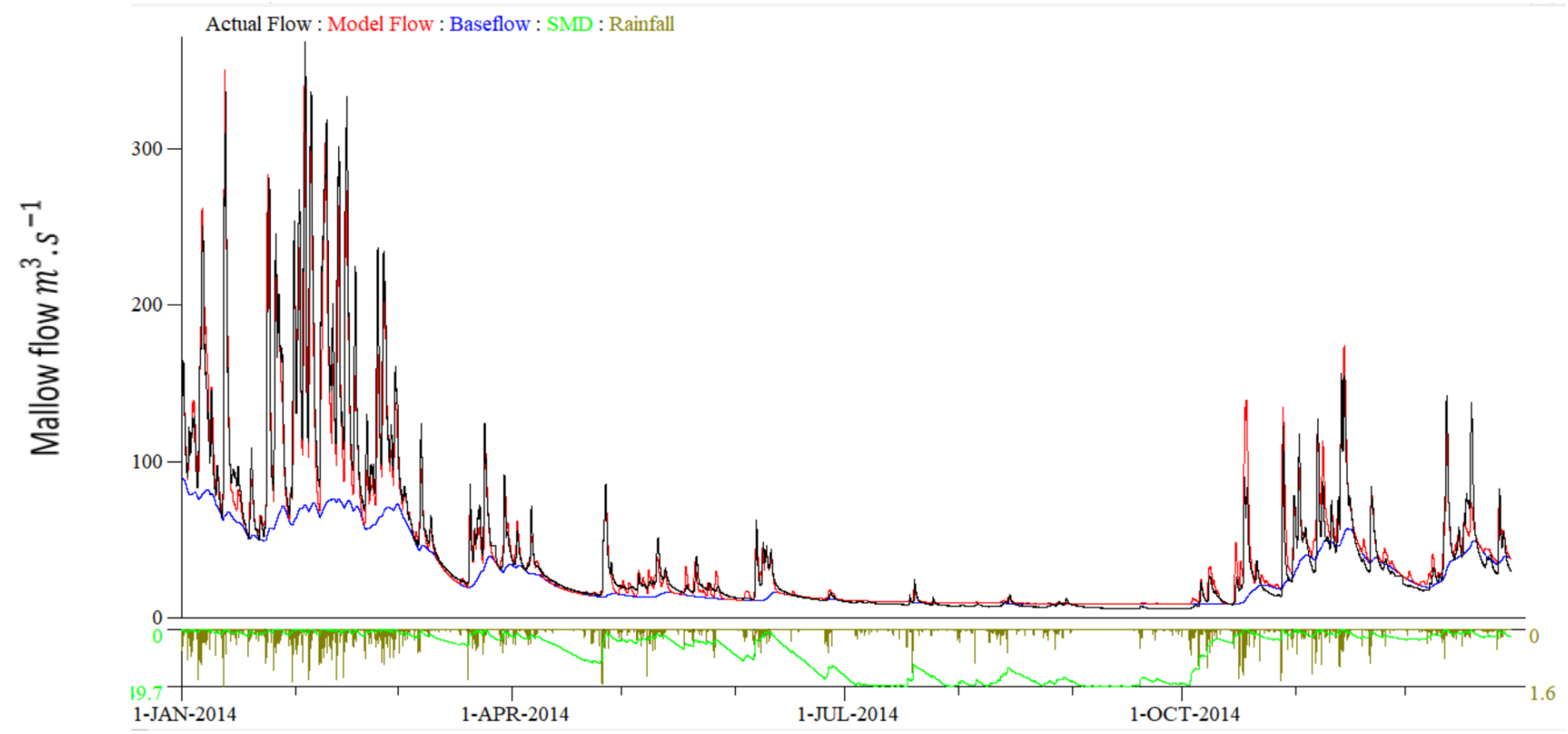


Figure 6.6 (a) 15-Minute precipitation & PDM simulation for 2014 at Mallow Rail BR - observed versus modelled riverflow after manual and automatic optimisation, $R^2 = 0.9558$

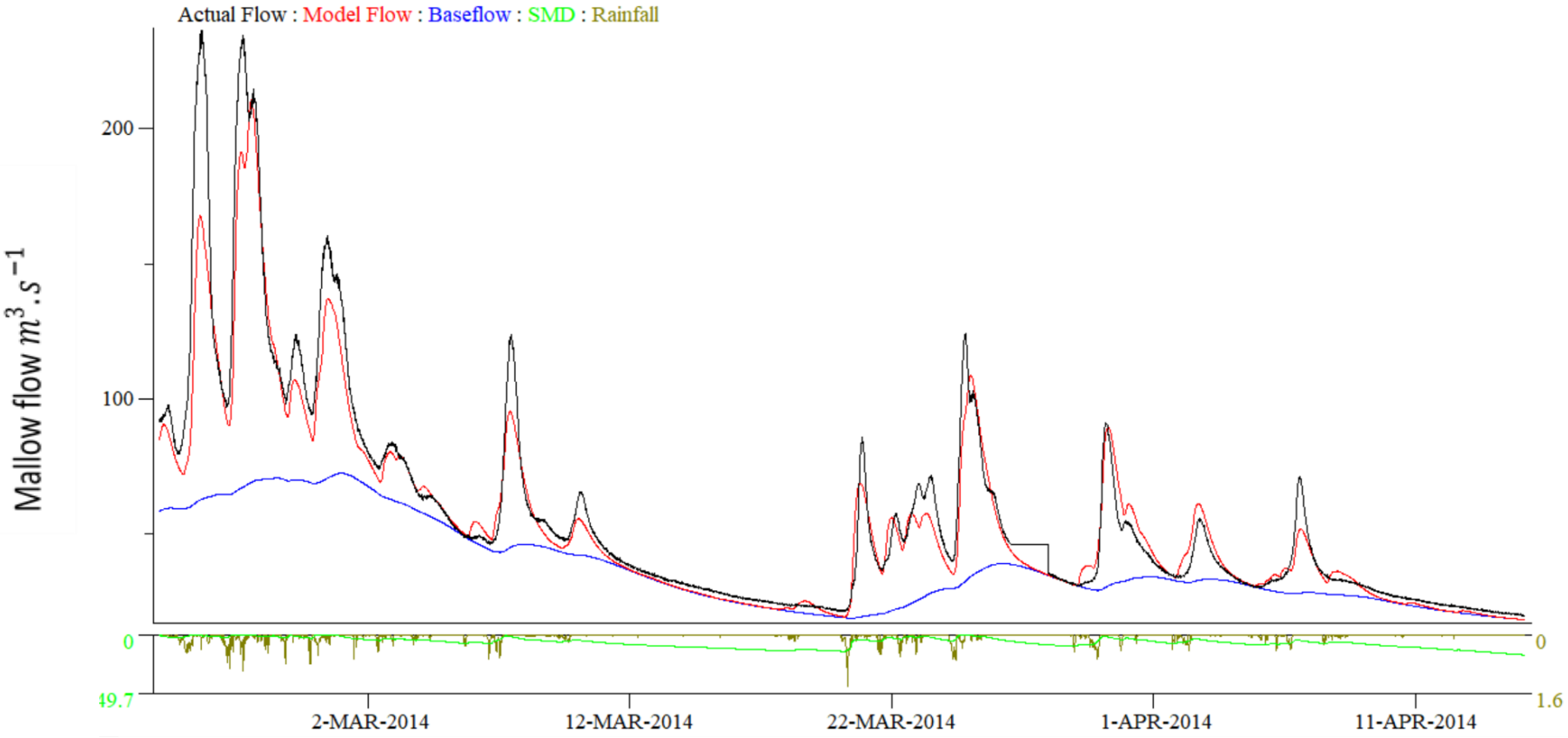


Figure 6.6 (b) 15-Minute precipitation & PDM simulation from March to April 2014 at Mallow Rail BR - observed versus modelled riverflow after manual parameter adjustment and automatic optimisation $R^2 = 0.9558$.

```
PDM: BlackWater at Killavullen(Munster)
 objtyp  maxfun  konvge    objtol
  11     1000        4     8.00000
analys    qmin       qmax         pkmin    pkrge
   2      0.00      1000.0          8.00    48
 prtfnf  prtopt   common file name of plot files
   0       3     'USER_EVPLOT'
indx xs     smpstp        smptol       xs name
1  1.0000   5.00000E-02  1.00000E-02  'rainfac'
1  0.54017  2.97896E-02  1.11309E-03  'cmin'
1  73.682  -0.73682      8.00000E-02  'cmax'
1  0.22553  3.24543E-03  3.64016E-04  'b'
1  719.20  -7.1920       1.00000E-02  'be'
1  18.148   0.18148      9.37700E-03  'k1'
1  5.4563  -5.45631E-02  1.09126E-02  'k2'
1  126.24   1.2624       7.19672E-02  'kb'
1  273.08  -4.0969       0.29627      'kg'
1  48.322   0.48322      2.00000E-03  'St'
1  1.6378  -1.63779E-02  2.00000E-03  'bg'
1  4.7218   4.72178E-02  1.65727E-03  'qconst'
1  3.3467  -3.34669E-02  2.00000E-03  'tdly'
 /
PROBABILITY DISTRIBUTED LOSS MODEL
 Flow    Rainfall   Climate-station   Daily Rain
'kil_flow_1'  'MULTIPLE'  'sin curve    1.40000'  'NULL'
loss  runoff  surface  base  tdly    i_sdp  nsdr  sqmodel  oap
   1      0     22      3    -13      0      0    'NONE'  'FLOW'
Multiple gauges (number)
  17
'BMET_80711'  'BMET_80712'  'BMET_80716'  'BMET_80713'  'BMET_80705'
'BMET_80704'  'BMET_80717'  'BMET_80714'  'BMET_80710'  'BMET_80720'
'BMET_80708'  'BMET_80718'  'BMET_80702'  'BMET_80701'  'BMET_80715'
'BMET_80721'  'BMET_80732'
 Number rain zones and gauge weights
   1
0.0438  0.0674  0.0717  0.0441  0.0654  0.0591  0.0648  0.055   0.1165
0.0547  0.0638  0.0778  0.0615  0.0281  0.0626  0.0487  0.015
 Catchment area (km**2)
1248.32
Event Begin       Event End       Noi     End 15min Rain
'00:00  01 Jan 2014'  '23:45 29 Dec 2014'    /
/
```

Figure 6.7 PDM input file for Killavullen subcatchment after manual and automatic optimisation.

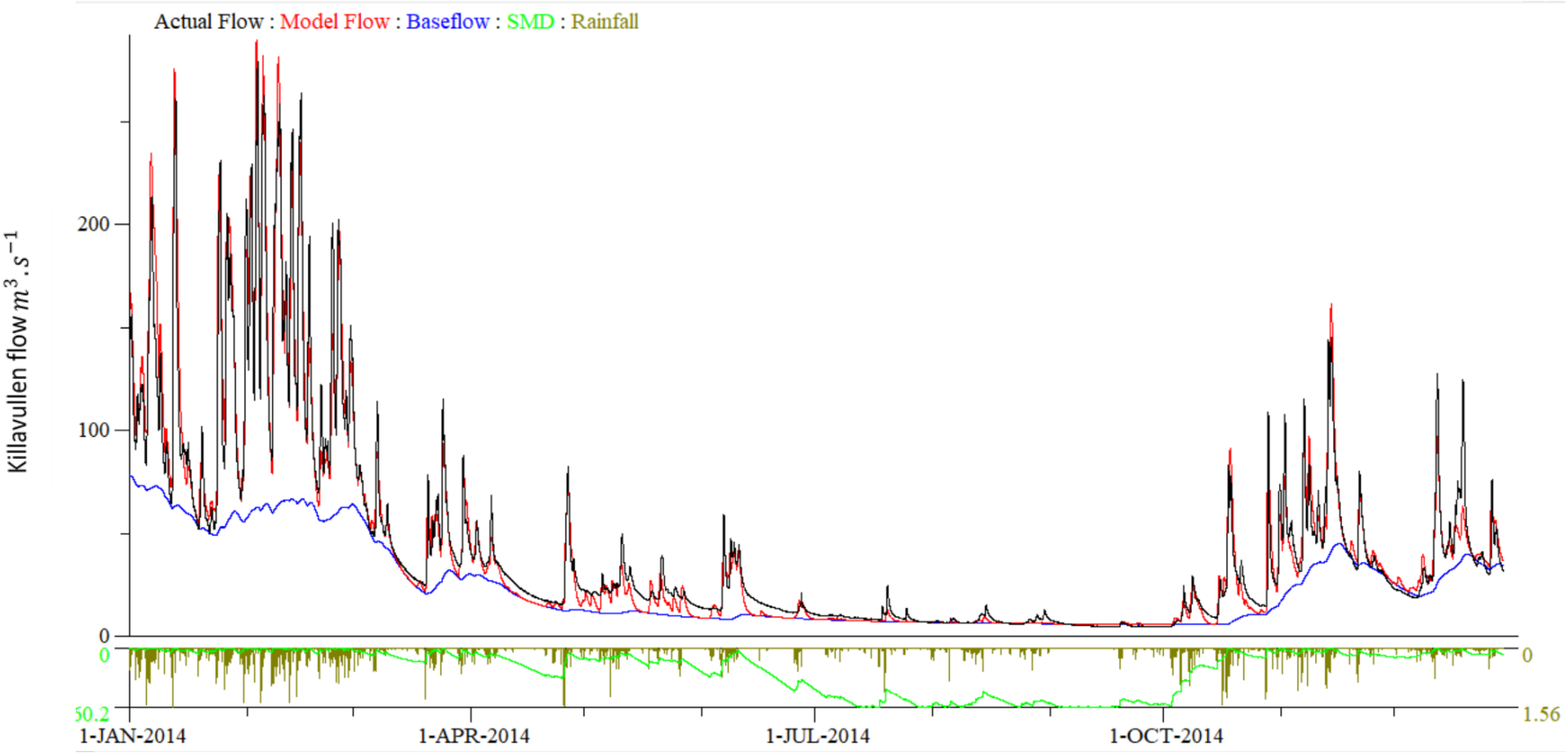


Figure 6.8 15-Minute precipitation & PDM simulation for 2014 at Killavullen - observed versus modelled riverflow. Calibration, $R^2 = 0.9641$

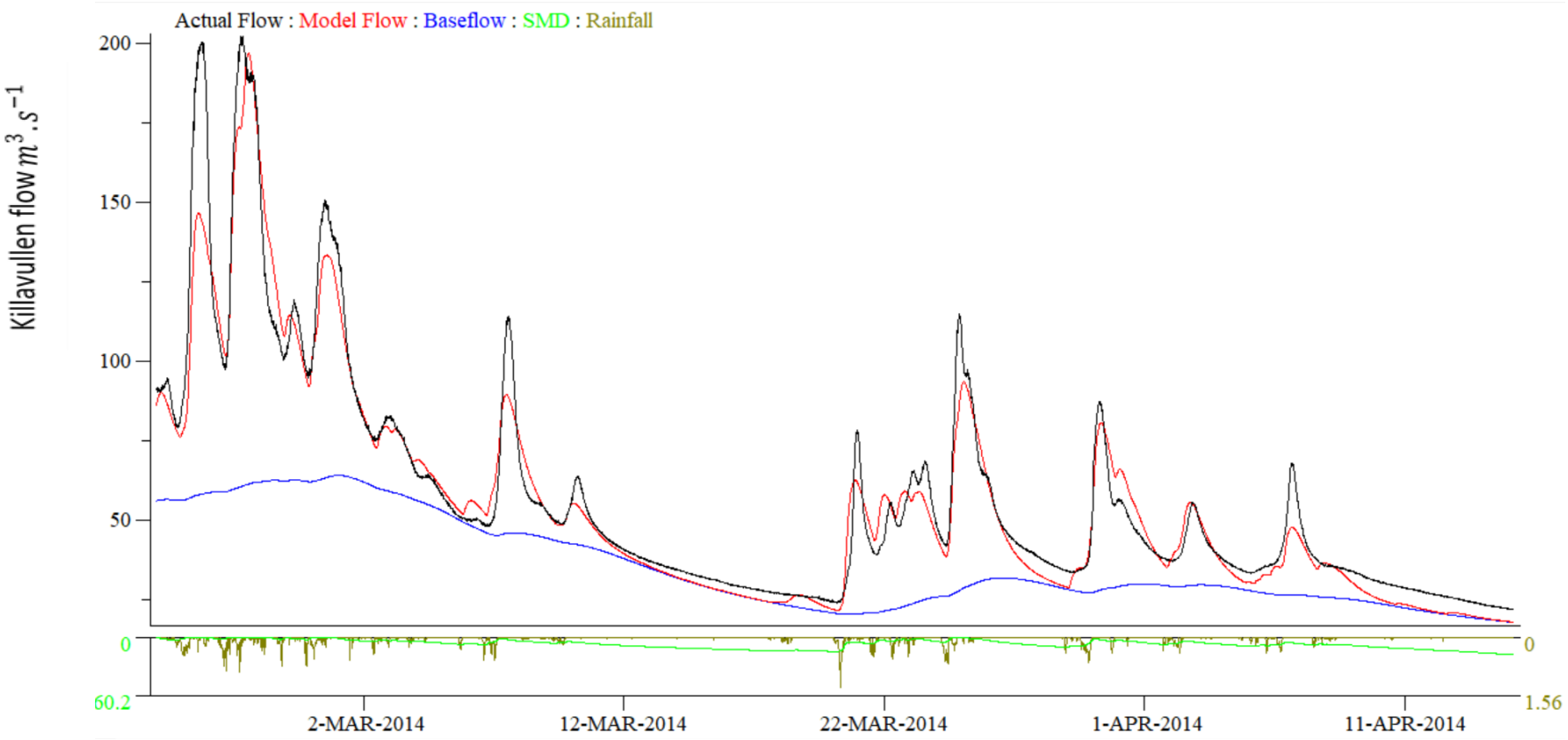


Figure 6.8 (b) 15-Minute precipitation & PDM simulation from March to April 2014 at Killavullen - observed versus modelled riverflow after manual parameter adjustment and automatic optimisation $R^2 = 0.9641$.

Table 6.3 shows the summary of the models for Mallow Rail BR and Killavullen. The high values of R^2 of more than 90% for both model runs indicate the models can account for a high proportion of the variability in the observed flows. The value of potential evaporation (PE) is 511 and actual evaporation (AE) as computed by the models are 446.25 and 452.85 for Mallow Rail BR and Killavullen. These values of AE across Ireland range from about 88 % to 95 % of the potential evaporation. The value of observed flow and computed flow for both model runs are similar. This confirms that the PDM model with these adjusted parameters can account for about 90% of the variability of observed flows. This completes the calibration process. Now we can use PDM with these parameters for validation and forecasting. Appendix A shows the performance and summary of the results of the PDM model runs for three subcatchments over the period of 2010 to 2017.

Table 6.3 Summary of PDM Model outputs after manual and automatic optimisation in two subcatchment for 2014

	R^2	Rainfall (mm)	Computed baseflow (mm)	Observed flow (mm)	Computed flow (mm)	Potential evaporation (mm)	Actual evaporation (mm)	Net rainfall (mm)
Mallow Rail BR	0.9558	1374.93	738.01	1173.01	1170.85	511	446.25	928.68
Killavullen	0.9641	1358.19	595.56	1072.15	1059.77	511	452.85	905.33

6.3 Model Validation

6.3.1 Model validation for a year

After calibration, the PDM model (without changing any parameters) was applied for each year between 2010 to 2017 on three subcatchments for validation. Table 6.4 shows the summary of the performance of each model run for 2015 as validation and Figures 6.9 to Figure 6.11 show the resulting hydrograph for each of the three subcatchments for 2015. Apendix A contains the summary and hydrograph for the period of 2010 to 2017. These indicate the PDM model runs can account for about 90% of the variability of observed flows in different subcatchments over a long period of time especially when a high standard of data quality are available, for example in 2015.

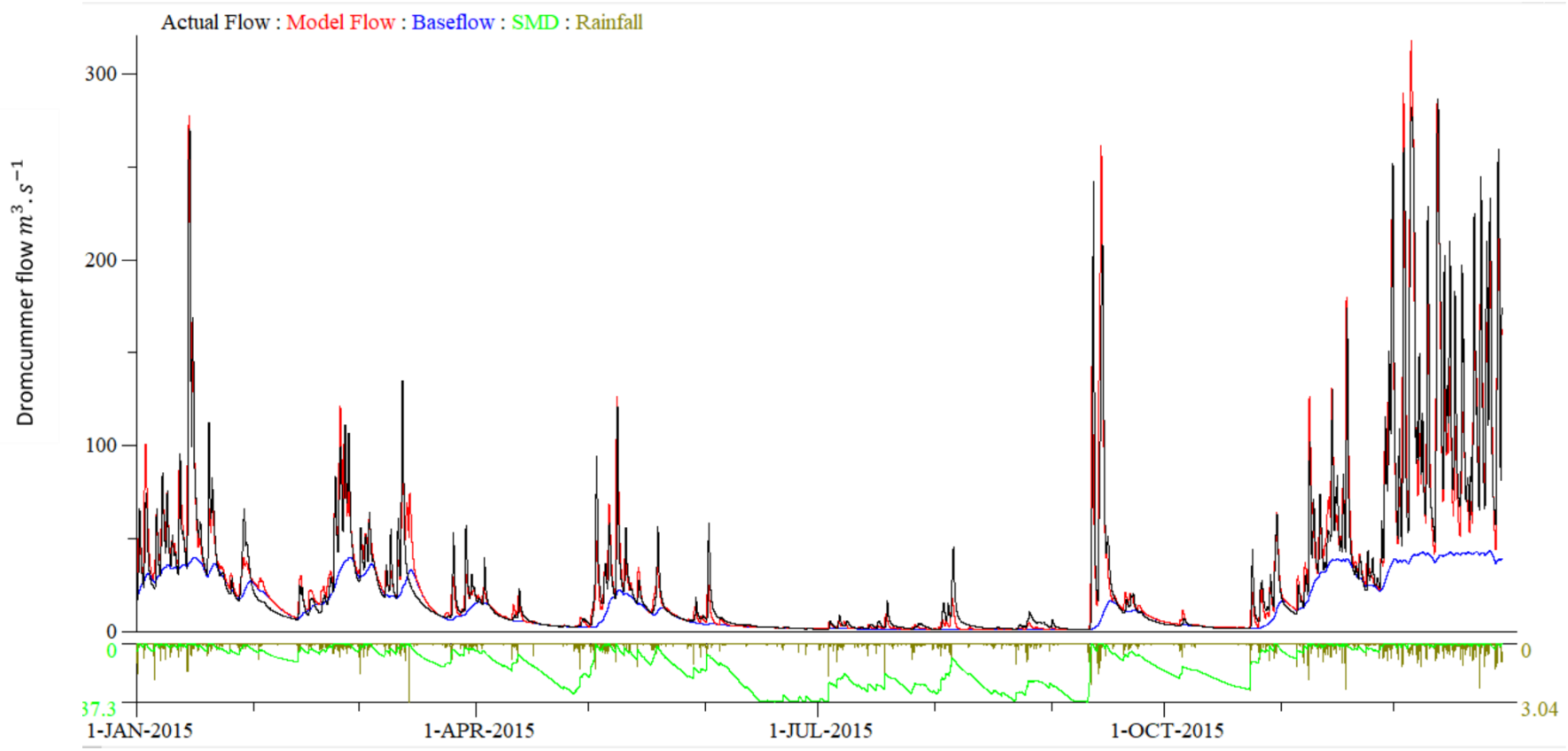


Figure 6.9 (a) 15-Minute precipitation & PDM simulation for 2015 at Dromcummer - observed versus modelled riverflow. Validation, $R^2 = 0.9213$

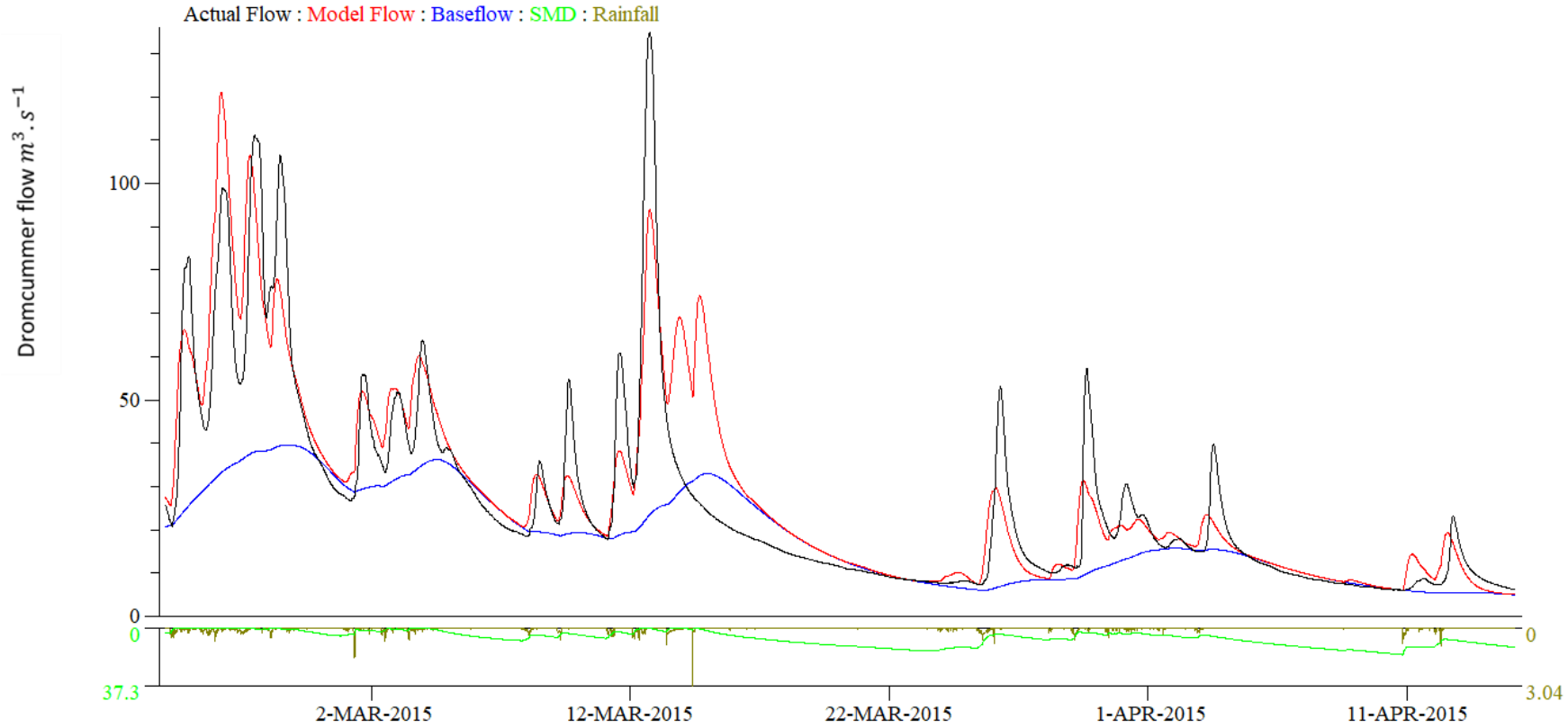


Figure 6.9 (b) 15-Minute precipitation & PDM simulation for March to April 2015 at Dromcummer - observed versus modelled riverflow. Validation, $R^2 = 0.9213$

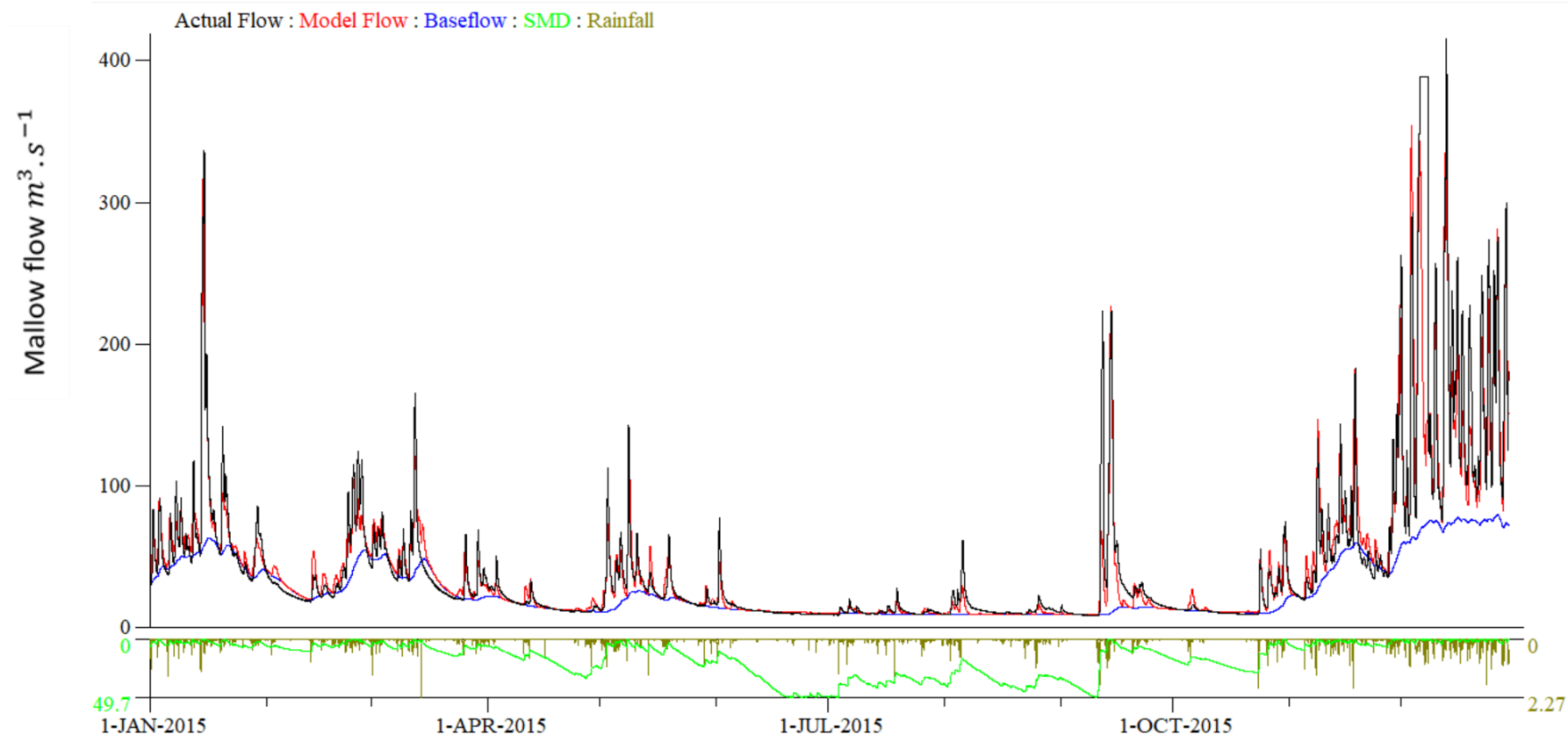


Figure 6.10 (a) 15-Minute precipitation & PDM simulation for 2015 at Mallow Rail BR - observed versus modelled riverflow .Validation $R^2 = 0.8756$

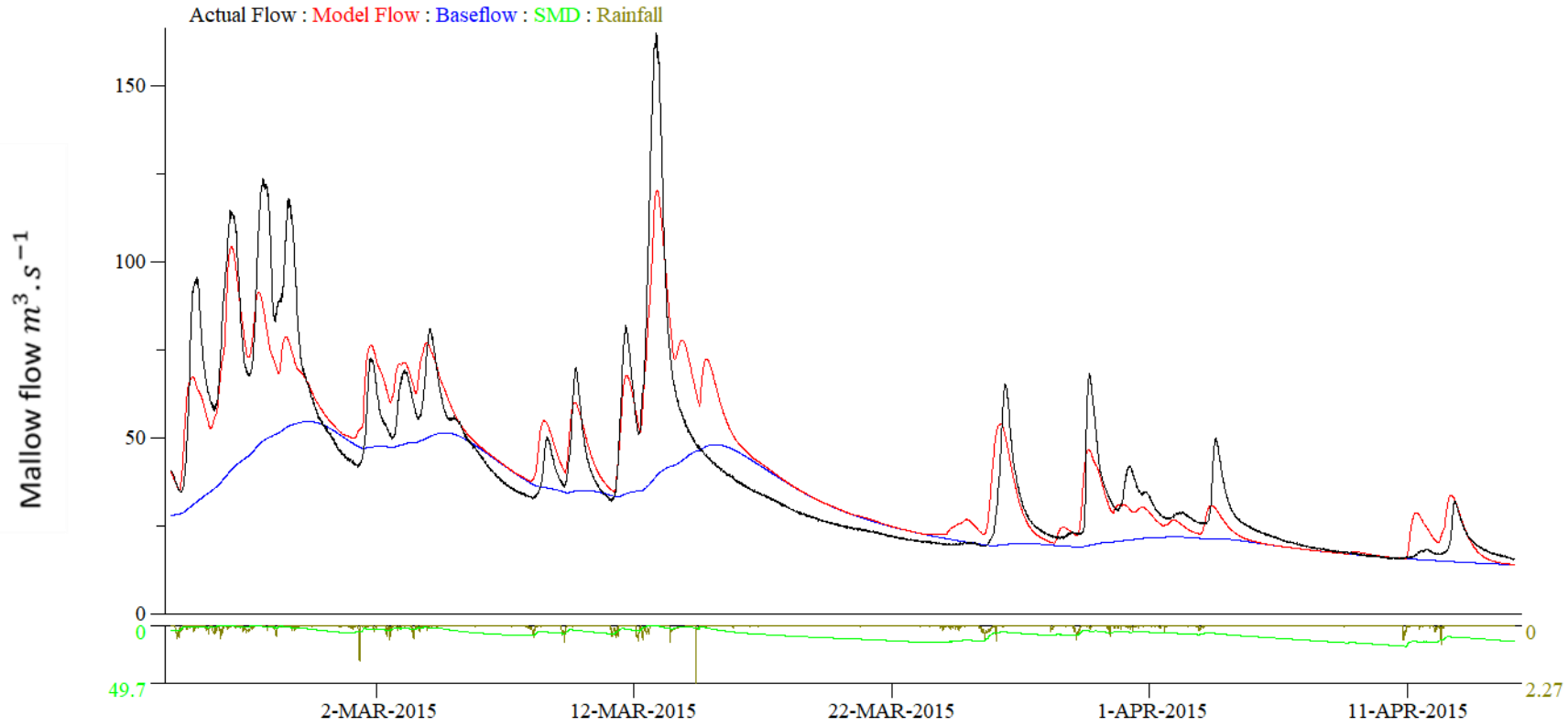


Figure 6.10 (b) 15-Minute precipitation & PDM simulation from March to April 2015 at Mallow Rail BR - observed versus modelled riverflow .Validation $R^2 = 0.8756$

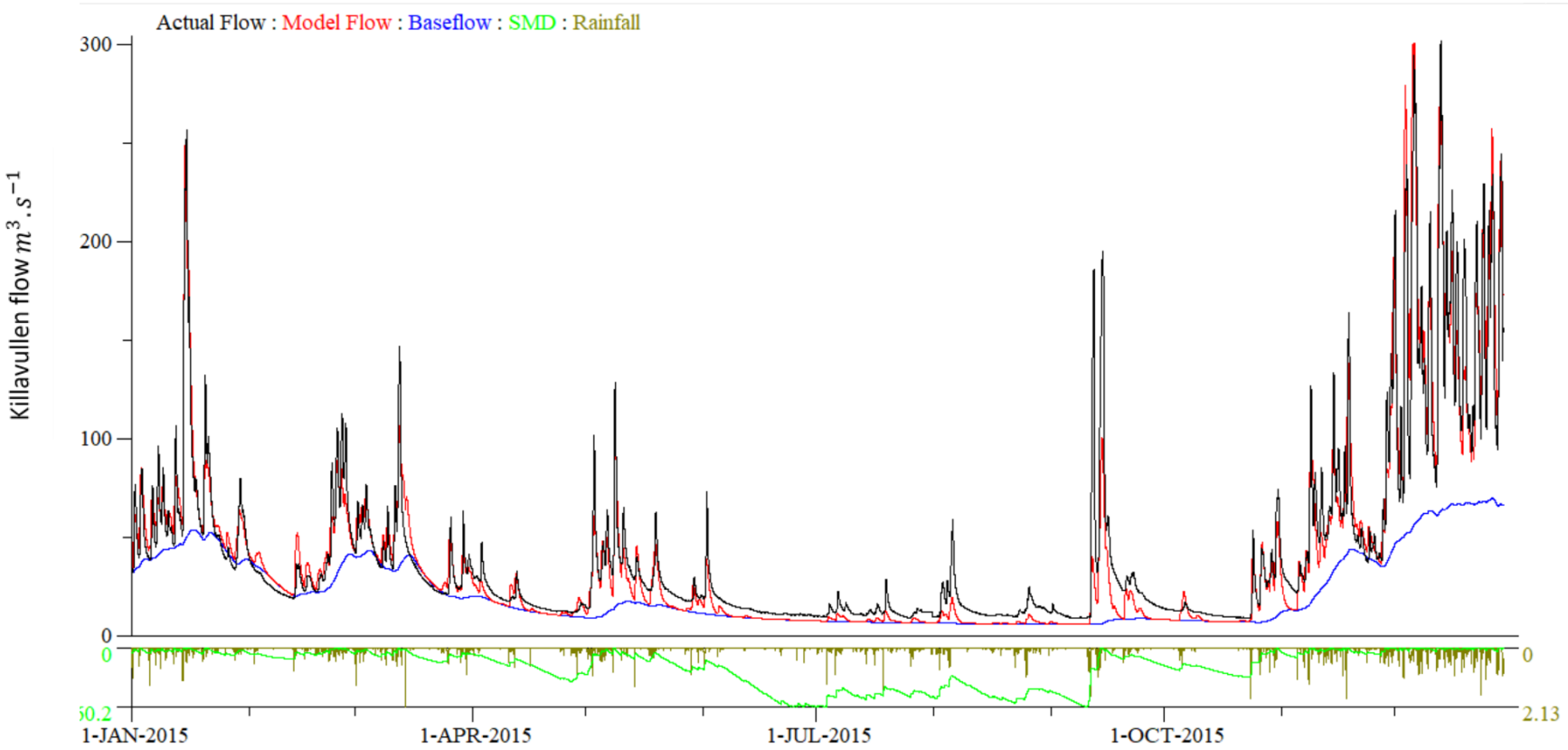


Figure 6.11 (a) 15-Minute precipitation & PDM simulation 2015 Killavullen - observed versus modelled riverflow. Validation, $R^2 = 0.9224$

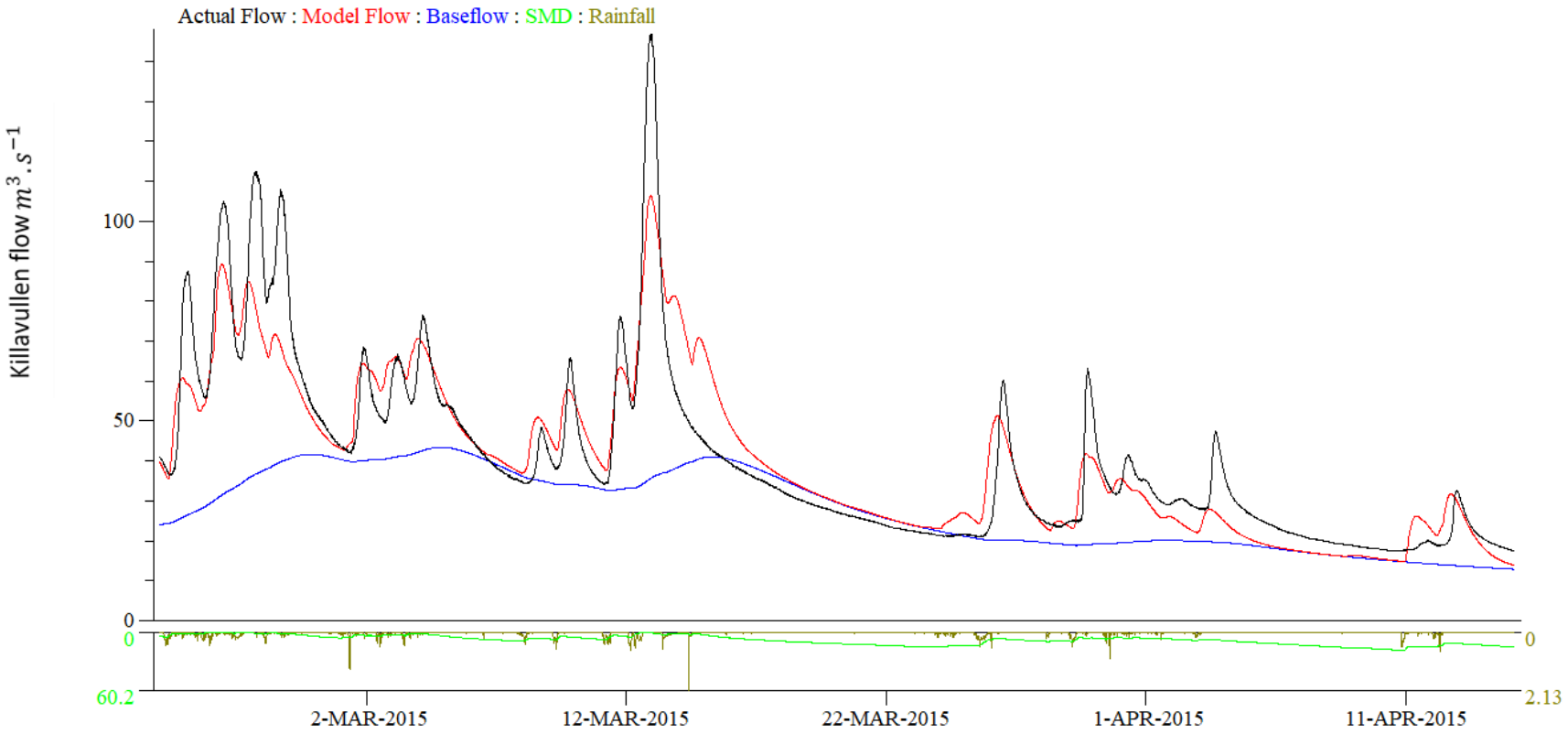


Figure 6.11 (b) 15-Minute precipitation & PDM simulation from March to April 2015
Killavullen - observed versus modelled riverflow. Validation, $R^2 = 0.9224$

Table 6.4 Summary of validation of PDM Model for 2015. Note that the observed and computed flows are close.

	R^2	Rainfall (mm)	Computed baseflow (mm)	Observed flow (mm)	Computed flow (mm)	Potential evaporation (mm)	Actual Evaporation (mm)	Net rainfall (mm)
Dromcummer	0.9213	1525.42	505.90	1061.41	1066.79	511	473.84	1051.58
Mallow Rail BR	0.8756	1436.90	691.06	1197.20	1141.02	511	486.81	950.09
Killavullen	0.9224	1413.35	543.62	1074.12	1009.72	511	493.79	919.55

6.3.2 River Flow Comparison

Comparison between observed river flows at three subcatchments for each year between 2010 to 2017, Figure 6.12 (a) to (h) indicate the river flow at Dromcummer (area 862 km^2) is less than the river flow at Mallow Rail BR (area 1168 km^2) as expected. While river flow at Mallow Rail BR is more than river flow at Killavullen (area 1248 km^2) despite that Killavullen has a catchment area 7 % greater than Mallow Rail BR.

The magnitude and duration of the flood depends on the catchment characteristics (see Figure 3.4 and Figure 3.5) such as catchment area, precipitation, stream slope, soil type, topography, geology and land cover, and extent of lakes and peatland within the catchment (Kiely, Leahy, Fenton, & Donovan, 2008). The river flows at Killavullen being less than Mallow Rail BR's may be explained by

1. Current inaccuracies of either or both rating curves.
2. By the fact that as the river approaches Mallow Rail BR and Killavullen, there is a wide flood plain for flood attenuation.
3. By high flows underestimated at Killavullen as flow bypasses at the side of the bridge at Killavullen and may not be correctly measured.
4. Also karstic caves and depressions on the floodplain downstream of Mallow and near Killavullen may cause additional losses to groundwater when the ground is not saturated (Donovan, 2015).

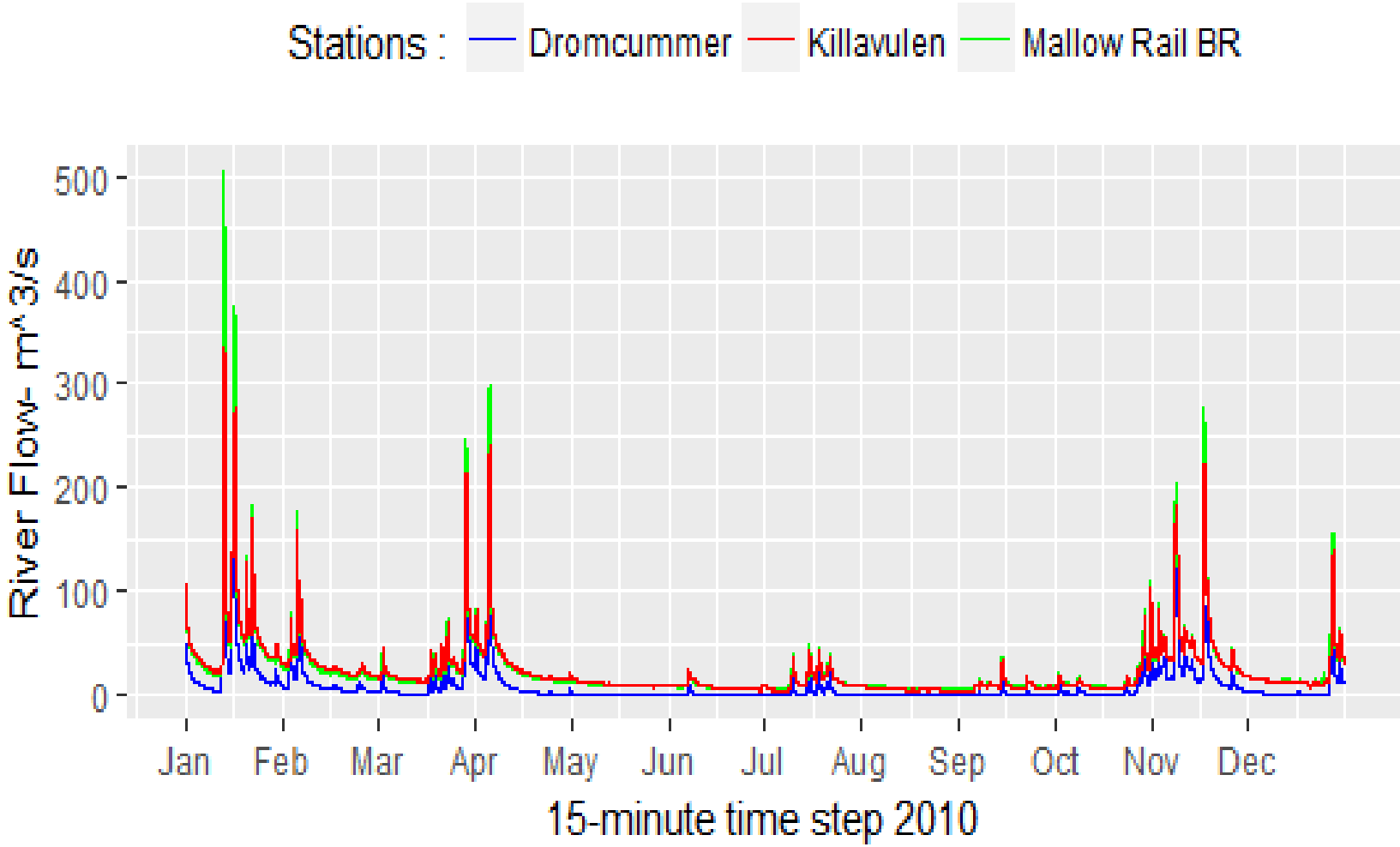


Figure 6.12 (a) Comparison of observed river flow in three subcatchments for 2010

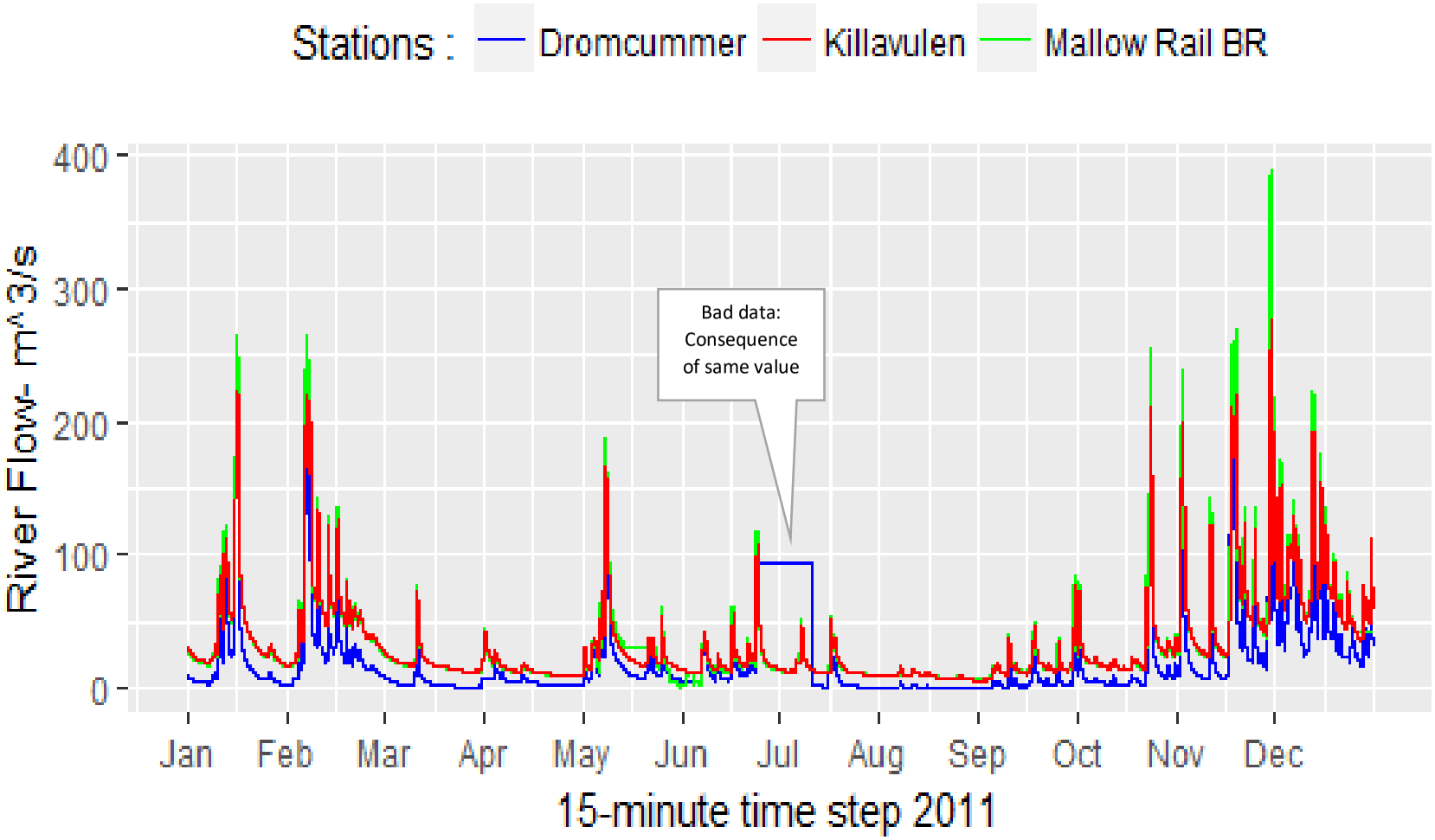


Figure 6.12 (b) Comparison of observed river flow in three subcatchments for 2011

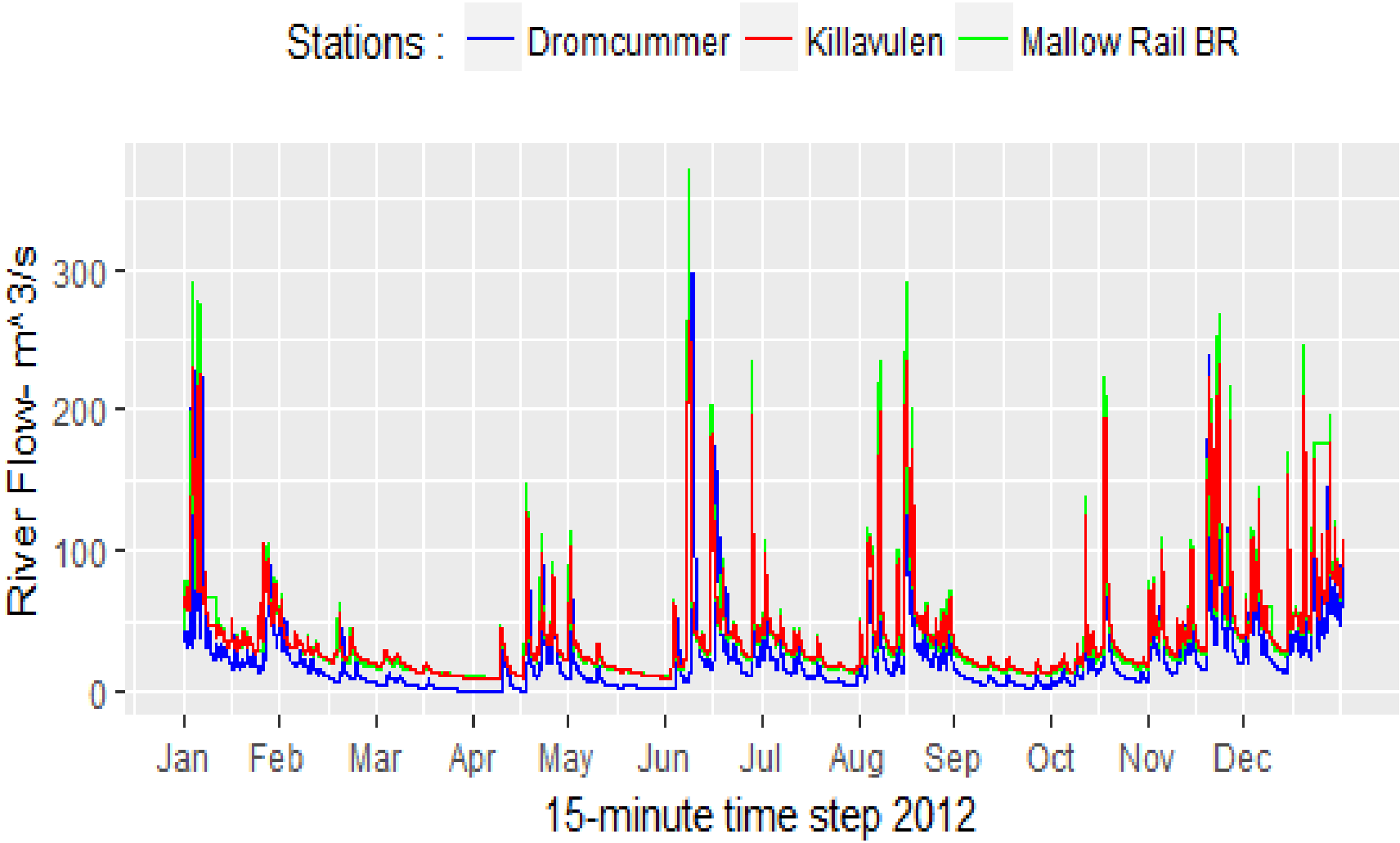


Figure 6.12 (c) Comparison of observed river flow in three subcatchments for 2012

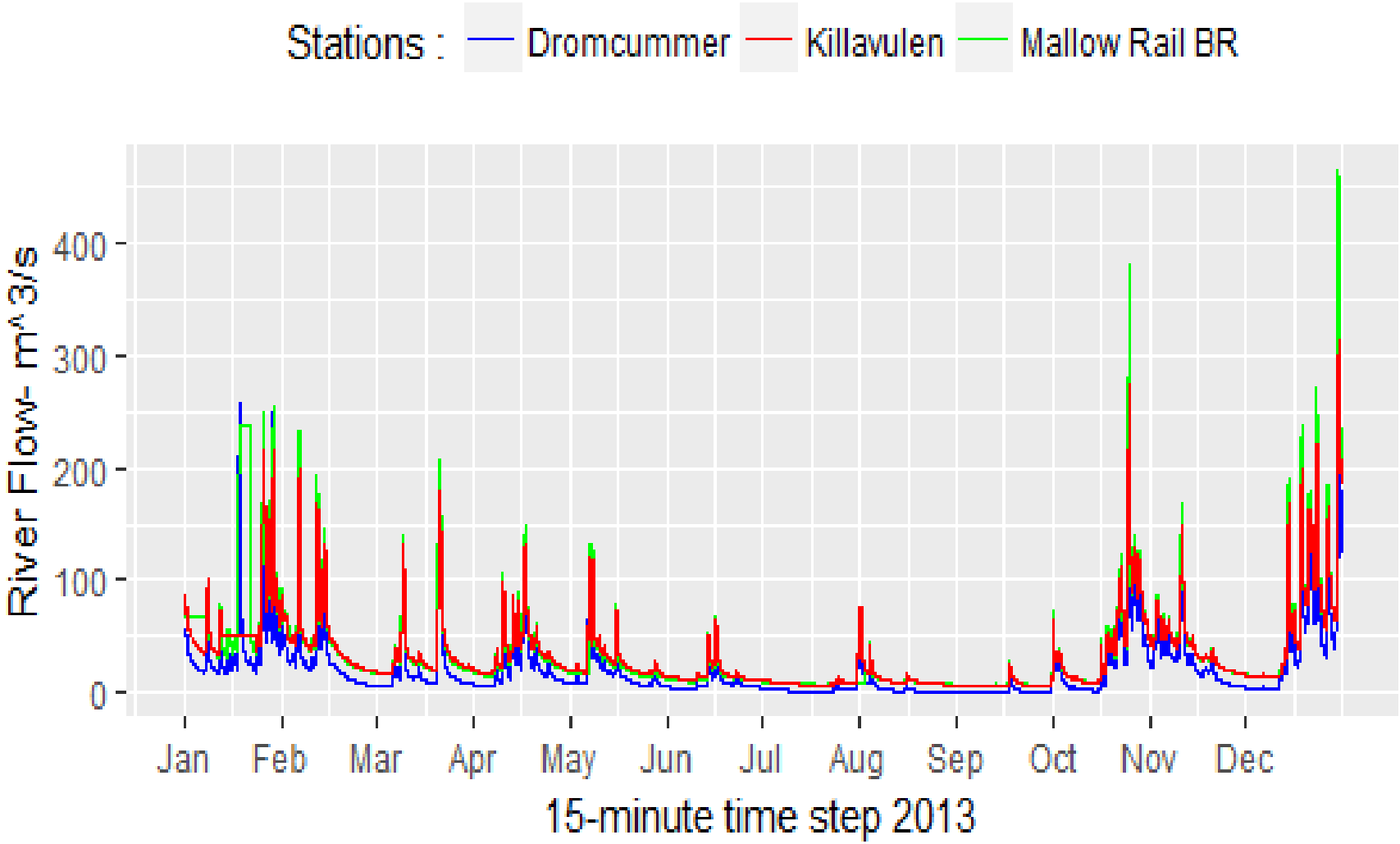


Figure 6.12 (d) Comparison of observed river flow in three subcatchments for 2013

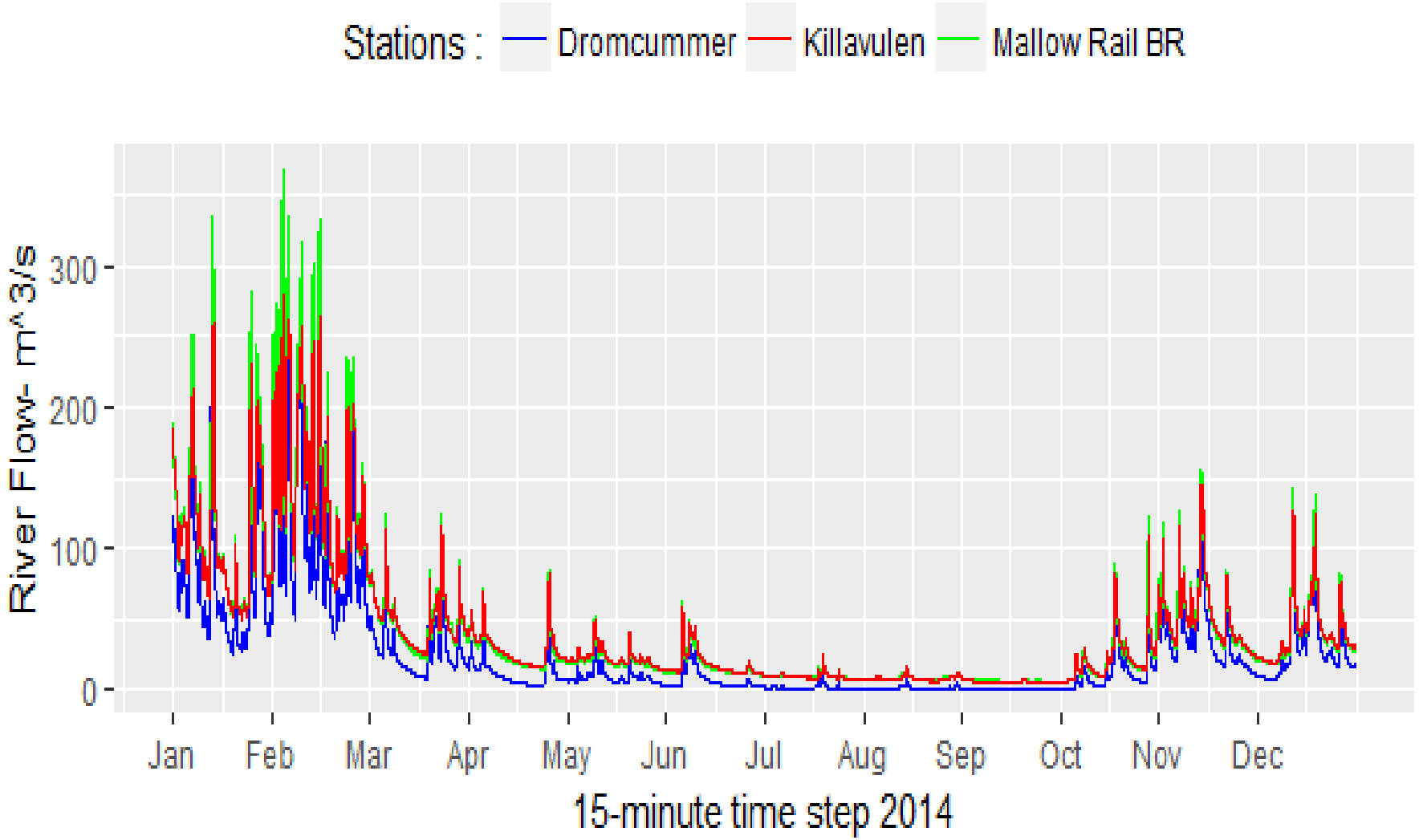


Figure 6.12 (e) Comparison of observed river flow in three subcatchments for 2014

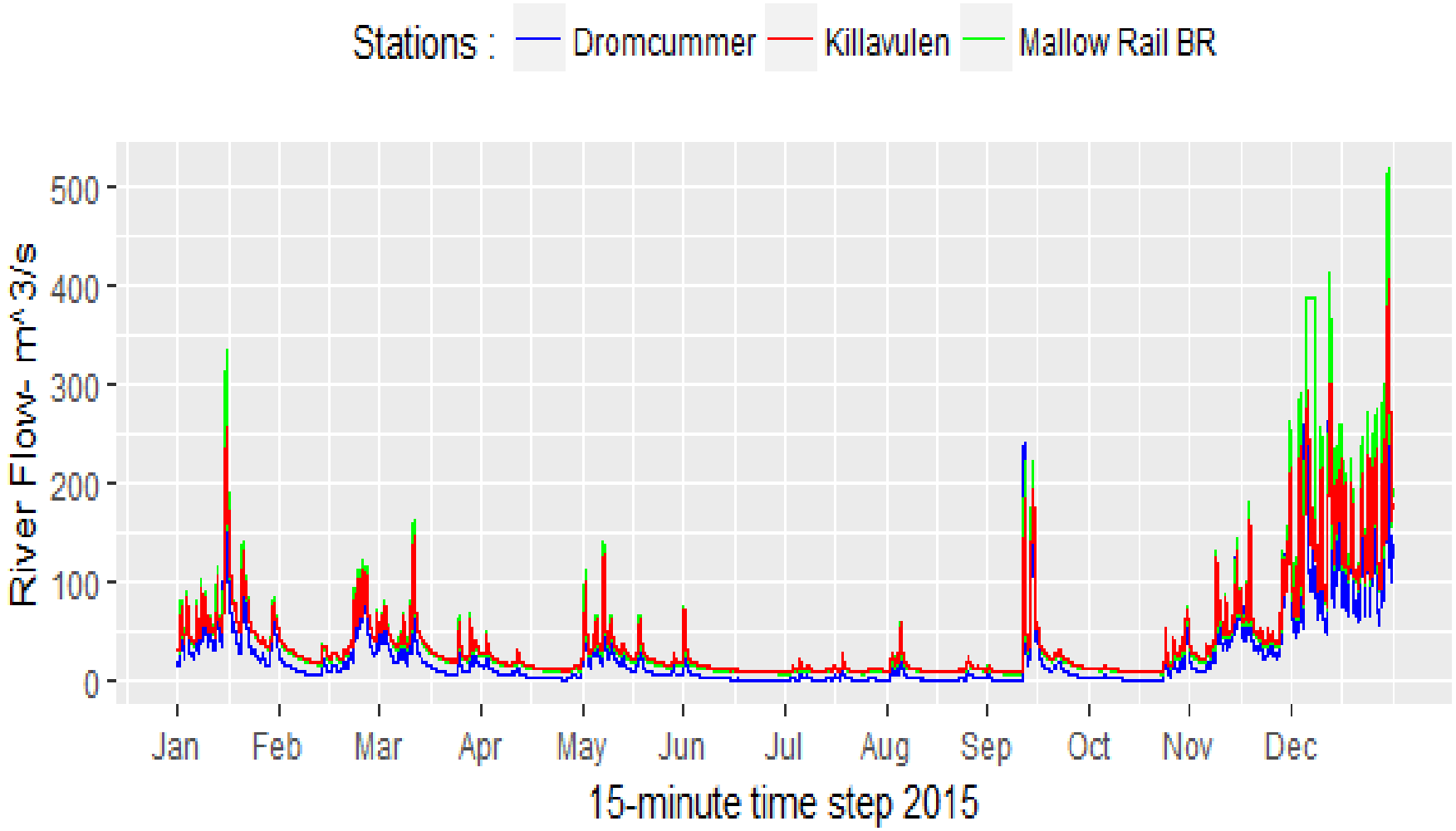


Figure 6.12 (f) Comparison of observed river flow in three subcatchments for 2015

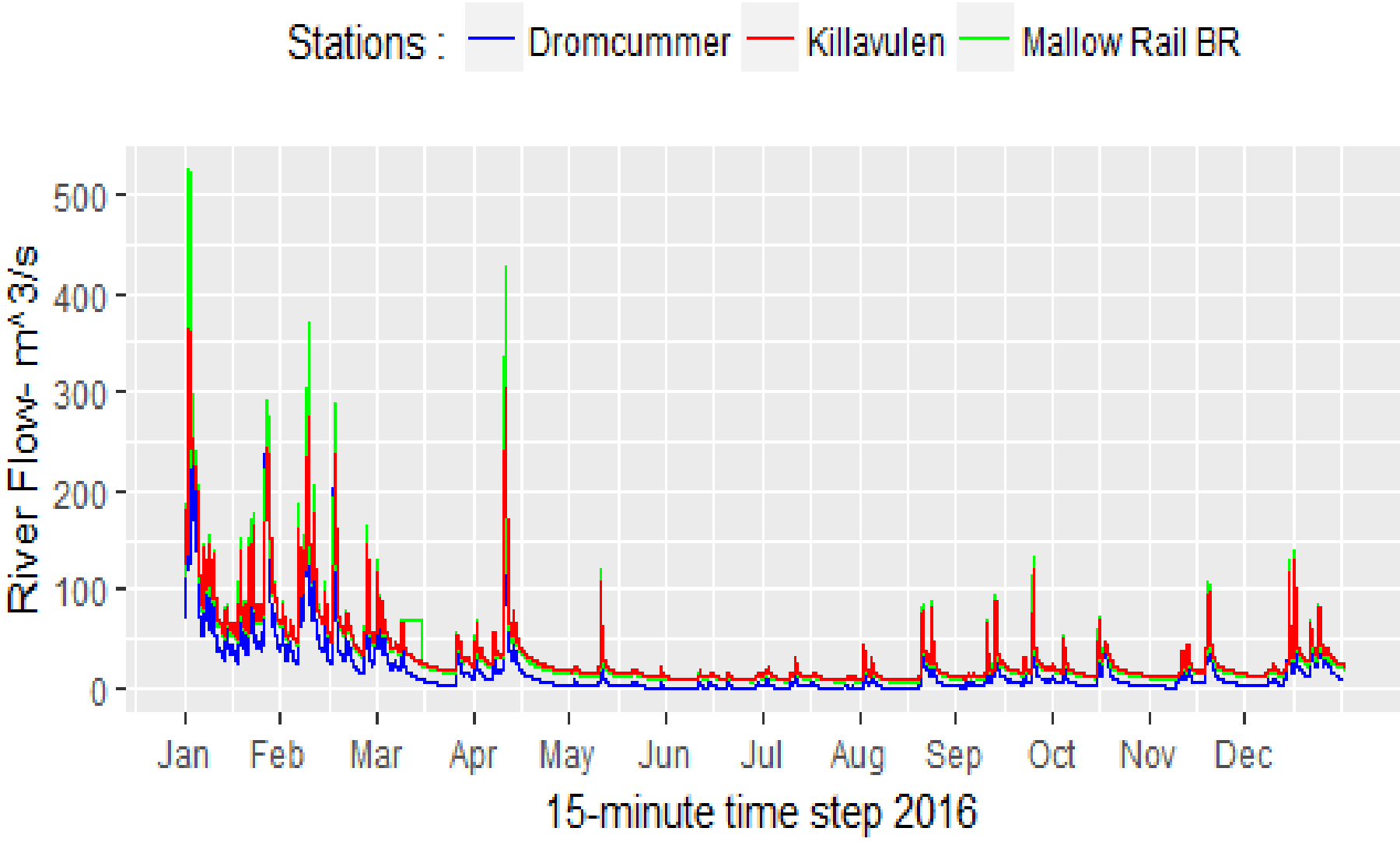


Figure 6.12 (g) Comparison of observed river flow in three subcatchments for 2016

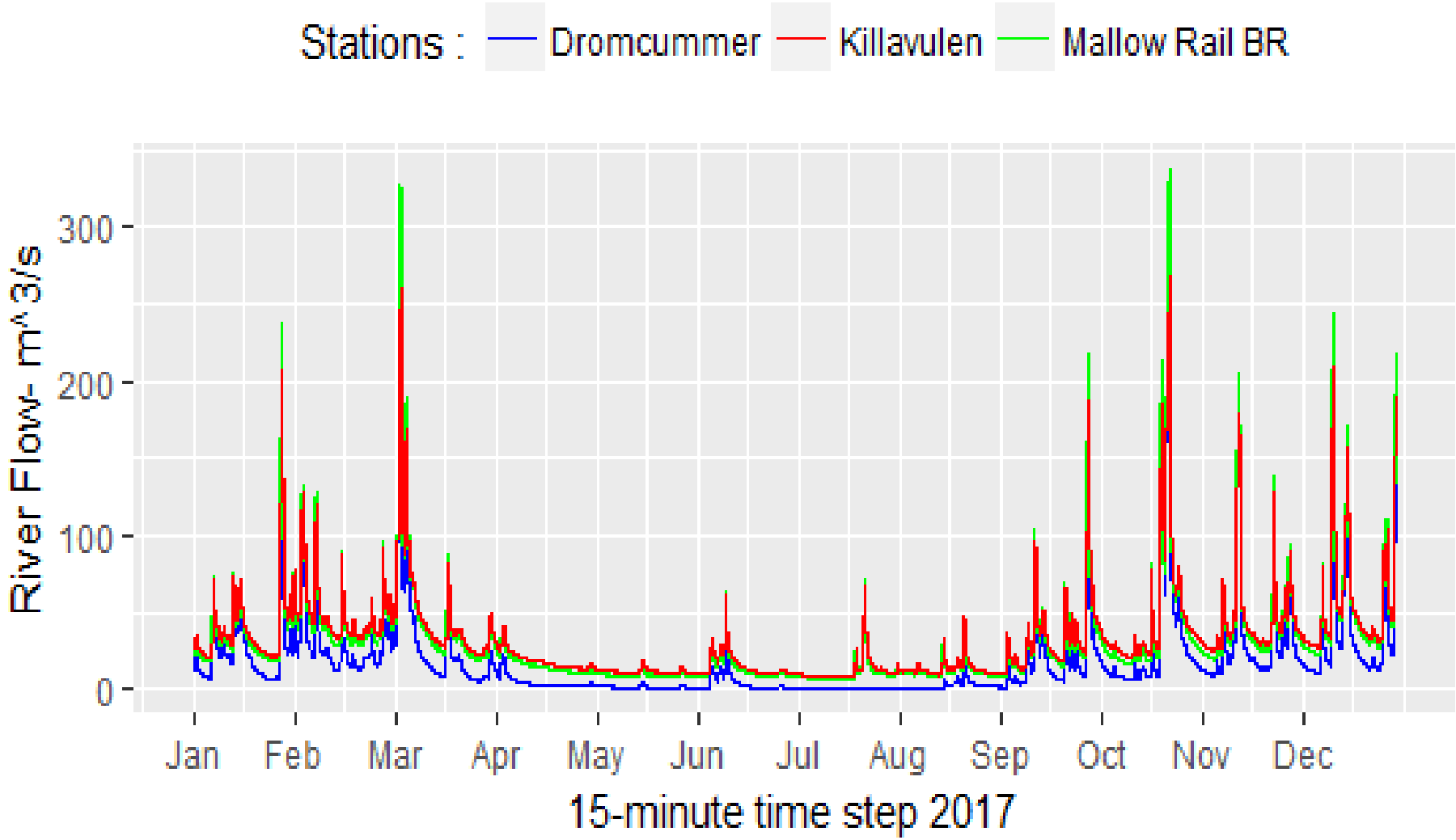


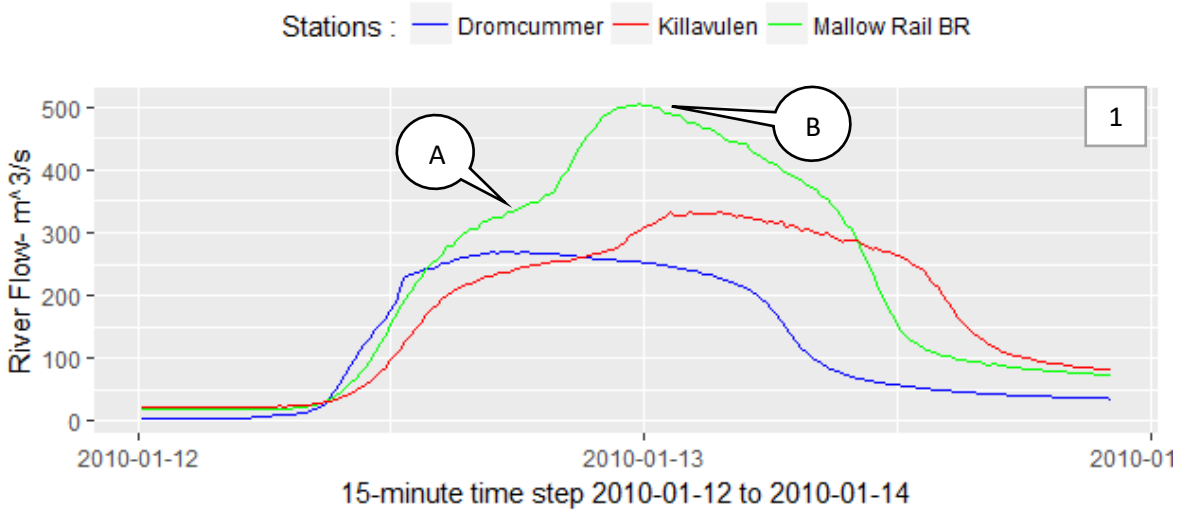
Figure 6.12 (h) Comparison of observed river flow in three subcatchments for 2017

A comparison between observed river flow for the three subcatchments for the largest event of each year (Figure 6.13 (a) to (d)) indicate the shape of hydrograph in each subcatchment for each event is very similar. In Dromcummer the rate of increasing observed river flow during the event to reach the peak is slightly similar to the rate of decreasing observed river flow after the event and the hydrograph is wide during the peak time. While the shape of hydrographs in Mallow show different patterns, there are sometimes two successive flood waves, the first one is smaller than the second one. Several tributaries join the Blackwater river in different locations (Figure 3.5), so unsynchronised flood waves from these tributaries may cause the two successive waves in the hydrograph. The rate of increase of the hydrograph is slightly less than the rate of decrease in Mallow.

The shapes of the observed hydrographs In Killavullen are wide, the rate of increase being less than the rate of decrease. River flow is lowest at Dromcummer and highest in Mallow despite Killavullen being farthest downstream. Table 6.6 shows each events peak time and peak flow at each subcatchment.

Table 6.6 Comparison of observed flood events peak time and observed peak flow ($m^3.s^{-1}$) at three subcatchments . Event No.3 shows the peak flow at Killavullen is greater than the peak flow at Dromcummer. Note that Figure 6.13 (b) (N0.3) shows the peak flow at Dromcummer is wide and the maximum flow occurs at the end of the flood event.

	Dromcummer (area 862 km^2)		Mallow Rail BR (area 1168 km^2)		Killavullen (area 1248 km^2)	
Event No.	Peak Time	Peak Flow $m^3.s^{-1}$	Peak Time	Peak Flow $m^3.s^{-1}$	Peak Time	Peak Flow $m^3.s^{-1}$
1	2010/01/12 18:15	268.96	2010/01/13 00:30	504.03	2010/01/13 03:00	333.68
2	2011/11/29 17:45	254.24	2011/11/30 01:15	388.15	2011/11/30 05:00	277.14
3	2012/08/16 00:45	240.34	2012/08/16 03:30	291.1	2012/08/16 06:45	233.81
4	2013/12/30 07:00	282.32	2013/12/30 14:30	463.9	2013/12/30 17:00	314.15
5	2014/02/03 12:15	261.64	2014/02/03 19:30	367.83	2014/02/03 23:30	279.04
6	2015/12/30 09:00	311.6	Unreliable data	Unreliable data	2015/12/30 15:15	406.16
7	2016/01/02 09:30	294.03	2016/01/02 07:45	524.02	2016/01/02 10:15	362.96
8	2017/10/22 01:00	284.54	2017/10/22 07:00	336.34	2017/10/22 11:00	268.8



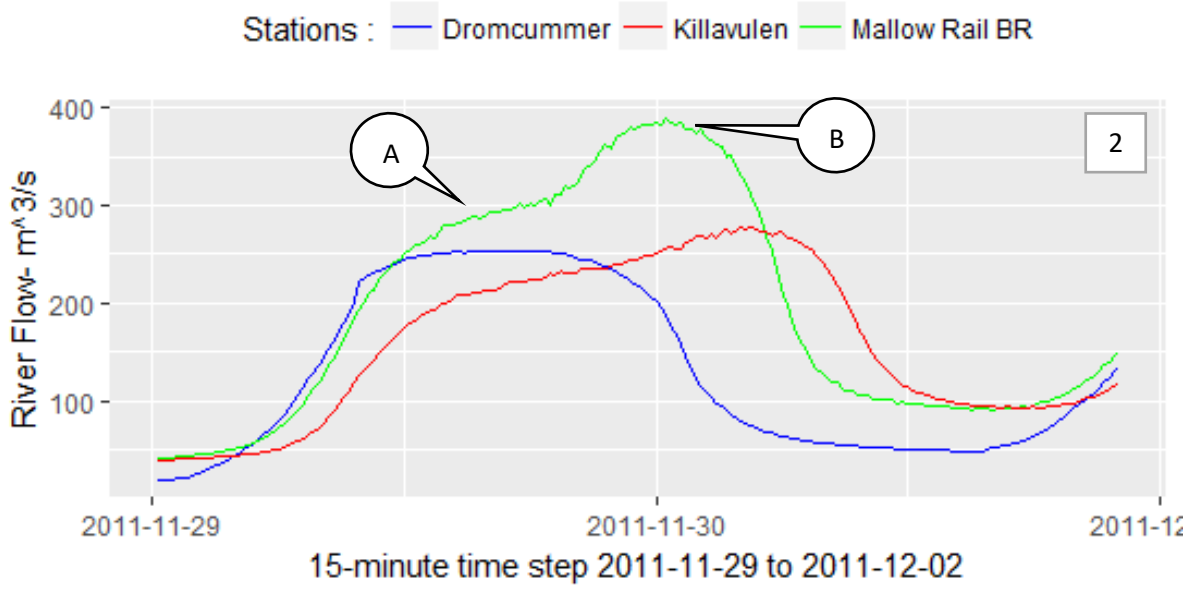
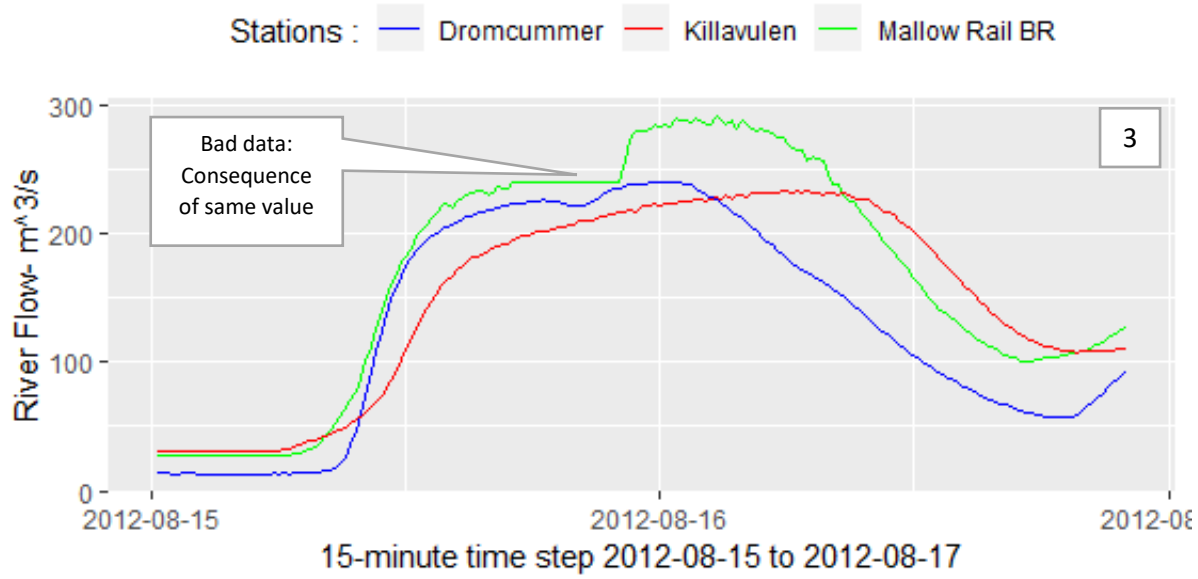


Figure 6.13 (a) Comparison of observed river flow in three subcatchments for the largest event in 2010 and 2011. Note that the hydrograph at Mallow shows the culmination of two flood waves, the flood wave A and the second wave B superimposed on flood wave A.



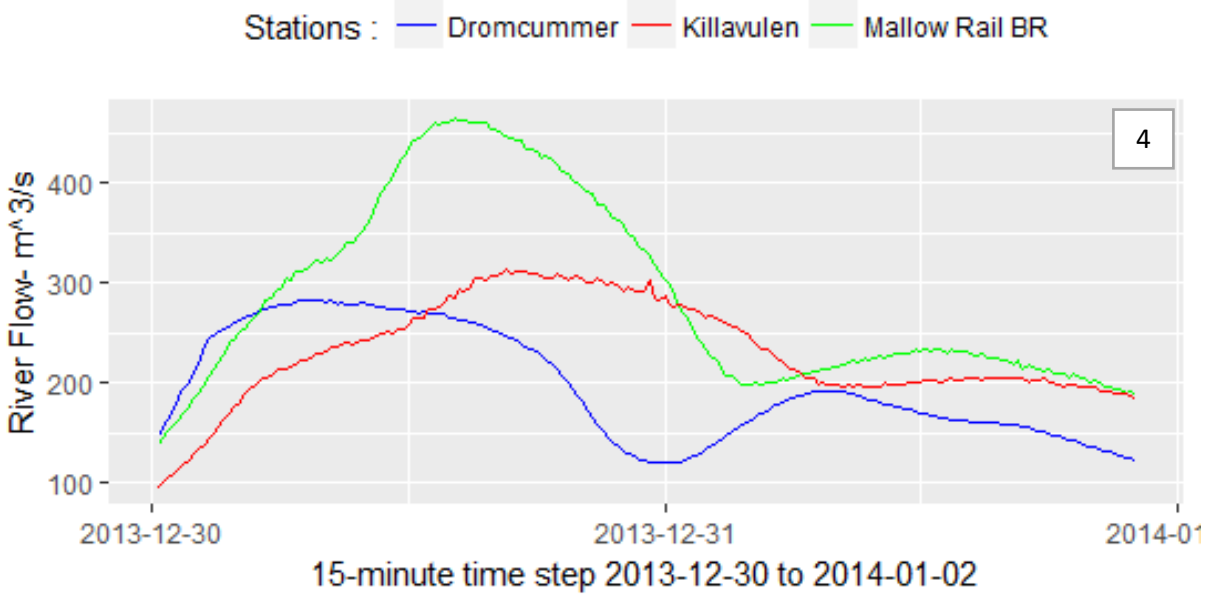
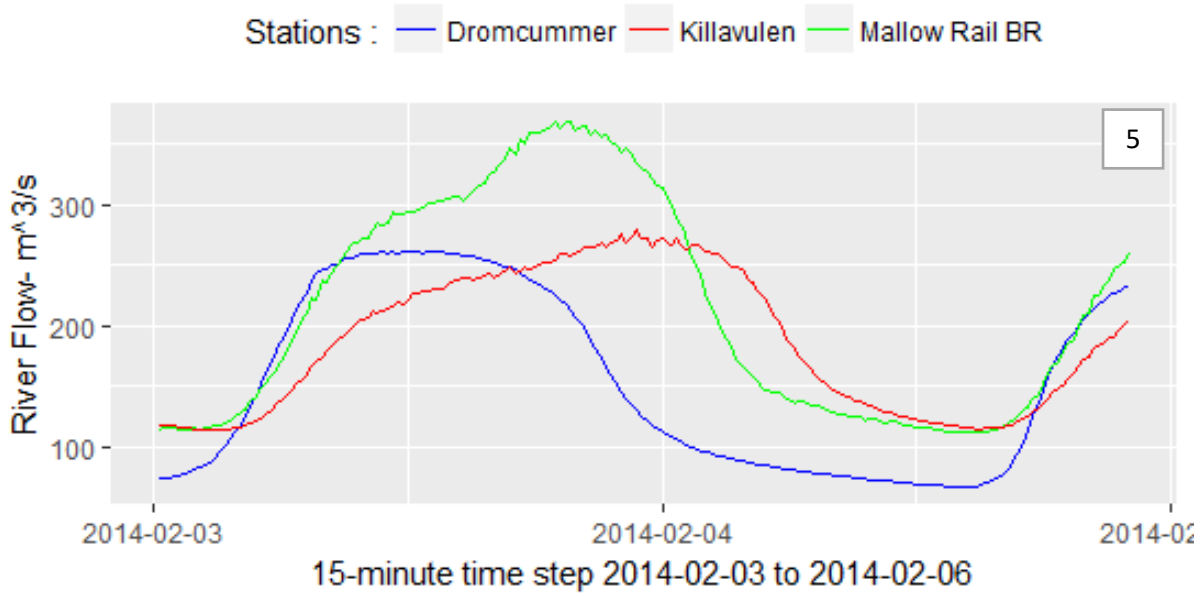


Figure 6.13 (b) Comparison of observed river flow in three subcatchments for the largest event in 2012 and 2013



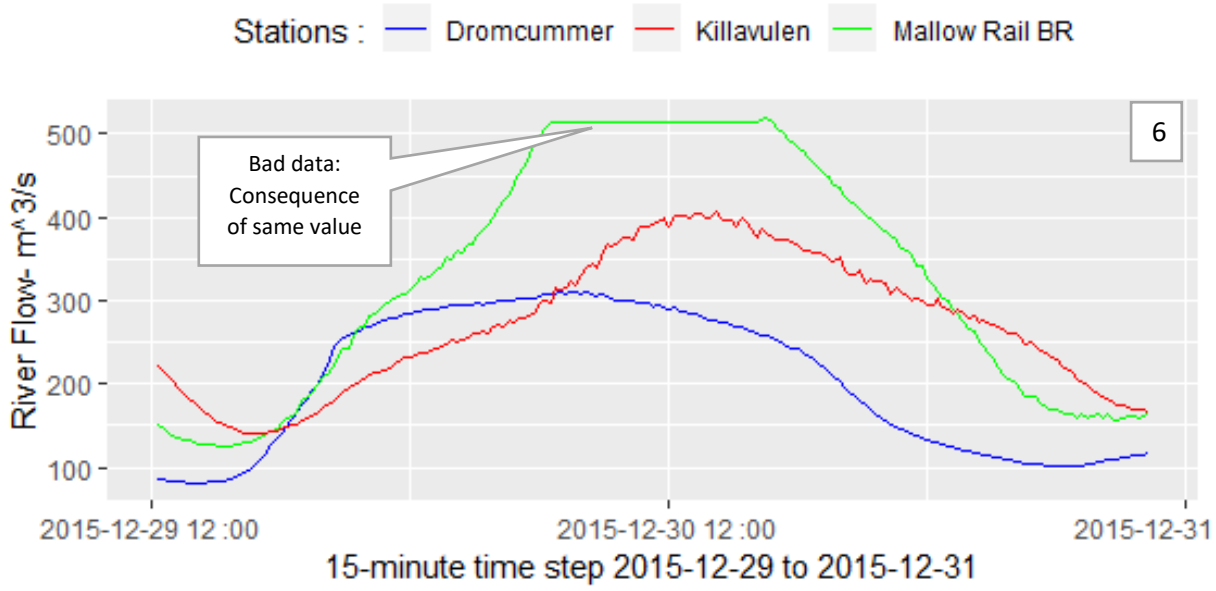
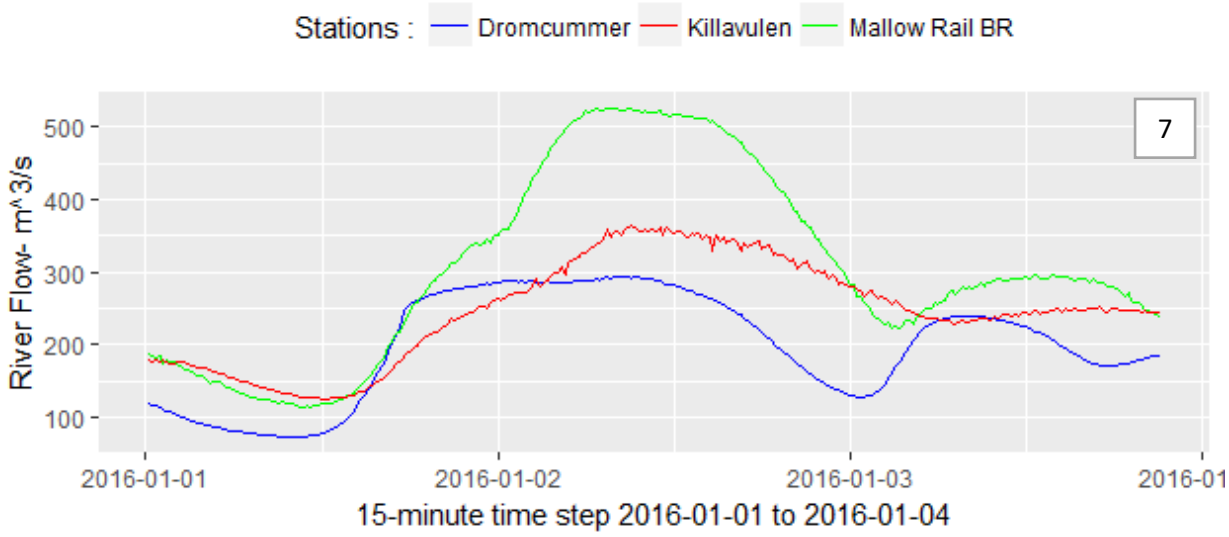


Figure 6.13 (c) Comparison of observed river flow in three subcatchments for the largest event in 2014 and 2015



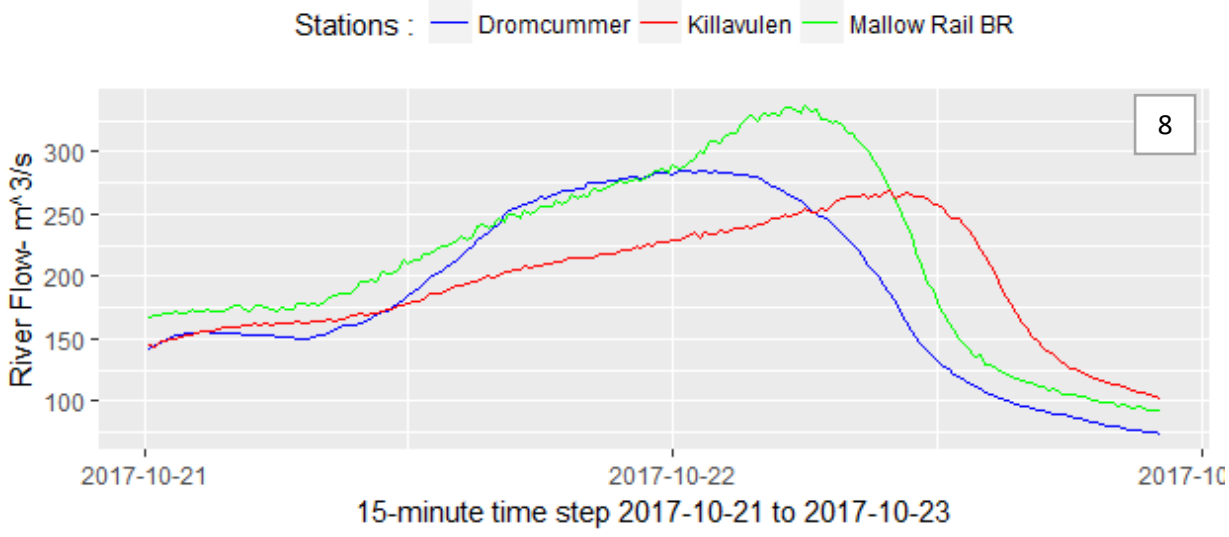


Figure 6.13 (d) Comparison of observed river flow in three subcatchments for the largest event in 2016 and 2017

Table 6.6 and Figure 6.13 (c) (N0.5) show the highest flood event at the three subcatchments in 2014. The highest peak flow of 261.64 $m^3.s^{-1}$ occurred on 2014/02/03 at 12:15 at Dromcummer, then at 19:30 (7 hours 15 minutes later) at Mallow Rail BR with 367.83 $m^3.s^{-1}$ flow and after another four hours (at 23:30) at Killavullen with 279.03 $m^3.s^{-1}$ flow. At Dromcummer the river flow grew from 73.72 $m^3.s^{-1}$ to 261.64 $m^3.s^{-1}$ on 2014/02/02 at 23:45 after approximately 12 hours that means the growth rate of river flow is 15.66 $m^3.s^{-1}.h^{-1}$. While the growth rate of river flow at Mallow Rail BR and Killavullen are 12.5 $m^3.s^{-1}.h^{-1}$ and 6.8 $m^3.s^{-1}.h^{-1}$ respectively. This indicates the river flow grows more rapidly and decreases faster at Dromcummer than at the two other subcatchments. This is due to the fact that the gradient of the river is greater upstream (above Dromcummer) than the slope at the downstream locations of Mallow and Killavullen. This can be examined for more events to find a probable pattern to estimate the time of arrival floods in different subcatchments.

6.3.3 Model Validation for an Event

The accuracy and suitability of PDM for long periods (yearly) were examined in section 6.3.1. In this section we will study PDM's use for short period floods events (e.g. a few days) in more detail. There were several flood events (observed flow more than 320 $m^3.s^{-1}$) during 2010 to 2017. The largest flood event at Mallow Rail BR and the associated events in the two other subcatchments (as the largest event in Mallow Rail BR is not necessarily the largest event in the two other subcatchments) for each year were selected (see Table 6.5). Comparison between river flows in the three subcatchments was made to find the pattern of the peak flows and to select different peak flood events.

Table 6.5 Mallow Rail BR's annual maxima river flow in each year during 2010 - 2017

Time	Observed River flow ($m^3.s^{-1}$)
2010/01/13 00:30	504.03
2011/11/30 1:15	388.15
2012/06/08 10:30	369.38
2013/12/30 14:30	463.9
2014/02/03 19:30	367.83
2015/12/30 17:45	518.3
2016/01/02 07:45	524.02
2017/10/22 07:00	336.34

6.3.4 Case 1: Event from 3^{th} Feb 2014 to 5^{th} Feb 2014

The PDM model was run (with calibration parameters) for the event between 3^{th}Feb 2014 to 5^{th} Feb 2014 in three subcatchments as validation (i.e. no change to the parameters as used in the calibration run). To obtain the correct initial states in the start of the event two ways can be applied:

- Determine the number of time step (Noi) for initialization at the start of an event. In input file (*.inp) after start and ending event's dates, adding this number. For example:

```
Event Begin      Event End      Noi     End 15min Rain
'00:00  03 Feb 2014'     '23:45 05 Feb 2014'   2594  /
```

- Consider the whole period of records as a two continues events. The first event starts from the beginning of the available record or calibration period to the just a time step before the beginning of an intended event and the second event starts from the beginning of intended event to the end of intended event. For example:

```
Event Begin      Event End      Noi     End 15min Rain
'00:00  01 Jan 2014'     '23:45 02 Feb 2014'    /
'00:00  03 Feb 2014'     '23:45 05 Feb 2014'    /
```

Table 6.7 shows the summary of each model run performance. The resulting hydrographs are illustrated in Figure 6.14 to Figure 6.16. They indicate the shape of the computed flow in Dromcummer is similar to the shape of the observed flow with the observed peak flows seeming to be wider and decreasing more rapidly than the modeled flow. The accuracy of the model is 93 %, which indicates the estimation of the magnitude of the computed flows, are close to the observed flows for most of the time.

Table 6.7 At the three river stations, a summary of validation of the PDM Model for the Event from 03/02/2014 to 05/02/2014

	R^2	Rainfall (mm)	Computed baseflow (mm)	Observed flow (mm)	Computed flow (mm)	Potential evaporation (mm)	Actual evaporation (mm)	Net rainfall (mm)
Dromcummer	0.9376	37.04	8.67	30.46	29.34	0.51	0.51	36.53
Mallow Rail BR	0.8089	35.52	12.70	31.48	29.68	0.51	0.51	35.01
Killavullen	0.7834	35.97	10.54	25.15	27.15	0.51	0.51	35.46

Note that observed flows and computed flows are the total flows during the event. The R^2 Efficiency is a function of the sum of differences of observed flow and computed flow in each time step. When R^2 Efficiency is small, it means the computed flow in a specific time is not a good estimation of observed flow. So observed flow could be close to computed flow but R^2 Efficiency being small that indicates the overall performance of the model is not sufficient (see Figure A.4 (c) in Appendix A). The PDM model is sufficient when the R^2 Efficiency is large and observed flow is close to computed flow. Actual evaporation is about 90% of Potential evaporation.

The model performance at Mallow Rail BR is not as good as the model's performance at Dromcummer but is better than the model's performance at Killavullen. The shape of the computed hydrograph is similar to the shape of the observed hydrograph, but does not show the two successive peaks. In general, the computed flow is underestimated.

The model performance at Killavullen is also not as good as the model's performance at the two other subcathments. The growth rate of computed flow is bigger than the growth rate of observed flow so the peak value was estimated earlier and the rate of decrease of computed flow is less than the rate of decrease of observed flow. In general, the computed flow is overestimated at Killavullen.

Figure 6.14 to Figure 6.16 show the modelled versus observed highest flood event at the three river stations in 2014. The highest modelled peak flow was estimated as 265.035 $m^3.s^{-1}$ on 2014/02/03 at 11:15, one hour earlier than the observed peak flow of 261.64 $m^3.s^{-1}$ in Dromcummer (Table 6.6 and Table 6.8).

At Mallow Rail BR the highest modelled peak flow was estimated as 340.862 $m^3.s^{-1}$at 15:00, four hours earlier than the observed peak flow of 367.83 $m^3.s^{-1}$. Modelled peak flow in Mallow was estimated four hours later than the modelled peak flow in Dromcummer. The modelled peak flow in Killavullen was estimated one hour thirty minutes later than Mallow at 16:30 with modelled flows of 280.376 $m^3.s^{-1}$ by comparison with the observed peak flow of 279.04 $m^3.s^{-1}$. The highest modelled flow estimated well but seven hours earlier than the observed flow in Killavullen.

At Dromcummer the modelled river flow growth from 74.0 $m^3.s^{-1}$ to 265.03 $m^3.s^{-1}$ at 11:15 after approximately 11 hours so the growth rate of modelled river flow is 17.66 $m^3.s^{-1}.h^{-1}$. While the growth rate of modelled river flow at Mallow Rail BR and Killavullen are 20.75 $m^3.s^{-1}.h^{-1}$ and 4.01$m^3.s^{-1}.h^{-1}$.the growth rate of modelled flow slightly different from growth rate of observed flows particularly for Mallow which growth rate of modelled flows was estimated 20.7 $m^3.s^{-1}.h^{-1}$ while the growth rate of observed flows was 12.5 $m^3.s^{-1}.h^{-1}$. This indicates the modelled river flows grow faster than observed flows. Table 6.8 shows the modelled peak flow time, magnitude

and the rising rate and falling rate at three subcatchments indicates that the modelled flow falling limb is more than twice as rising limb. Further research might address the possibility of using the model to generate wider hydrographs.

Table 6.8 Comparison of modelled peak time, flow, rising and falling limb at three river stations.

	Time peak	Flow peak $m^3.s^{-1}$	Rising rate $m^3.s^{-1}.h^{-1}$	Falling rate $m^3.s^{-1}.h^{-1}$	Rising limb hours	Falling limb hours
Dromcummer	2014/02/03 11:15	265.035	17.36	7.25	11:15	28:30
Mallow Rail BR	2014/02/03 15:00	340.862	20.75	8.14	11:15	27:45
Killavullen	2014/02/03 16:30	280.376	4.015	5.29	12:00	26:30

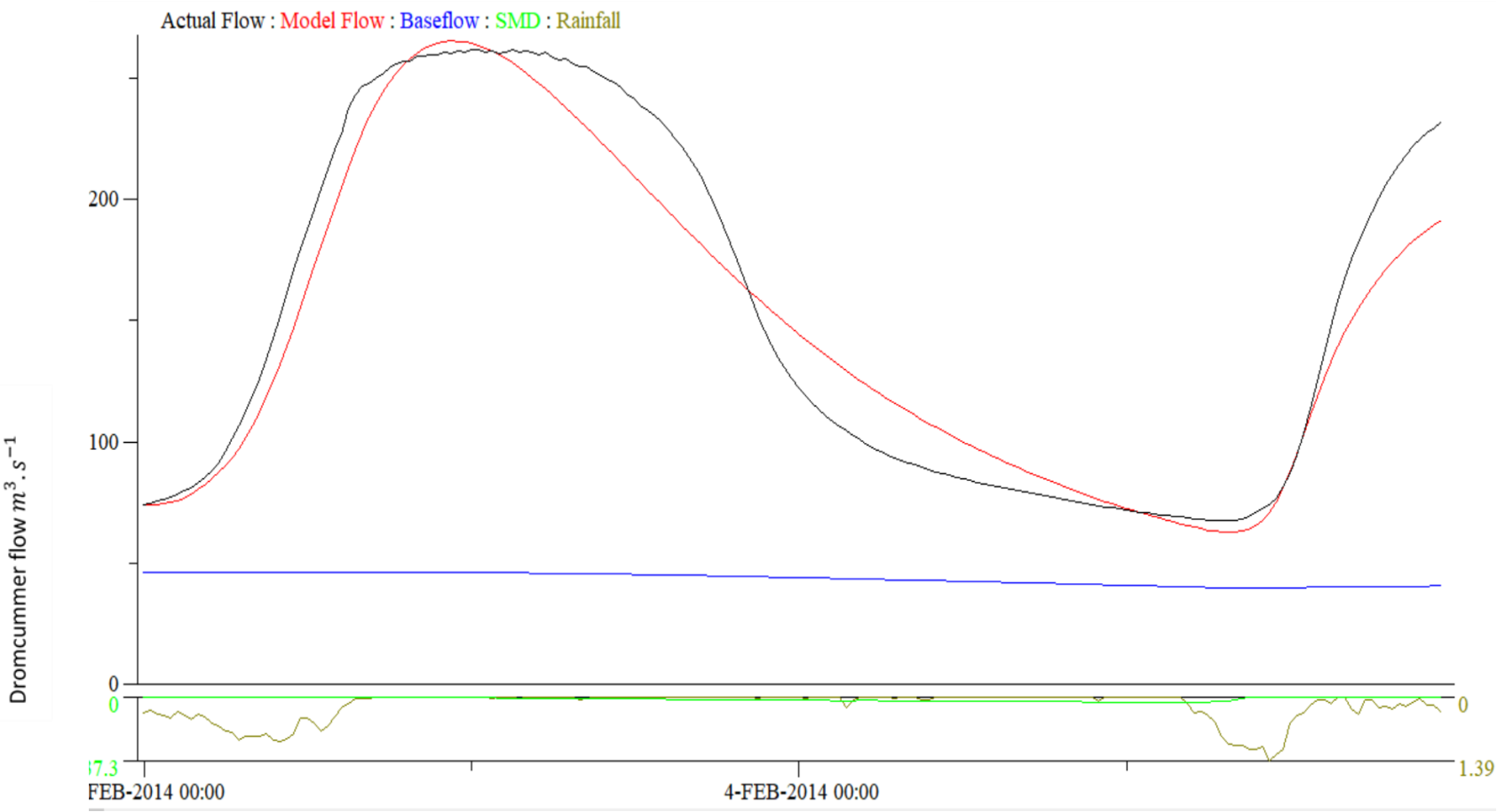


Figure 6.14 PDM simulation for validation, event 03/02/2014 to 05/02/2014 in Dromcummer, $R^2 = 0.9376$

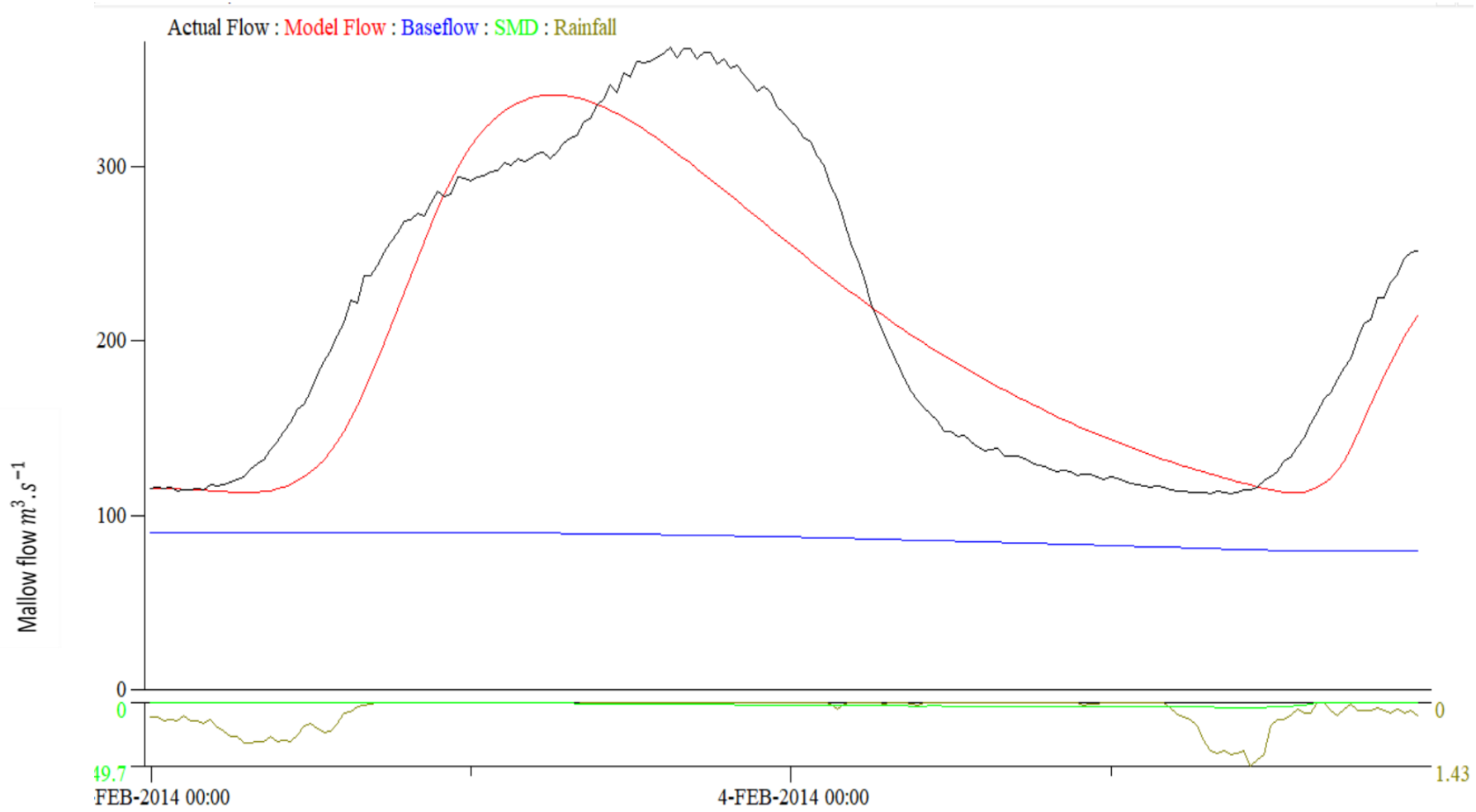


Figure 6.15 PDM simulation for validation, event 03/02/2014 to 05/02/2014 in Mallow Rail BR, $R^2 = 0.8089$

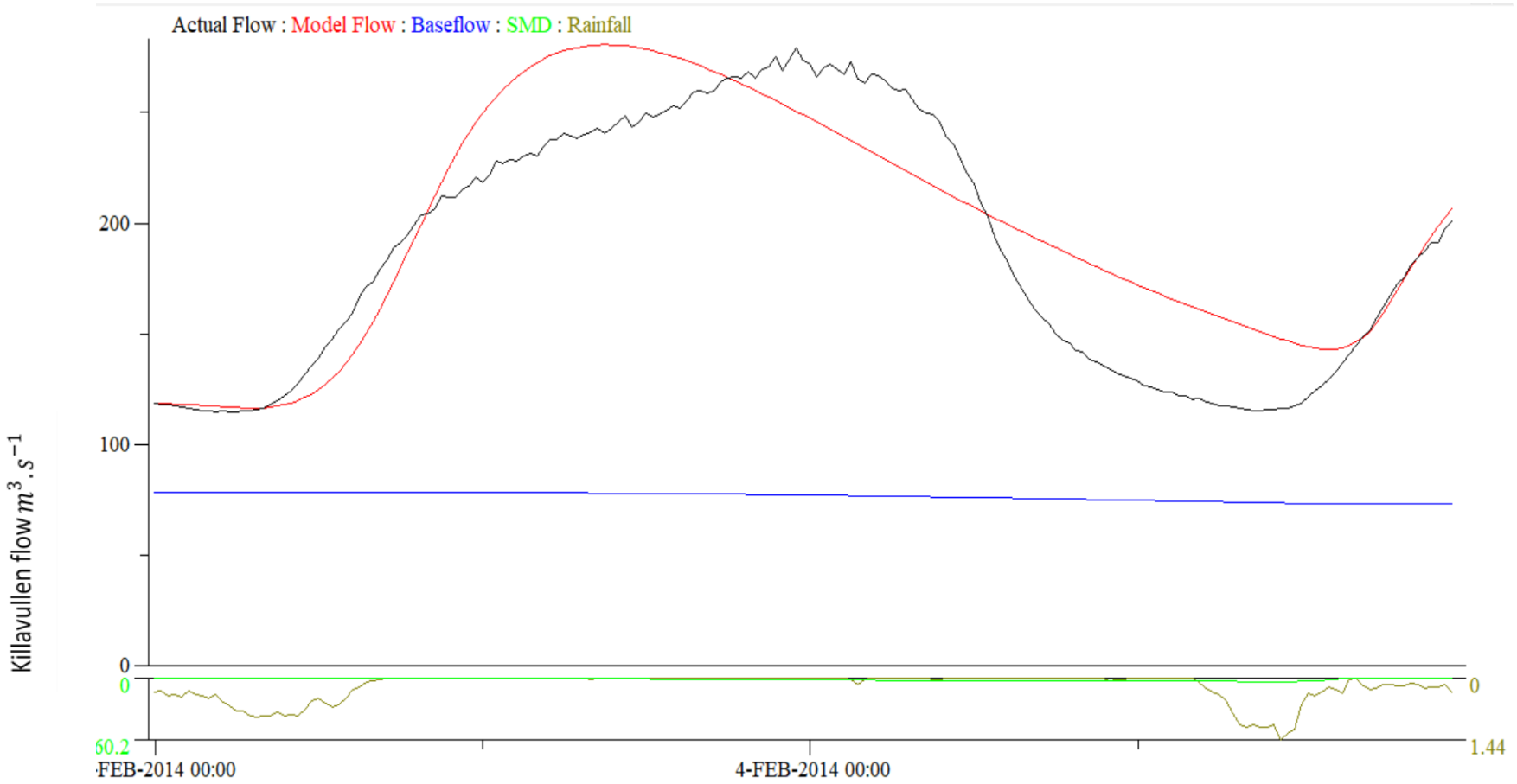


Figure 6.16 PDM simulation for validation, event 03/02/2014 to 05/02/2014 in Killavullen, $R^2 = 0.7834$

Chapter 7 Precipitation Streamflow Sensitivity Modelling

7.1 Precipitation Streamflow Sensitivity Modelling

Floods can be the result of several factors with rain being the most important. Generally increasing intensive and heavy rain can generate flooding. Trend analysis on the Killavullen subcatchment with area 1248 km^2 confirmed that there are strong trends in both temperature and rainfall and also showed a significant positive persistent trend in the flows, most of which is attributable to the unprecedented large rainfall amounts (Nicholson, 2012). While using global data sets for non-urban catchments in different geographical zones shows that for catchments with areas greater than 1000 km^2, the increase in rainfall from a warming world does not seem to increase the magnitude of extreme flood events (Wasko & Sharma, 2017). However for catchments with areas less than 1000 km^2, the reverse seems to be true (Wasko & Sharma, 2017). This latest study indicates precipitation-temperature sensitivity is positive while streamflow-temperature sensitivity is negative. That means precipitation increases with higher temperature while the streamflow paradoxically decreases with higher temperature. This can be explained using the term of "hydrological losses". Considering the fact that all precipitation is not converted to streamflow, some precipitation infiltrates to ground, some evaporates and the rest which is less than the total precipitation will generate streamflow. The trend analysis was applied in different catchment sizes which indicates the streamflow increases with higher temperature for smaller catchments with less precipitation losses (they define small catchment with area less than 1000 km^2 and large catchment with area larger than 1000 km^2). If we consider the Killavullen subcatchment as a small catchment, both studies (Nicholson, 2012 and Wasko & Sharma, 2017) represent the same result.

The PDM Model's validation in chapter 6 confirms that the PDM model can be used to account for the variability of observed flows in the Munster Blackwater in different situations over both long and short periods of time especially when a high standard of data quality is available. The next step would be to test the PDM model to predict the computed flows under possible future climate change conditions. One of these climatic change scenarios is the growth of precipitation as is suggested for Ireland (Fealy, Bruyére, & Duffy, 2011; Gleeson, McGrath, & Treanor, 2013). In this chapter, the relation between growth of rainfall and flows will be examined. First, the PDM models have been run for the year 2014 with the annual precipitation increased linearly by 10%, 15%, 20%, 25% and 30%. Comparison of computed flows of these PDM model runs for each subcatchment is illustrated in Figure 7.1(a) to 7.1(c). This indicates that there are different patterns of growth of computed flows.

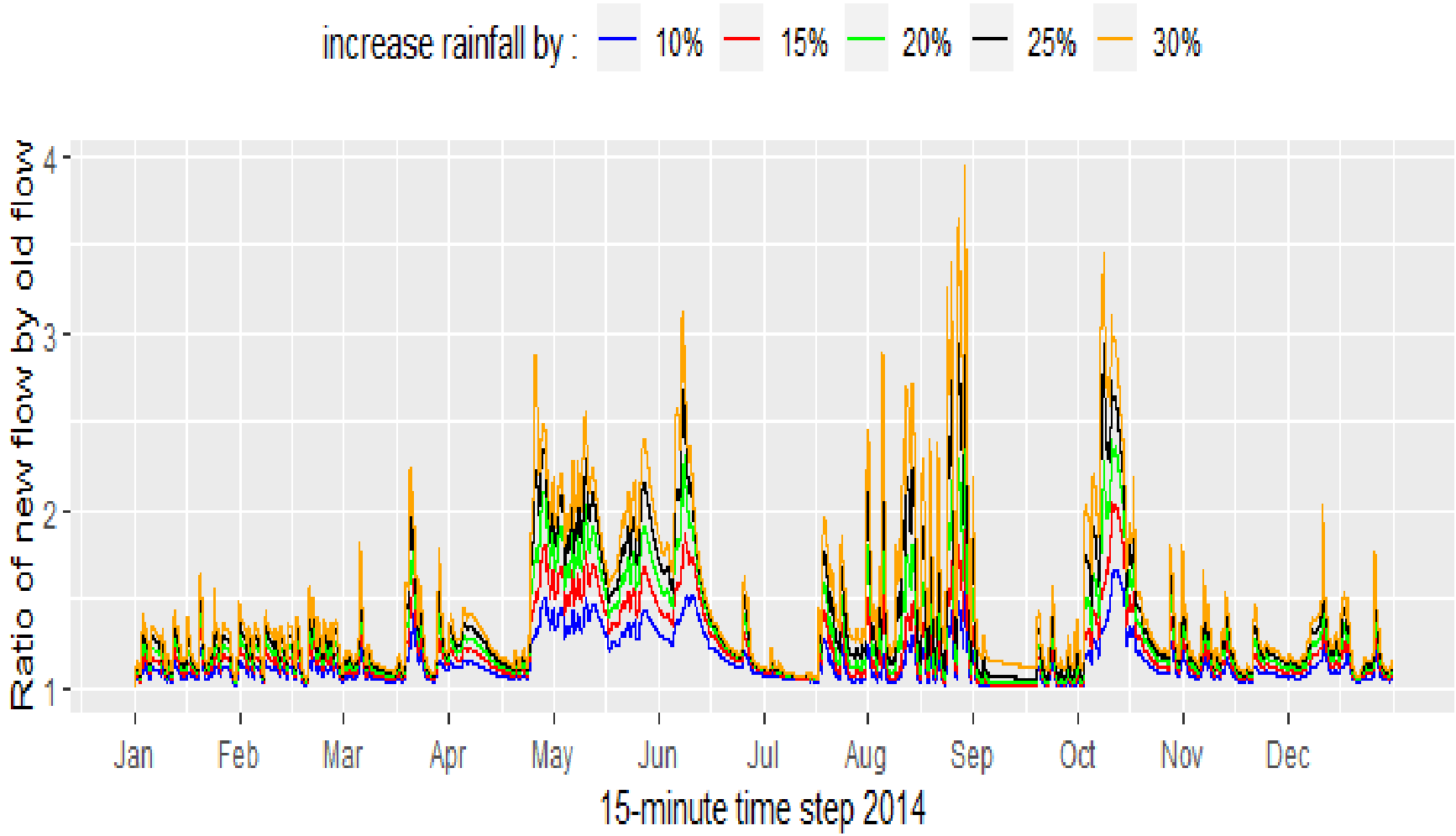


Figure 7.1(a) Comparison between the ratio of new flow (computed flow with rainfall increased by 10, 15,.. 30 %) to old flow (computed flows with actual rainfall) in Dromcummer for 2014

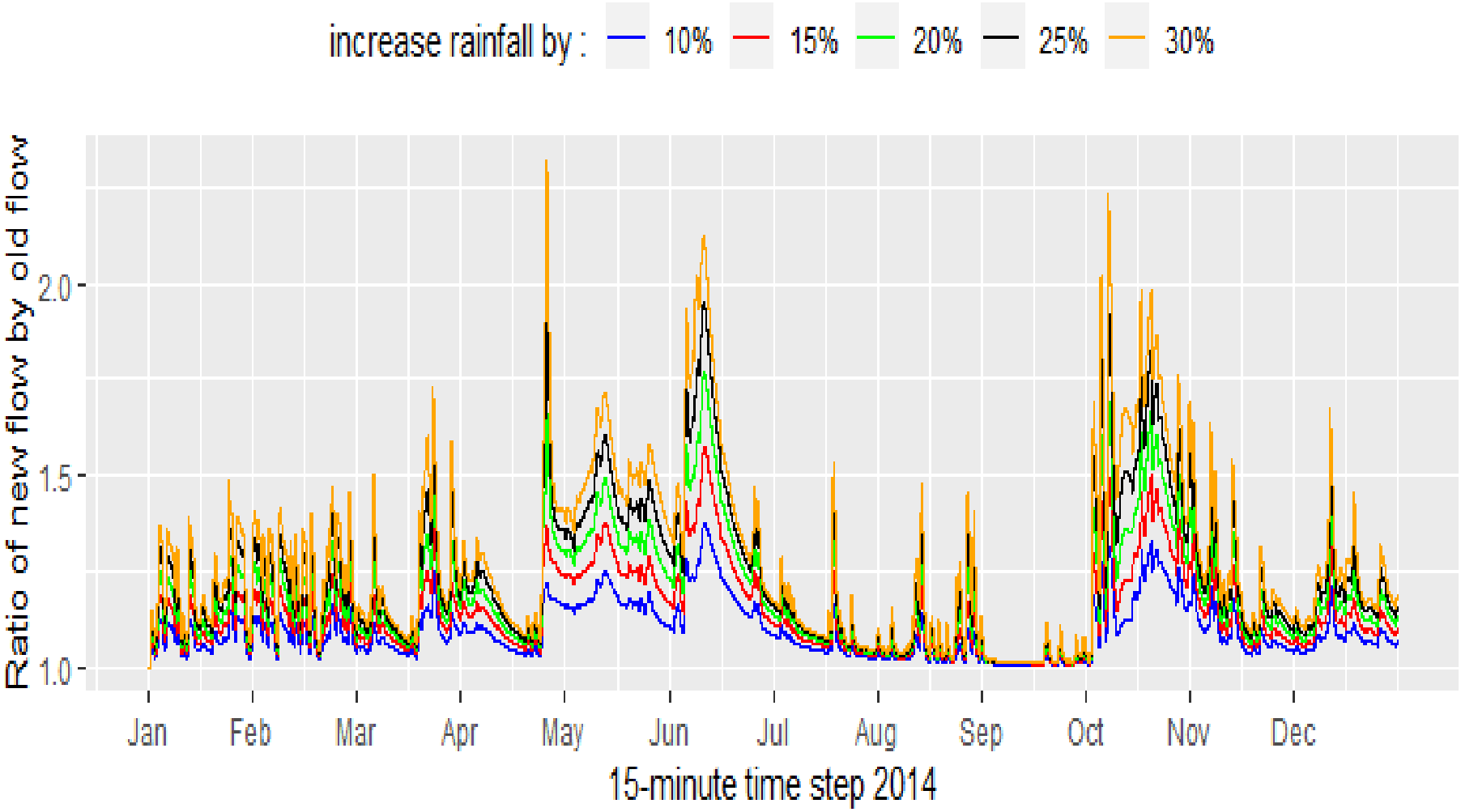


Figure 7.1(b) Comparison between the ratio of new flow (computed flow with rainfall increased by 10, 15,.. 30 %) to old flow (computed flows with actual rainfall) in Mallow Rail BR for 2014

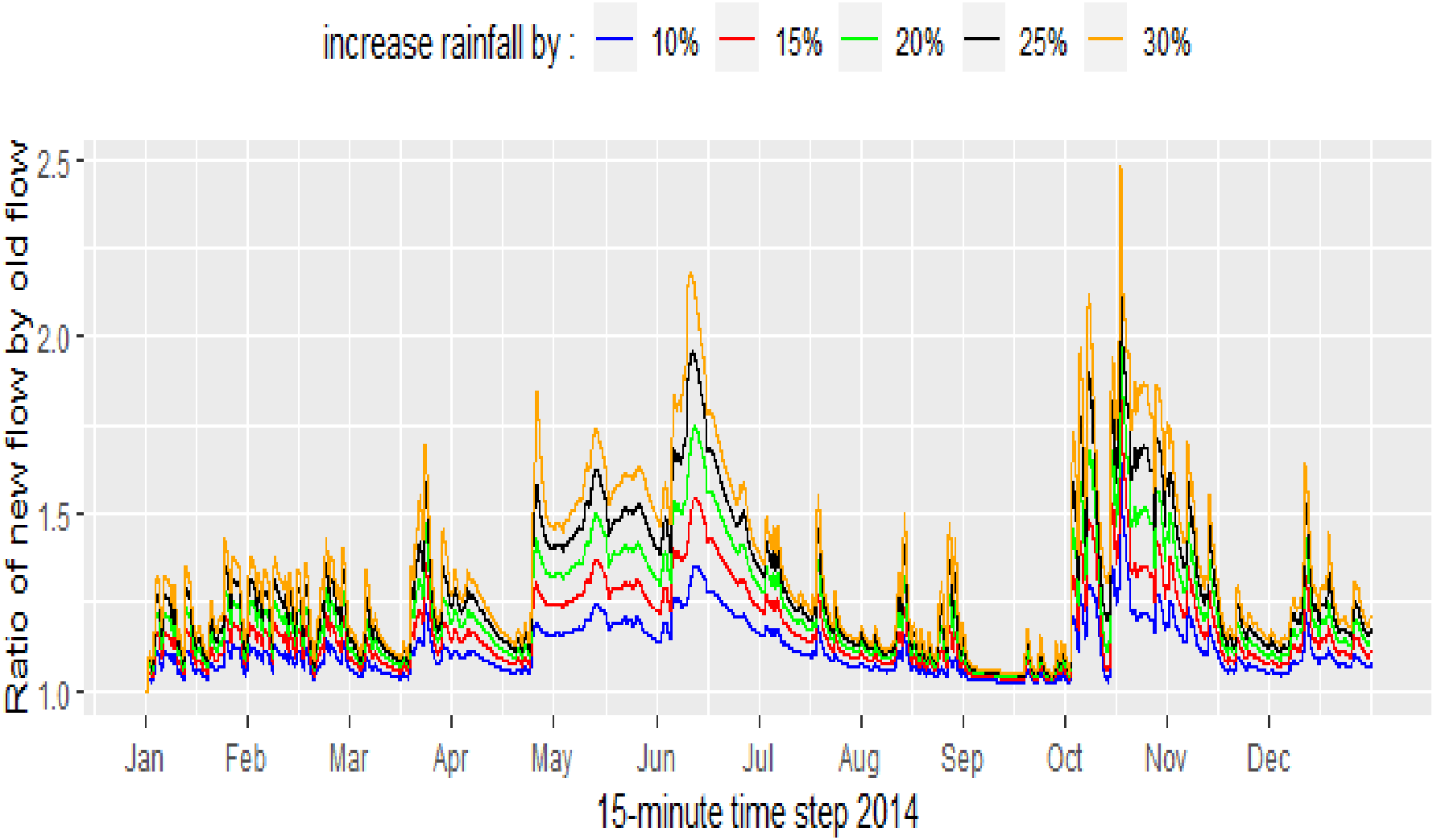


Figure 7.1(c) Comparison between the ratio of new flow (computed flow with rainfall increased by 10, 15,.. 30 %) to old flow (computed flows with actual rainfall) in Killavullen for 2014

Figure 7.2 shows a shorter period of time (between 2014/09/30 to 2014/10/04). There is a small percentage increase of computed flow for low flows and there is greater percentage increase of computed flow for higher flows. To examine these two patterns, two periods were selected: one contains maximum observed flow and the second contains the low observed flows.

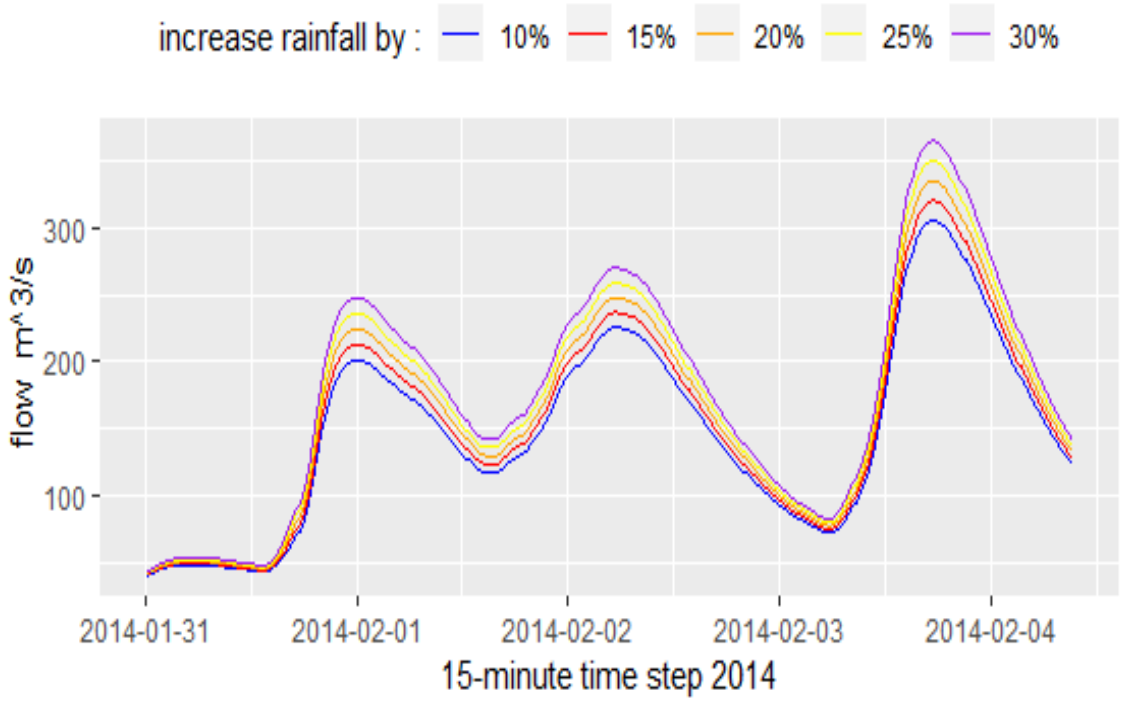


Figure 7.2 Comparison between the computed flows when rainfall increases by (10, 15,.. 30 %) in Dromcummer for period of 31^{th}Jan 2014 to 4^{th} Feb 2014

7.1.1 Pattern 1: period containing the largest event (high flow)

The PDM model was run for the case study event in each subcatchment when precipitation was increased by 10%, 15%, 20%, 25% and 30% (presented in section 6.3.1). A comparison of computed flows of these PDM model runs for the Dromcummer subcatchment is shown in Figure 7.3. Comparison of the ratio of the computed flow when rainfall is increased by 10, 15,20..30 % to computed flow with actual rainfall (Figure 7.4) indicates that there is a strong positive linear correlation between increases in computed flows and increases in rainfall when the observed flow is large.

The linear regression models (M1, M2 and M3) for three subcatchments (Dromcummer, Mallow Rail BR and Killavulllen) can be written as follows:

M1: Computed flow growth% = 1.06 * rainfall growth% + 0.079 Eq (7.1)

M2: Computed flow growth% = 1.04 * rainfall growth% + 0.013 Eq (7.2)

M3: Computed flow growth% = 1.02 * rainfall growth% + 0.09 Eq (7.3)

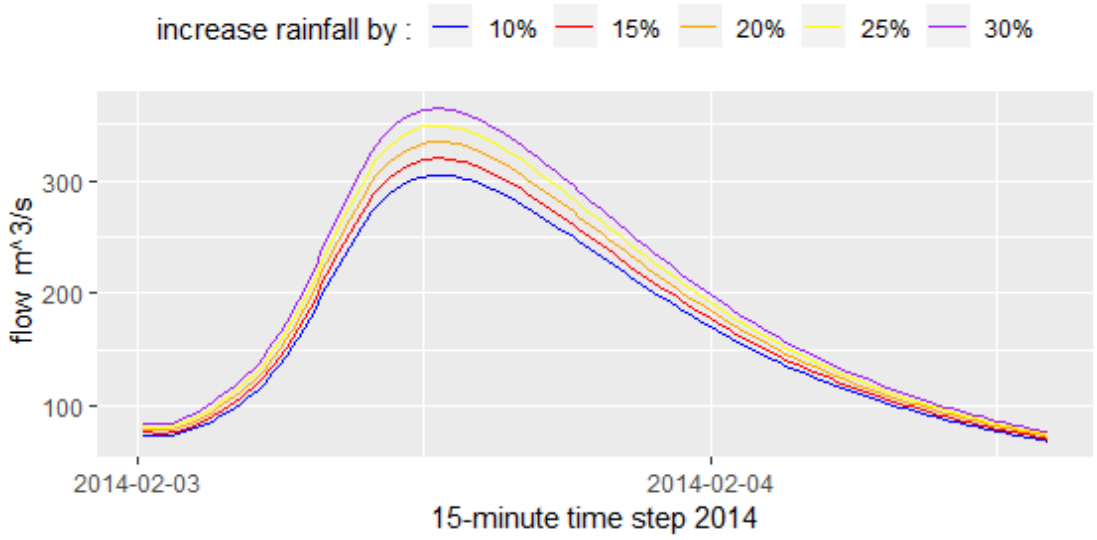


Figure 7.3 Comparison between the computed flows when rainfall increases by (10, 15,.. 30 %) in Dromcummer for the period 3^{th}Feb 2014 to 5^{th} Feb 2014 (high flow events)

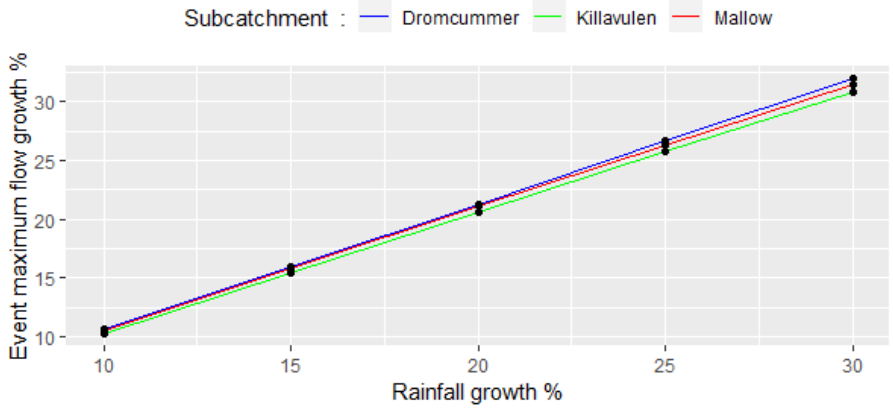


Figure 7.4 Comparison between the ratio of new flow (computed flow with rainfall increases by 10, 15,.. 30 %) to old flow (computed flows with actual rainfall) for largest flow (3^{th}Feb 2014 to 5^{th} Feb 2014) for three subcatchments. The equation for high flow in Dromcummer is:

Computed flow growth % = 1.06 * rainfall growth % + 0.079

Table 7.1 shows the summary of each linear regression model for each subcatchment. P-values are very small, less than 0.05 (the significance level) indicating the linear regression model fits the data better than the model with no independent variables.

This means that for every additional 1 % growth in rainfall there is 1.06 % growth in computed flow at Dromcummer. Similarly, increasing rainfall by 10% results is an increase in the flow by 10.6% or increasing the rainfall by 20% results is a flow increase of 21.2 %. These linear models can be used for flow prediction in each subcatchment when flows are high (e.g. after heavy rainfall events). This model exprement assumes there is no change in evaporation. However with increasing rainfall (as a result of increase in temperature) we expect an increase in evaporation. There may also be an increase or a decrease in soil moisture. If soil moisture increases, then it is possible that flows increases will be less than the values given below in Table 7.1.

Table 7.1 Summary of each linear regression model for each subcatchment.
M1: Computed flow growth% = 1.06 * rainfall growth% + 0.079
M2: Computed flow growth% = 1.04 * rainfall growth% + 0.013
M3: Computed flow growth% = 1.02 * rainfall growth% + 0.09

Model	Intercept	Slope	R-square	P-value
M1 for Dromcummer	0.079	1.06	1.00	$2.105 * 10^{-10}$
M2 for Mallow	0.13	1.04	1.00	$2.472 * 10^{-9}$
M3 for Killavullen	0.09	1.02	1.00	$1.861 * 10^{-9}$

PDM describe the variation of soil moisture storage capacity at different points of the catchment by a probability distribution. PDM introduces the distribution of storage depth over the catchment, which ranges from a minimum depth C_{min} to a maximum depth C_{max}. So C_{min} and C_{max} are two parameters to control the soil moisture in PDM.

Soil tension storage capacity (S_t) is another parameter which controls the soil moisture and evaporation to prevent complete drainage of soil moisture store and maintain evaporation (Table 5.3). So in further research by changing these three parameters in PMD input file (*.inp) and keeping other parameters unchanged the impact of soil moisture on flow can be examined when rainfall increases.

7.1.2 Pattern 2: period containing the smallest event (low flow)

The PDM model was run for the period which contains low observed flow in each subcatchment when precipitation increases by 10%, 15%, 20%, 25% and 30%. Comparison of computed flows of these PDM model runs in the period of (2014/09/17 to 2014/09/19) for the Dromcummer subcatchment (Figure 7.5) indicates the growth of computed flow is not linear when precipitation

increases by 10%, 15%, 20%, 25% and 30%. Figure 7.6 illustrates the nonlinear relation between the ratio of the computed flow when rainfall increases by 10, 15, 20, 25, 30 % by computed flow with actual rainfall and rainfall growth. Table 7.2 shows the computed flow when rainfall increases by 10, 15,20, 25, 30% . When rainfall increases by 10% computed flow increases only 1.150% ((0.966/0.955) *100 -100 ≈ 1.15). For 20% increase of rainfall computed flow increase by around 3.14% ((0.985/0.955)*100 – 100 ≈ 3.14). While 30% increases of rainfall produce around 12% increase of computed flow ((1.069/0.955)*100 – 100 ≈ 11.93). Further investigation is needed to obtain the nonlinear model for each subcatchment.

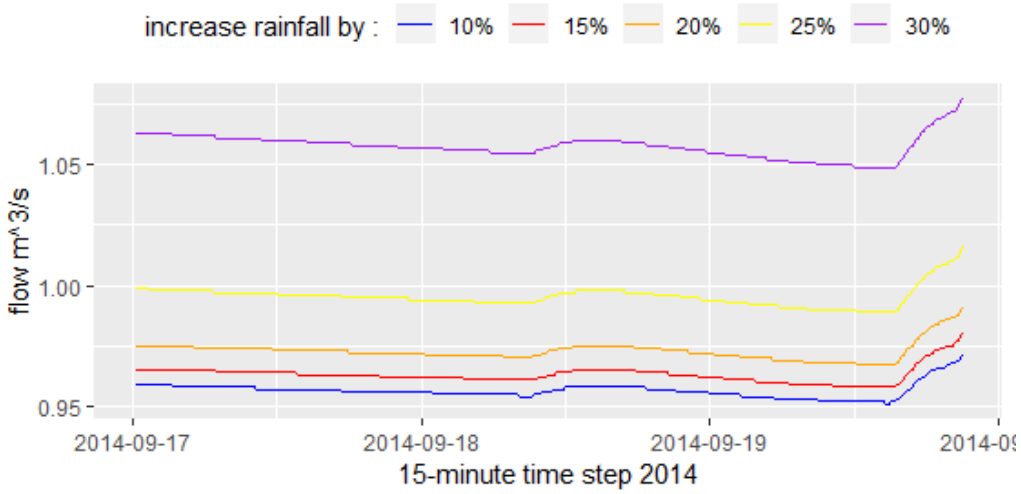


Figure 7.5 Comparison between the computed flows when rainfall increases by (10, 15,.. 30 %) in Dromcummer for period of 2014-09-17 to 2014-09-19.(low flow event)

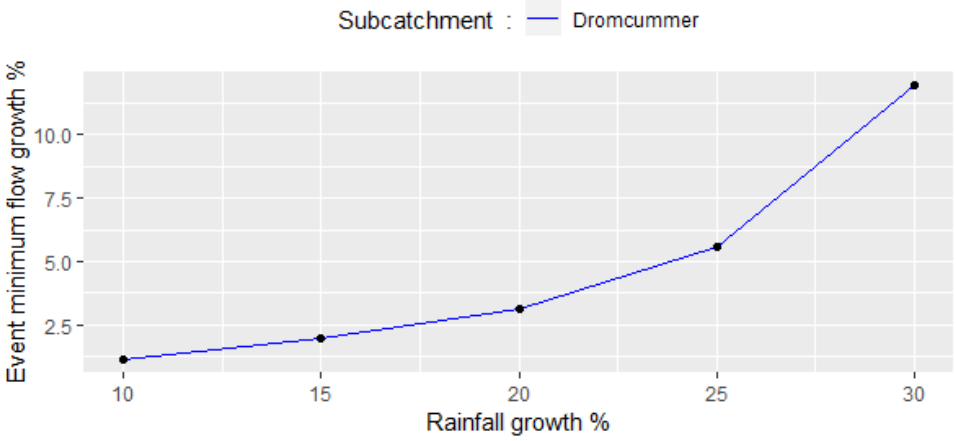


Figure 7.6 Comparison between the ratio of new flow (computed flow with rainfall increases by (10, 15,.. 30 %) to old flow (computed flows with actual rainfall) for lowest flow(2014-09-17 to 2014-09-19) in the Dromcummer subcatchment.

Table 7.2 High flows and low flows increase when rainfall increases by 10, .. 30% in Dromcummer

Rainfall Increases	Event Type: High Flow Flows ($(m^3.s^{-1})$	Event Type: low Flow Flows ($(m^3.s^{-1})$
0 %	272.692	0.955
10	301.821	0.966
15	316.325	0.974
20	330.792	0.985
25	345.249	1.008
30	359.692	1.069

7.2 Discussion

The nonlinear growth of flows when flow is low can be explained as follows: when the flow is low, some of the incoming water (rainfall growth) infiltrates to ground until the ground will be saturated so the growth of flow is slow. After ground saturation, all incoming water (rainfall growth) will produce runoff so in this stage the flow grows very fast. After this stage, pattern 1 occurs, meaning rainfall growth will produce runoff so the flows will grow linearly.

A limitation of these sensitivity analyses is that there is no change in potential/actual evapotranspiration for any rainfall increases. As reported in IPCC reports, Ireland is expected to undergo temperature increases which will increases AE. However in Ireland, there is little evapotranspiration during short term flood events (e.g. one day).

Chapter 8 Discussion and Conclusions

8.1 Discussion

The aim of this research has been to simulate the outflow hydrograph of storm events (rather than long-term flows) for the Munster Blackwater catchment. Several hydrological models from four main categories: lumped, semi-distributed, fully distributed and artificial neural networks have been reviewed to choose a suitable hydrological model for the Munster Blackwater catchment. The PDM hydrological lumped model (Moore, 2007) developed at the UK Center for Ecology & Hydrology was chosen for this study primarily because it does not require detailed data on the catchment topography, land cover and soils. It requires 15-minute rainfall, river flows and potential evaporation. With various parameterisations, PDM is then able to produce the riverflow hydrograph time series.

Precipitation data and river level/ flow data from different sources (OPW, EPA and Met Éireann) for three subcatchments of the Munster Blackwater catchment was examined to obtain the best possible input data suitable for PDM.

Primary analysis of rainfall and river flow data revealed a significant amount of missing values and outliers especially in precipitation data from the OPW. Several gap filling methods were developed and applied to reduce the amount of missing data and outliers. A machine learning method, Kmeans clustering, was applied to cluster the rain gauges based on their precipitation similarity. Then a gap filling method was applied in each cluster to fill the missing values and modify the outliers of each of the 23 rain gauges with values from its cluster's members. This method reduces significantly the amount of missing values and outliers when there are more than two members in each cluster and also reduces the chance of long sequences of missing values and outliers. A combination of several gap filling methods was applied to obtain maximum quality controlled data. There is a potential risk in gap filling the long sequences of missing values with their nearest records to create a long sequence of the same values in the data, so further examination is required in this case.

The validation of the calibrated PDM model for three subcatchments (Dromcummer, Mallow Rail BR and Killavullen) of the Munster Blackwater catchment for both long (annual) and short (flood event) periods of time was examined. PDM produced good results of river flow output during the calibration period for the year 2014, particularly when the model's automatic optimisation was used.

8.1.1 Validation for Long period of time (annually)

Tables A.1 (a) to (c) in Appendix A contain summaries of annual PDM Model runs in three subcatchments and Figures A.3 (a) to (c) show the resulting hydrographs.

The comparison of these tables and figures can be summarised as follows:

- The accuracy (R^2) for 2015 is greater than 90% in each subcatchment.
- The worst accuracy is in 2013 for all three subcatchments due to some remaining bad data which were not detected by the gap filling and outliers correction methods (high amount of rainfall but all less than 15 mm and not sequential, see Figure 4.17). Total computed flows during the year are overestimated in all subcatchments. There are a few periods when computed flows are highly overestimated, between April and May. At the beginning of Oct 2013 when computed flow peaks are very high, greater than 300 $m^3.s^{-1}$ and wide (long duration), the observed flows are around 100 $m^3.s^{-1}$ and narrow (short duration). These events need more investigation as it might be due to a timing error in the data. The rest of the peak observed flows are underestimated. (See Figures A.3(a) to (c) N0.4)
- The PDM model performance shows better accuracy at Mallow Rail BR and Killavullen, with R^2 more than 80 %.
- Overall computed flows per year are overestimated in each year except 2012 and 2015.

8.1.2 Validation for short period of time (event)

The largest events in each year between 2010 to 2017 at Dromcummer and the associated events in two other subcatchments were selected and the PDM model run for each subcatchment to examine the behaviour of the event in each consecutive subcatchment. The summary of these model runs is shown in Tables A.2 (a) to (c) and Figures A.4 (a) to (c), Appendix A illustrate the resulting hydrograph. The following results derive from study of these Figures and Tables:

- The peak timing is very accurate for all events in all subcatchments except the event of 2017 when the computed flows predicted reached their peaks earlier than observed flow peaks. Observed flows grow smoothly and declines sharply while computed flows behaves vice versa.
- The shapes of computed flow hydrographs for each event at each subcatchment are similar (see section 6.3.4, similar to case study 1).
- The overall performance of the model is worst at Killavulen for each event.
- In most events the overall performance of the models is best at Mallow Rail BR, the accuracy in some cases is greater than 90 %.
- The best peak estimates are at the Mallow Rail BR (R^2 = 0.9461).

- In most cases peak flows will estimated well, or overestimated in Dromcummer and Killavullen.

8.1.3 Conclusion

Preliminary analysis of rainfall and river flow data shows significant amount of missing values and outliers especially in precipitation data from the OPW. This justifies the use of several gap-filling methods particularly using Machin learning, i.e. Kmeans clustering method to reduce these significant amounts of missing and outliers to an acceptable level prior to run the PDM Model. (PDM omits the period of missing value in flow data and needs catchment average rainfall time series, which can be obtained by Thiessen polygon prior to run the PDM model).

The validation of the calibrated PDM model for three subcatchments (Dromcummer, Mallow Rail BR and Killavullen) of the Munster Blackwater catchment for both long (annual) and short (flood event) periods produced good results of river flow output and confirmed that PDM is a suitable hydrological model for modelling riverflow in the Munster Blackwater.

Although the sensitivity analysis was explored just by impact of increasing precipitation on riverflow (in absent of other weather features such as temperature) for both long (annual) and short (flood event) period. The results show general agreement with studies discussed in section 7.1 (Nicholson, 2012; Wasko & Sharma, 2017) by considering topography and catchment's size (the area of three Blackwater's subcatchments are less or around 1000 km). The results also show the high flows (period 03/02/2014 to 05/02/2014, consider the high flow events mostly occurs in winter and spring) increased at a little greater percentage than the rainfall and the low flows (period 17/09/2014 to 19/09/2014, the low flow events mostly occurs in summer and autumn) increased at a much lower rate than the rainfall increases (further research is needed for reduction of rainfall). This confirms the previous studies results discussed in section 1.3 (Charlton et al., 2001; Steele-Dunne et al., 2008) which predicted that reductions in riverflows are likely for autumn months, accompanied by higher flows in winter and spring.

The following conclusion derive from this study:

- PDM is a suitable hydrological model for modelling riverflows in the Munster Blackwater.
- Once calibrated and parameters are optimised by PDM, the streamflow is modelled well.
- Good simulations were achieved in validation by using the same parameter list as were used during calibration.
- Peak flows and low flows were modelled well (e.g R^2 = 0.9461).

- While the peak flow magnitudes were modelled well, the timing of the peaks was not as good as expected. This may be due to the fact that PDM produced hydrographs that tended to be short duration (narrow) almost bell shaped, while the observed peaks tended to be much longer duration (wide). Some observed peaks at Mallow and Killavullen lasted as long as 12 hours while PDM modelled peaks tended to be shorter about a few hours.
- In a sensitivity analysis when rainfall was increased by 10 to 30 %, the high flows increased at a little greater percentage than the rainfall. For instance for 10% increase in rainfall, the riverflow high peaks increased by 10.6 %. However the low flows increase at a much lower rate than the rainfall increases.

8.1.4 Recommendations for Further Research

- The quality of PDM models performance in both long and short periods of time is dependent on the data quality of both precipitation and river level/ flows. River level/ flow data in particular needs more examination to reduce the long sequences of missing values on the Munster Blackwater.
- Use other PDM models options to improve accuracy of the models: for example PDM extension for groundwater catchments.
- Apply the PDM models for different events including high, medium and low flows to examine the accuracy of models in these situations to find out the suitability of PDM models for different situations.
- More study of climatic conditions, examining the PDM models in different events with high, medium and low flows when the rainfall increases by 10, 15 , .. % and find the relationship between growth of rainfall and flow in different subcatchments to get better conclusions and find linear and/or nonlinear models to define this relation which can be used for flow prediction.
- Use PDM to change the soil moisture states. In other words reduce the soil moisture and see what is the impact on flood flows. Repeat for increases in soil moisture.
- Use PDM to examine the relation between growth of potential evaportation and flows when potential evaporation increases by 10%, 15%, 20%, 25% and 30%.

References:

Agrawal, R. K. R. A. (2013). An Introductory Study on Time Series Modeling and Forecasting. *ArXiv Preprint ArXiv:1302.6613*. https://doi.org/10.13140/2.1.2771.8084

Ambroise, B., Beven, K., & Freer, J. (1996). Toward a generalization of the TOPMODEL concepts: Topographic indices of hydrological similarity. *Water Resources Research*, *32*(7), 2135–2145. https://doi.org/10.1029/95WR03716

Ardabili, S., Mosavi, A., Dehghani, M., & Varkonyi-koczy, A. R. (2019). Deep learning and machine learning in hy- drological processes climate change and earth systems a systematic review, (August), 1–24. https://doi.org/10.20944/preprints201908.0166.v1

Asanka, P. (2017). Finding the optimal number of clusters for K-Means through Elbow method using a mathematical approach compared to graphical approach. Retrieved 27 August 2018, from https://www.linkedin.com/pulse/finding-optimal-number-clusters-k-means-through-elbow-asanka-perera

Beven, K. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, *16*(1), 41–51. https://doi.org/10.1016/0309-1708(93)90028-E

Botros, F. E., Onsoy, Y. S., Ginn, T. R., & Harter, T. (2012). Richards Equation–Based Modeling to Estimate Flow and Nitrate Transport in a Deep Alluvial Vadose Zone. *Vadose Zone Journal*, *11*(4), 0. https://doi.org/10.2136/vzj2011.0145

Brikowski, T. (2012). GEOS 4430 Lecture Notes : Darcy ' s Introduction.

Burton DeWilde. (2012). Classification of Hand-written Digits (3). Retrieved 5 September 2018, from http://bdewilde.github.io/blog/blogger/2012/10/26/classification-of-hand-written-digits-3/

Castrounis, A. (2016). Artificial Intelligence, Deep Learning, and Neural Networks, Explained. Retrieved 5 September 2018, from https://www.kdnuggets.com/2016/10/artificial-intelligence-deep-learning-neural-networks-explained.html

CEH. (2012). A Practical user guide to the PDM. (CEH) Centre for Ecology & Hydrology Maclean Building Crowmarsh Gifford Wallingford Oxfordshire UK.

Charlton, R., Moore, S., Sweeney, J., & Fealy, R. (2001). Climate change and water resources in Ireland : initial investigation using downscaled GCMS and hydrological modelling techniques. *Cuadernos de Investigación Geográfica*, *27*(0), 125. https://doi.org/10.18172/cig.1118

Chen, Y. L., Gao, J. X., Han, Y. W., & Chen, Y. M. (2009). Response of climate change to final energy-CO2 emission in Beijing. *2009 International Conference on Energy and Environment Technology, ICEET 2009*, *3*, 101–105. https://doi.org/10.1109/ICEET.2009.491

Corcoran, G. (2004). *Development and Examinations of Flood Warning Systems for the Munster Blackwater at Mallow. MEng.Sc. Thesis.University College Cork, 2004*.

Creating Thiessen polygons. (n.d.). Retrieved 7 September 2018, from http://resources.esri.com/help/9.3/arcgisdesktop/com/gp_toolref/coverage_toolbox/creating_thiessen_polygons.htm

Dabrowska, K., Zielinska, Budzynska, M., Gatkowska, M., Kowalik, W., Bartold, M., & Kiryla, M.

(2017). Importance of grasslands monitoring applying optical and radar satellite data in perspective of changing climate. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *2017-July*, 5782–5785. https://doi.org/10.1109/IGARSS.2017.8128322

Devia, G. K., Ganasri, B. P., & Dwarakish, G. S. (2015). A Review on Hydrological Models. *Aquatic Procedia*, *4*, 1001–1007. https://doi.org/10.1016/j.aqpro.2015.02.126

Donovan, T. (2015). *South western RBD CFRAM study*.

Doswell, C. A. (2003). Flooding. *Elsevier Science Ltd,University of Oklahoma, Norman, OK, USA*, 769–776.

Durand, P., Robson, A., & Neal, C. (1992). Modelling the hydrology of submediterranean montane catchments (Mont-Lozère, France) using TOPMODEL: initial results. *Journal of Hydrology*, *139*(1–4), 1–14. https://doi.org/10.1016/0022-1694(92)90191-W

Endrizzi, S., Dall'Amico, M., Gruber, S., & Rigon, R. (2017). GEOtop Users Manual, 124.

EPA, E. P. A. (2007). What is climate change ? *Environmental Protection Agency (EPA)*. Retrieved from http://www.epa.ie/climate/communicatingclimatescience/whatisclimatechange/

epaHydroNet. (2016). EPA - WISKI Web Public. Retrieved 7 September 2018, from http://www.epa.ie/hydronet/#Water Levels

Fealy, R., Bruyére, C., & Duffy, C. (2011). *Regional Climate Model Simulations for Ireland for the 21st Century. EPA Report No.244*. Retrieved from www.epa.ie

Fiore, S., Plociennik, M., Doutriaux, C., Palazzo, C., Boutte, J., Zok, T., … Aloisio, G. (2016). Distributed and cloud-based multi-model analytics experiments on large volumes of climate change data in the earth system grid federation eco-system. *Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016*, 2911–2918. https://doi.org/10.1109/BigData.2016.7840941

Gleeson, E., McGrath, R., & Treanor, M. (2013). *Ireland's Climate: the road ahead*. Retrieved from papers2://publication/uuid/DD5D3CBF-4FD6-488F-9503-D36BC1C5EF99

Gleeson, E., Whelan, E., & Hanley, J. (2017). Met Éireann high resolution reanalysis for Ireland. *Advances in Science and Research*, *14*, 49–61. https://doi.org/10.5194/asr-14-49-2017

Govindaraju, R. (2000). Artificial Neural Networks in Hydrology. I: Preliminary Concepts. *Journal of Hydrologic Engineering*, *5*(2), 115–123. https://doi.org/10.1061/(ASCE)1084-0699(2000)5:2(115)

Gupta, D., & Ghose, U. (2015). A comparative study of classification algorithms for forecasting rainfall. In *2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICRITO.2015.7359273

Hoppe, H., & Kiely, G. (1999). *Precipitation Over Ireland-Observed Change Since 1940*. *Phys. Chem. Earrh (E)* (Vol. 24).

Jazwinski. (1972). Stochastic processes and filtering theory. *IEEE Transactions on Automatic Control Volume: 17 , Issue: 5*. https://doi.org/10.1109/TAC.1972.1100136

K-means Clustering Tutorial-Machine Learning. (n.d.). Retrieved 11 September 2018, from https://www.dezyre.com/data-science-in-r-programming-tutorial/k-means-clustering-techniques-tutorial

Kiely, G., Leahy, P., Fenton, M., & Donovan. (2008). *FLOOD STUDIES UPDATE PROGRAMME WORK*

GROUP 3. FLOOD HYDROGRAPH ANALYSIS WORK PACKAGE 3.2 Flood Event Analysis- OPW Report. Retrieved from http://opw.hydronet.com/data/files/UCC_FEA_3_2_July_2008.pdf

Kiely, G., Leahy, P., Ludlow, L., Stefanini, B., Reilly, E., Monk, M., & Harris, J. (2007). *Extreme Weather , Climate and Natural Disasters in Ireland*. *Environmental Protection Agency*.

Leahy, P., & Kiely, G. (2010). Short Duration Rainfall Extremes in Ireland: Influence of Climatic Variability. *Water Resources Management (2011) 25(3) 987-1003*. https://doi.org/10.1007/s11269-010-9737-2

Leahy, P., Kiely, G., & Corcoran, G. (2008). Structural optimisation and input selection of an artificial neural network for river level prediction. *Journal of Hydrology*, *355*(1–4), 192–201. https://doi.org/10.1016/j.jhydrol.2008.03.017

Lewis, C. (2011). Measurement and modelling of soil hydrological properties for use in the distributed rainfall runoff model – GEOtop, (October), 218.

Li, G., Harrison, S. P., Bartlein, P. J., Izumi, K., & Prentice, I. C. (2013). Precipitation scaling with temperature in warm and cold climates: An analysis of CMIP5 simulations. *Res. Lett*, *40*, 4018–4024. https://doi.org/10.1002/grl.50730

Lin, K., Zhang, Q., & Chen, X. (2010). An evaluation of impacts of DEM resolution and parameter correlation on TOPMODEL modeling uncertainty. *Journal of Hydrology*, *394*(3–4), 370–383. https://doi.org/10.1016/j.jhydrol.2010.09.012

Liu, K. Q., Liang, Y. T., Zhou, J. L., & Liu, M. (2013). Drought and flood variations and rationality of land use in Honghu Lake during the last 50 years. *Proceedings of the 2013 3rd International Conference on Intelligent System Design and Engineering Applications, ISDEA 2013*, 612–617. https://doi.org/10.1109/ISDEA.2012.148

Mac Carthaigh, M. (1996). *An Assessment of the 1995 Drought Including a comparison with other known drought years*. Retrieved from https://www.epa.ie/pubs/reports/water/flows/EPA_Assessment_of_1995_Drought2.pdf

Mandych, A. F. (2009). *CLASSIFICATION OF FLOODS*. Retrieved from http://www.eolss.net/sample-chapters/c01/E4-06-02-05.pdf

Met Éireann. (2018a). climate change Met Éireann - The Irish Meteorological Service - Google Search. Retrieved 12 December 2018, from https://www.google.ie/search?q=climate-change

Met Éireann. (2018b). *Recent Warm Weather Report*. Retrieved from https://www.met.ie/cms/assets/uploads/2018/07/WarmWx062018-4.pdf

Met Éireann. (2018c). Weather Extreme Records for Ireland - Met Éireann - The Irish Meteorological Service. Retrieved 18 November 2018, from https://www.met.ie/climate/weather-extreme-records

Met, O. (2017). Data Assimilation - Met Office. Retrieved 18 November 2018, from https://www.metoffice.gov.uk/research/weather/data-assimilation

Moličova, H., Grimaldi, M., Bonell, M., & Hubert, P. (1997). Using topmodel towards identifying and modelling the hydrological patterns within a headwater, humid, tropical catchment. *Hydrological Processes*, *11*(9), 1169–1196. https://doi.org/10.1002/(SICI)1099-1085(199707)11:9<1169::AID-HYP551>3.0.CO;2-W

Moore, R. J. (2007). The PDM rainfall-runoff model. *Hydrology and Earth System Sciences*, *11*(1), 483–499. https://doi.org/10.5194/hess-11-483-2007

Moore, R. J., & Bell, V. a. (2002). Incorporation of groundwater losses and well level data in rainfall-runoff models illustrated using the PDM. *Hydrology and Earth System Sciences*, *6*(1), 25–38. https://doi.org/10.5194/hess-6-25-2002

Nandigam, V., & Arrowsmith, J. R. (2018). OpenTopography. Retrieved 27 August 2018, from https://opentopography.org/

Napolitano, G. (2011). An exploration of neural networks for real-time flood forecasting PHD Thesis. The University of Leeds School of Geography.

Nicholson, O. (2012). An Analysis of Changes in the Long Term Characteristics of River Flows in the Munster Blackwater Catchment, (June). Retrieved from http://core.kmi.open.ac.uk/download/pdf/11527068.pdf

Nolan, P. (2015). *Ensemble of regional climate model projections for Ireland- Report No. 159*. EPA, Wexford, Ireland. Retrieved from www.epa.ie

Nourani, V., Roughani, A., & Gebremichael, M. (2011). Topmodel capability for rainfall-runoff modeling of the Ammameh watershed at different time scales using different terrain algorithms. *Journal of Urban and Environmental Engineering*, *5*(1), 1–14. https://doi.org/10.4090/juee.2011.v5n1.001014

O' Brien, J. (2014). *Ballyguyroe Landfill Site Annual Environmental Report Cork County Council Cork County Council-Ballyguyroe Landfill Annual Environmental Report BALLYGUYROE LANDFILL SITE ANNUAL ENVIRONMENTAL REPORT*. EPA Wexford, Ireland. Retrieved from http://www.epa.ie/licences/lic_eDMS/090151b2804dc51a.pdf

OPW. (2004). Report of the flood policy review group, 24–25.

OPW HYDRO website. (2013). Office of Public Works: Station Search. Retrieved 16 November 2018, from http://waterlevel.ie/hydro-data/disclaimer.html

Pina, R. D., Ochoa-Rodriguez, S., Simões, N. E., Mijic, A., Marques, A. S., & Maksimović, Č. (2016). Semi- vs. Fully-distributed urban stormwater models: Model set up and comparison with two real case studies. *Water (Switzerland)*, *8*(2). https://doi.org/10.3390/w8020058

Probability density function. (2015). 1 Engineering Computation Part 6. 2 Probability density function. - ppt download. Retrieved 26 September 2018, from https://slideplayer.com/slide/5016926/

Quinn, P., Beven, K., & Chevallier, P. (1991). The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrological Processes*, *5*(October 1990), 59–79. https://doi.org/10.1002/hyp.3360050106

Santoalla, D. (2016). Packages - GRIB-API - ECMWF Confluence Wiki. Retrieved 16 November 2018, from https://confluence.ecmwf.int/display/GRIB/Packages#Packages-Python

Sattler, K.-U. (2016). Data Quality Dimensions. *Encyclopedia of Database Systems*, 1–5. https://doi.org/10.1007/978-1-4899-7993-3_108-2

Scanlon, T. M., Kiely, G., & Xie, Q. (2004). A nested catchment approach for defining the hydrological controls on non-point phosphorus transport. *Journal of Hydrology*, *291*(3–4), 218–231. https://doi.org/10.1016/j.jhydrol.2003.12.036

SciLine. (2017). Attribution Science: Climate Change & Extreme Weather. Retrieved 19 September 2018, from https://www.sciline.org/evidence-blog/climate-attribution

Sharif, M., & Burn, D. H. (2006). Simulating climate change scenarios using an improved K-nearest

neighbor model. *Journal of Hydrology*, *325*(1–4), 179–196. https://doi.org/10.1016/j.jhydrol.2005.10.015

Sitterson, J., Knightes, C., Parmar, R., Wolfe, K., Muche, M., & Avant, B. (2017). An Overview of Rainfall-Runoff Model Types. *U.S. Environmental Protection Agency*, (September), 0–29.

Sozollosi-Nagy, A. (1976). Introductory Remarks on the State Space Modeling of Water Resource Systems. *International Institute for Applied Systems Analysis*, (RM-76-73), 81. Retrieved from http://pure.iiasa.ac.at/id/eprint/611/1/RM-76-073.pdf

Steele-Dunne, S., Lynch, P., McGrath, R., Semmler, T., Wang, S., Hanafin, J., & Nolan, P. (2008). The impacts of climate change on hydrology in Ireland. *Journal of Hydrology*, *356*(1–2), 28–45. https://doi.org/10.1016/j.jhydrol.2008.03.025

Steinmann, E. (2005). *An Investigation of Flood Forecasting using a Physically-Based Rainfall-Runoff Model*. MEng.Sc. Thesis. University College Cork. Retrieved from http://nina.ucc.ie/hydromet/files/Steinmann.2005.pdf

Tarboton, D. G. (2003). On doing better hydrological science. In *Rainfall-Runoff Processes*. https://doi.org/10.1002/hyp.7108

Trenberth, K. E., Cheng, L., Jacobs, P., Zhang, Y., & Fasullo, J. (2018). Hurricane Harvey Links to Ocean Heat Content and Climate Change Adaptation. *Earth's Future*, *6*(5), 730–744. https://doi.org/10.1029/2018EF000825

Treut, L., Somerville, R., Cubasch, U., Ding, Y., Mauritzen, C., Mokssit, A., … Tignor, M. (2007). Historical Overview of Climate Change Science.pdf. *Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.*

Trevino, A. (2016). Introduction to K-means Clustering. Retrieved 6 September 2018, from https://www.datascience.com/blog/k-means-clustering

Vivoni, V. Y., & Ivanov, R. L. B. tRIBS user manual Massachusetts Institute of Technology. (2002).

Wasko, C., & Sharma, A. (2017). Global assessment of flood and storm extremes with increased temperatures. *Scientific Reports*, *7*(1), 1–8. https://doi.org/10.1038/s41598-017-08481-1

Wheater, H. S., Jakeman, A. J., & Beven, K. J. (1993). Progress and directions in rainfall-runoff modelling. *Modelling Change in Environmental Systems*. https://doi.org/citeulike-article-id:13481151

Wikle, C. K., & Berliner, L. M. (2007). A Bayesian tutorial for data assimilation. *Physica D: Nonlinear Phenomena*, *230*(1–2), 1–16. https://doi.org/10.1016/j.physd.2006.09.017

Wolock, D. M. (1995). Effects of Subbasin Size on Topographic Characteristics and Simulated Flow Paths in Sleepers River Watershed, Vermont. *Water Resources Research*, *31*(8), 1989–1997. https://doi.org/10.1029/95WR01183

Xu, C. (2002). *Hydrologic Models University of Uppsala*. *Report* (Vol. 72).

Yamada, I. (2017). Thiessen polygons. *In International Encyclopedia of Geography: People, the Earth, Environment and Technology (Pp. 1–6). Oxford, UK: John Wiley & Sons, Ltd.* https://doi.org/10.1002/9781118786352.wbieg0157

Zakšek, K., Oštir, K., & Kokalj, Ž. (2011). Sky-view factor as a relief visualization technique. *Remote Sensing*, *3*(2), 398–415. https://doi.org/10.3390/rs3020398

Zhang, Y., Mazza, A., & Bompard, E. (2019). Data-driven Feature Description of Heat Wave Effect on

Distribution System. *2019 IEEE Milan PowerTech*, 1–6.

Zwiers, F. (2004). *Detection and Attribution of External Influences on the Climate System: A Review of Recent Advances*. Retrieved from https://www.geos.ed.ac.uk/homes/ghegerl/0401-jan-06-idag.pdf

Appendix A Annual and Event Validation- 2010 to 2017

A.1 Instructions On How To Run PDM Model

After preparing 15-minutes precipitation data and river level/ flow in suitable format for PDM (Figure 5.10), the standard PDM model with multiple rainfall data is calibrated for each subcatchment as follows:

1. Create a folder for each subcatchment (e.g Drom2010)
2. Copy all rainfall data and river flow data associated with a subcatchment in the subcatchment's folder
3. Create the input file (*.inp) (e.g Drom2010.inp) with recommended initial values (e.g Figure 6.1)
4. Run the PDM model by clicking on c:/CEH Model Calibration and navigate to a folder (e.g Drom2010) and select Drom2020.inp
5. The dummy data interface file (Drom2010.flx) is automatically produced , copy and save it as a Drom2010.fli
6. Open Data interface file (Drom2010.fli) which contains several data blocks, we should complete them with the filename of rainfall data and river flow data (Figure A.1) by adding
   6.1. **sfile** rainfall data filename /river flow data filename (e.g BMET_80720.dat)
   6.2. **close**
7. Save and close the file
8. Run the PDM model again, this time PDM model produces a hydrograph of observed and modelled flow in the Model calibration shell. Observed flow is in black, modelled flow is in red and base flow is in blue.
9. For manual parameter adjustment enter 'W' and then the number of the parameter you wish to modify (this is a trial and error process and some knowledge about catchment can help to select suitable value, also the shape of produced hydrograph can help to adjust the parameters). The manual adjustment can be applied for parameters with 'indx' set to 1 (Figure 6.1). This can be applied for as many parameters you wish to modify.
10. Select 'X' from the calibration menu to stop the PDM and create an output file (e.g Drom2010.out) and also a copy of input file (Drom2010.inp_rev01), 'Y' will stop the model without creating the output file.
11. Examine the output file (water balance section) to find the PDM model's performance result and its accuracy (R^2)
12. If the actual evaporation (AE) is overestimated/underestimated then the value of parameter 'be' should be decreased/increased.

13. Run the model again, and select ‘O’ from the calibration menu to apply automatic optimisation, the parameters with ‘indx’ set to 1 will change automatically in each time step.
14. Repeated the step 9 to 13 several times to get the best accuracy.
15. Set the value of ‘indx’ to 0 in input file (Drom2010.inp) to avoid any change of the parameters, now the PDM model is ready for validation.
16. In input file (Drom2010.inp) modify the date of events (it could be a single event or multiple events) and run the model again.

PDM has some useful tools for creating interactive and informative hydrographs. By selecting ‘G’ from calibration menu, the graphic submenu is activated, which has several useful options for example:

- Zoom: zoom in a particular part of hydrograph to examine more details.
- L: display the lower plot contains the rainfall and Soil moisture Deficit (SMD) plots.
- I: change x-axis scale labelling

PDM also produced an output file containing the result of the model in each time step (15 minutes) which can be used for more investigation. Figure (A.2) shows a part of this output. This file was used in chapter 7 to examine the result of modified climatic condition.

```
NEXT data-series required is for site-id=<Drom_flow_10_17>
! data-type = river flows
! time-step =     15 minutes! this is event-data
! Requirements for event river flows are
! instantaneous values at date-time given
NAME not known
! above should be a descriptive site name, if known
! Data should cover at least the following period
! First time-point 24:00 31 DEC 2009
! Last  time-point 24:00 31 DEC 2010
! Minumum number of data-points in file    35041
sfile Drom_flow_10_17.dat
close
FIRST 24:00 31 DEC 2009
LAST  24:00 31 DEC 2010
DATA    <<Drom_flow_10_17>>
END     <<Drom_flow_10_17>>
! ------------------------------------------------------------------------
! NEXT data-series required is for site-id=<BMET_80711>
! data-type = rainfall
! time-step =     15 minutes! this is event-data
! Requirements for event rainfall are
! totals date-labelled by end of interval
!
NAME not known
! above should be a descriptive site name, if known
! Data should cover at least the following period
! First time-point 24:00 31 DEC 2009
! Last  time-point 24:00 31 DEC 2010
! Minumum number of data-points in file    35041
sfile BMET_80711.dat
close
FIRST 24:00 31 DEC 2009
LAST  24:00 31 DEC 2010
DATA    <<BMET_80711>>
END     <<BMET_80711>>
```

Figure A.1 Part of PDM data interface file for Dromcummer subcatchment

time(hrs)	observed	computed	error	Baseflow	SMD	Rainfall	qauxconst
0.000	141.750	141.685	0.065	45.967	0.057	0.514	0.847
0.250	144.330	141.498	2.832	45.965	0.113	0.277	0.847
0.500	146.920	141.352	5.568	45.961	0.000	0.085	0.847
0.750	148.230	141.417	6.813	45.961	0.000	0.191	0.847
1.000	150.860	141.559	9.301	45.961	0.000	0.375	0.847
1.250	152.180	141.678	10.502	45.961	0.000	0.399	0.847

Figure A.2 Part of PDM output file for Dromcummer subcatchment

A.2 Example of annual runs of PDM for each subcatchment

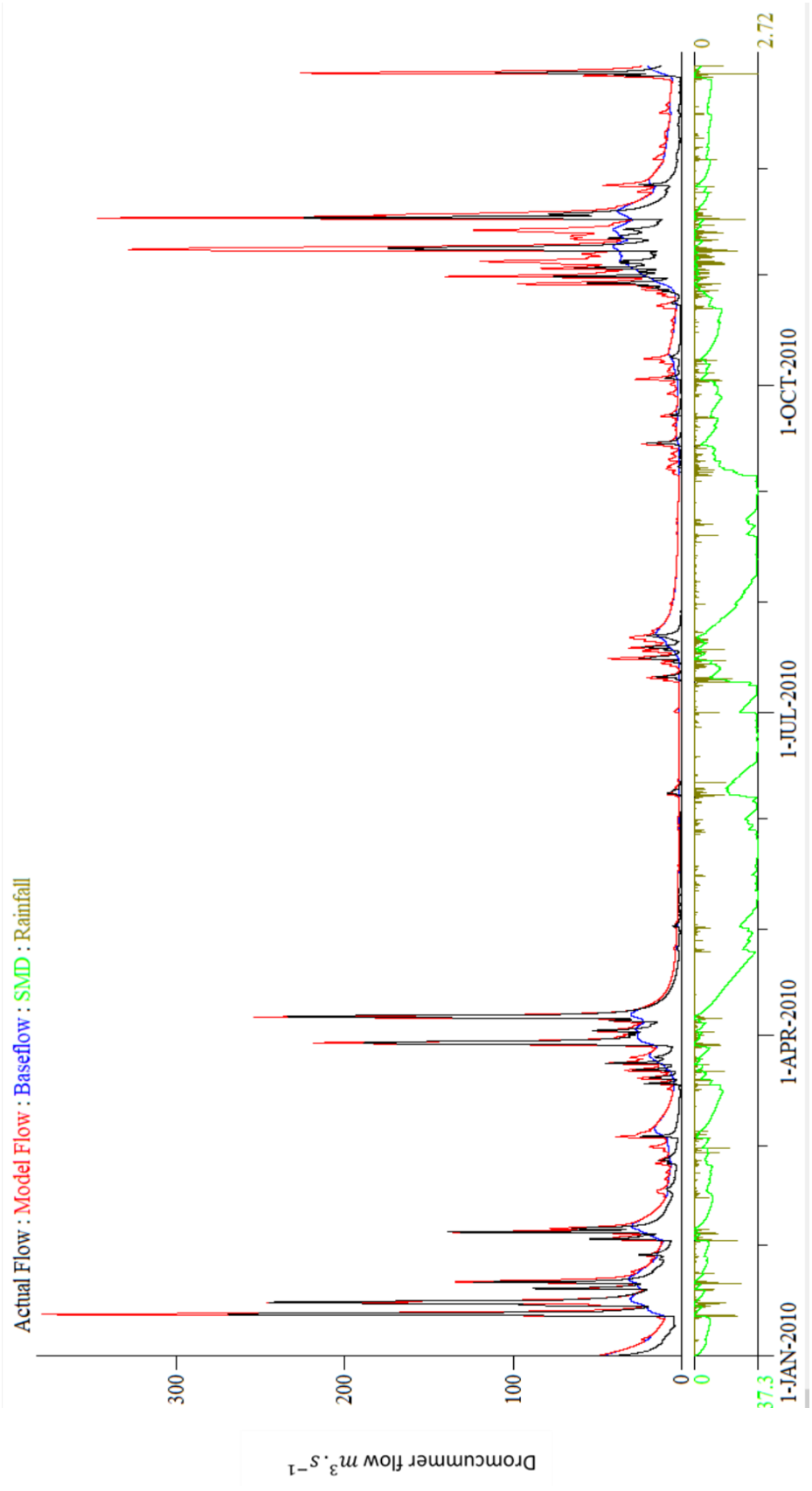


Figure A.3 (a).1 15-Minute precipitation & PDM simulation for 2010 at Dromcummer - observed versus modelled streamflow.

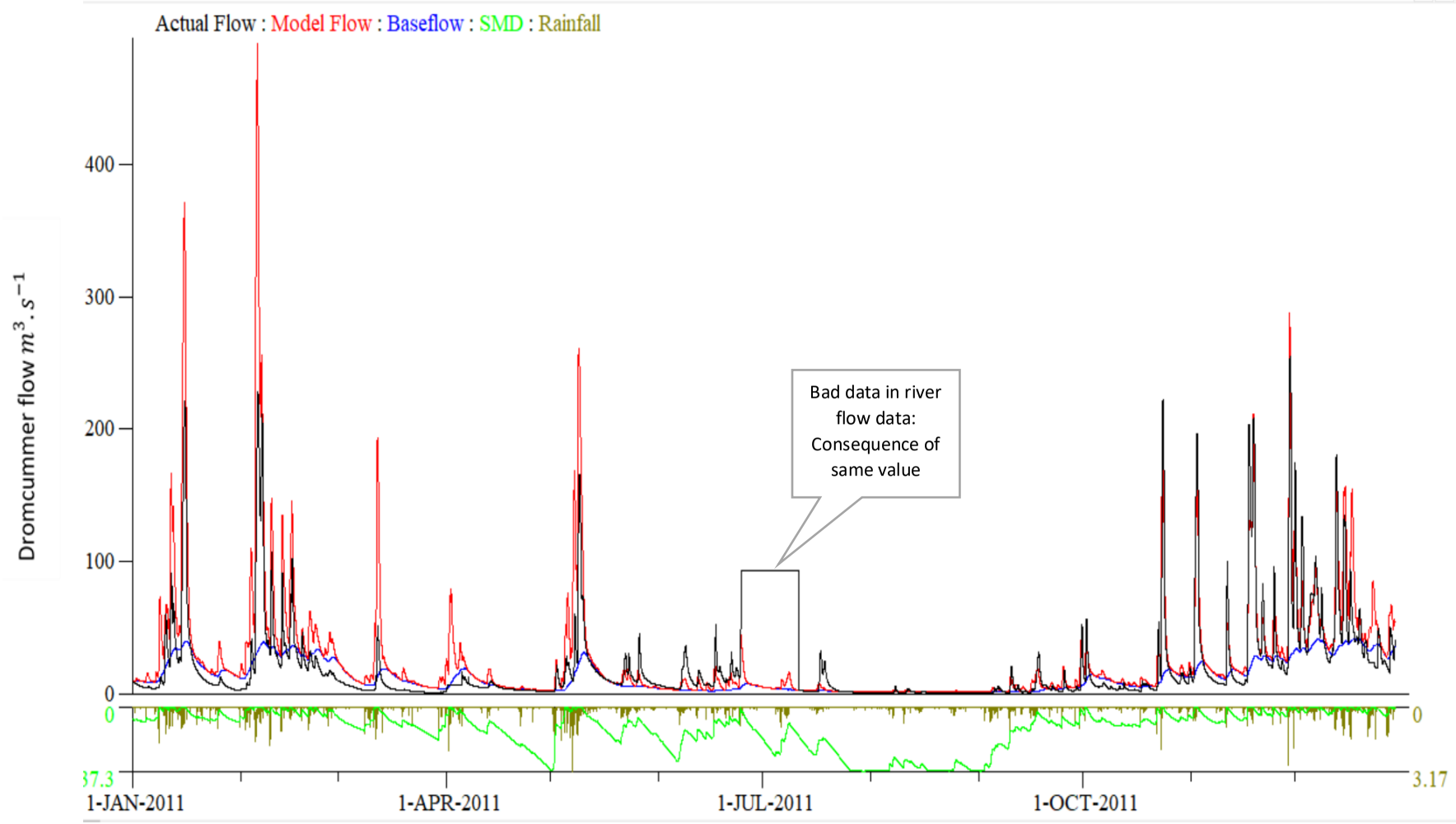


Figure A.3 (a).2 15-Minute precipitation & PDM simulation for 2011 at Dromcummer - observed versus modelled streamflow.See Figure 6.12 (b)

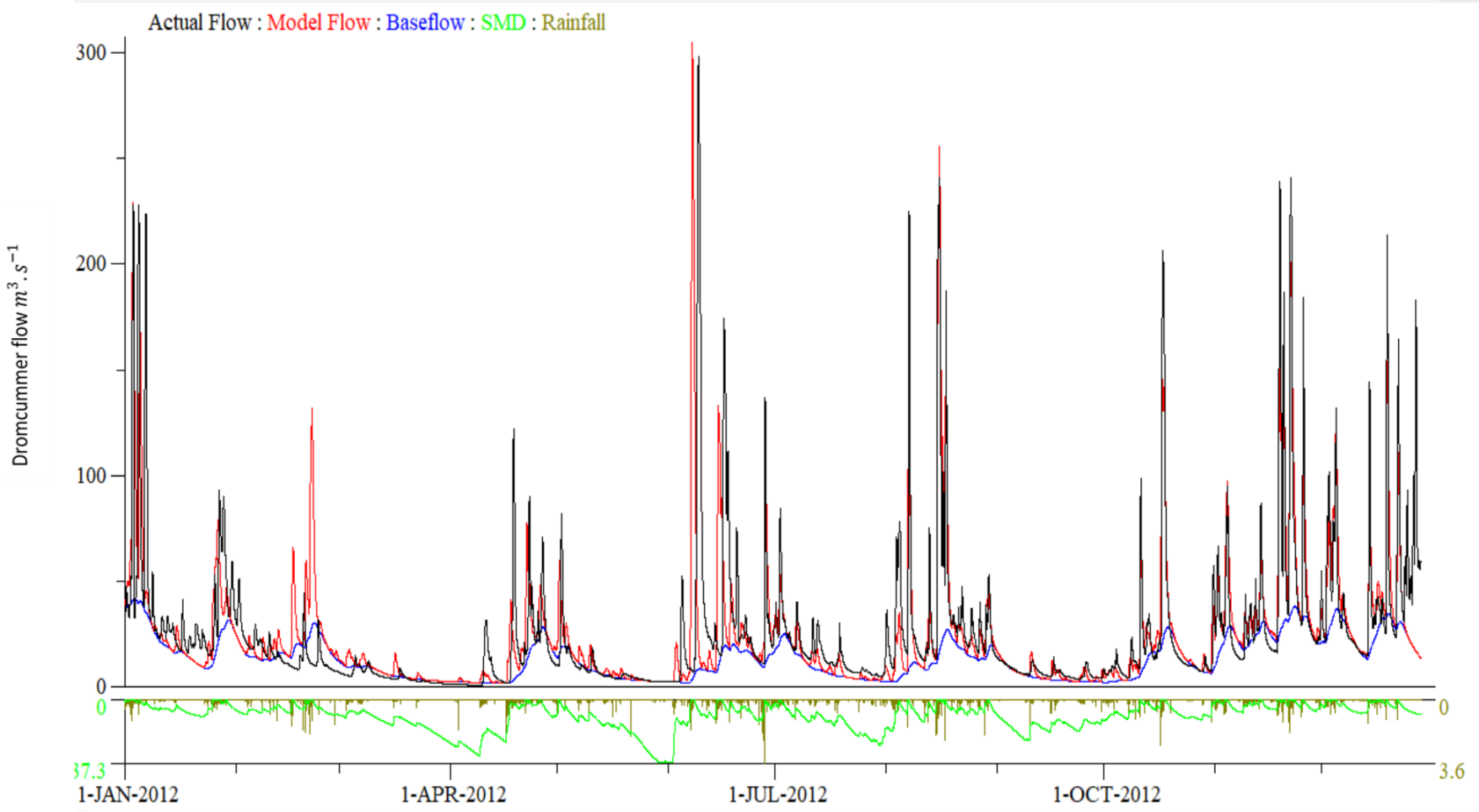


Figure A.3 (a).3 15-Minute precipitation & PDM simulation for 2012 at Dromcummer - observed versus modelled streamflow.

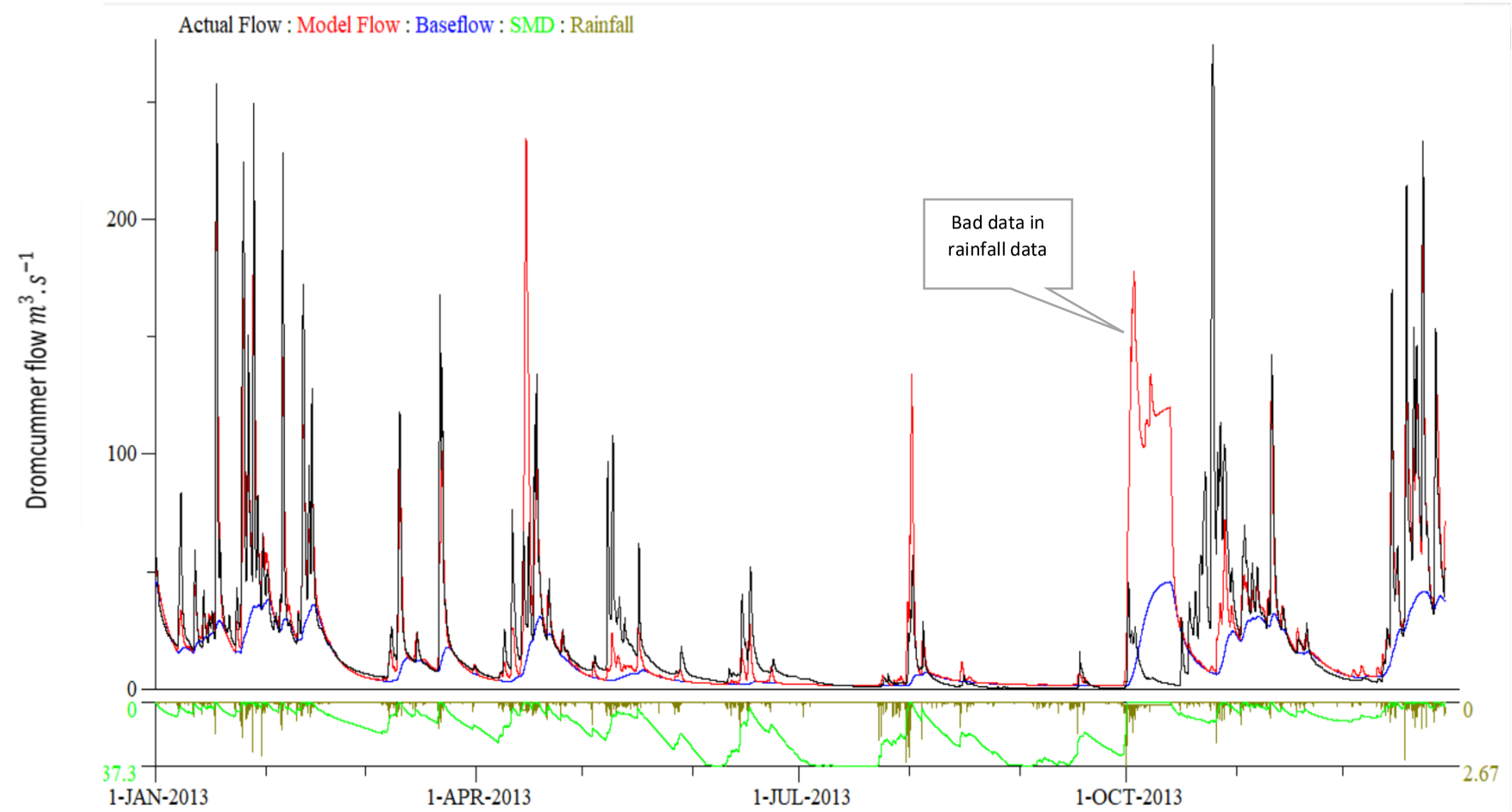


Figure A.3 (a).4 15-Minute precipitation & PDM simulation for 2013 at Dromcummer - observed versus modelled streamflow. . See Figure 4.17 example of bad data in rainfall data in year 2013

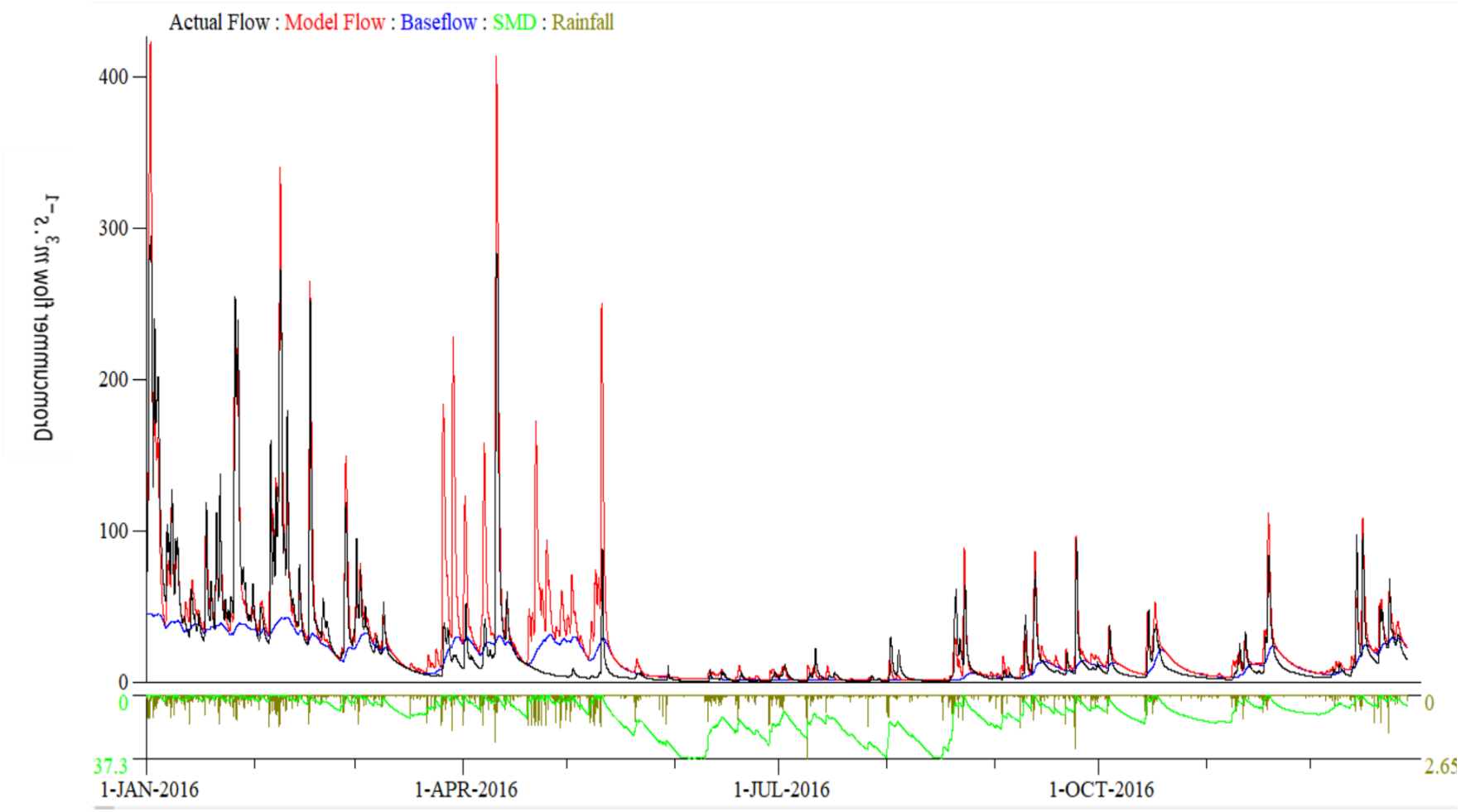


Figure A.3 (a).5 15-Minute precipitation & PDM simulation for 2016 at Dromcummer - observed versus modelled streamflow.

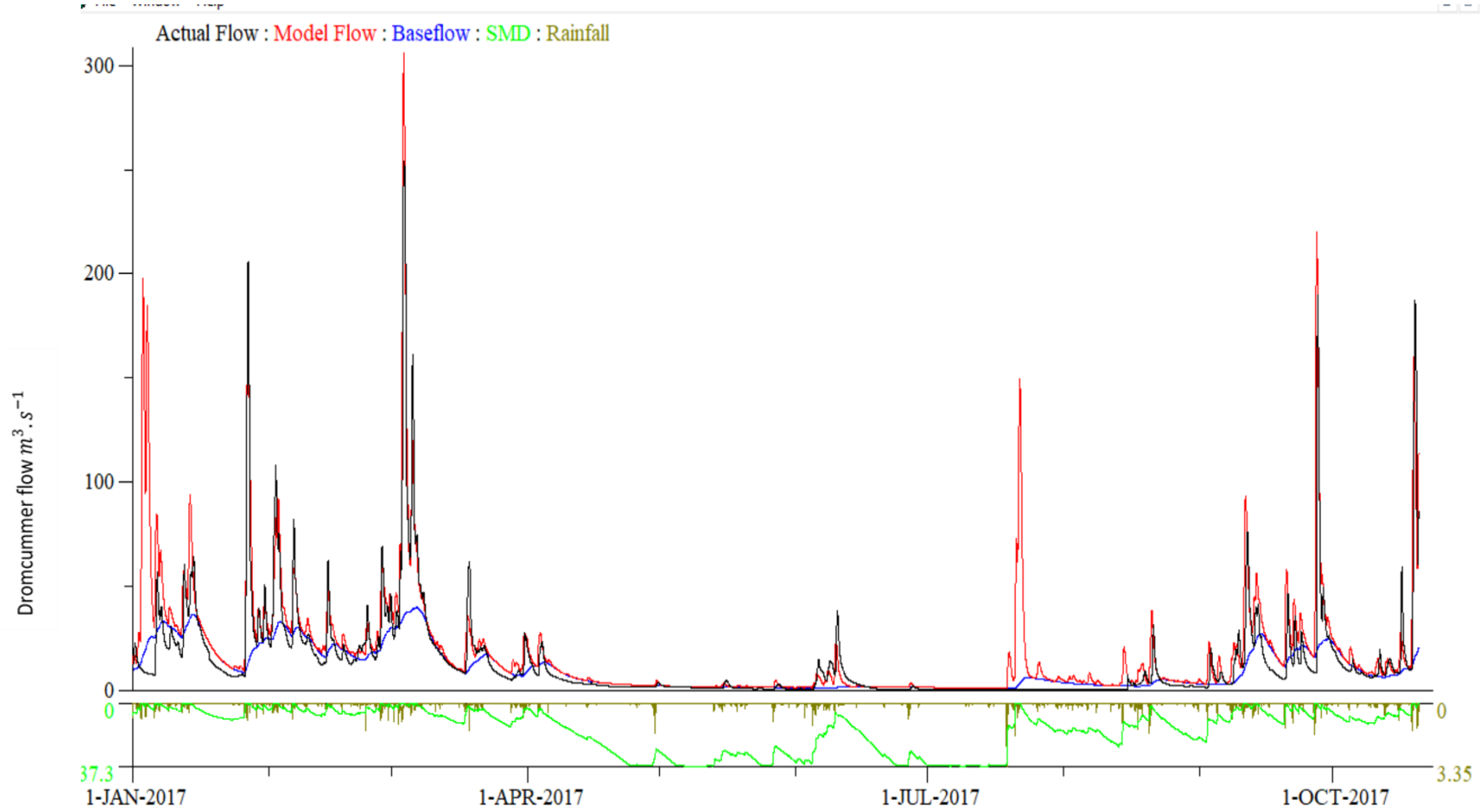


Figure A.3 (a).6 15-Minute precipitation & PDM simulation for 2017 at Dromcummer - observed versus modelled streamflow.

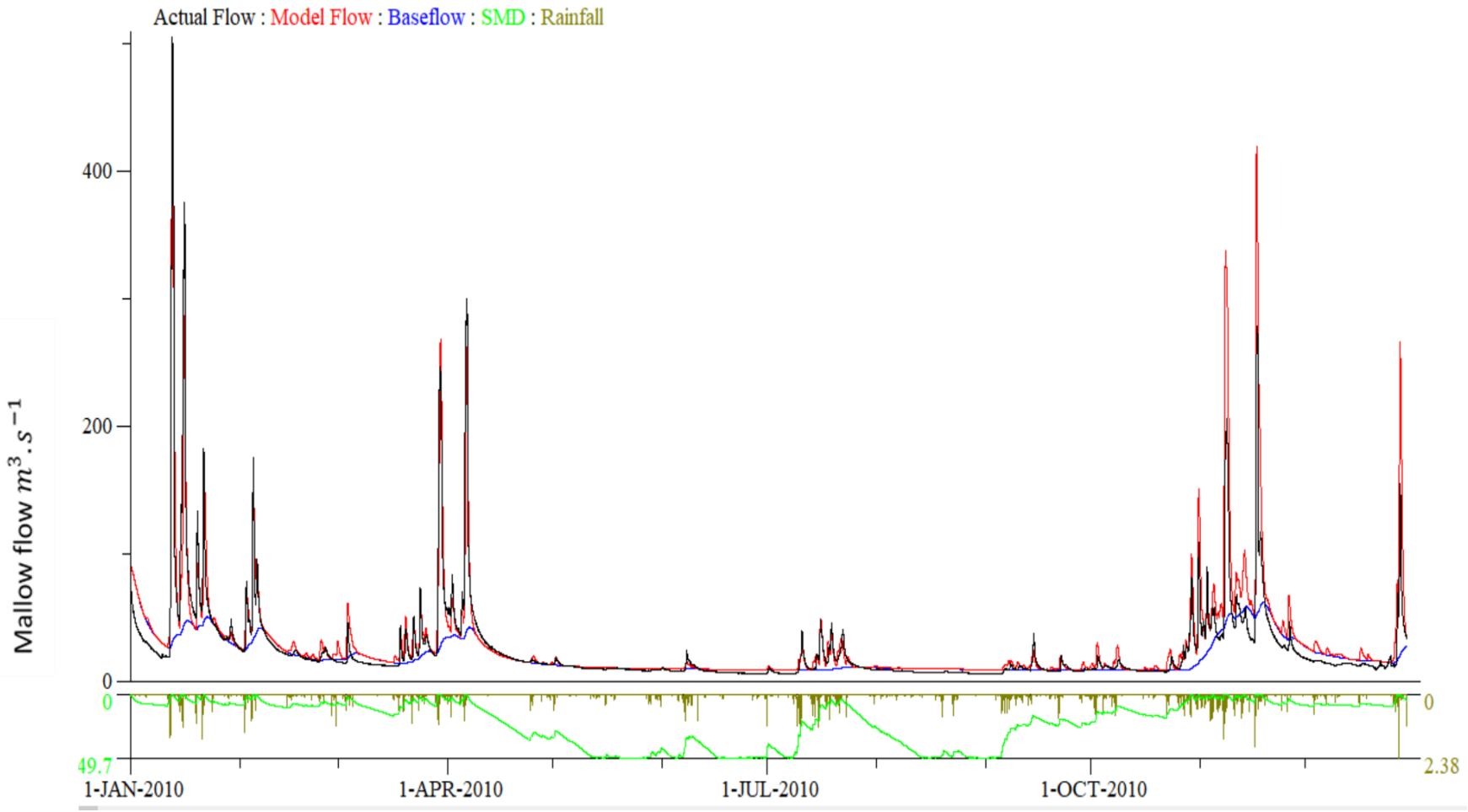


Figure A.3 (b).1 15-Minute precipitation & PDM simulation for 2010 at Mallow Rail BR - observed and modelled streamflow.

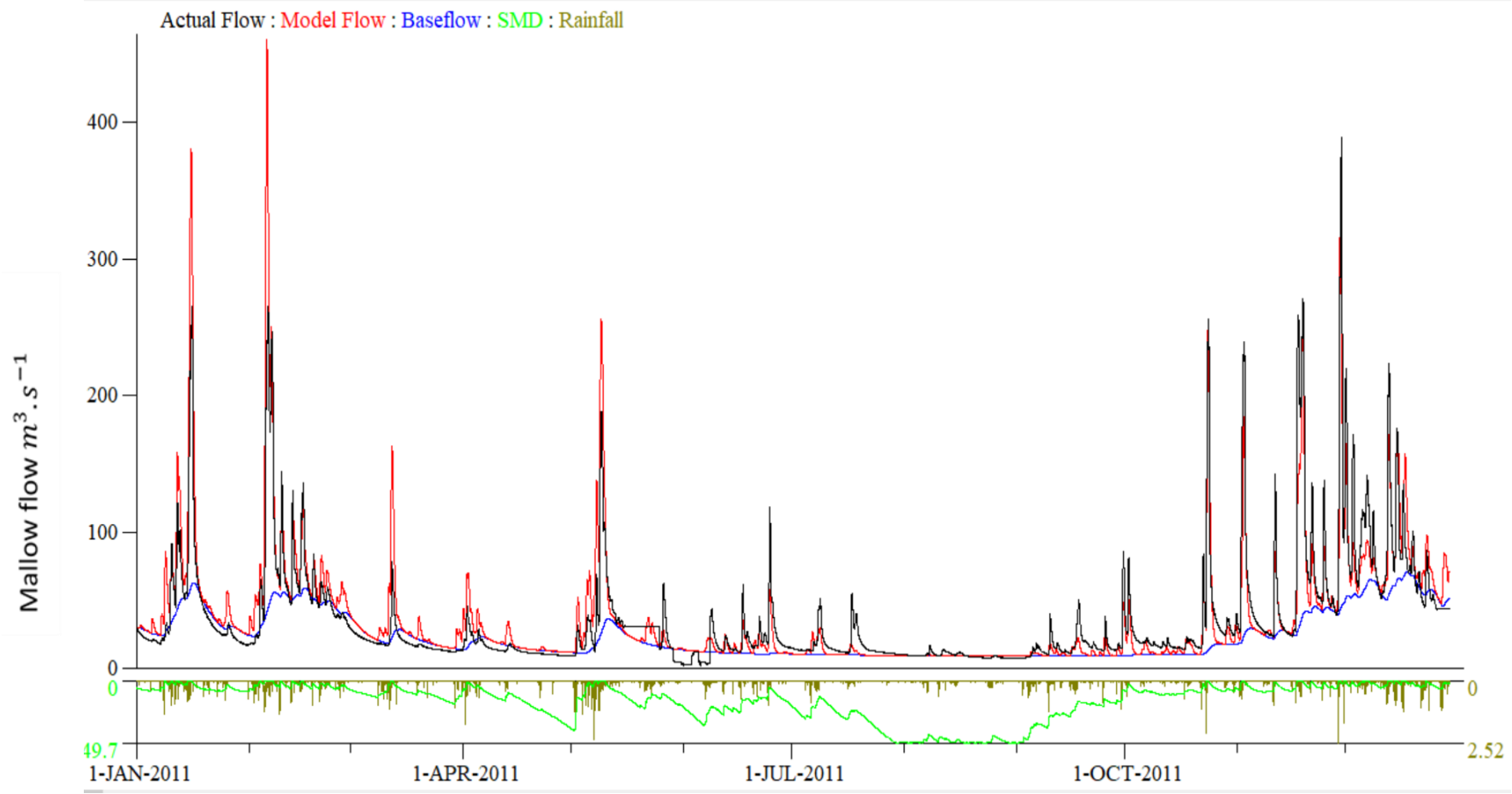


Figure A.3 (b).2 15-Minute precipitation & PDM simulation for 2011 at Mallow Rail BR - observed and modelled streamflow.

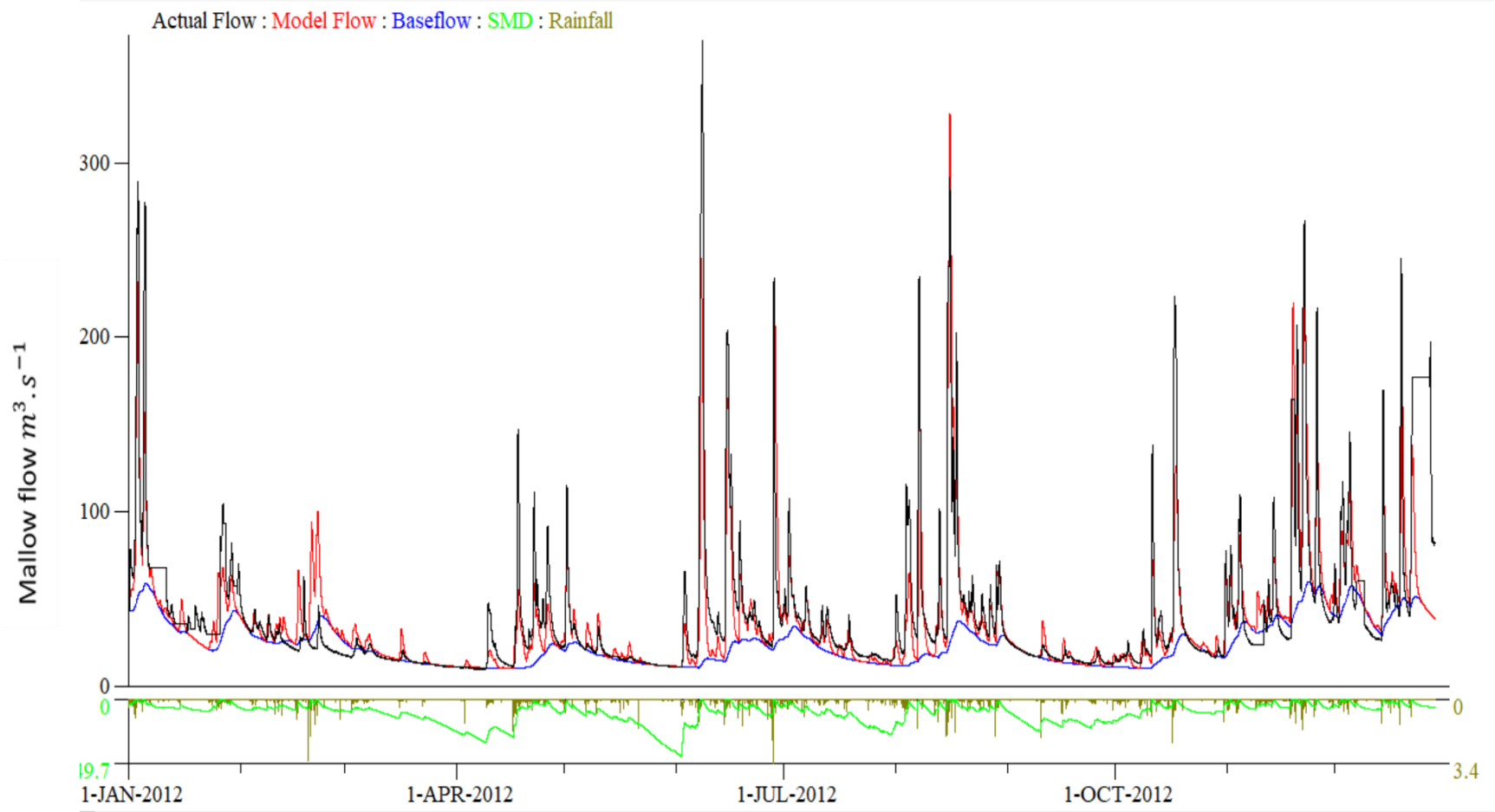


Figure A.3 (b).3 15-Minute precipitation & PDM simulation for 2012 at Mallow Rail BR - observed and modelled streamflow.

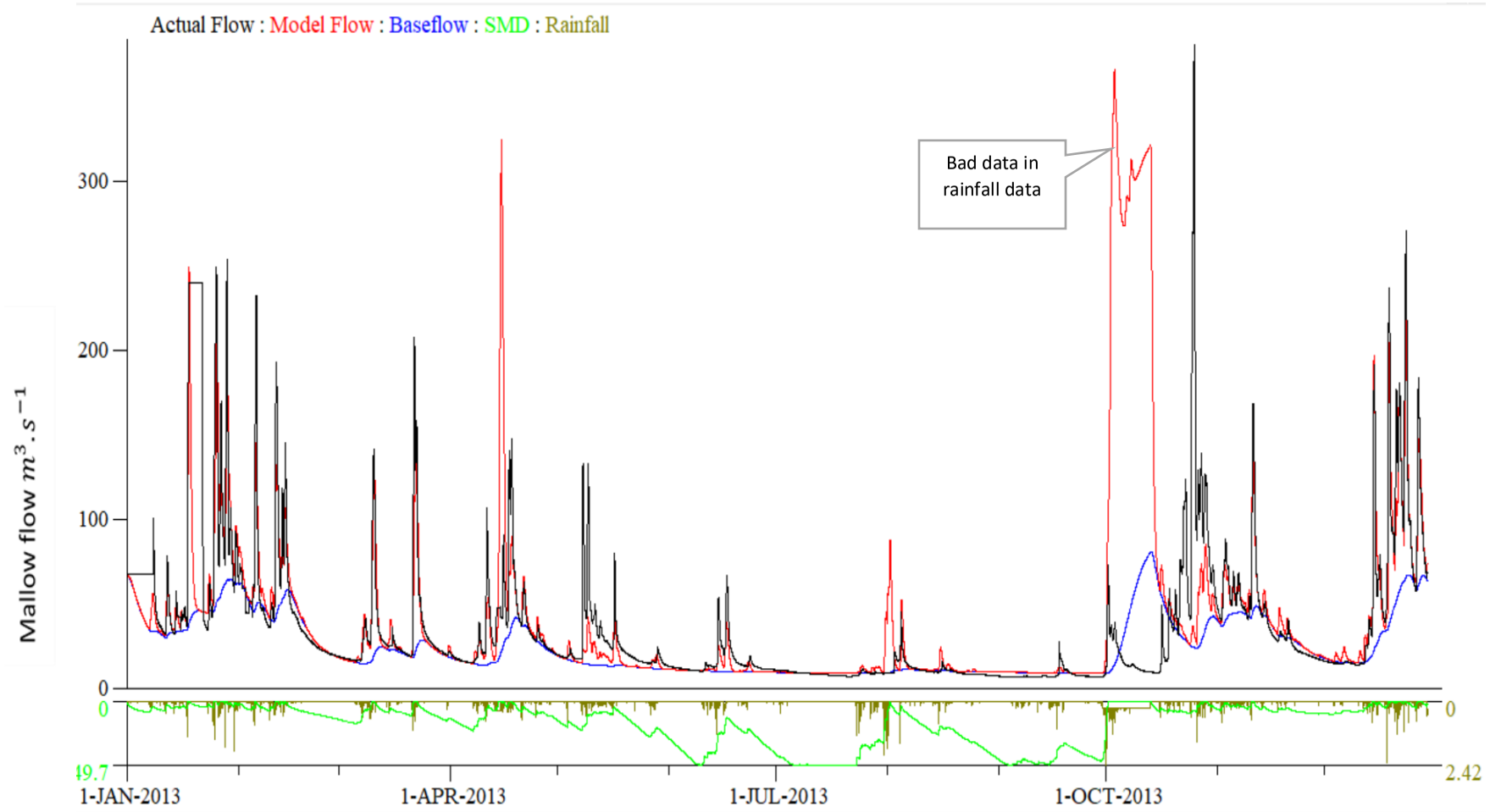


Figure A.3 (b).4 15-Minute precipitation & PDM simulation for 2013 at Mallow Rail BR - observed and modelled streamflow. See Figure 4.17 example of bad data in rainfall data in year 2013

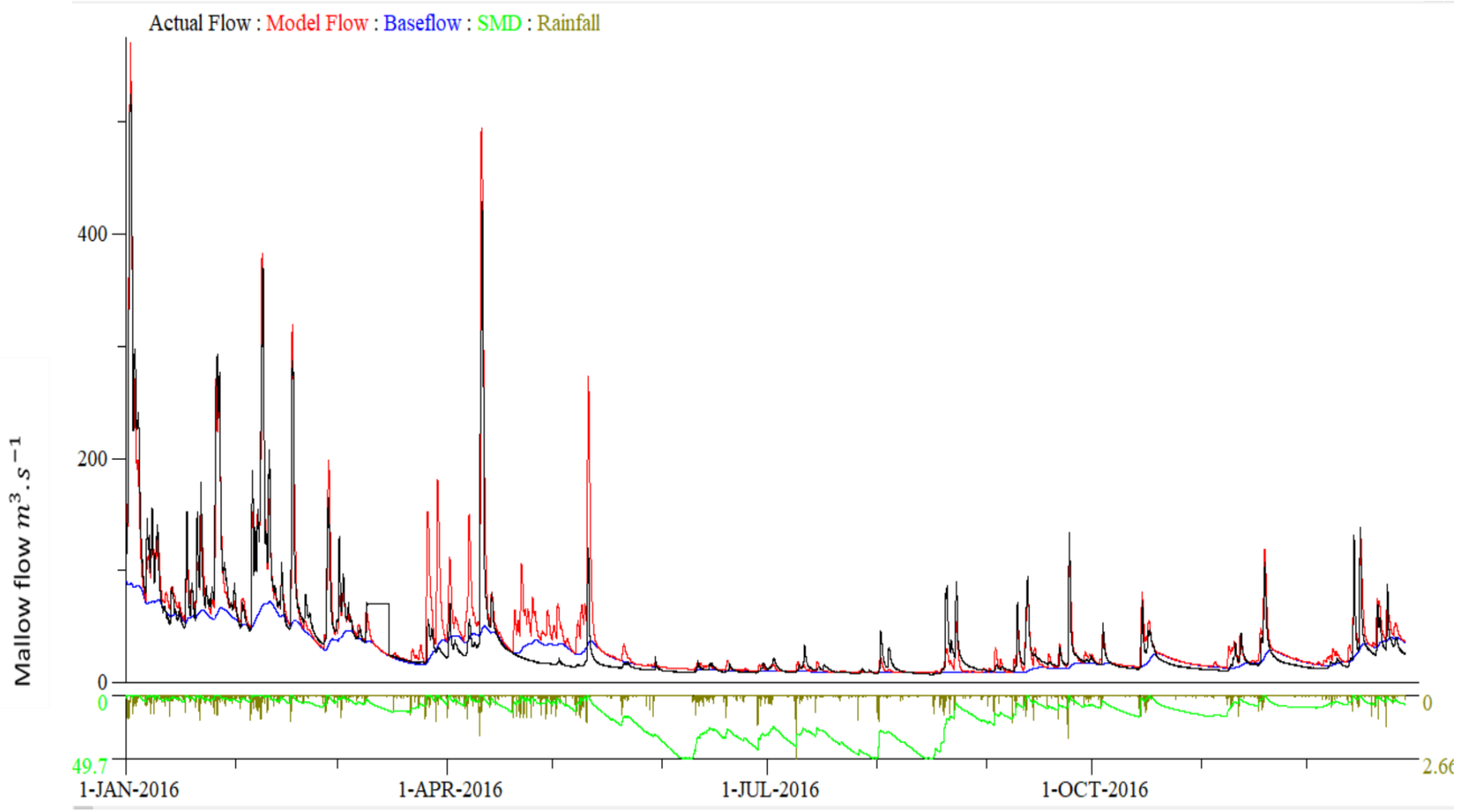


Figure A.3 (b).5 15-Minute precipitation & PDM simulation for 2016 at Mallow Rail BR - observed and modelled streamflow.

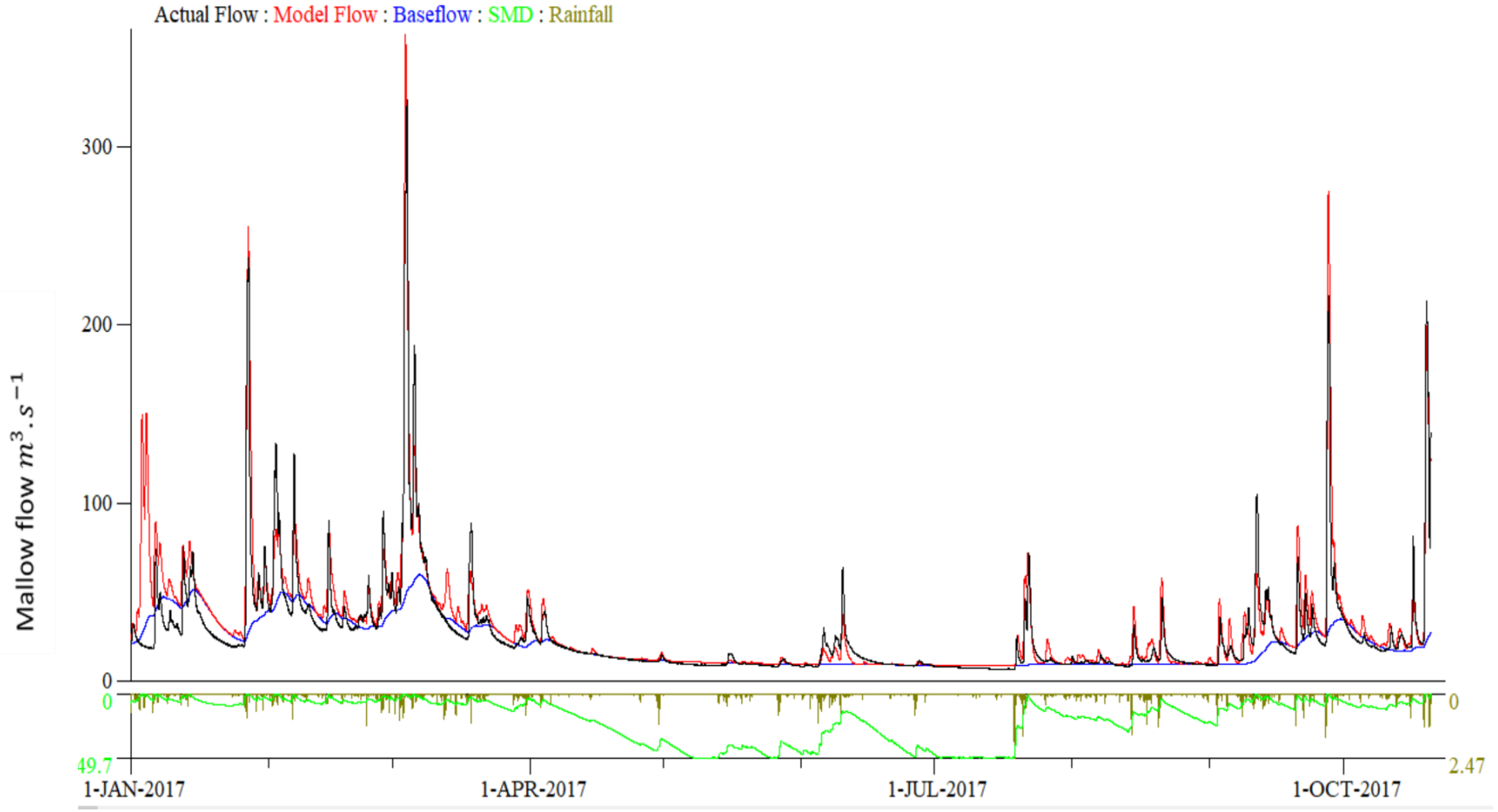


Figure A.3 (b).6 15-Minute precipitation & PDM simulation for 2017 at Mallow Rail BR - observed and modelled streamflow.

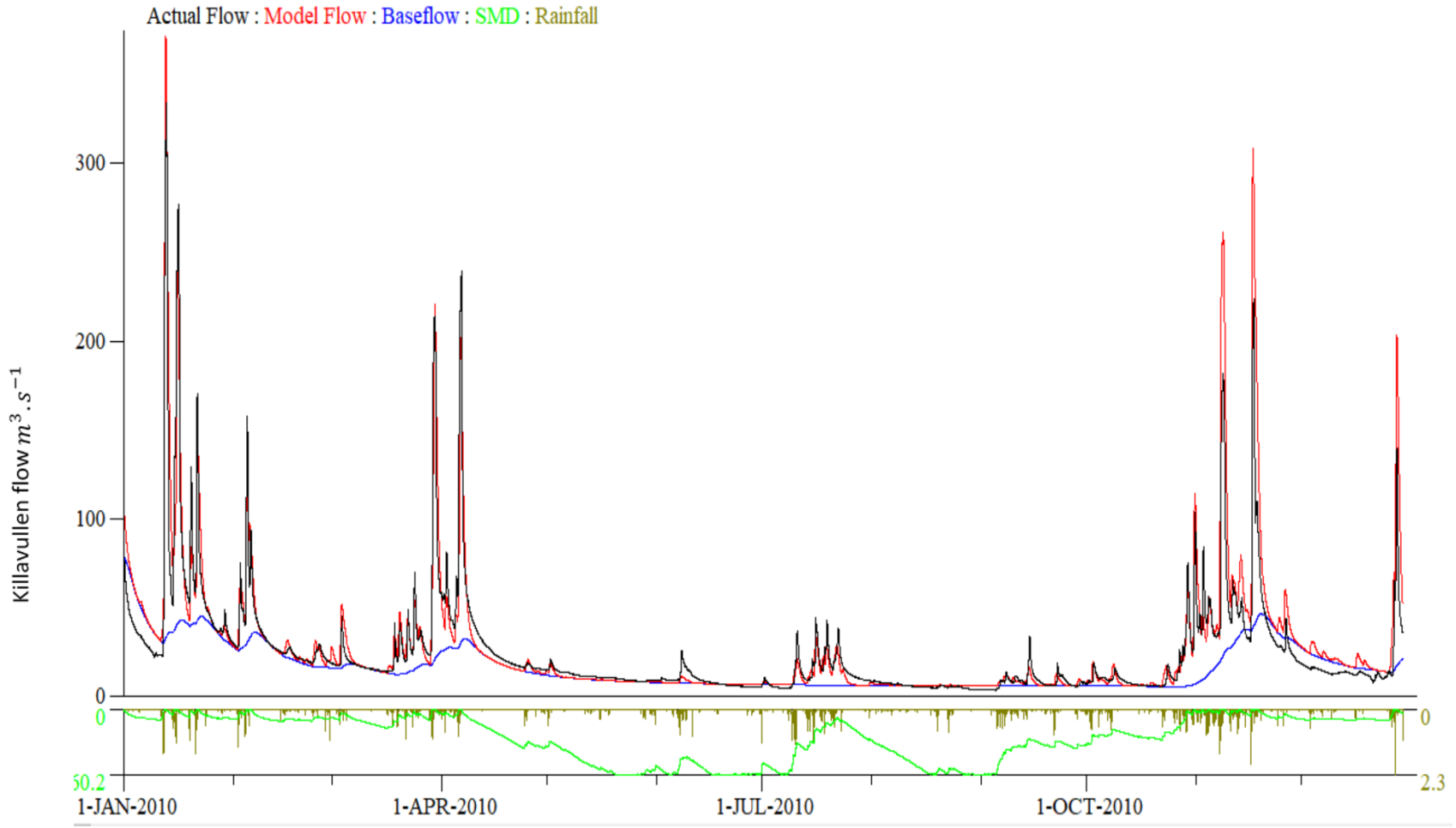


Figure A.3 (c) .1 15-Minute precipitation & PDM simulation for 2010 at Killavullen - observed and modelled streamflow.

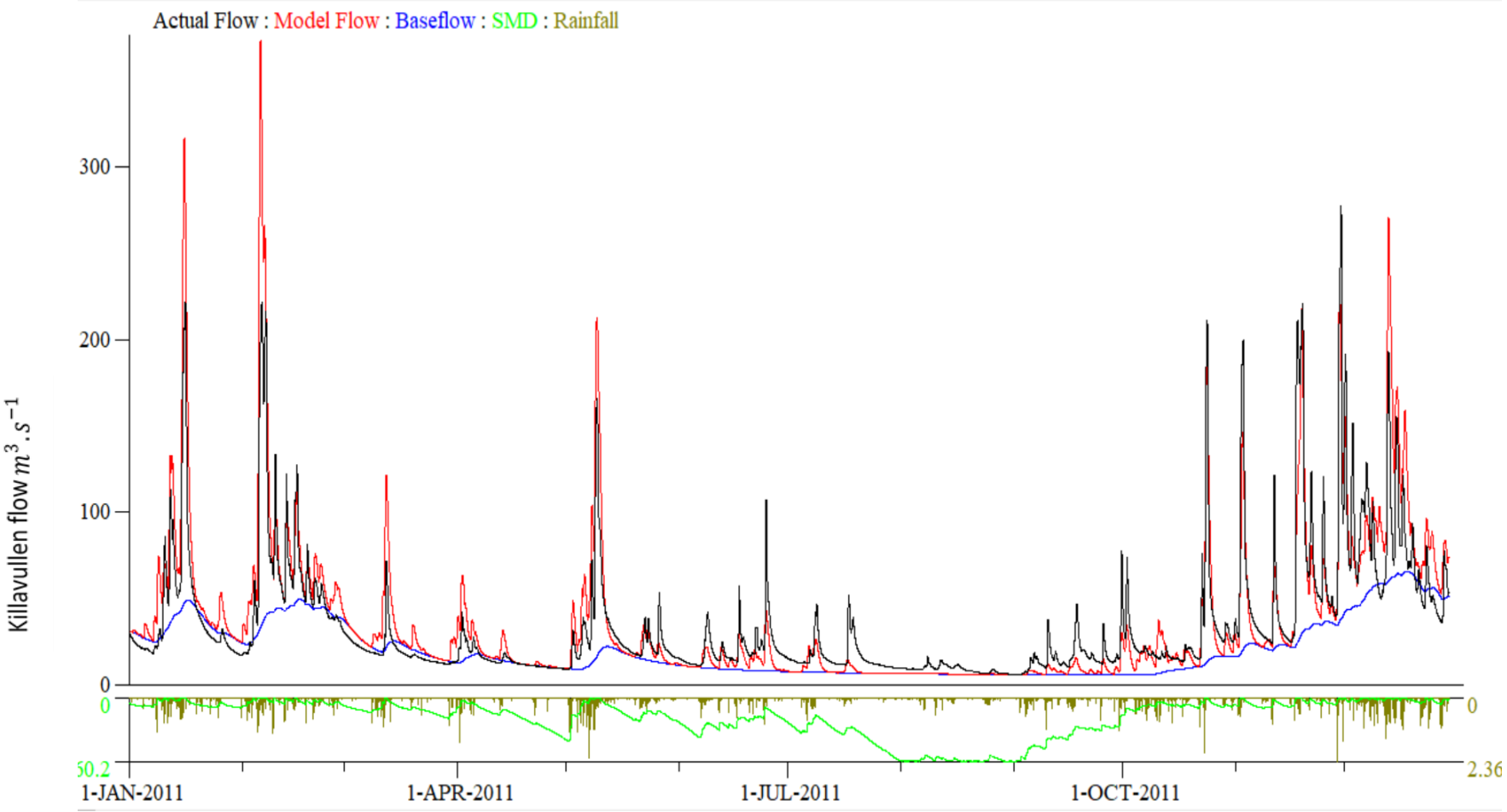


Figure A.3 (c) .2 15-Minute precipitation & PDM simulation for 2011 at Killavullen - observed and modelled streamflow.

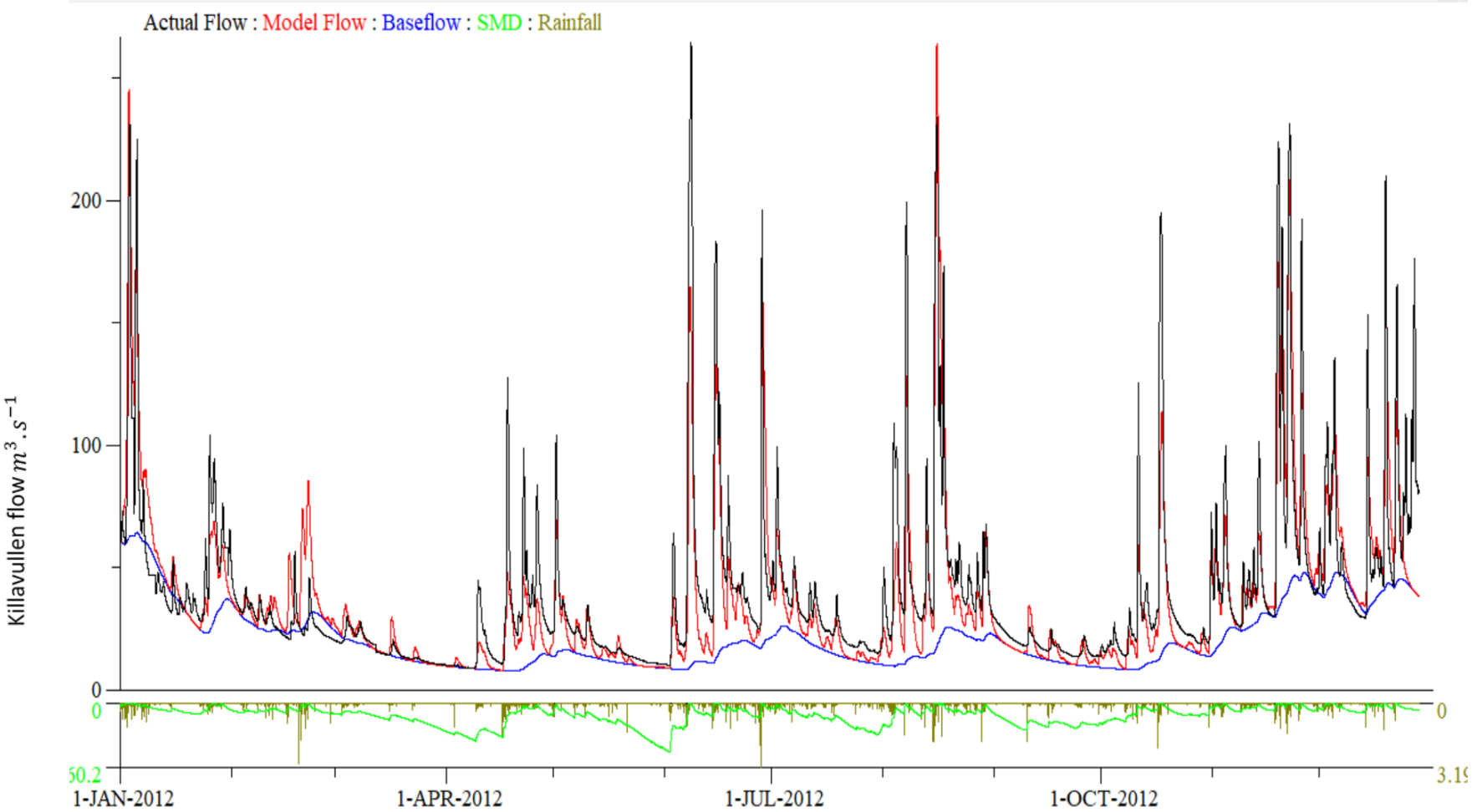


Figure A.3 (c) .3 15-Minute precipitation & PDM simulation for 2012 at Killavullen - observed and modelled streamflow.

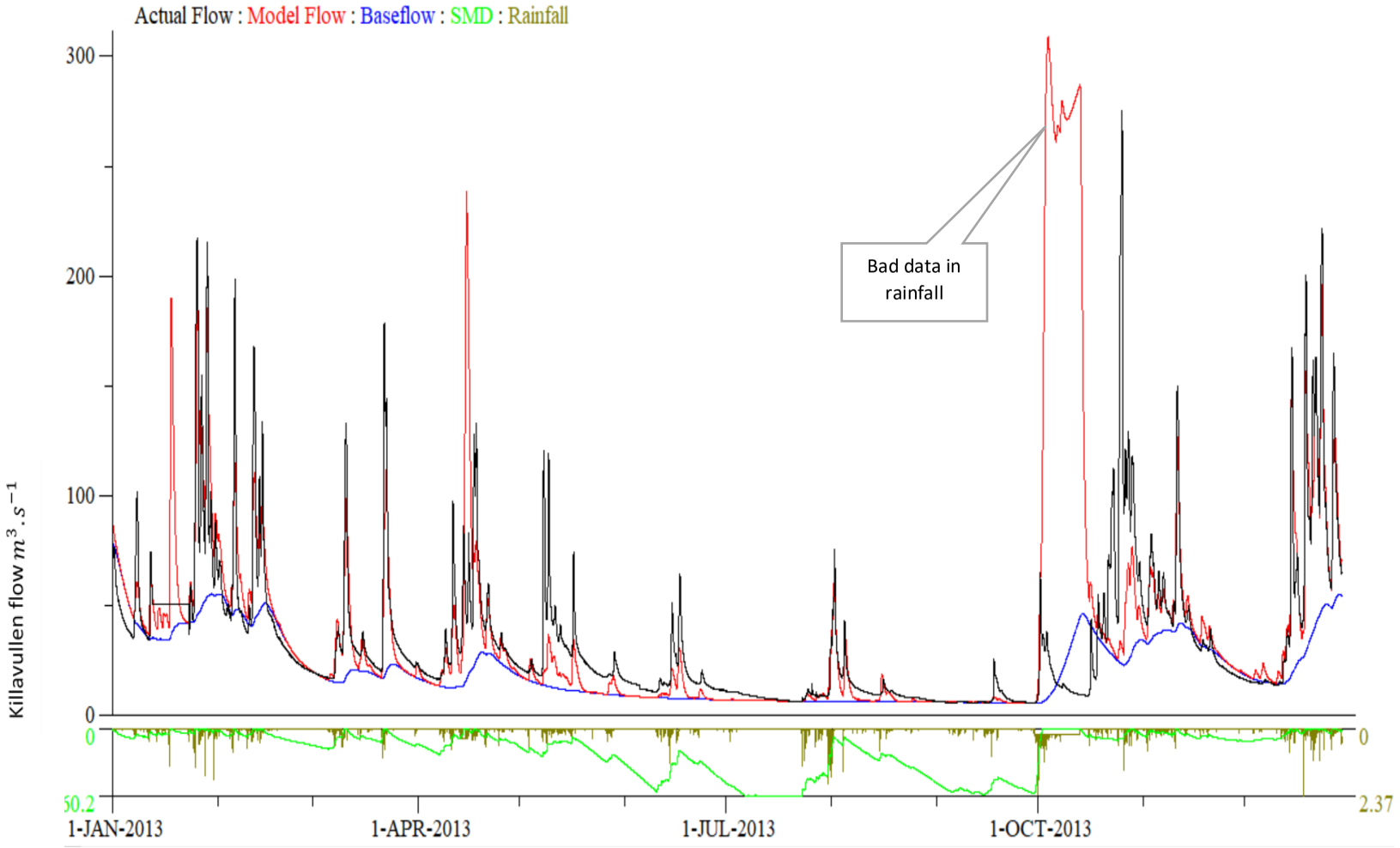


Figure A.3 (c) .4 15-Minute precipitation & PDM simulation for 2013 at Killavullen - observed and modelled streamflow. See Figure 4.17 example of bad data in rainfall data in year 2013

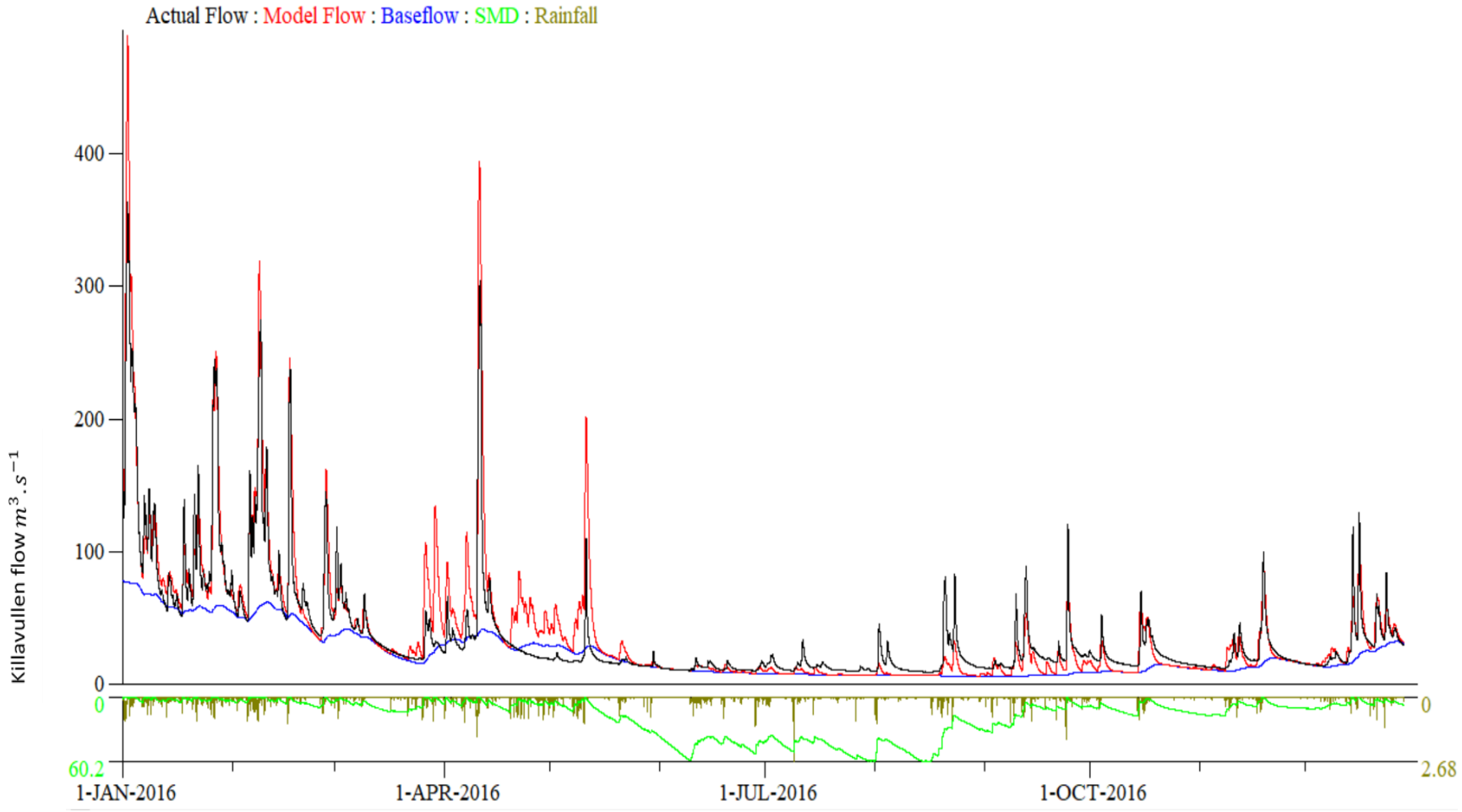


Figure A.3 (c) .5 15-Minute precipitation & PDM simulation for 2016 at Killavullen - observed and modelled streamflow.

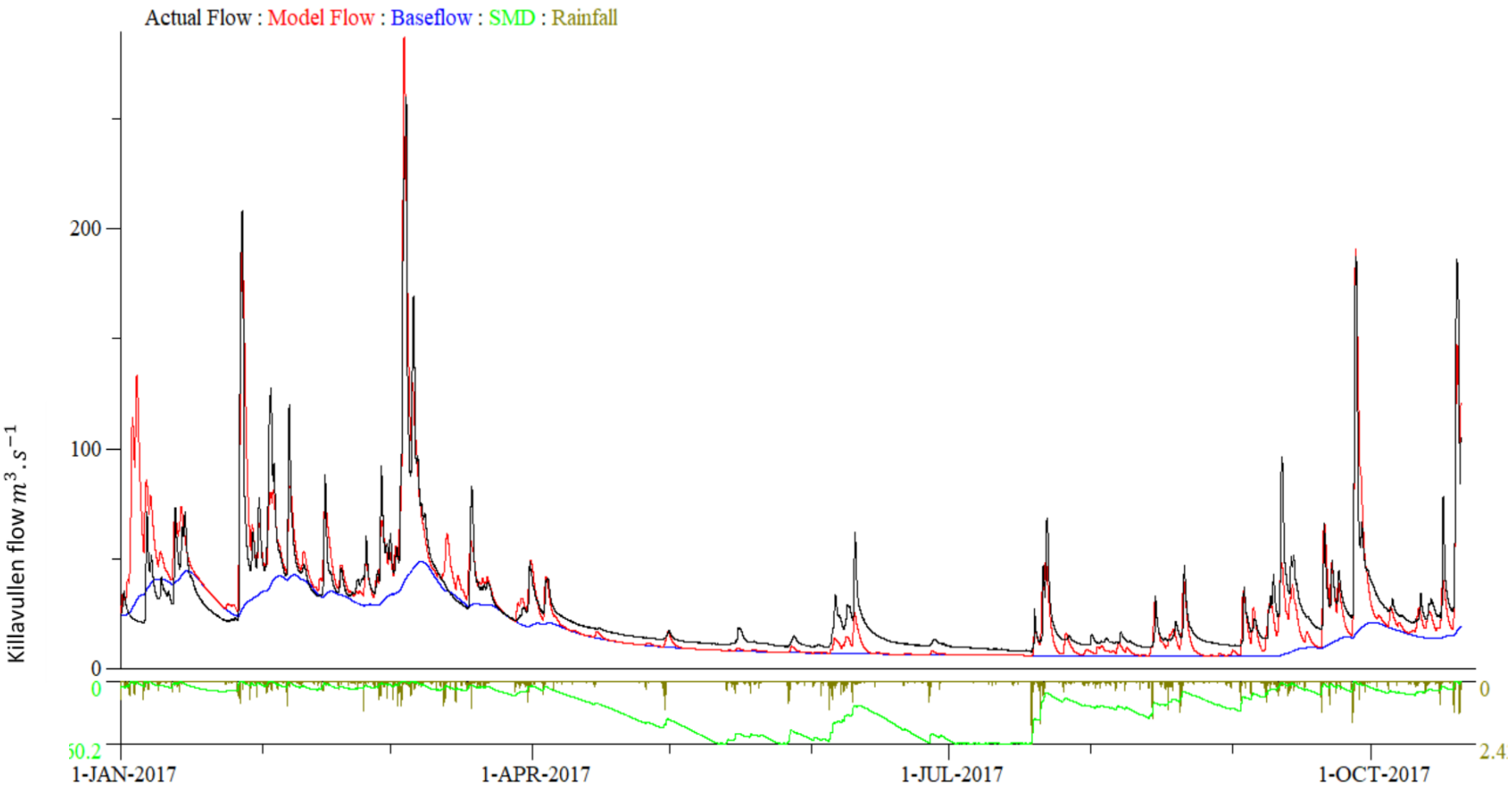


Figure A.3 (c) .6 15-Minute precipitation & PDM simulation for 2017 at Killavullen - observed and modelled streamflow.

Table A.1 (a) Summary of annual validation at Dromcummer- 2010, 2011, 2012, 2013, 2016 and 2017

No	Year	R^2	Rainfall (mm)	Computed baseflow ($m^3.s^{-1}$)	Observed flow ($m^3.s^{-1}$)	Computed flow ($m^3.s^{-1}$)	Potential evaporation (mm)	Actual evaporation (mm)	Net rainfall (mm)
1	2010	0.5264	1028.13	339.21	364.83	674.51	511	390.79	637.33
2	2011	0.3104	1429.69	456.19	812.66	976.55	511	465.62	964.08
3	2012	0.3935	1331.84	474.52	930.79	873.62	511	503.68	828.16
4	2013	0.3126	1290.21	433.83	822.85	888.66	511	427.19	863.02
5	2016	0.6600	1489.98	515.60	762.96	1042.40	511	486.81	1003.18
6	2017	0.6054	966.46	300.42	431.50	566.11	491.31	409.28	557.18

Table A.1 (b) Summary of annual validation at Mallow Rail BR- 2010, 2011, 2012, 2013, 2016 and 2017

No	Year	R^2	Rainfall (mm)	Computed baseflow ($m^3.s^{-1}$)	Observed flow ($m^3.s^{-1}$)	Computed flow ($m^3.s^{-1}$)	Potential evaporation (mm)	Actual evaporation (mm)	Net rainfall (mm)
1	2010	0.8132	963.67	508.28	670.55	793.12	511	407.40	556.27
2	2011	0.7916	1270.48	616.58	947.23	1001.34	511	470.15	800.33
3	2012	0.7123	1293.10	645.95	1108.33	1002.61	511	511	782.10
4	2013	-0.5814	1443.21	656.47	1016.18	1200.82	511	449.97	993.24
5	2016	0.8506	1366.93	693.75	981.26	1120.47	511	491.06	875.87
6	2017	0.8148	938.49	437.54	587.85	666.02	491.20	429.38	509.11

Table A.1 (c) Summary of annual validation Killavulen- 2010, 2011, 2012, 2013, 2016 and 2017

No	Year	R^2	Rainfall (mm)	Computed baseflow ($m^3.s^{-1}$)	Observed flow ($m^3.s^{-1}$)	Computed flow ($m^3.s^{-1}$)	Potential evaporation (mm)	Actual evaporation (mm)	Net rainfall (mm)
1	2010	0.8239	955.02	392.73	619.28	684.17	511	423.14	531.88
2	2011	0.7354	1300.38	503.64	865.08	929.35	511	475.33	825.05
3	2012	0.7657	1289.07	533.13	1008.52	916.34	511	511	778.07
4	2013	-0.8656	1390.56	520.39	880.91	1045.16	511	464.13	926.44
5	2016	0.8268	1338.26	557.48	918.90	997.69	511	498.45	839.82
6	2017	0.8173	924.65	339.47	596.30	572.11	491.21	438.91	485.75

A.2 PDM Event Hydrograph examples

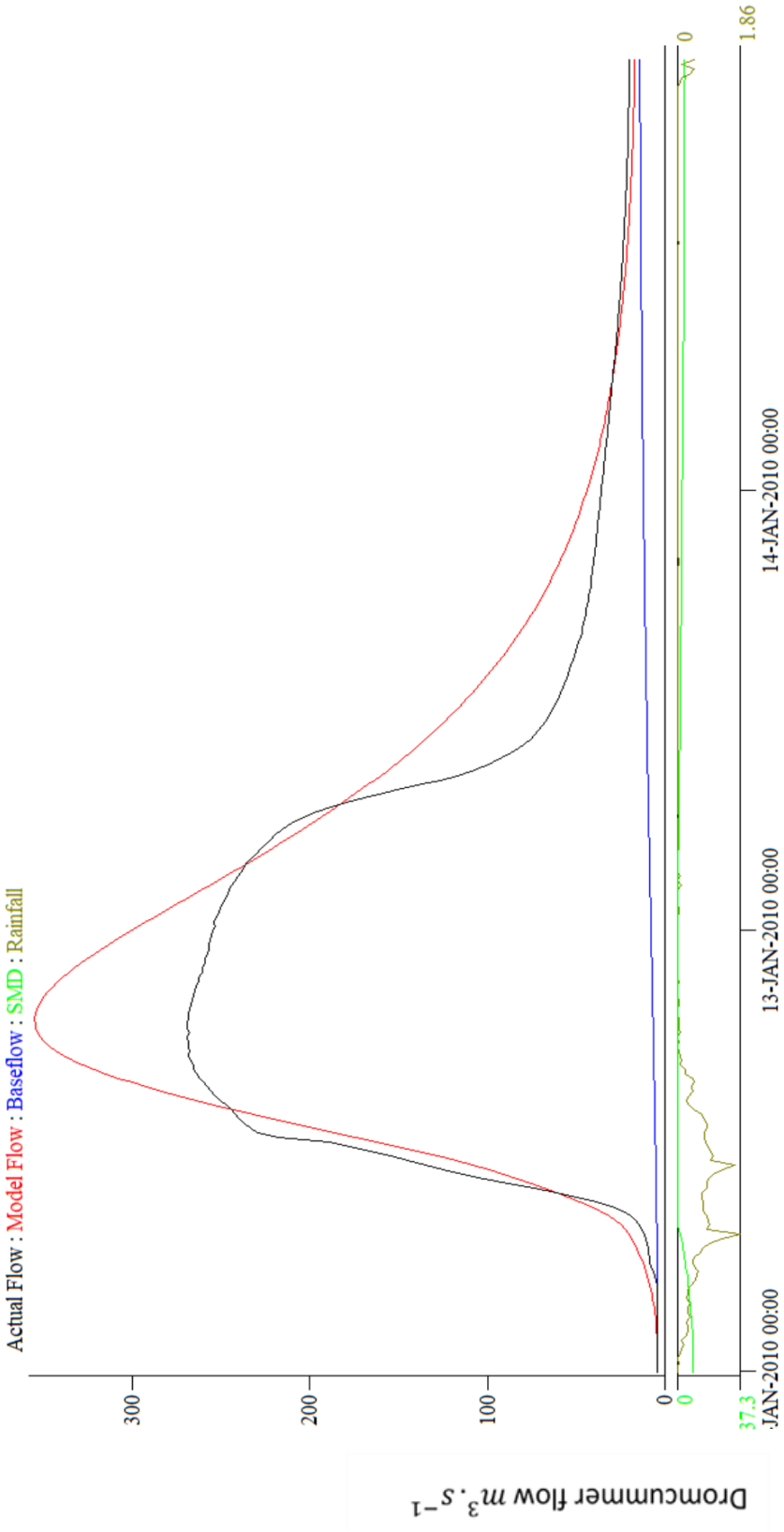


Figure A.4 (a).1 PDM simulation for validation, event in 2010 in Dromcummer.

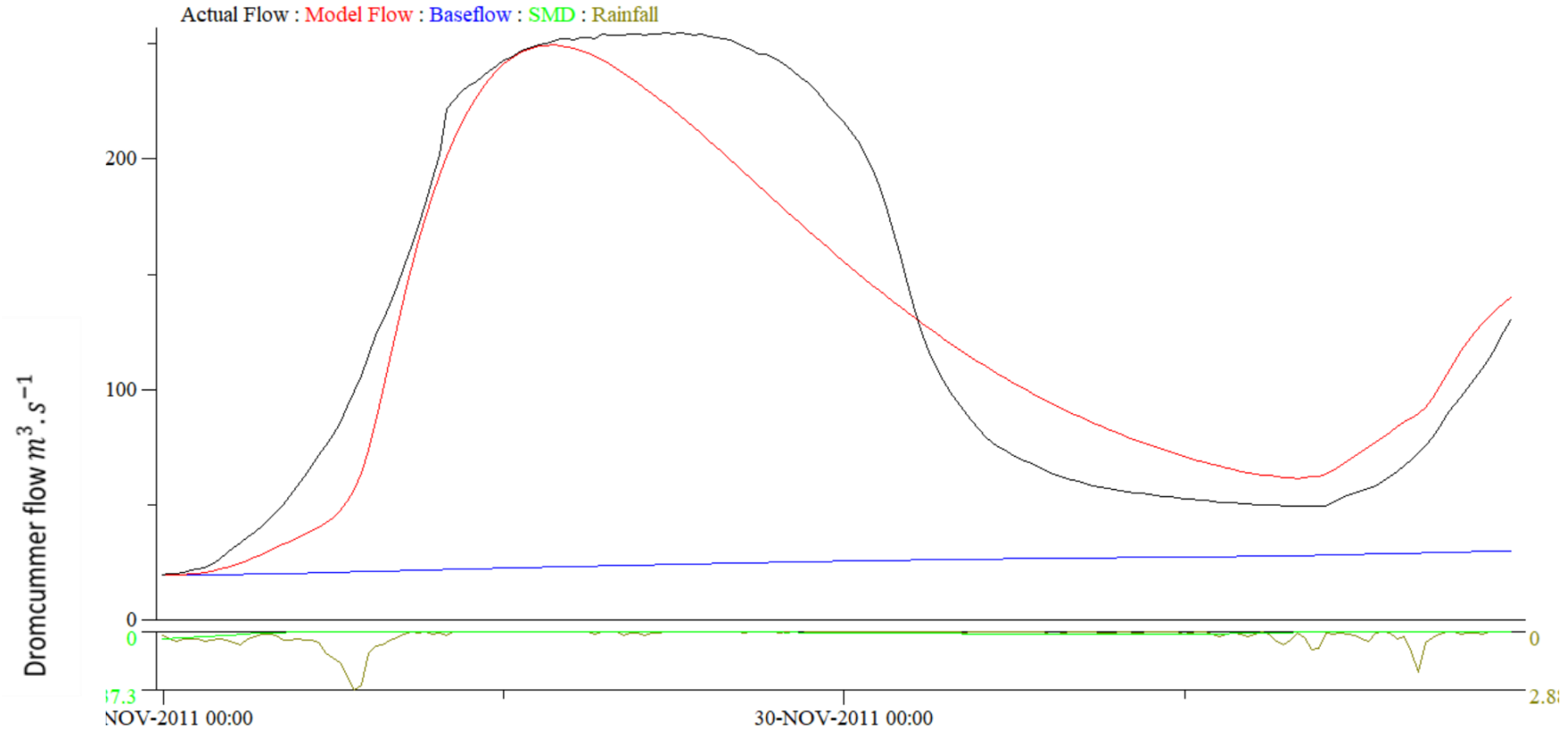


Figure A.4 (a).2 PDM simulation for validation, event in 2011 in Dromcummer.

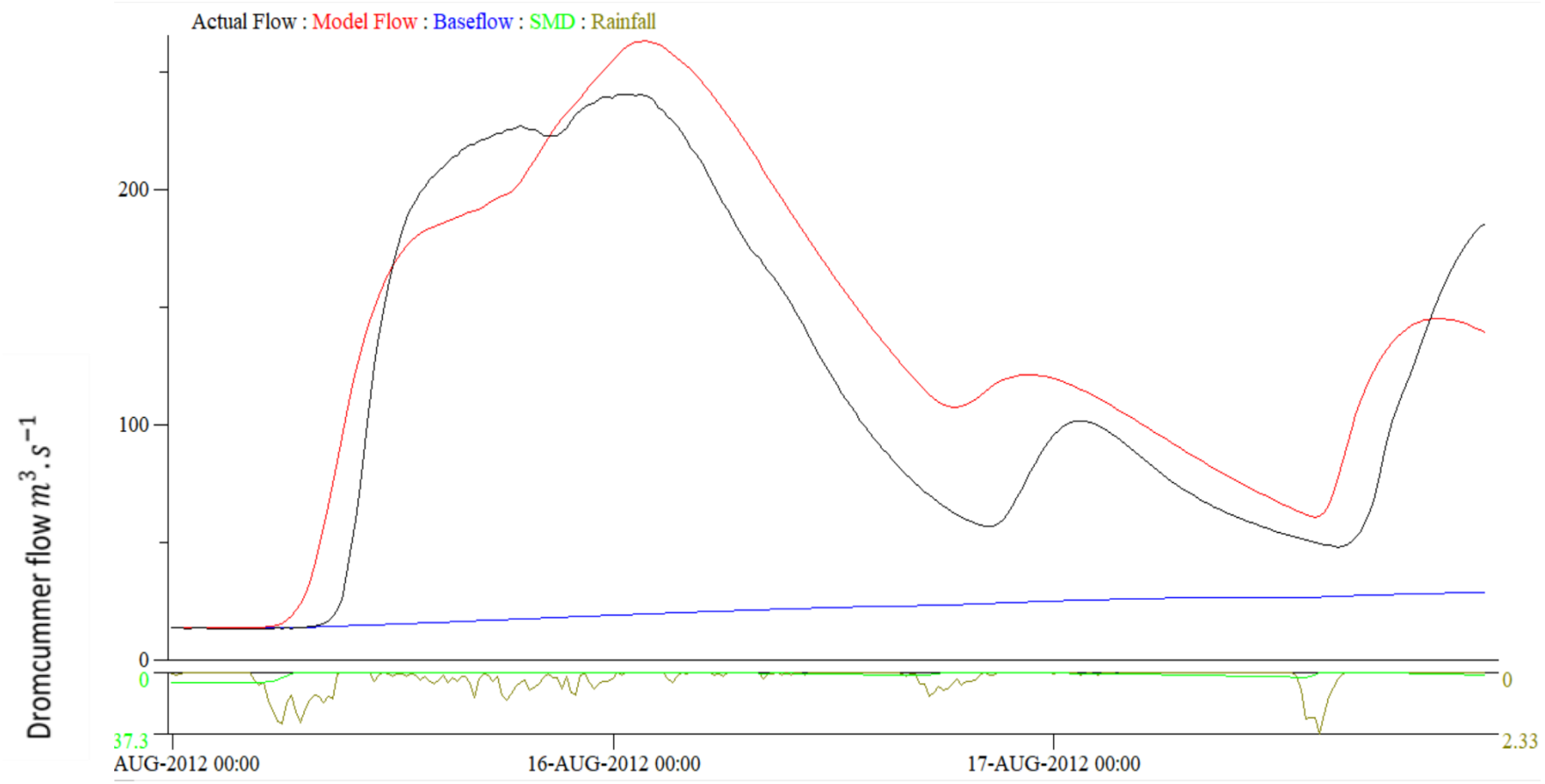


Figure A.4 (a).3 PDM simulation for validation, event in 2012 in Dromcummer.

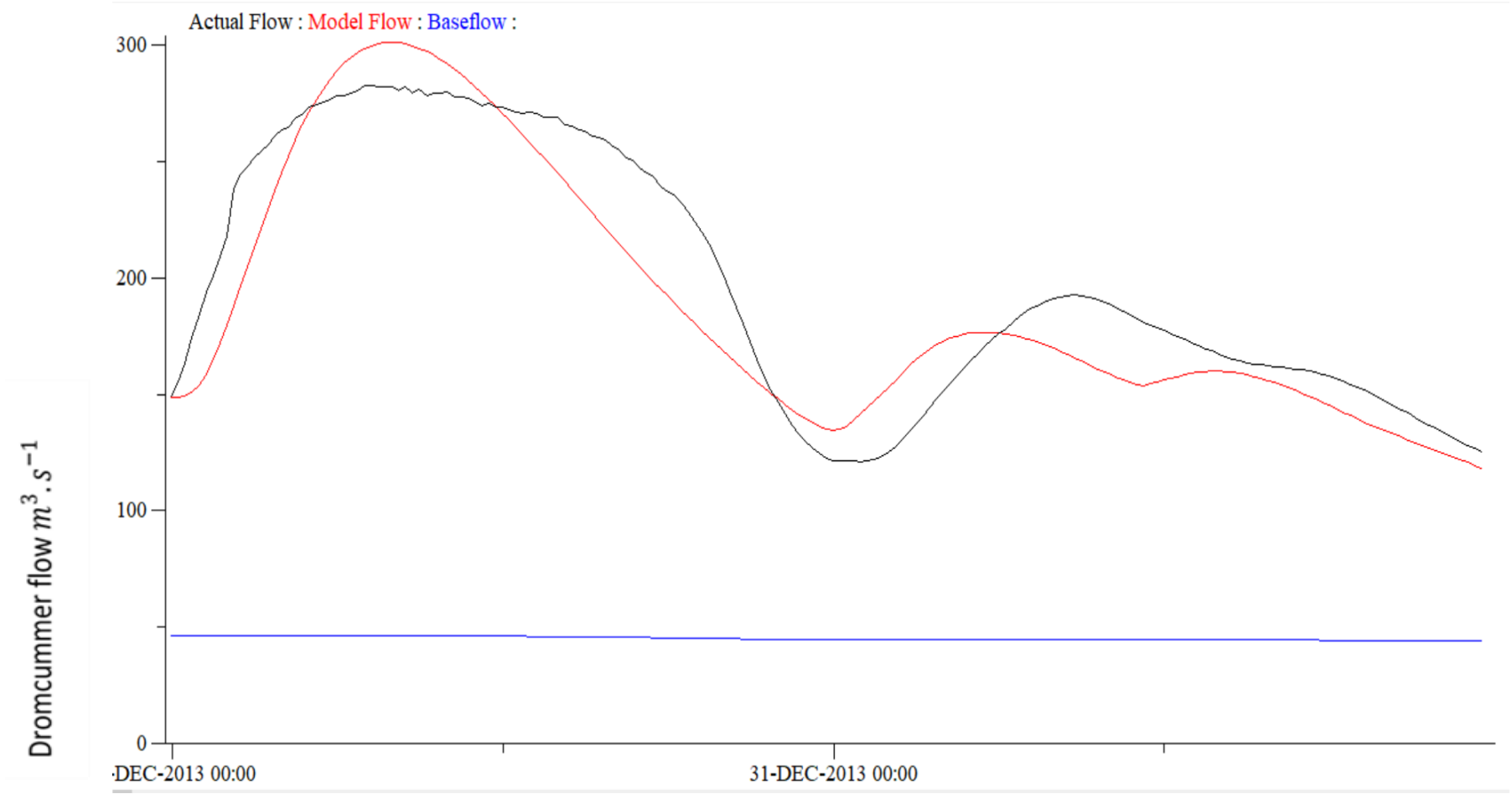


Figure A.4 (a).4 PDM simulation for validation, event in 2013 in Dromcummer.

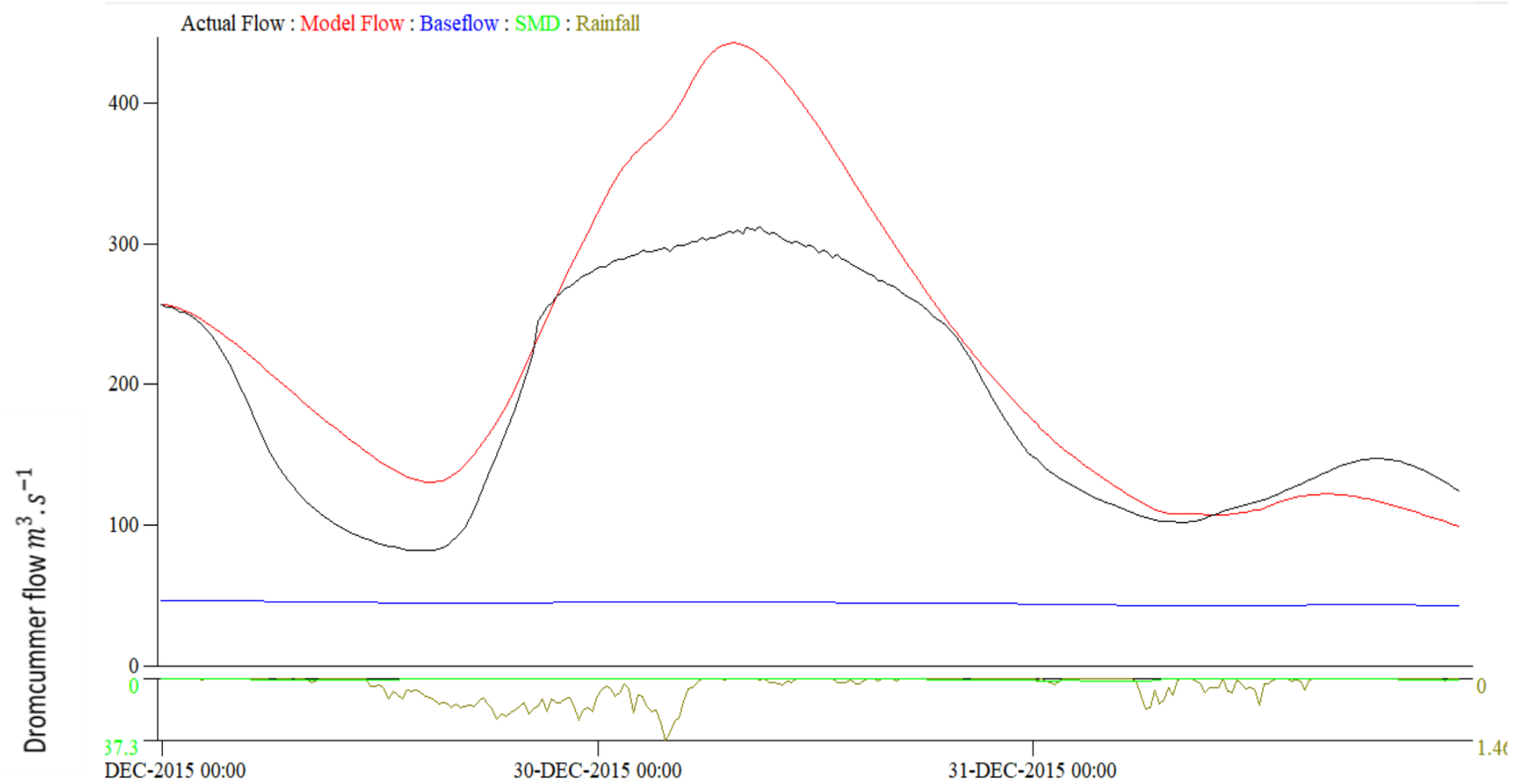


Figure A.4 (a).5 PDM simulation for validation, event in 2015 in Dromcummer.

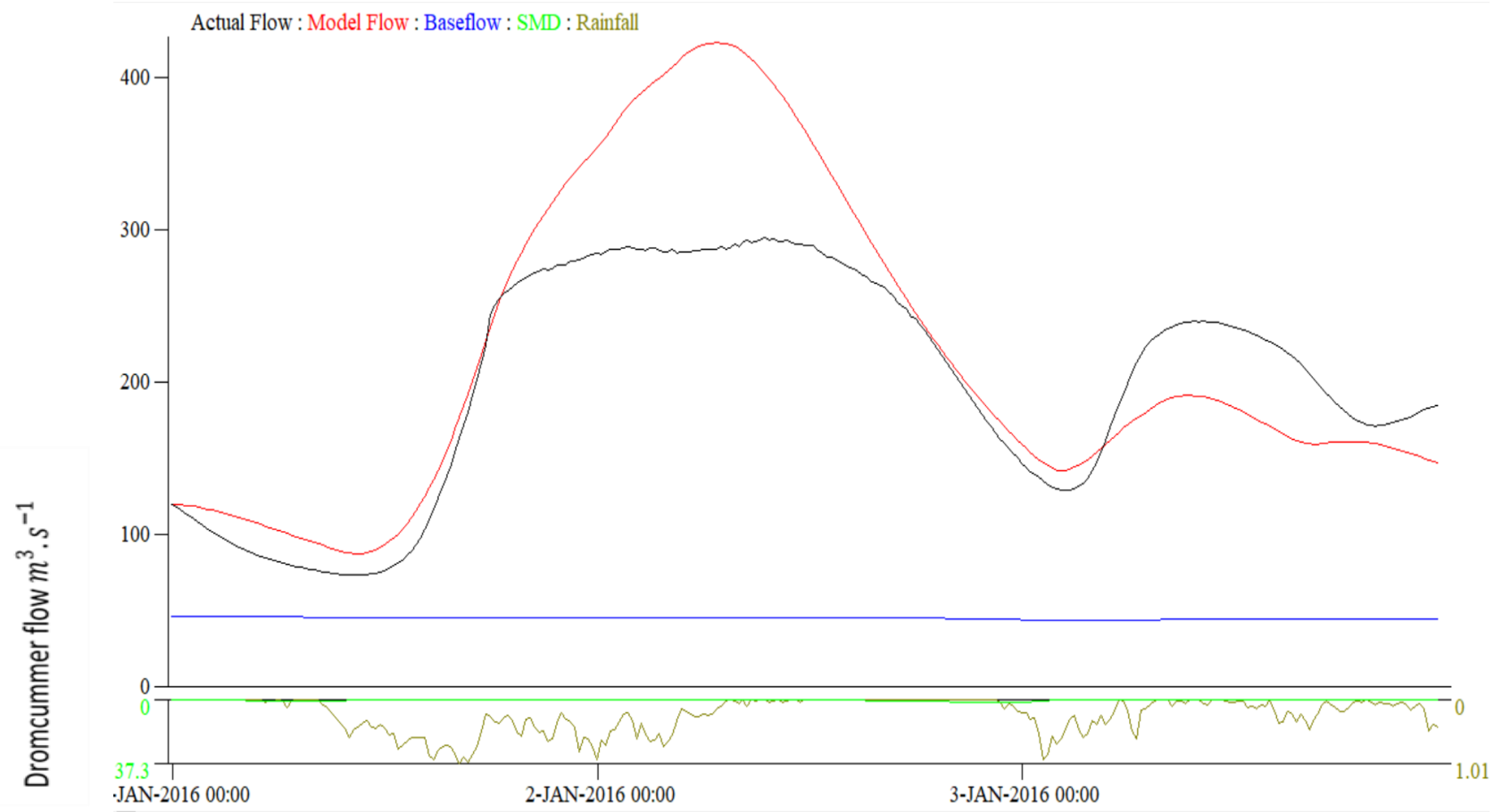


Figure A.4 (a).6 PDM simulation for validation, event in 2016 in Dromcummer.

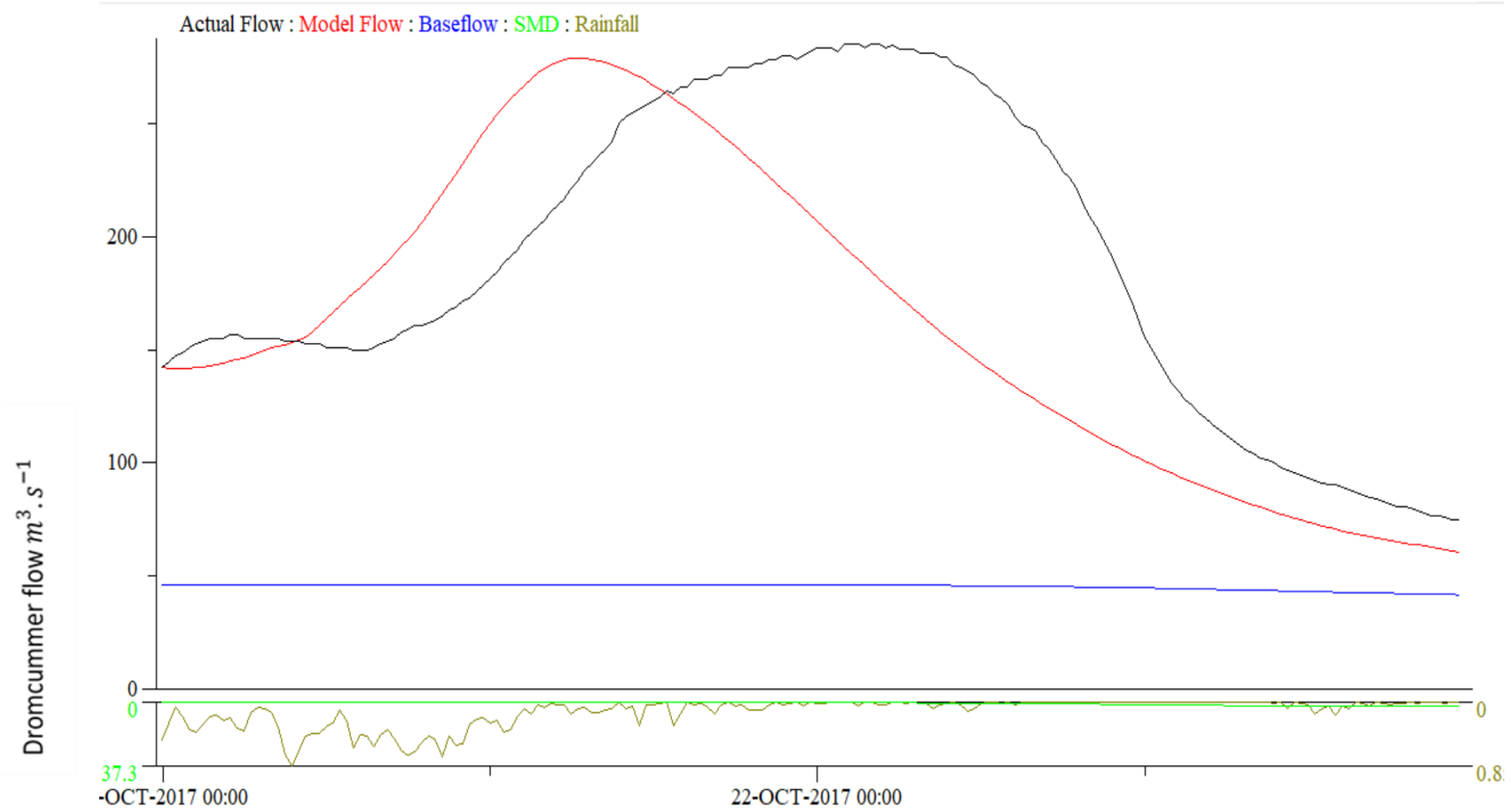


Figure A.4 (a).7 PDM simulation for validation, event in 2017 in Dromcummer.

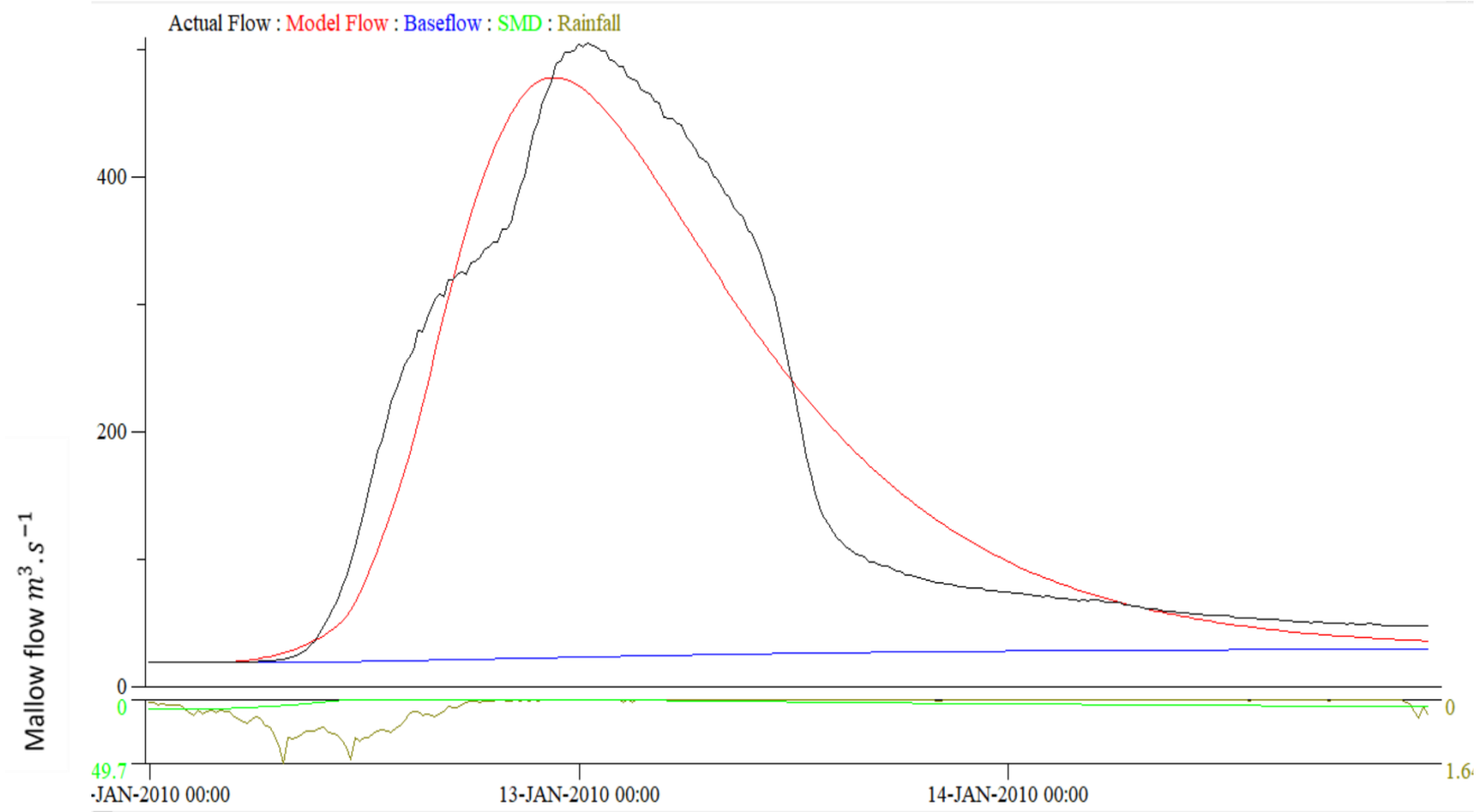


Figure A.4 (b).1 PDM simulation for validation, eventin 2010 in Mallow Rail BR.

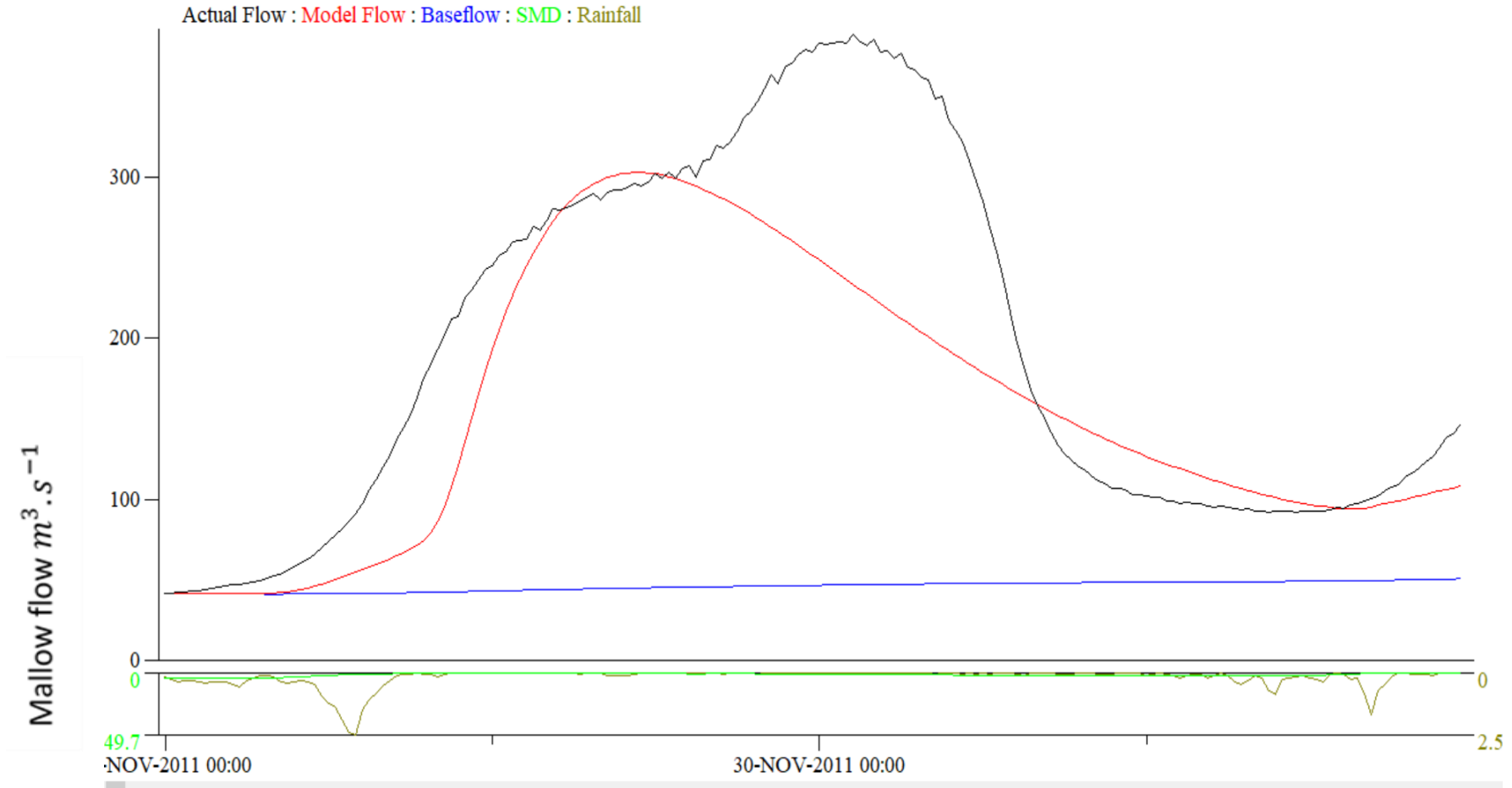


Figure A.4 (b).2 PDM simulation for validation, eventin 2011 in Mallow Rail BR.

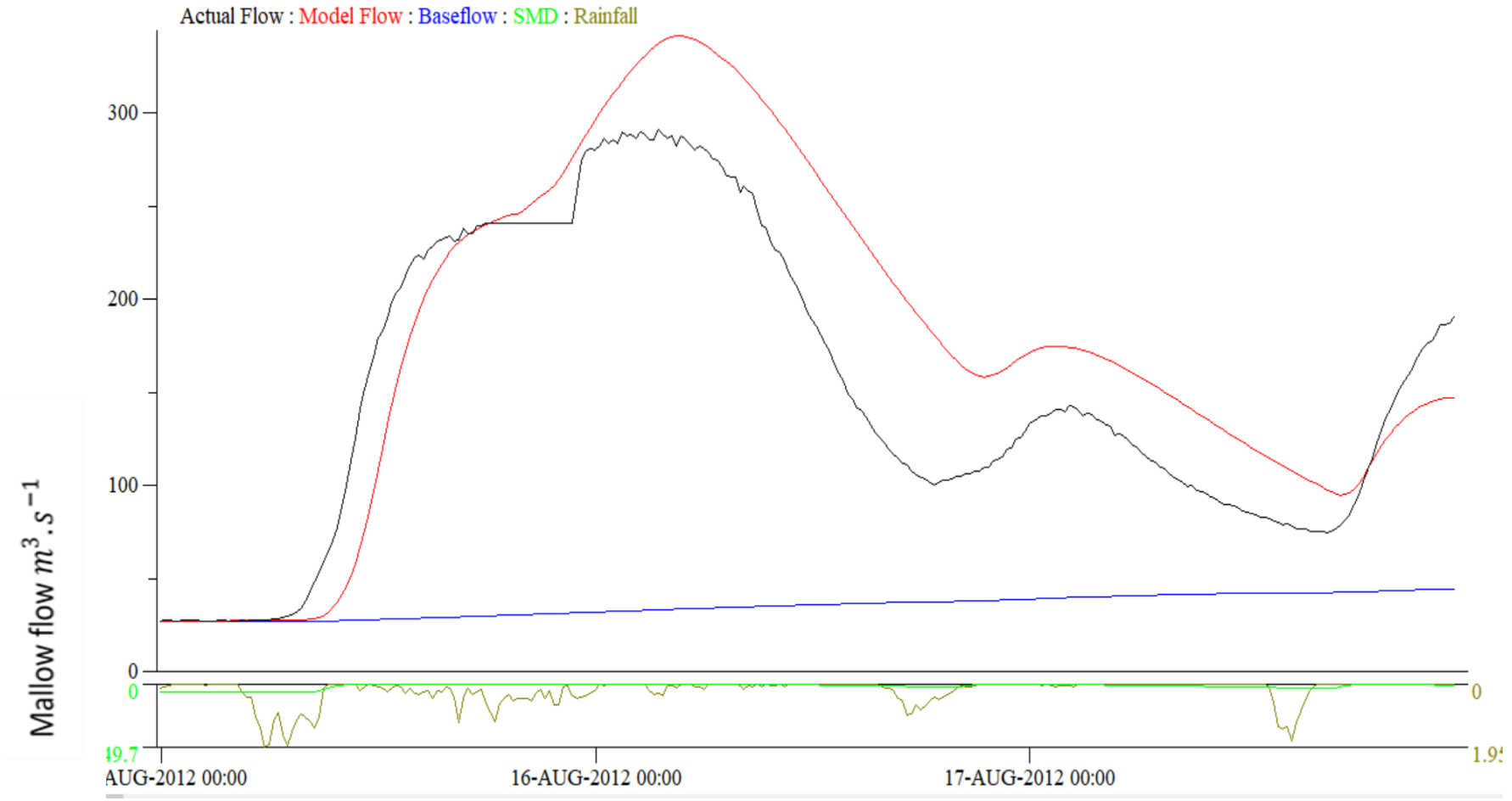


Figure A.4 (b).3 PDM simulation for validation, eventin 2012 in Mallow Rail BR.

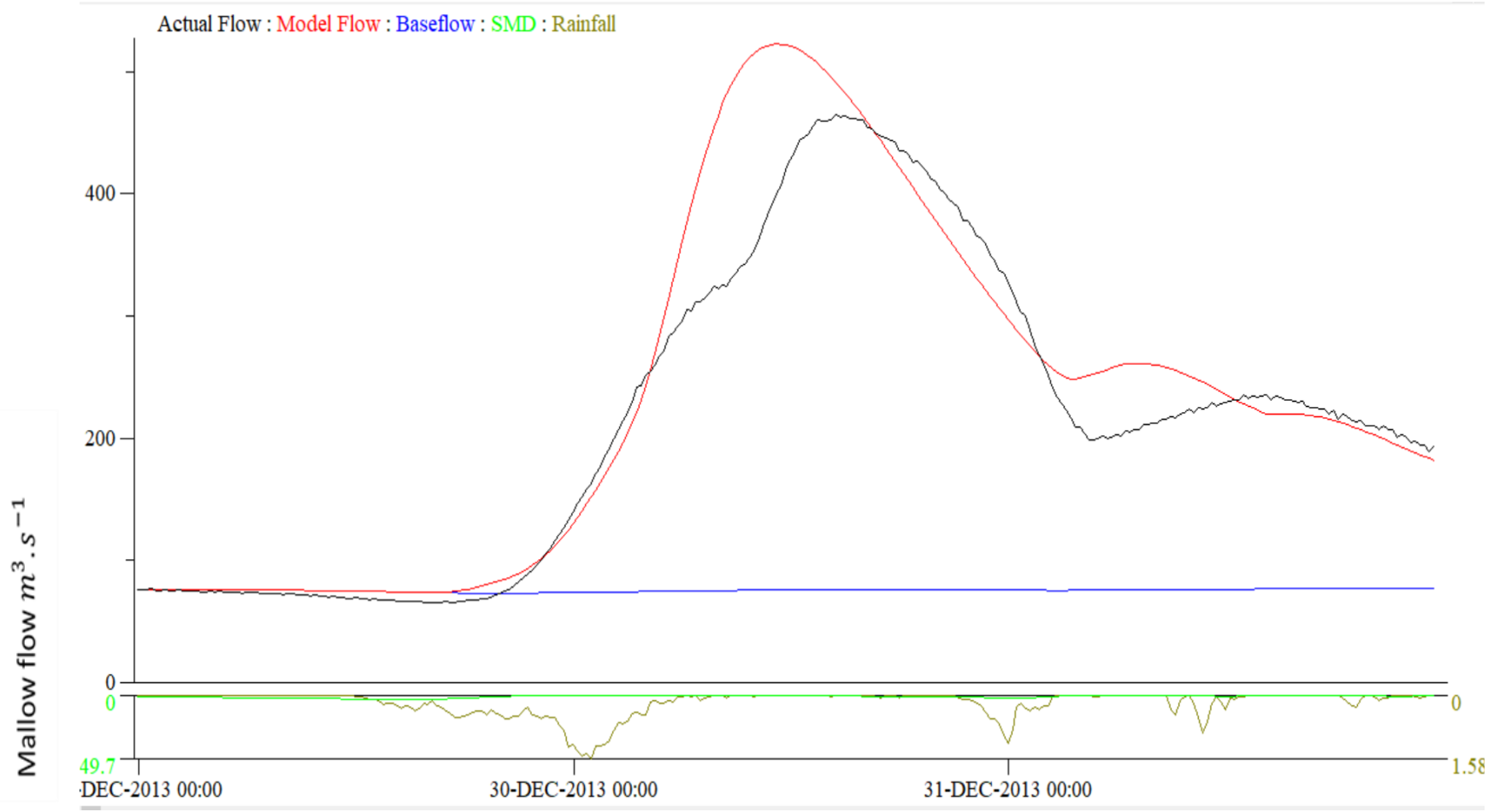


Figure A.4 (b).4 PDM simulation for validation, eventin 2013 in Mallow Rail BR.

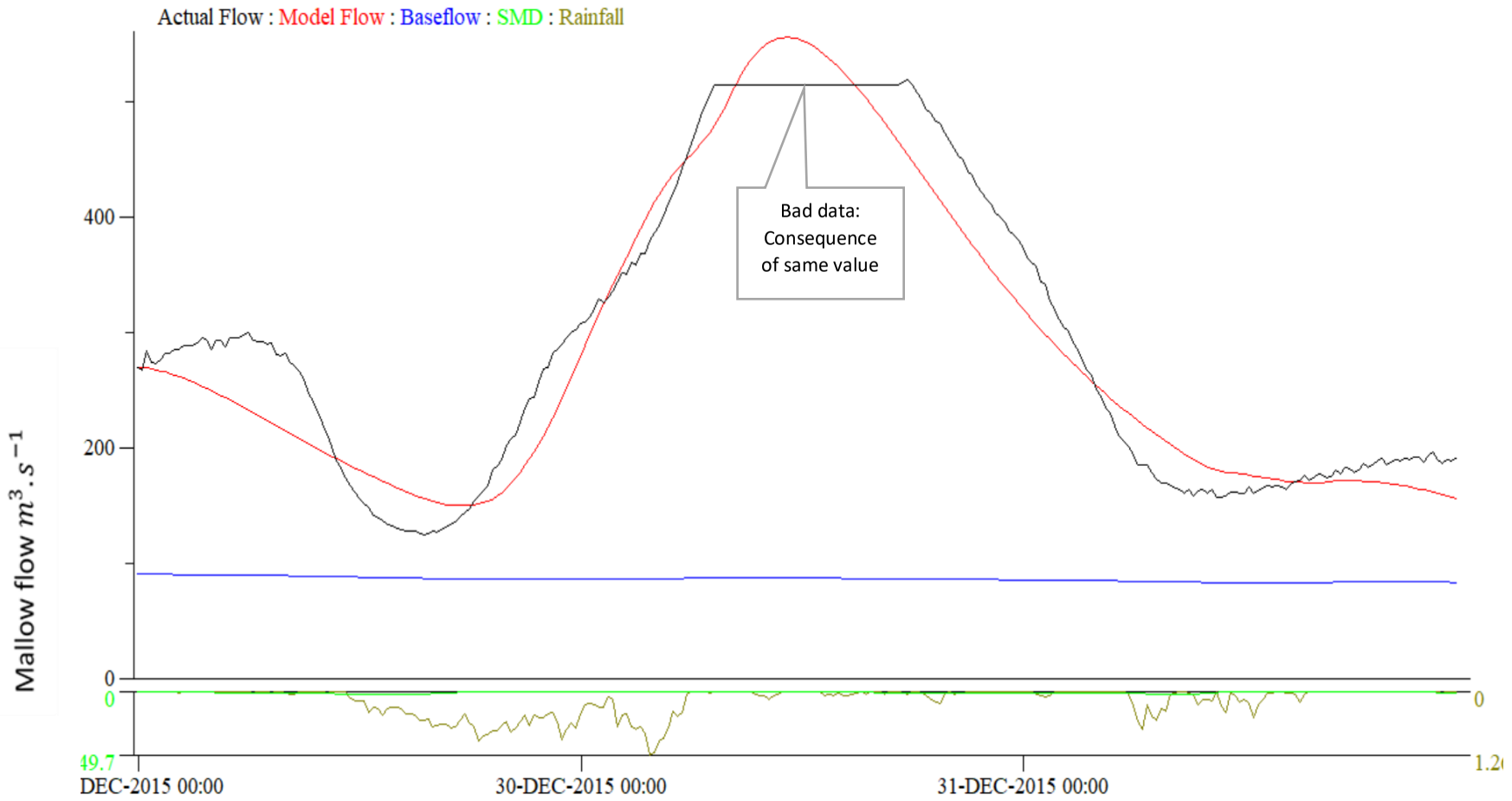


Figure A.4 (b).5 PDM simulation for validation, eventin 2015 in Mallow Rail BR.

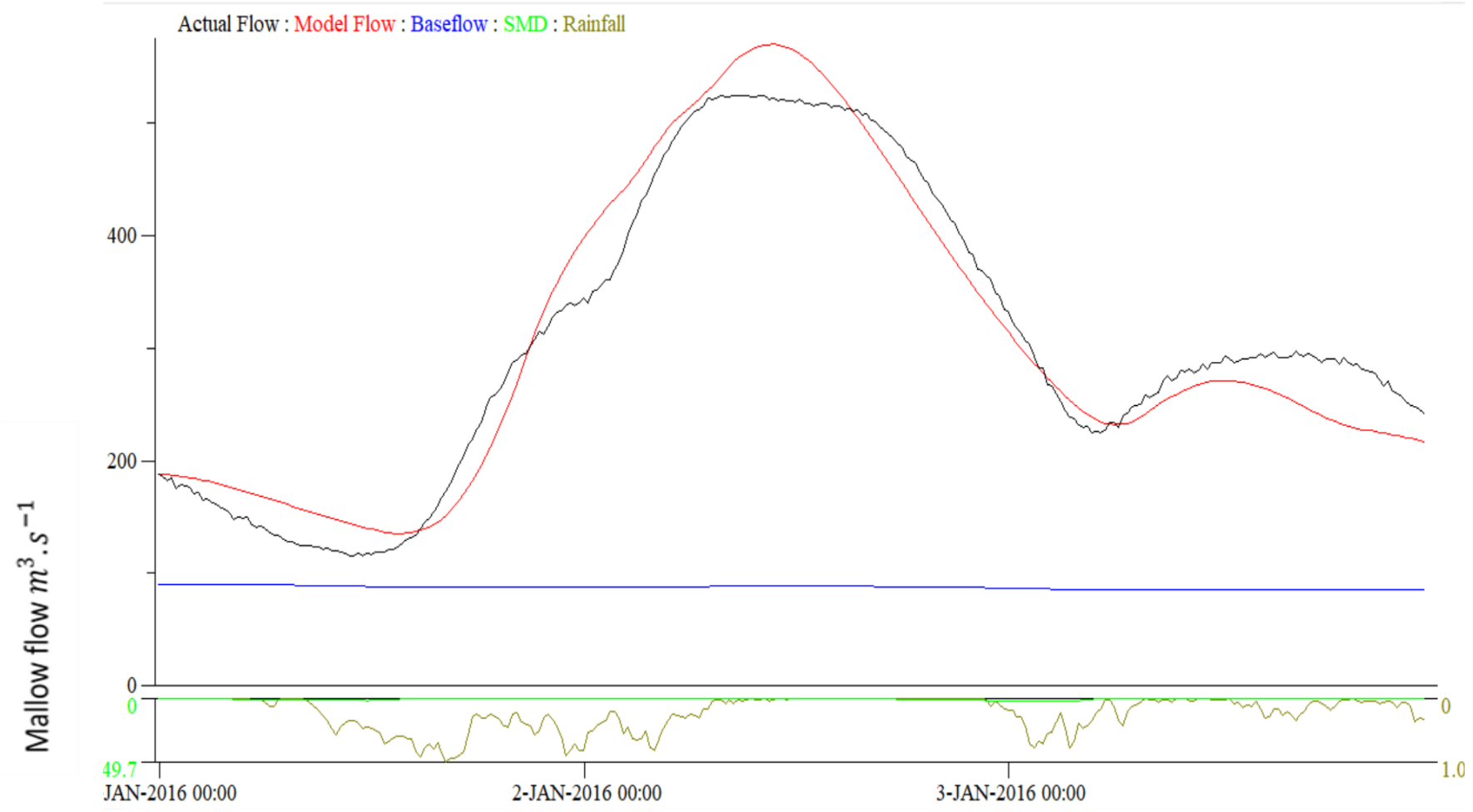


Figure A.4 (b).6 PDM simulation for validation, eventin 2016 in Mallow Rail BR.

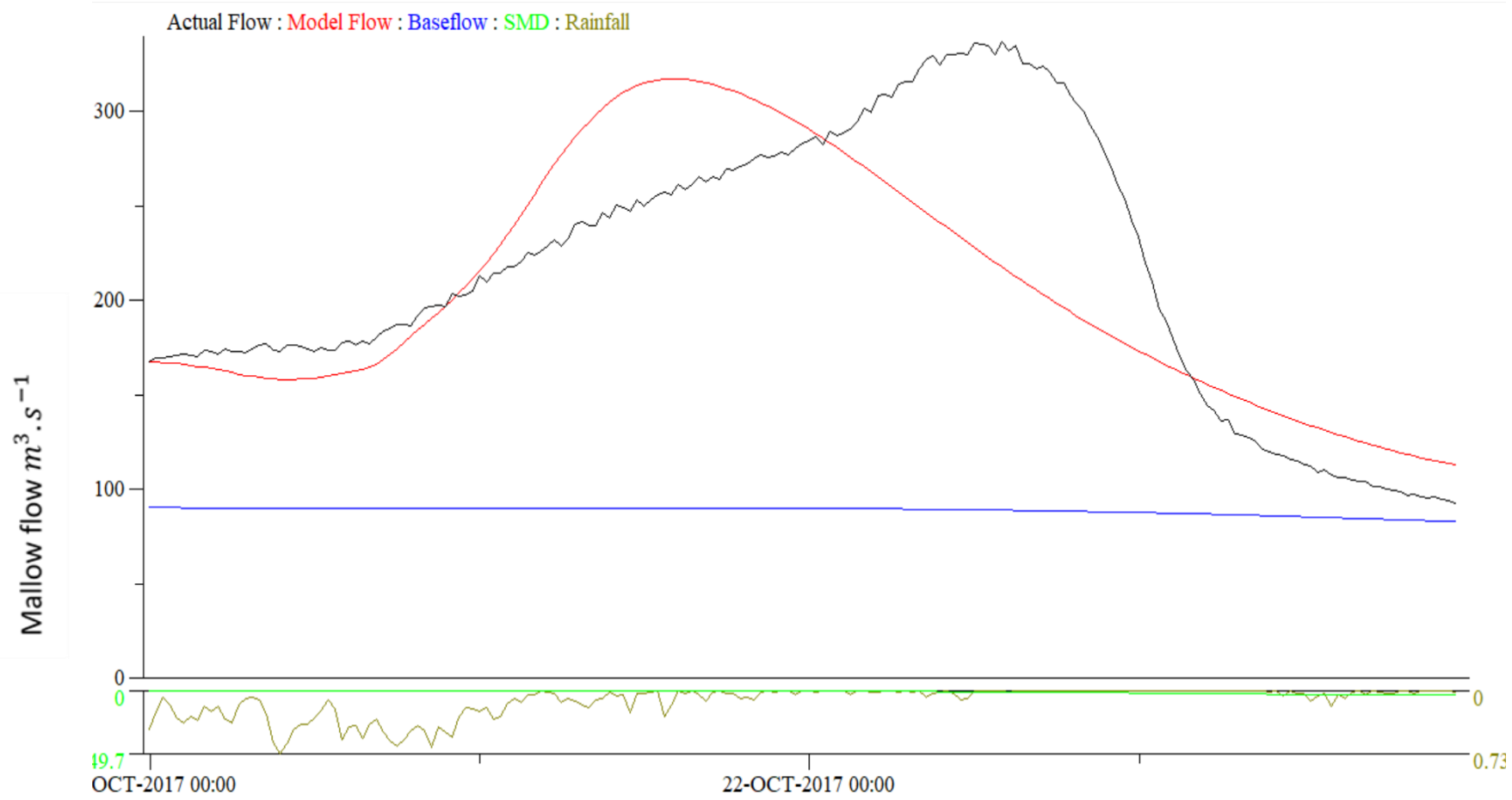


Figure A.4 (b).7 PDM simulation for validation, eventin 2017 in Mallow Rail BR.

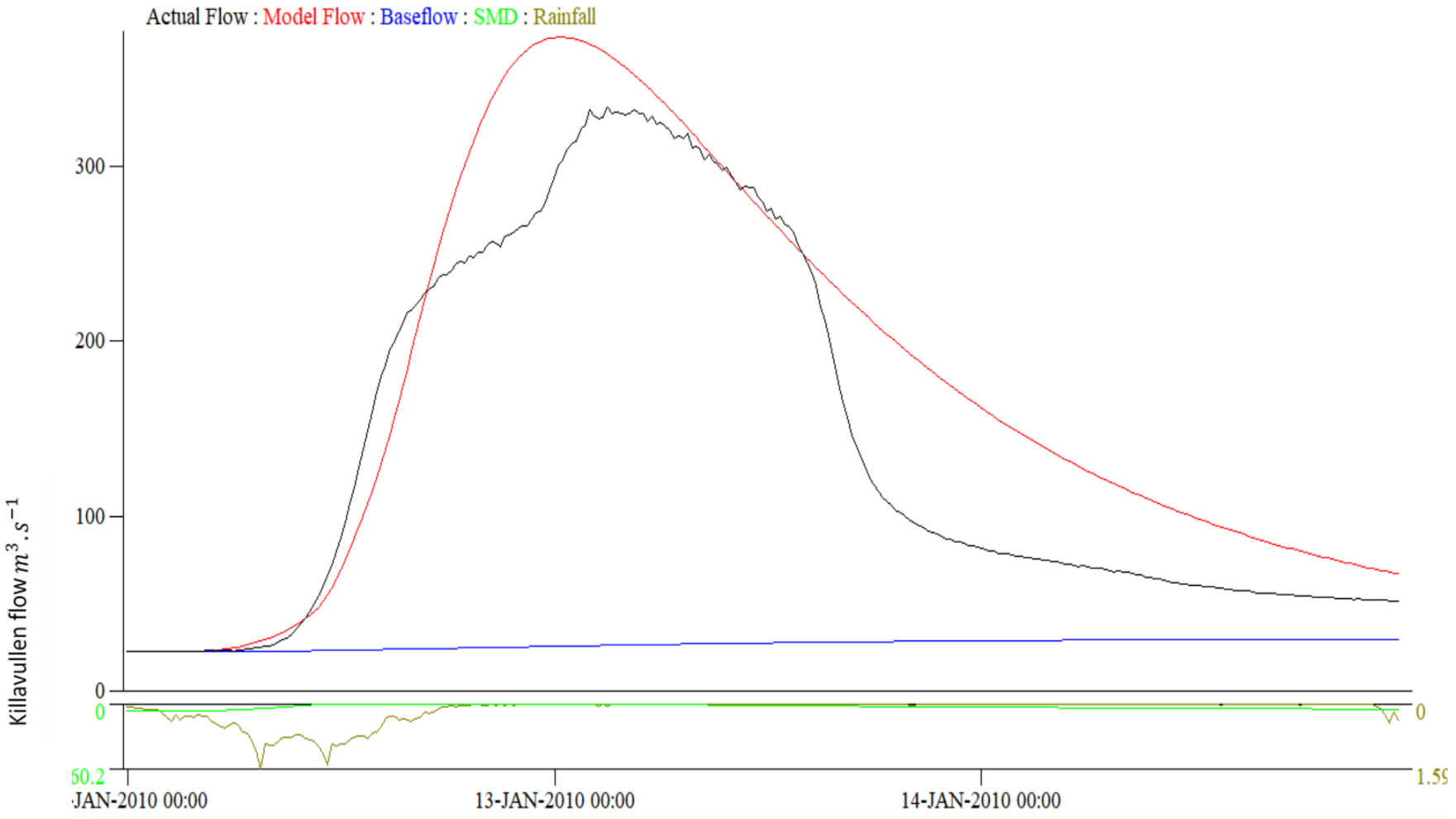


Figure A.4 (c).1 PDM simulation for validation, eventin 2010 in Killavullen.

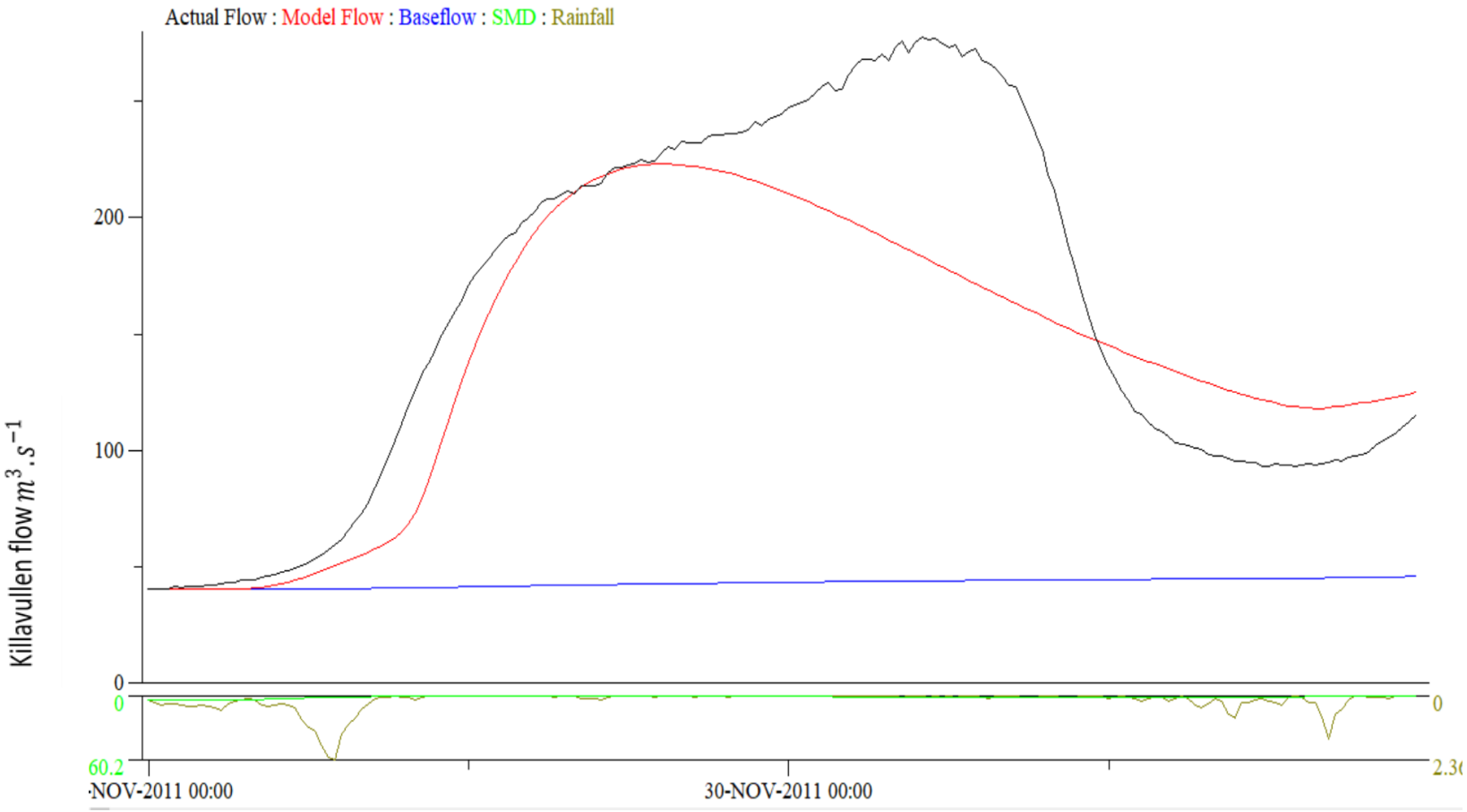


Figure A.4 (c).2 PDM simulation for validation, eventin 2011 in Killavullen.

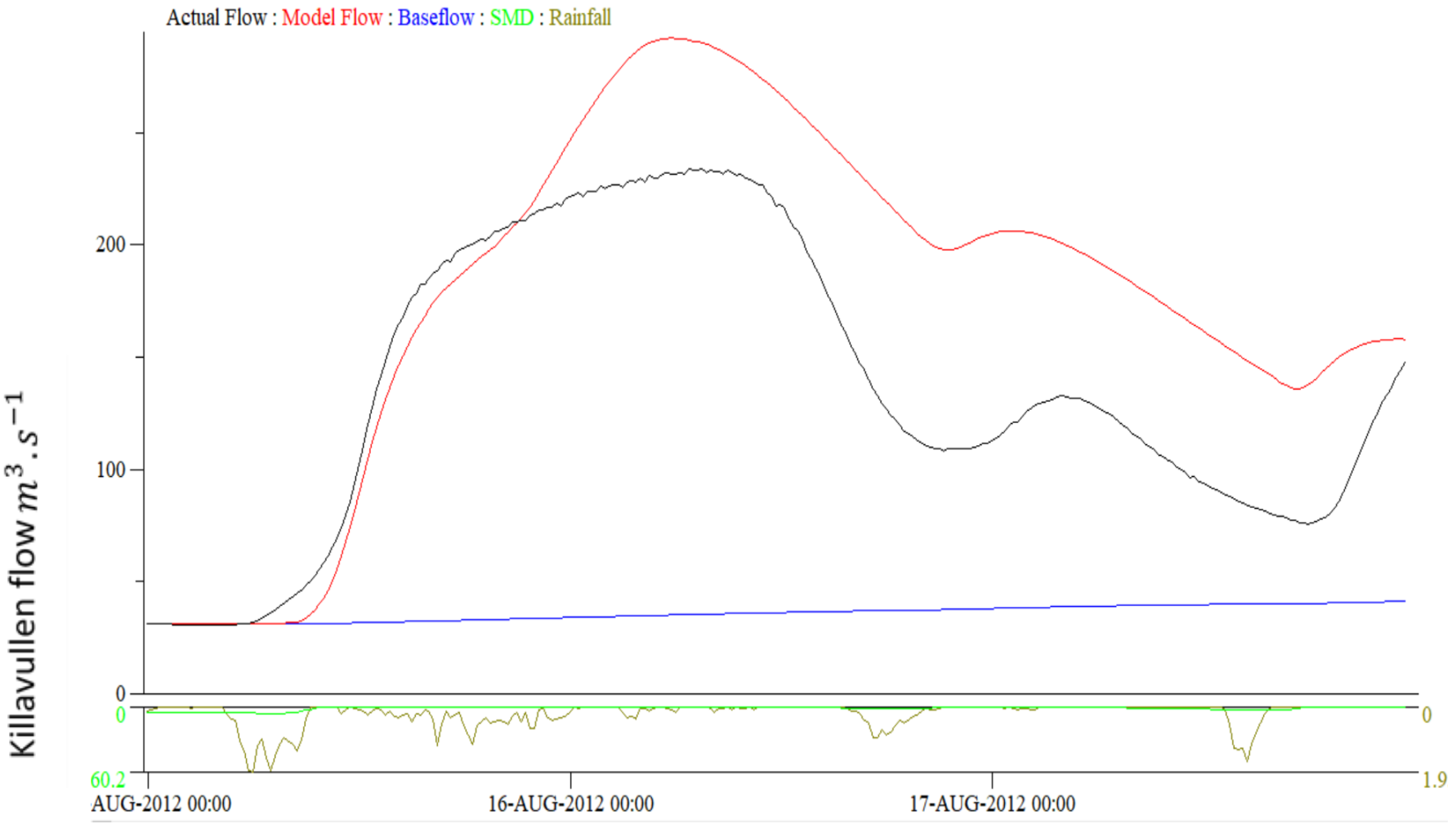


Figure A.4 (c).3 PDM simulation for validation, eventin 2012 in Killavullen.

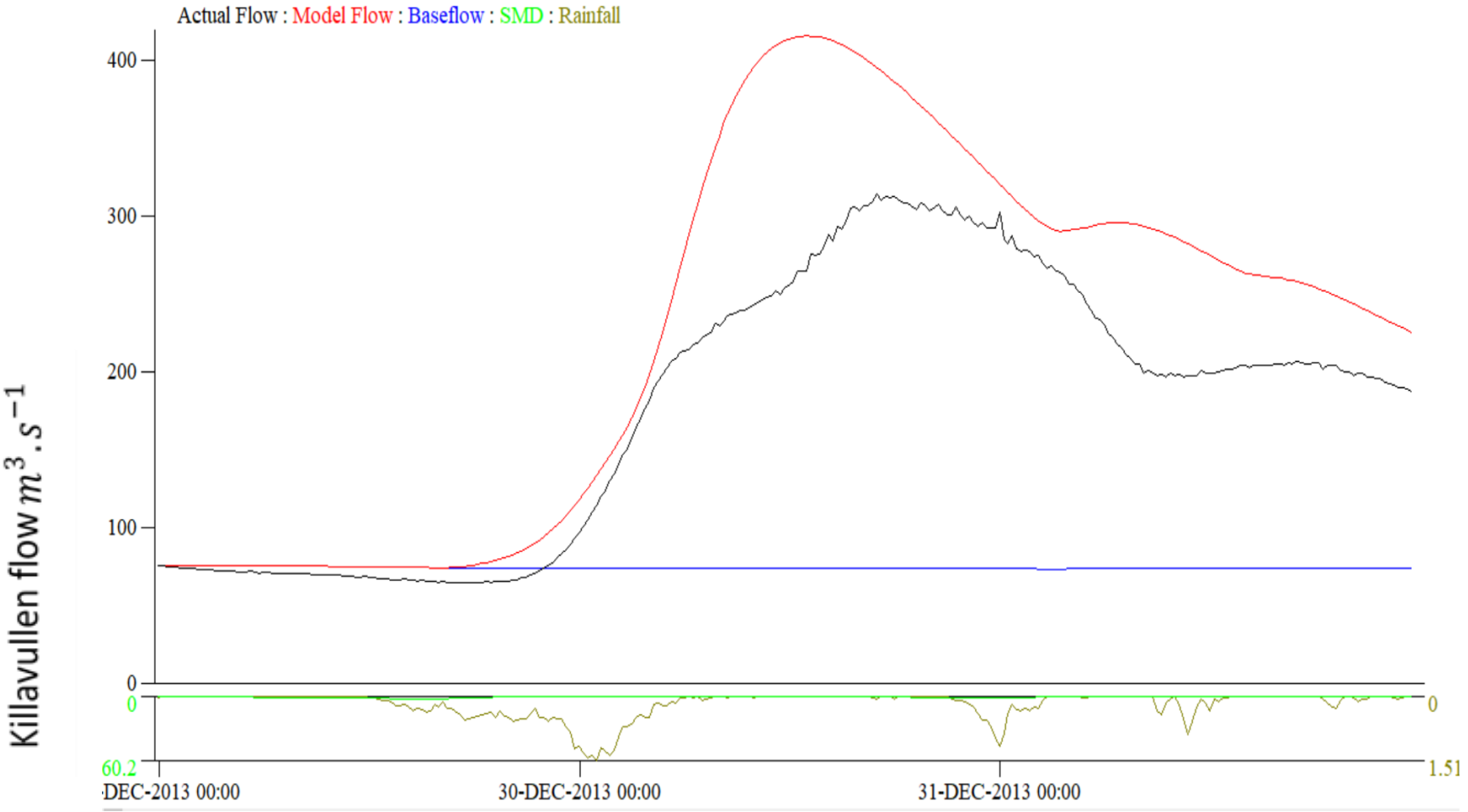


Figure A.4 (c).4 PDM simulation for validation, eventin 2013 in Killavullen.

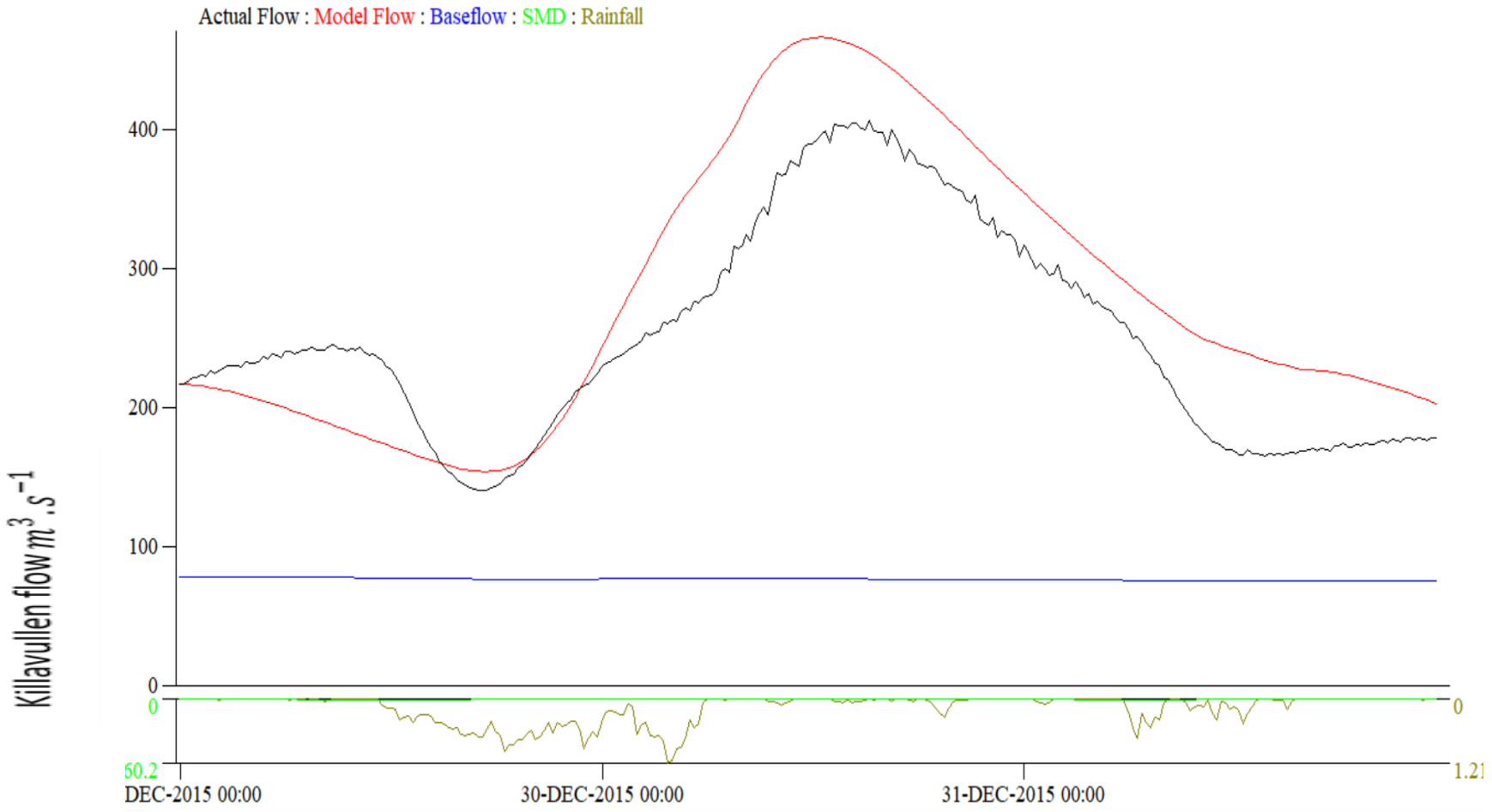


Figure A.4 (c).5 PDM simulation for validation, eventin 2015 in Killavullen.

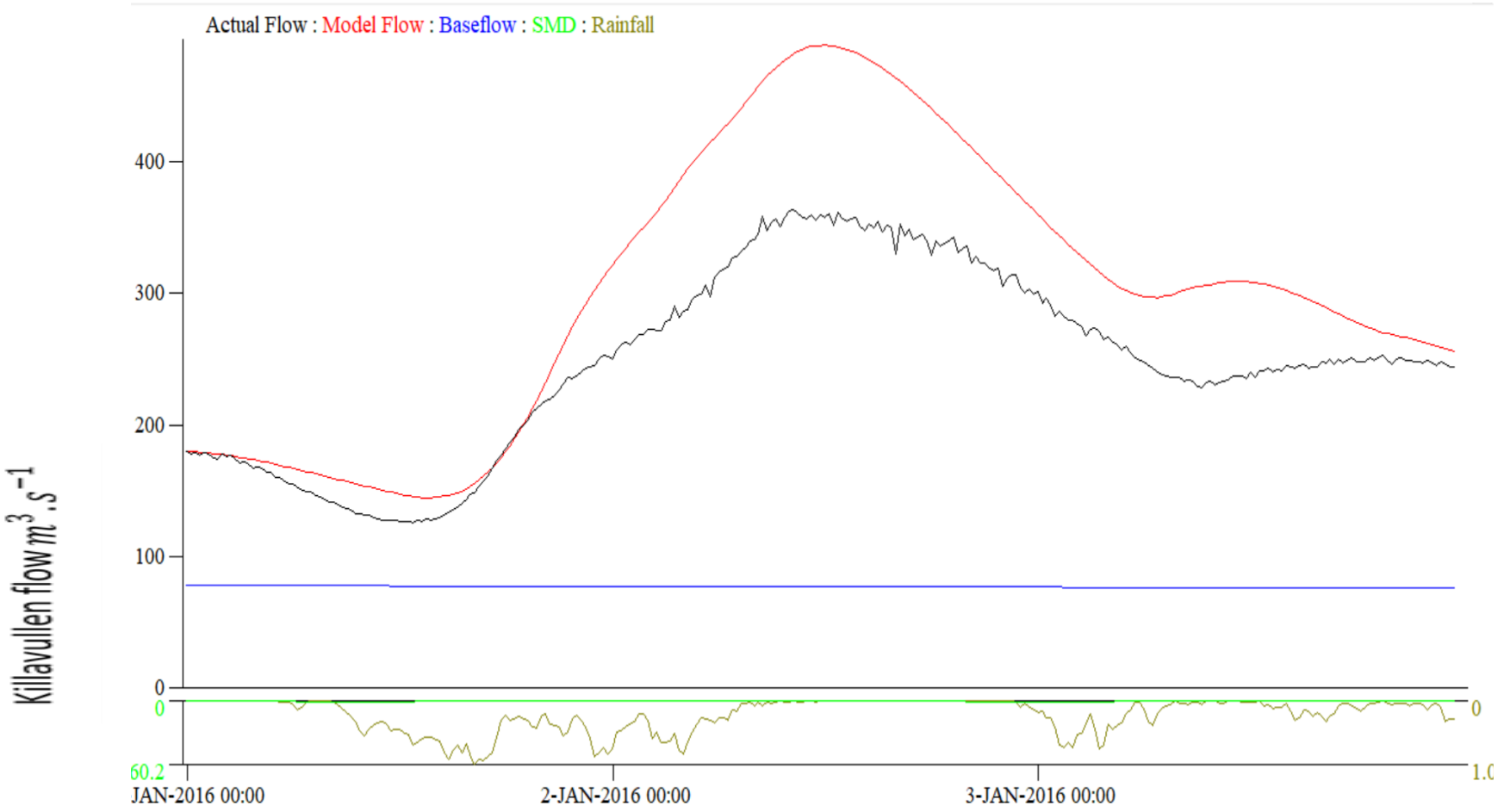


Figure A.4 (c).6 PDM simulation for validation, eventin 2016 in Killavullen.

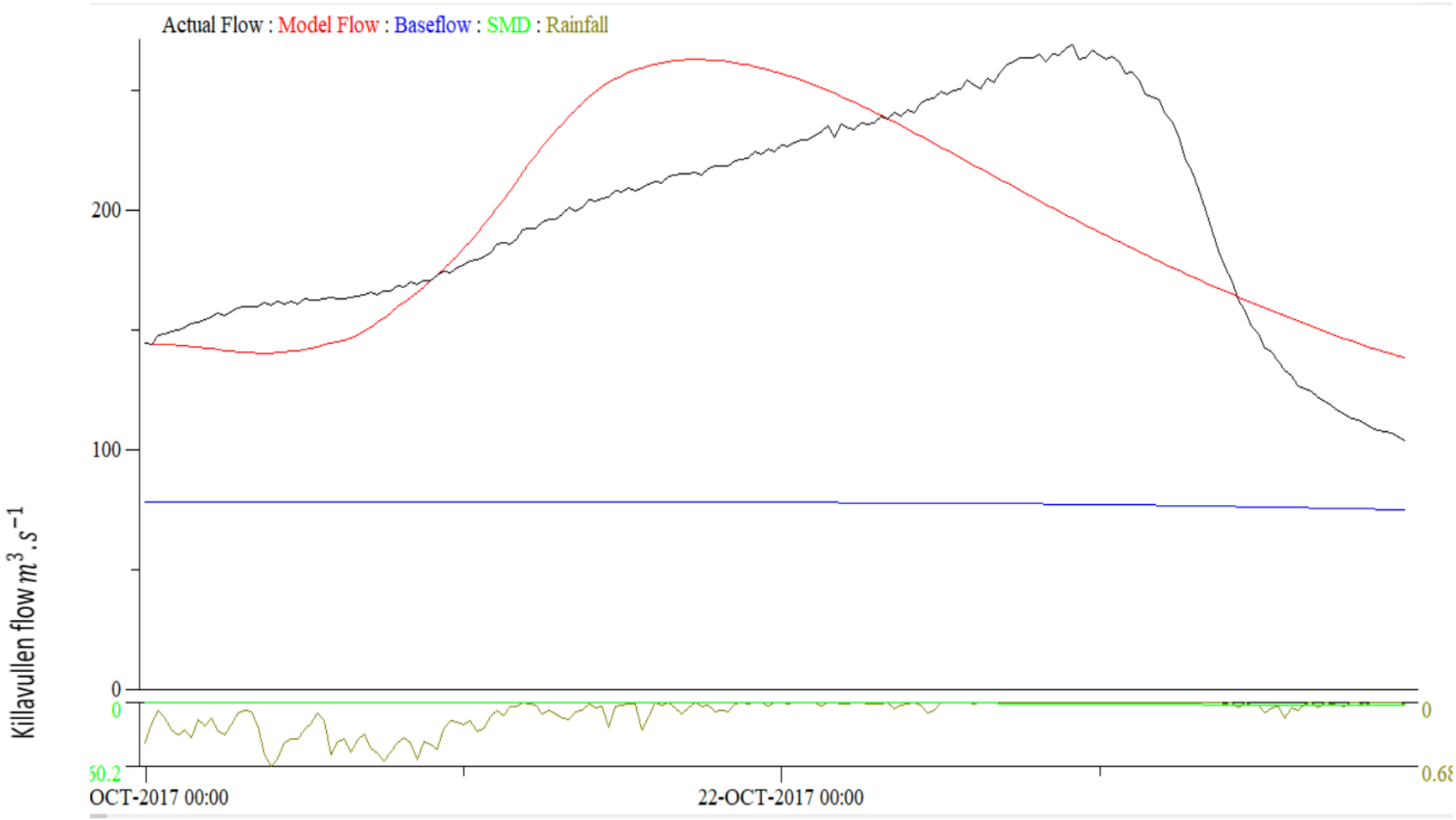


Figure A.4 (c).7 PDM simulation for validation, eventin 2017 in Killavullen.

Table A.2 (a) Summary of events validation at Dromcummer- 2010 to 2017

No	Year	R^2	Rainfall (mm)	Computed baseflow ($m^3.s^{-1}$)	Observed flow ($m^3.s^{-1}$)	Computed flow ($m^3.s^{-1}$)	Potential evaporation (mm)	Actual evaporation (mm)	Net rainfall (mm)
1	2010	0.8878	45.48	2.81	27.88	32.76	0.13	0.13	45.36
2	2011	0.8870	40.48	4.94	25.73	24.64	0.37	0.37	40.11
3	2012	0.8228	62.36	6.4	34.99	40.56	7.06	7.06	55.30
4	2013	0.8332	34.06	8.99	39.48	37.59	0	0	34.06
5	2015	0.5341	53.51	13.28	55.65	65.35	0	0	53.51
6	2016	0.4741	66.44	13.41	59.49	65.29	0.01	0.01	59.29
7	2017	0.2048	25.88	9.02	37.72	32.63	1.17	1.17	24.12

Table A.2 (b) Summary of events validation at Mallow Rail BR- 2010 to 2017

No	Year	R^2	Rainfall (mm)	Computed baseflow ($m^3.s^{-1}$)	Observed flow ($m^3.s^{-1}$)	Computed flow ($m^3.s^{-1}$)	Potential evaporation (mm)	Actual evaporation (mm)	Net rainfall (mm)
1	2010	0.9309	44.03	5.46	37.10	36.46	0.12	0.12	43.91
2	2011	0.6583	35.96	6.74	28.37	23.09	0.38	0.38	28.37
3	2012	0.6717	61.81	7.75	33.39	38.82	7.07	7.07	54.74
4	2013	0.8729	53.91	16.65	47.86	51.25	0	0	53.91
5	2015	0.9316	52.54	19.04	65.16	62.69	0	0	52.54
6	2016	0.9461	65.54	19.27	67.79	67.75	0.01	0.01	65.54
7	2017	0.5177	23.35	13.04	31.79	30.48	1.77	1.77	21.58

Table A.2 (c) Summary of events validation at Killavullen- 2010 to 2017

No	Year	R^2	Rainfall (mm)	Computed baseflow ($m^3.s^{-1}$)	Observed flow ($m^3.s^{-1}$)	Computed flow ($m^3.s^{-1}$)	Potential evaporation (mm)	Actual evaporation (mm)	Net rainfall (mm)
1	2010	0.7933	44.05	5.49	29.07	35.66	0.12	0.12	43.92
2	2011	0.7245	34.77	5.91	22.03	19.50	0.37	0.37	34.40
3	2012	0.2849	60.65	7.40	28.08	36.57	7.07	7.07	53.59
4	2013	0.4471	53.35	15.34	36.79	46.73	0	0	53.35
5	2015	0.5422	52.53	15.81	51.50	57.78	0	0	52.53
6	2016	0.0596	64.64	15.91	51.16	62.55	0.01	0.01	64.64
7	2017	0.3858	22.77	10.70	27.04	26.87	1.77	1.77	20.99

Appendix B Python & R Code

B.1 Python Code MERA data using GRIBapi Package:

```
#-------------------------------------------------------------------------------------------------------------------------------
#    Monthly Rainfall
#    1)  inpu1:  file contain raingauges lat,lon and elevation
#    2) input : high resolution rainfall data from MERA for each month of 2010
#    2) output: file contain lat, lon, total monthly rainfall of nearest point to raingues' location
#-------------------------------------------------------------------------------------------------------------------------------
import traceback
from  gribapi import *
import pyproj
import numpy as np
import pandas as pd
from Tkinter import Tk
from tkFileDialog import askopenfilename

Tk().withdraw()                                 # we don't want a full GUI, so keep the root window from appearing
file_grb = askopenfilename()                    # show an "Open" dialog box and return the path to the selected file
path = file_grb[:file_grb.find('MERA')]+'MERA/MERA_Month/       '# path for output , all monthly rainfall data
fout = path + file_grb[file_grb.find('_2010'):file_grb.find('_61')] +'.csv'    # define the name of each output file
VERBOSE=1
filename = '/media/sf_Geotopshare/stationLL.csv'          # input file contains raingauges' lon,lat and elevation
# define function to read raingaues' data and find the each raingauge's nearest point from MERA
#calculate monthly rainfall of these points
#write each a file contains raingauge's nearest point lat, lon and monthly rainfall
def  monthly_rain():
   df_s1 = pd.DataFrame(columns=['seq','rain'])
   fn = open(filename, 'r')
   df_station =    pd.read_csv(fn,names=
                   ['FID','Stn_No_','Station_Na','Catchment','Easting','Northing','Elevation','lon','lat'],header = 0)
   df_station['nlat'] = 0.0
   df_station['nlon'] = 0.0
   df_station['total_rain'] = 0.0
   f1 = open(file_grb)
   mcount = grib_count_in_file(f1)
   gid_list = [grib_new_from_file(f1) for i in range(mcount)]
   tag = True
   s = 1
   for gid in gid_list:
    for i in range(len(df_station.index)):
       nearest = grib_find_nearest(gid,float(df_station.lat[i]),float(df_station.lon[i]))[0]
       if tag:
         df_station.nlat[i] = nearest.lat
         df_station.nlon[i] = nearest.lon-360
       if (s % 3 == 0) :
          df_station.total_rain[i] += nearest.value
    tag = not tag
    s += 1
   df_station.to_csv(fout,sep =',',ignore_index = True)

monthly_rain()
```

```
#-------------------------------------------------------------------------------------------------------------------------------------------
#    Annual Rainfall
#    1) inputs:  12 files  contain lat, lon, total monthly rainfall of nearest point to each rain gauge's location
#    2) output: flie contains lat, lon and total annual rainfall of nearest point to each rain gauge's location
#-------------------------------------------------------------------------------------------------------------------------------------------

months = ['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec']
path = '/media/sf_Geotopshare/MERA/MERA_Month/'  # path of monthly data produce by monthly_rain function
path1 = '/media/sf_Geotopshare/MERA/'
listfile = os.listdir(path)   # create list of 12 months monthly rainfall data
fout = path + 'annual_2010.csv' # define output file
colnames =
['FID','Stn_No_','Station_Na','Catchment','Easting','Northing','Elevation','lon','lat','nlat','nlon','total_rain']
# number of raingues
stn = ('80701','80702','80703','80704','80705','80706','80707','80708','80709','80710',
      '80711','80712','80713','80714','80715','80716','80717','80718','80719','80720',
      '80721','80722','80723','80724','80725','80726','80727','80728','80729','80730',
      '80731','80732')
y_pos = np.arange(len(stn))
tag = True
m = 0
i = 0
mn_rain = np.zeros((12,32))
# Read each monthy rainfall data to calculates annual rainfall
for fn in listfile:
    df = pd.read_csv(path + fn,names = colnames,header = 0)
    mn_rain[m,:] = df.total_rain
    if tag :
       df_annual = df
       tag = not tag
    else :
      df_annual.total_rain += df.total_rain
    m += 1
df_annual.to_csv(fout,sep= ',', ignore_index=True)



#-------------------------------------------------------------------------------------------------------------------------------------------
#    Plots(Bar plot , Box plot)
#    1) inputs:  12 file monthly rainfall
#    2) input:  annual rainfall
 #-------------------------------------------------------------------------------------------------------------------------------------------
def Bar_MERA():    # plot annual rainfall for each rainggauge's nearest point
   opann_rain_10 = genfromtxt(path1 + 'ann_rain_10.csv', delimiter=' ')
   plt.bar(df_annual.FID,df_annual.total_rain,width = 0.004, align = 'center' )
   plt.xticks(y_pos,stn,rotation = 90,fontsize = 12,fontname = 'Calibri')
   plt.xlabel('Staion Number')
   plt.ylabel('Rainfall: mm')
   plt.show()
Bar_MERA()
```

```
#-------------------------------------------------------------------------------------------------------------------------------
#   Plot:  Compare Monthly rainfall from MERA with OPW
#-------------------------------------------------------------------------------------------------------------------------------
def compare_MERA_OPW():
  opann_rain_10 = genfromtxt(path1 + 'ann_rain_10.csv', delimiter=' ')
  print opann_rain_10
  plt.bar(df_annual.FID,df_annual.total_rain,color = 'g',width = 0.3, align = 'center' )
  plt.bar(df_annual.FID +0.3,opann_rain_10,color = 'r',width = 0.3, align = 'center' )
  plt.xticks(y_pos,stn,rotation = 90,fontsize = 12,fontname = 'Calibri')
  plt.xlabel('Staion Number')
  plt.ylabel('Rainfall: mm')
  plt.legend(['MERA','OPW'])
  plt.show()
#compare_MERA_OPW()


#-------------------------------------------------------------------------------------------------------------------------------
#   Plot:  Scatter plot of Monthly rainfall from MERA
#-------------------------------------------------------------------------------------------------------------------------------
def scatter_MERA():
 for i in range(1,32):
    plt.scatter(range(12), mn_rain[:,i], i , label="StN='{0}'".format(stn[i])  )
 plt.legend(numpoints =1, bbox_to_anchor=(1.02, 0.8),loc = 0,
 bbox_transform=plt.gcf().transFigure, fontsize = 6)
 plt.xticks(range(12),months,rotation = 45,fontsize = 8,fontname = 'Calibri')
 plt.xlabel('Month')
 plt.ylabel('Rainfall: mm')
 plt.show()
#scatter_MERA()

#-------------------------------------------------------------------------------------------------------------------------------
#   Plot:  Boxplot of Monthly rainfall from MERA
#-------------------------------------------------------------------------------------------------------------------------------
mn_rain = np.zeros((32,12))
for fn in listfile:
 df = pd.read_csv(path + fn,names = colnames,header = 0)
 mn_rain[:,m] = df.total_rain
 m += 1

def Box_MERA():
   plt.boxplot(mn_rain)
   plt.xticks(range(1,13),months,rotation = 45,fontsize = 8,fontname = 'Calibri')
   plt.xlabel('Month')
   plt.ylabel('Rainfall: mm')
   plt.show()
Box_MERA()
```

B.2 R Code for River level/ River flow data:

```
#-----------------------------------------------------------------------------------------------------------------------------------
#   create a datframe with PDM format contains 15-minute time step
#   from 2010 to 2017 to fill unrecorded values in river level/river flow data
#-----------------------------------------------------------------------------------------------------------------------------------
from_opw <- as.POSIXlt("2010/01/01 00:00:00", tz = "GMT")
to_opw <- as.POSIXlt("2017/12/30 00:00:00", tz = "GMT")

my_data_opw <- data.frame(X1 = seq.POSIXt(from_opw, to_opw , by = "15 mins", format = "%Y/%m/%d
%H:%M:%S" ))
my_data_opw <- separate(my_data_opw,"date", c("Date", "Time"), sep = " ")
my_data_opw <- separate(my_data_opw,"Date", c("Year", "mm", "dd"), sep = "/")
my_data_opw <- separate(my_data_opw,"Time", c("hh", "mn", "ss"), sep = ":")
my_data_opw$Year <- as.integer(my_data_opw$Year)
my_data_opw$mm <- as.integer(my_data_opw$mm)
my_data_opw$dd <- as.integer(my_data_opw$dd)
my_data_opw$hh <- as.integer(my_data_opw$hh)
my_data_opw$mn <- as.integer(my_data_opw$mn)
my_data-opw$ss <- as.integer(my_data_opw$ss)

nrow(my_data_opw)


#-----------------------------------------------------------------------------------------------------------------------------------
#   function to find missing, unrecorded and outliers of river level data
#   input filename and thershold of outliers
#-----------------------------------------------------------------------------------------------------------------------------------
total_row <- nrow(my_data_opw)
summary_2fn <- function(fn,tr){
 gap <- total_row - nrow(fn)
 missing <- nrow(fn[is.na(fn$V2),])
 outlier5 <-  nrow(fn[fn$V2 > tr & !is.na(fn$V2),])
 percent <- round((gap + missing + outlier5) * 100 / total_row,2)
 tprint <- paste(  '  ', gap , '  ', missing ,'   ', outlier5,'   ',percent  )
 print (tprint)
}

summary_2fn(dro_lvl_2010_2017,3) # summary of Dromcummer river levle data outlier less than 3 m


#-----------------------------------------------------------------------------------------------------------------------------------
#   function to find missing, unrecorded and outliers of river flow data
#   input filename and thershold of outliers
#-----------------------------------------------------------------------------------------------------------------------------------
summary_1fn <- function(fn,tr){
 gap <- total_row - nrow(fn)
 missing <- nrow(fn[is.na(fn$flow),])
 outlier5 <-  nrow(fn[fn$flow > tr & !is.na(fn$flow),])
 percent <- round((gap + missing + outlier5) * 100 / total_row,2)
 tprint <- paste(  '  ', gap , '  ', missing ,'   ', outlier5,'   ',percent  )
 print (tprint)
}
```

```
#-------------------------------------------------------------------------------------------------------------------------------
#    convert flow data from OPW to PDM format (for Killavullen)
#-------------------------------------------------------------------------------------------------------------------------------
Pdm_flow_format<- function(fn){
 # tmp_file <- fn[substr(fn$V1,1,4) == as.character(year),]
 tmp_file <- fn
 num_row <- nrow(tmp_file)
 print(num_row)
 fn_out_n <- paste("kil_flow_1",".dat",sep = '')
 fn_out <- data.frame (Date = character(num_row), Time = character(num_row) ,flow = double(num_row) )
 tmp_file$X1 <- as.POSIXct(tmp_file$X1)
 tmp_file$X2 <- as.numeric(tmp_file$X2)
 #tmp_file$V3 <- as.numeric(tmp_file$V3)
 fn_out$Date <- format( tmp_file$X1 ,format = "%Y/%m/%d")
 fn_out$Time <- format(tmp_file$X1 ,format = "%H,%M,%S")
 fn_out <- separate(fn_out,"Date", c("Year", "mm", "dd"), sep = "/")
 fn_out <- separate(fn_out,"Time", c("hh", "mn", "ss"), sep = ",")
 fn_out$flow <- tmp_file$X2
 fn_out$colz <- 0
 fn_out$Year <- as.integer(fn_out$Year)
 fn_out$mm <- as.integer(fn_out$mm)
 fn_out$dd <- as.integer(fn_out$dd)
 fn_out$hh <- as.integer(fn_out$hh)
 fn_out$mn <- as.integer(fn_out$mn)
 fn_out$ss <- as.integer(fn_out$ss)
 #fn_out <- fn_out[fn_out$mn == 0 & fn_out$ss == 0,]
 index <- with(fn_out, order(Year,mm,dd,hh,mn,ss))
 flow_sort <- fn_out[index,]
 write.table(fn_out,fn_out_n,sep = ", ",col.names = FALSE, row.names = FALSE)
 return(fn_out)
}
#-------------------------------------------------------------------------------------------------------------------------------
#      function to fill the gap (missing value) with two nearest record
#      input:  15-minutes time step data and filename
#-------------------------------------------------------------------------------------------------------------------------------
average_gap_filling <- function(mdata,fn){
 tmp_fn  <- merge(mdata,fn,by = "X1" , all.x=TRUE)
 len <- nrow(tmp_fn)-1
 if (is.na(tmp_fn$X2[1]) & !is.na(tmp_fn$X2[2]))
  tmp_fn$X2[1] <- tmp_fn$X2[2]
 for(i in 2:len){
  b <- i -1
  a <- i+1
  if (is.na(tmp_fn$X2[i]) & !is.na(tmp_fn$X2[b])) {
   if (!is.na(tmp_fn$X2[a]))
    tmp_fn$X2[i] <- (tmp_fn$X2[b] + tmp_fn$X2[a]) /2
   else
    tmp_fn$X2[i] <- tmp_fn$X2[b]
   #print (paste(i,tmp_fn$V1[i],tmp_fn$V2[b], tmp_fn$V2[i],tmp_fn$V2[a]))
  }
 }
 if (is.na(tmp_fn$X2[len+1])){
  tmp_fn$X2[len+1] <- tmp_fn$X2[len]
  print (paste(i,tmp_fn$X1[i],tmp_fn$X2[b], tmp_fn$X2[i],tmp_fn$X2[a]))
 }
 return(tmp_fn)
}
```

```
#-------------------------------------------------------------------------------------------------------------------------------
#     Function to find outliers greater than the thershold tr
#     and correct them with average of two nearest record
#     input: filename and thershold
#-------------------------------------------------------------------------------------------------------------------------------
average_outler_filling <- function(fn,tr){
 #tmp_fn  <- merge(my_data,fn,by = "V1" , all=TRUE)
 tmp_fn <- fn
 len <- nrow(tmp_fn)-1
 if (tmp_fn$X2[1] > tr & tmp_fn$X2[2] <= tr)
  tmp_fn$X2[1] <- tmp_fn$X2[2]
 for(i in 2:len){
  b <- i -1
  a <- i+1
  if (!is.na(tmp_fn$X2[i])  & !is.na(tmp_fn$X2[b]) ){
   if (tmp_fn$X2[i] > tr & tmp_fn$X2[b] <= tr){
    if (tmp_fn$X2[a] <= tr)
     tmp_fn$X2[i] <- (tmp_fn$X2[b] + tmp_fn$X2[a]) /2
    else
     tmp_fn$X2[i] <- tmp_fn$X2[b]
   }
  }
 }
 if (tmp_fn$X2[len+1] > tr & !is.na(tmp_fn$X2[len+1])){
  tmp_fn$X2[len+1] <- tmp_fn$X2[len]
  print (paste(i,tmp_fn$X1[i],tmp_fn$X2[b], tmp_fn$X2[i],tmp_fn$X2[a]))
 }
 return(tmp_fn)
}

#-------------------------------------------------------------------------------------------------------------------------------
#     function to calculate river flow using rating curve coefficient(a & b) and river level data
#     input: river level data file and a & b
#-------------------------------------------------------------------------------------------------------------------------------
Pdm_calflow_1coef <- function(fn,a1,b1){
 # tmp_file <- fn[substr(fn$V1,1,4) == as.character(year),]
 tmp_file <- fn
 num_row <- nrow(tmp_file)
 print(num_row)
 fn_out_n <- paste("Drom_flow_1",".dat",sep = '')
 fn_out <- data.frame (Date = character(num_row), Time = character(num_row) ,flow = double(num_row) )
 tmp_file$X1 <- as.POSIXct(tmp_file$X1)
 tmp_file$X2 <- as.numeric(tmp_file$X2)
 #tmp_file$V3 <- as.numeric(tmp_file$V3)
 fn_out$Date <- format( tmp_file$X1 ,format = "%Y/%m/%d")
 fn_out$Time <- format(tmp_file$X1 ,format = "%H,%M,%S")

 fn_out <- separate(fn_out,"Date", c("Year", "mm", "dd"), sep = "/")
 fn_out <- separate(fn_out,"Time", c("hh", "mn", "ss"), sep = ",")
 #fn_out$flow <- round(a1 * (tmp_file$X2 ^ b1 ),2)
 for (j in 1:nrow(tmp_file)){
  if (tmp_file$X2[j] > 0)
   fn_out$flow[j] <- round(a1 * (tmp_file$X2[j] ^ b1 ),2)
  else
   fn_out$flow[j] <- 0
 }
```

```
 fn_out$colz <- 0

 fn_out$Year <- as.integer(fn_out$Year)
 fn_out$mm <- as.integer(fn_out$mm)
 fn_out$dd <- as.integer(fn_out$dd)

 fn_out$hh <- as.integer(fn_out$hh)
 fn_out$mn <- as.integer(fn_out$mn)
 fn_out$ss <- as.integer(fn_out$ss)
 #fn_out <- fn_out[fn_out$mn == 0 & fn_out$ss == 0,]

 write.table(fn_out,fn_out_n,sep = ", ",col.names = FALSE, row.names = FALSE)
 return(fn_out)
}


#-------------------------------------------------------------------------------------------------------------------------------------------
#      function to find the suitable confident in Duarrigle based of the river level
#      input: height of river level
#-------------------------------------------------------------------------------------------------------------------------------------------
find_coff <- function(value){
 # print(value)
 range <- c(0,0.201,1.592,2.385)
 indx <- 3
 for (i in 2:4){
  if (value >= range[i-1] & value < range[i] )
    indx <-  i-1
 }
 return(indx)
}


#-------------------------------------------------------------------------------------------------------------------------------------------
#      function to calculate river flow using rating curve coefficient(different a & b for Duarrigle) and river level #
#     data
#      input: river level data file
#-------------------------------------------------------------------------------------------------------------------------------------------
Pdm_calflow <- function(fn){
 # tmp_file <- fn[substr(fn$V1,1,4) == as.character(year),]
 tmp_file <- fn
 num_row <- nrow(tmp_file)
 print(num_row)
 # 3 different value for coefficient a & b for Duarrigle
 a1 <- c(8.857,22.6327,30.9276)
 b1 <- c(1.25151,1.83451,1.16186)
 # fn_out_n <- paste("Dur_flow_",as.character(year),".dat",sep = '')
 fn_out_n <- paste("Dur_flow_nf",".dat",sep = '')
 fn_out <- data.frame (Date = character(num_row), Time = character(num_row) ,flow = double(num_row) )
 tmp_file$V1 <- as.POSIXct(tmp_file$V1)
 tmp_file$V2 <- as.numeric(tmp_file$V2)
 tmp_file$V3 <- as.numeric(tmp_file$V3)
 fn_out$Date <- format( tmp_file$V1 ,format = "%Y/%m/%d")
 fn_out$Time <- format(tmp_file$V1 ,format = "%H,%M,%S")

 fn_out <- separate(fn_out,"Date", c("Year", "mm", "dd"), sep = "/")
 fn_out <- separate(fn_out,"Time", c("hh", "mn", "ss"), sep = ",")
```

```
 for (j in 1:nrow(tmp_file)){
  tmp_lvl <- tmp_file$V2[j]
  k <- find_coff(tmp_lvl)
  fn_out$flow[j] <- round(a1[k]* (tmp_lvl ^ b1[k] ),2)
 }
 fn_out$colz <- 0

 fn_out$Year <- as.integer(fn_out$Year)
 fn_out$mm <- as.integer(fn_out$mm)
 fn_out$dd <- as.integer(fn_out$dd)

 fn_out$hh <- as.integer(fn_out$hh)
 fn_out$mn <- as.integer(fn_out$mn)
 fn_out$ss <- as.integer(fn_out$ss)
 #fn_out <- fn_out[fn_out$mn == 0 & fn_out$ss == 0,]

 write.table(fn_out,fn_out_n,sep = ", ",col.names = FALSE, row.names = FALSE)
 return(fn_out)
}

#-----------------------------------------------------------------------------------------------------------------------------------
#   apply gap filling and outliers correction and create PDM format flow data for Killavullen
#   write the PDM format data in a Kil_flow_1.dat
#-----------------------------------------------------------------------------------------------------------------------------------
fgap_kil_flow_10_17 <- average_gap_filling(my_data_opw,kil_flow_2010_2017)
foutl_kil_flow_10_17 <- average_outler_filling(fgap_kil_flow_10_17,450)
kil_flow_10_17 <- subset(foutl_kil_flow_10_17[!duplicated(foutl_kil_flow_10_17$X1),])
kil_flow1017 <- Pdm_flow_format(kil_flow_10_17)

#-----------------------------------------------------------------------------------------------------------------------------------
#       flow data for Mallow from river level data
#       write the PDM format in Mal_flow_10_17.dat
#-----------------------------------------------------------------------------------------------------------------------------------
fgap_mal_lvl_10_17 <- average_gap_filling(my_data_opw,mal_lvl_2010_2017)
foutl_mal_lvl_10_17 <- average_outler_filling(fgap_mal_lvl_10_17,4.5)

mal_lvl_10_17 <- subset(foutl_mal_lvl_10_17[!duplicated(foutl_mal_lvl_10_17$X1),])
nrow(durv3_lvl_10_13)
mal_flow1017 <- Pdm_calflow_1coef (mal_lvl_10_17,27.266,1.967)
index <- with(mal_flow1017, order(Year,mm,dd,hh,mn,ss))
flow_sort <- mal_flow1017[index,]
write.table(flow_sort,'Mal_flow_10_17.dat',sep = ", ",col.names = FALSE, row.names = FALSE)

#-----------------------------------------------------------------------------------------------------------------------------------
#       flow data for DromCummer from river level data
#       write the PDM format in Drom_flow_10_17.dat
#-----------------------------------------------------------------------------------------------------------------------------------
fgapdr_lvl_10_17 <- average_gap_filling(my_data_opw,dro_lvl_2010_2017)
foutl_dr_lvl_10_17 <- average_outler_filling(fgapdr_lvl_10_17,3.)

dro_lvl_10_17 <- subset(foutl_dr_lvl_10_17[!duplicated(foutl_dr_lvl_10_17$X1),])
nrow(durv3_lvl_10_13)
dro_flow1017 <- Pdm_calflow_1coef(dro_lvl_10_17,48.419,1.703)
index <- with(dro_flow1017, order(Year,mm,dd,hh,mn,ss))
Dro_flow_sort <- dro_flow1017[index,]
write.table(Dro_flow_sort,'Drom_flow_10_17.dat',sep = ", ",col.names = FALSE, row.names = FALSE)
```

```
#-------------------------------------------------------------------------------------------------------------------------------------------------------
#       flow data for Duarrigle from river level data
#       write the PDM format in Dur_flow_10_13.dat
#-------------------------------------------------------------------------------------------------------------------------------------------------------
fgapd_lvl_10_13 <- average_gap_filling(dur_lvl_2010_2013)
foutl_d_lvl_10_13 <- average_outler_filling(fgapd_lvl_10_13,3.8)

durv2_lvl_10_13 <- subset(foutl_d_lvl_10_13[!duplicated(foutl_d_lvl_10_13$V1),])
nrow(durv3_lvl_10_13)
dur_flow1013 <- Pdm_calflow(durv2_lvl_10_13)
index <- with(dur_flow1013, order(Year,mm,dd,hh,mn,ss))
flow_sort <- dur_flow1013[index,]
write.table(flow_sort,'Dur_flow_10_13.dat',sep = '', '',col.names = FALSE, row.names = FALSE)

#-------------------------------------------------------------------------------------------------------------------------------------------------------
```

B.3 R Code for plot River level/ River flow data:

```
#-------------------------------------------------------------------------------------------------------------------------------------------------------
#    plot to compare observed river flow at different subcatchment in each year                                                                        #
input: year                                                                                                                                          #--
-------------------------------------------------------------------------------------------------------------------------------------------------------

pl_flow <- function(year){
  nx <- nrow(mal_flow1017[mal_flow1017$Year== year,])
  p1 <- ggplot() +
    #geom_line(data = dur_flow1013[dur_flow1013$Year== year,] , aes(x=1:nx, y = flow, colour= 'Duarrigle')) +
    geom_line(data = dro_flow1017[dro_flow1017$Year== year,] , aes(x=1:nx, y = flow, colour= 'Dromcummer'))
+
    geom_line(data = mal_flow1017[mal_flow1017$Year == year,], aes(x=1:nx, y= flow, colour='Mallow Rail BR'))
+
    geom_line(data = kil_flow1017[kil_flow1017$Year== year,], aes(x=1:nx, y= flow, colour='Killavulen')) +
    xlab(paste("15-minute time step", year)) + ylab("River Flow- m^3/s") +
    scale_x_continuous(breaks = c(0,2920,5840,8760,11680,14600,17520,20440,23360,26280,29200,32120),
                       labels = c('Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','Sep','Oct','Nov','Dec') ) +
    scale_colour_manual("Stations :",
                        values = c(  "Dromcummer" = "blue", "Mallow Rail BR" = "green", "Killavulen" = "red")) +
    theme(legend.position = "top")

  return(p1)
}
pl_flow(2014)

#-------------------------------------------------------------------------------------------------------------------------------------------------------
#   plot time series of river level from 2010 to 2017 at Dromcummer
#-------------------------------------------------------------------------------------------------------------------------------------------------------
date_l <- c('Jan 2010','Jan 2011','Jan 2012','Jan 2013','Jan 2014','Jan 2015','Jan 2016','Jan 2017')
break_l <- 364 * 24 * 4 * c(0:7)
plot_lvl <- ggplot(data = dro_flow1017,aes(x = 1:nrow(dro_flow1017),dro_flow1017$flow)) +
  geom_bar(sta = "Identity",width = 0.1, color = "gray")
plot_lvl + xlab("Date") + ylab("River Flow- m^3/s") +
  scale_x_continuous(breaks = break_l, labels = date_l )

plot_lvl <- ggplot(data = dro_lvl_10_17,aes(x = 1:nrow(dro_lvl_10_17),dro_lvl_10_17$X2)) +
  geom_bar(sta = "Identity",width = 0.1, color = "gray")
plot_lvl + xlab("Date") + ylab("River Level- m") +
  scale_x_continuous(breaks = break_l, labels = date_l )

#-------------------------------------------------------------------------------------------------------------------------------------------------------
#    plot to compare  observed flow for an event at 3 subcatchment
#    input: event start time(year,month, day and hour) and event duration
#-------------------------------------------------------------------------------------------------------------------------------------------------------
pl_flow_event_r <- function(year,m,d,h,dur){
  ind <- which(mal_flow1017$Year == year & mal_flow1017$mm == m & mal_flow1017$dd == d &
mal_flow1017$hh == h & mal_flow1017$mn == 15)
  ind_end <- ind + (dur * 24 *4)
  nx <- ind_end - ind +1
  print (nx)
  l24 <- as.Date( paste(mal_flow1017[ind,]$Year,mal_flow1017[ind,]$mm,mal_flow1017[ind,]$dd , sep='-'))
```

```
  u24 <-
as.Date(paste(mal_flow1017[ind_end,]$Year,mal_flow1017[ind_end,]$mm,mal_flow1017[ind_end,]$dd,
sep='-'))
  date_l <- as.Date(l24: u24,origin = '1970-01-01')
  print (date_l)
  ll <- length(date_l) -1
  break_l <- 100* c(0:ll)
  print(break_l)
  p1 <- ggplot() +
    geom_line(data = dro_flow1017[ind:ind_end,] , aes(x=1:nx, y = flow, colour= 'Dromcummer')) +
    geom_line(data = mal_flow1017[ind:ind_end,], aes(x=1:nx, y= flow, colour='Mallow Rail BR')) +
    geom_line(data = kil_flow1017[ind:ind_end,], aes(x=1:nx, y= flow, colour='Killavulen')) +
    xlab(paste("15-minute time step",l24, 'to',u24)) + ylab("River Flow- m^3/s") +
    scale_x_continuous(breaks = break_l, labels = paste(date_l,h,':00') ) +
    scale_colour_manual("Stations :",
                        values = c(  "Dromcummer" = "blue", "Mallow Rail BR" = "green", "Killavulen" = "red")) +
    theme(legend.position = "top")

  return(p1)
}
pl_flow_event_r(2012,8,15,10,2)



#-------------------------------------------------------------------------------------------------------------------------------------------
#    two functions to find the largest event in a year or spesific time
#    input: filename, year or  event starttime and duration
#-------------------------------------------------------------------------------------------------------------------------------------------
flow_event_time <- function(fn,year,m,d,h,dur){
  ind <- which(mal_flow1017$Year == year & mal_flow1017$mm == m & mal_flow1017$dd == d &
mal_flow1017$hh == h & mal_flow1017$mn == 15)
  ind_end <- ind + (dur * 24 *4)
  l_temp <- fn[ind:ind_end,]
  print (l_temp[which.max(l_temp$flow),])
}

larg_event <- function(fn,year){
  l_temp  <- fn[fn$Year == year,]
  print (l_temp[which.max (l_temp$flow),])
}
larg_event(dro_flow1017,2012)



#-------------------------------------------------------------------------------------------------------------------------------------------
#  Climatic condition growth of rainfall by 10,15,20,0030 % in 2014
#  import PDM output when rainfall increase by 10,15,.. 30 % for each subcatchment for spesific year e.g 2014
#  make a list of each subcatchment (e.g afd is a list of Dromcummer PDM output when rainfall increases by %)
#-------------------------------------------------------------------------------------------------------------------------------------------
afd <- list(Dro14_1,Dro14_11, Dro14_15 , Dro14_20,Dro14_25, Dro14_30)
afm <- list(Mal14_1,Mal14_11, Mal14_15 , Mal14_20,Mal14_25, Mal14_30)
afk <- list(kil14_1,kil14_11, kil14_15 , kil14_20,kil14_25, kil14_30)
```

```
#-----------------------------------------------------------------------------------------------------------------------------------------
# function to generate an annual plot for comparing ratio of computed flow
# for different amount of rainfall(10,15,..30%) by computed flow with actual rainfall
# input: filename(e.g afd,afm,afk) and year
#-----------------------------------------------------------------------------------------------------------------------------------------
plRatio_flow <- function(fn,year){
  nx <- nrow(fn[[1]])
  p1 <- ggplot() +
    #geom_line(data = dur_flow1013[dur_flow1013$Year== year,] , aes(x=1:nx, y = flow, colour= 'Duarrigle')) +
    geom_line(data = fn[[1]] , aes(x=1:nx, y = fn[[2]]$computed/fn[[1]]$computed, colour= '10%')) +
    geom_line(data = fn[[1]] , aes(x=1:nx, y = fn[[3]]$computed/fn[[1]]$computed, colour= '15%')) +
    geom_line(data = fn[[1]] , aes(x=1:nx, y = fn[[4]]$computed/fn[[1]]$computed, colour= '20%')) +
    geom_line(data = fn[[1]] , aes(x=1:nx, y = fn[[5]]$computed/fn[[1]]$computed, colour= '25%')) +
    geom_line(data = fn[[1]] , aes(x=1:nx, y = fn[[6]]$computed/fn[[1]]$computed, colour= '30%')) +
    scale_x_continuous(breaks = c(0,2920,5840,8760,11680,14600,17520,20440,23360,26280,29200,32120),
                       labels = c('Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','Sep','Oct','Nov','Dec') ) +
    xlab(paste("15-minute time step", year)) + ylab("Ratio of new flow by old flow") +
    scale_colour_manual("increase rainfall by :",
                        values = c(  "10%" = "blue", "15%" ="red","20%" ="green","25%" ="black" , "30%" ="orange" )) +
    theme(legend.position = "top")

  return(p1)
}
plRatio_flow(afd,2014)
plRatio_flow(afm,2014)
plRatio_flow(afk,2014)

#-----------------------------------------------------------------------------------------------------------------------------------------
#   calculate number of hours
#   input: number of month and day
#-----------------------------------------------------------------------------------------------------------------------------------------
n_day <- c(31 ,28 ,31,30 ,31 ,30,31,31,30,31,30,31)
n_hours <- function(mn,day){return( (sum(n_day[1:mn]) + day) * 24 )}
n_hours(12,30)

#-----------------------------------------------------------------------------------------------------------------------------------------
#function to generate a plot for comparing computed flow for different amount of rainfall(10,15,..30%)
# input: filename(e.g afd,afm,afk) and start and end time(hours)
#-----------------------------------------------------------------------------------------------------------------------------------------

pl_flow_between <- function(fn,l,u){
  fn1 <- fn[[1]][fn[[1]]$time.hrs.>l & fn[[1]]$time.hrs.< u,]
  nx <- nrow(fn1)
  l24 <- l/24 +1
  u24 <- u/24 +1
  date_l <- as.Date(l24: u24,origin = '2014-01-01')
  print (date_l)
  ll <- length(date_l) -1
  break_l <- 100* c(0:ll)
  nx <- nrow(fn1)
  p1 <- ggplot() +
    #geom_line(data = fn1 , aes(x=1:nx, y = fn1$observed, colour= 'Obs_flow')) +
    # geom_line(data = fn1 , aes(x=1:nx, y = fn1$computed, colour= 'computed flow for acual rainfal')) +
```

```
   geom_line(data = fn1 , aes(x=1:nx, y = fn[[2]][fn[[2]]$time.hrs.> l & fn[[2]]$time.hrs.< u,]$computed, colour=
'10%')) +
   geom_line(data = fn1 , aes(x=1:nx, y = fn[[3]][fn[[3]]$time.hrs.> l & fn[[3]]$time.hrs.< u,]$computed, colour=
'15%')) +
   geom_line(data = fn1 , aes(x=1:nx, y = fn[[4]][fn[[4]]$time.hrs.> l & fn[[4]]$time.hrs.< u,]$computed, colour=
'20%')) +
   geom_line(data = fn1 , aes(x=1:nx, y = fn[[5]][fn[[5]]$time.hrs.> l & fn[[5]]$time.hrs.< u,]$computed, colour=
'25%')) +
   geom_line(data = fn1 , aes(x=1:nx, y = fn[[6]][fn[[6]]$time.hrs.> l & fn[[6]]$time.hrs.< u,]$computed, colour=
'30%')) +
   scale_x_continuous(breaks = break_l, labels = date_l ) +
   xlab(paste("15-minute time step", 2014)) + ylab(" flow  m^3/s") +
   scale_colour_manual("increase rainfall by :", values =
     c("Obs_flow" ="black", 'computed flow for acual rainfal'= "green","10%" = "blue", "15%" ="red","20%"
="orange" , "25%" ="yellow", "30%" ="purple")) +
   theme(legend.position = "top")
  return(p1)
}
pl_flow_between(afd,790,830)

#-----------------------------------------------------------------------------------------------------------------------------------------
# function to generate an event plot for camparing ratio of computedflow
# for different amount of rainfall(10,15,..30%) by computed flow with actual rainfall
# input: filename(e.g afd,afm,afk) and start and end time(hours)
#-----------------------------------------------------------------------------------------------------------------------------------------
plRatio_flow_between <- function(fn,l,u){
  fn1 <- fn[[1]][fn[[1]]$time.hrs.>l & fn[[1]]$time.hrs.< u,]
  nx <- nrow(fn1)
  l24 <- l/24 +1
  u24 <- u/24 +1
  date_l <- as.Date(l24: u24,origin = '2014-01-01')
  ll <- length(date_l) -1
  break_l <- 100* c(0:ll)
  p1 <- ggplot() +
    geom_line(data = fn1 , aes(x=1:nx, y = fn[[2]][fn[[2]]$time.hrs.> l &
   fn[[2]]$time.hrs.< u,]$computed/fn1$computed, colour= '10%')) +
     geom_line(data = fn1 , aes(x=1:nx, y = fn[[3]][fn[[3]]$time.hrs.> l
    & fn[[3]]$time.hrs.< u,]$computed/fn1$computed, colour= '15%')) +
     geom_line(data = fn1 , aes(x=1:nx, y = fn[[4]][fn[[4]]$time.hrs.> l
    & fn[[4]]$time.hrs.< u,]$computed/fn1$computed, colour= '20%')) +
    geom_line(data = fn1 , aes(x=1:nx, y = fn[[5]][fn[[5]]$time.hrs.> l
   & fn[[5]]$time.hrs.< u,]$computed/fn1$computed, colour= '25%')) +
    geom_line(data = fn1 , aes(x=1:nx, y = fn[[6]][fn[[6]]$time.hrs.> l
    & fn[[6]]$time.hrs.< u,]$computed/fn1$computed, colour= '30%')) +
    scale_x_continuous(breaks = break_l, labels = date_l ) +
    xlab(paste("15-minute time step", 2014)) + ylab("Ratio of newflow / oldflow") +
    scale_colour_manual("increase rainfall by :",
                        values = c("10%" = "blue", "15%" ="red", "20%" ="orange" , "25%" ="yellow", "30%" ="purple")) +
    theme(legend.position = "top")

  return(p1)
}
plRatio_flow_between(afd,790,830)
```

```
#-------------------------------------------------------------------------------------------------------------------------------------------------------
#function to create a list of max or min or sum of computed flow with different rainfall(10,15,...30 %)
#input: list of PDM output file (afd,afm or afk) and operation(max,min or sum
#-------------------------------------------------------------------------------------------------------------------------------------------------------
df_sub <- function(afn,f){
  max_f <- which.max(afn[[1]]$observed)
  min_f <- which.min(afn[[1]]$observed)
  for ( i in 2:6)
    test_dro[i-1] <- switch(f,
                            max = 100 *(afn[[i]][max_f,]$computed / afn[[1]][max_f,]$computed ) -100,
                            sum = 100 * (sum(afn[[i]]$computed)/sum(afn[[1]]$computed)) -100,
                            min = 100 *(afn[[i]][min_f,]$computed / afn[[1]][min_f,]$computed ) -100
    )
  df <- data.frame(xl,test_dro)
  return(df)
}
mfd <- df_sub(afd,"min")
mfm <- df_sub(afm,"min")
mfk <- df_sub(afk,"min")

dfd <- df_sub(afd,"max")
dfm <- df_sub(afm,"max")
dfk <- df_sub(afk,"max")

sfd <- df_sub(afd,"sum")
sfm <- df_sub(afm,"sum")
sfk <- df_sub(afk,"sum")

#-------------------------------------------------------------------------------------------------------------------------------------------------------
# function to plot the relation between growth of rainfall withgrowth of (max, min or sum)of computed flow
#input : list of (max,min or sum of computedflow in different subcatchment) and operation type(max,min,sum)
#-------------------------------------------------------------------------------------------------------------------------------------------------------
pl_fun <- function(fn1,fn2,fn3,type){
  type_y <- switch(type,
                   sum = " (Total of new flow/ Total old flow) %",
                   max = "Event maximum flow growth %",
                   min = "Event minimum flow growth %"
  )
  pl <- ggplot() +
    geom_line(data = fn1, aes(x = xl, y = test_dro,colour= 'Dromcummer' )) +
    geom_point(data = fn1, aes(x = xl, y = test_dro)) +
    geom_line(data = fn2, aes(x = xl, y = test_dro , colour = 'Mallow')) +
    geom_point(data = fn2, aes(x = xl, y = test_dro)) +
    geom_line(data = fn3, aes(x = xl, y = test_dro , colour = 'Killavulen')) +
    geom_point(data = fn3, aes(x = xl, y = test_dro)) +
    xlab ("Rainfall grpwth %") +  ylab (type_y)+xlim(10,30) +
    scale_colour_manual("Subcatchment  :",
                        values = c(  "Dromcummer" = "blue", "Mallow" ="red" , "Killavulen" = "green")) +
    theme(legend.position = "top")
  return(pl)
}

pl_fun(dfd,dfm,dfk,"max")
pl_fun(mfd,mfm,mfk,"min")
```

```
#-------------------------------------------------------------------------------------------------------------------------------------------------------
# fit the linear model to event maximum flow growth
# input: filename (e.g dfd) list of maximum computed flow when rain fall growth by 10,...30 %
#-------------------------------------------------------------------------------------------------------------------------------------------------------
pl_linear_1 <- function(fn1){
  type_y <- " function of flow %"
  linmodel <- lm(test_dro~ xl, data=fn1)
  print(linmodel)
  print(summary(linmodel))
  pred_fn1 <- data.frame(fn1_pred = predict(linmodel, fn1), xl= fn1$xl)
  pl <- ggplot() +
    geom_point(data = fn1, aes(x = xl, y = test_dro, color = 'flow growth')) +
    geom_line (data = pred_fn1,aes(x = xl, y = fn1_pred)) +
    xlab ("Rainfall grpwth %") +  ylab (type_y)+xlim(10,30) +
  return(pl)
}
pl_linear(dfd)
```

B.4 R Code rainfall data, Kmeans clustering, gap filling:

```
setwd("D:/Blackwater15min/")                # set working directory contains all rainfall data from OPW
list.filenames <- list.files(pattern=".csv$")     # create a list of  raingauge's namefiles (32 file from OPW)

#-------------------------------------------------------------------------------------------------------------------------------------------------------
#   function to convert OPW rainfall data to PDM format from start time
#   input:  filename and start time
#-------------------------------------------------------------------------------------------------------------------------------------------------------
Pdm_fileformat <- function(fn,start_time){
 fn$Date <- format(as.POSIXct(fn$Date,"%d/%m/%Y",tz="") ,format = "%Y/%m/%d")
 tmp_file <- subset(fn,fn$Date >= start_time)
 num_row <- nrow(tmp_file)
 print(num_row)
 fn_out <- data.frame ( Date = character(num_row),Time =character(num_row) ,Iprec = double(num_row) )
 fn_out$Date <- tmp_file$Date
 fn_out$Time <- as.character(tmp_file$Time)
 fn_out <- separate(fn_out,"Date", c("Year", "mm", "dd"), sep = "/")
 fn_out <- separate(fn_out,"Time", c("hh", "mn", "ss"), sep = ":")
 fn_out$Iprec <- as.numeric(as.character(tmp_file$Value..mm.))
 fn_out$colz <- 0
 fn_out$Year <- as.integer(fn_out$Year)
 fn_out$mm <- as.integer(fn_out$mm)
 fn_out$dd <- as.integer(fn_out$dd)
 fn_out$hh <- as.integer(fn_out$hh)
 fn_out$mn <- as.integer(fn_out$mn)
 fn_out$ss <- as.integer(fn_out$ss)
 return (fn_out)
}



#-------------------------------------------------------------------------------------------------------------------------------------------------------
#   read each rainfall data convert to PDM format by call the above function
#   input: file name , start and end time (e.g 2010 to 2017)
#   output: list of PDM rainfall format
#-------------------------------------------------------------------------------------------------------------------------------------------------------
Read_fun <- function(lfile,start_time){
 ldata <-list()
 for (i in 1:length((list.filenames)))
 {
  test_file <-read.csv(file = list.filenames[i],skip = 15)[ ,c('Date', 'Time', 'Value..mm.', 'State.of.value')]
  ldata[[i]] <- Pdm_fileformat(test_file,i,start_time)
  rm(test_file)
 }
 return(ldata)
}

list.data <- Read_fun(list.filenames,start_time)
```

```
#-----------------------------------------------------------------------------------------------------------------------------------------
#  function to generate the summary of number of missing and outliers of each  #  rain gauges' data.
#   this function can be used before and after gap filling  methods
#-----------------------------------------------------------------------------------------------------------------------------------------
summary_2fn <- function(list_fn,mrain){
  for (i in 1:length(list_fn)){
    gap <- total_row - nrow(list_fn[[i]])
    missing <- nrow(list_fn[[i]][!complete.cases(list_fn[[i]]),])
    outlier5 <-  nrow(list_fn[[i]][list_fn[[i]]$Iprec > mrain & complete.cases(list_fn[[i]]),])
    percent <- round((gap + missing + outlier5) * 100 / total_row,2)
    tprint <- paste( i, '  ', gap , '  ', missing ,'   ', outlier5,'   ',percent  )
    print (tprint)

  }
}
Summary_2fn(list.data)      # summary before gap filling
summary_2fn(glist.data)    # summary after gap filling

#-----------------------------------------------------------------------------------------------------------------------------------------
# function to generate summary of missing value in each rain gauge for each year
#  this function can be applied before and after gap filling
#-----------------------------------------------------------------------------------------------------------------------------------------
miss_per_year <- function(gdata,stn){
  list.year <- c("2010","2011","2012","2013","2014","2015","2016","2017")
  for (year in list.year){
    print(paste(year, nrow(gdata[[stn]][!complete.cases(gdata[[stn]]) & gdata[[stn]]$Year == year  ,])))
  }
}


#-----------------------------------------------------------------------------------------------------------------------------------------
#  function to gap filling between two raingauges and outliers greater than a threshold in this case 15 mm
# input: 2 rainfall data and threshold
# ----------------------------------------------------------------------------------------------------------------------------------------
fill_gapping <- function (fn1,fn2,tr){
  tests <- merge(fn1,fn2,by = c("Year","mm","dd","hh","mn","ss"),all = TRUE)
  #****** sort testf by date and time ***********
  index <- with(tests, order(Year,mm,dd,hh,mn,ss))
  testf <- tests[index,]
  testf$Iprec.x[is.na(testf$Iprec.x)] <- testf$Iprec.y[is.na(testf$Iprec.x) ]
  testf$Iprec.y[is.na(testf$Iprec.y)] <- testf$Iprec.x[is.na(testf$Iprec.y)]
  testf$Iprec.x[testf$Iprec.x > tr & !is.na(testf$Iprec.x) ] <- testf$Iprec.y[testf$Iprec.x > tr & !is.na(testf$Iprec.x)
& !is.na(testf$Iprec.y)]
  testf$Iprec.y[testf$Iprec.y > tr & !is.na(testf$Iprec.y) ] <- testf$Iprec.x[testf$Iprec.y > tr & !is.na(testf$Iprec.y)
& !is.na(testf$Iprec.x)]
  fn1$Iprec <- testf$Iprec.x
  fn1$colz <- 0
  fn2$Iprec <- testf$Iprec.y
  fn2$colz <- 0
  return(list(fn1,fn2))
}
```

```
#-------------------------------------------------------------------------------------------------------------------------------------------------------
# Gap filling between members of a cluster
#input: list of name of  rainfall  data in a cluster or group
#input: list of rainfall data in a cluster and thershold
#-------------------------------------------------------------------------------------------------------------------------------------------------------
group_fill_gaping <- function(Glist,glist.data,tr){
 len <-length(Glist) -1
 for (t in 1: len){
  len1 <- length(Glist) - t
  for (i in 1:len1){
   ii <- i+ 1
   for (j in ii: length(Glist)){
    # print(paste('i:', i, 'j :',j,'807',Glist[i],'807',Glist[j]))
    test1 <- fill_gapping(glist.data[[Glist[i]]],glist.data[[Glist[j]]],tr)
    glist.data[[Glist[i] ]] <- test1[[1]]
    glist.data[[Glist[j] ]] <- test1[[2]]
   }
  }
 }
 return(glist.data)
}

#-------------------------------------------------------------------------------------------------------------------------------------------------------
#  run group gap filling  after Kmeans clustering
#-------------------------------------------------------------------------------------------------------------------------------------------------------
gMnlist <-list( c(1,17,18,27,28), c(2,4,5,13),c(6,7,14,19,22),c(11,12,16),c(15,21,32),c(8,10,20))
gMnlist.data <- list.data

for(k in 1:length(gMnlist.data))
{
 tmp_file <- merge(my_data,gMnlist.data[[k]],by = c("Year","mm","dd","hh","mn","ss") , all=TRUE)
 index <- with(tmp_file, order(Year,mm,dd,hh,mn,ss))       # sort testf by date and time
 gMnlist.data[[k]] <- tmp_file[index,]
}
for (k in 1:length(gMnlist))
{
 tr = 20
 gMnlist.data <- group_fill_gaping(gMnlist[[k]],gMnlist.data,tr)
}

#-------------------------------------------------------------------------------------------------------------------------------------------------------
# kmaeans clustring forall stations except those with high number of missing or outliers
#-------------------------------------------------------------------------------------------------------------------------------------------------------
del_stations <- c("80731","80730","80729", "80725","80709","80724","80703","80723","80726")
kannualene <- subset(annual_2010[ !(annual_2010$Stn_No_ %in% del_stations ), ])
# kmeans with 3 features easting, Northing, elevation  to find the nearest neighbors for each stations
set.seed(18)
clene <- kmeans(kannualene[,c(6,7,8)],6,nstart = 15)
#str(clrain)    create csv file adding cluster column
kannualene$cluster <- 0
for (i in 1:10){
 print(kannualene$Stn_No_[clene$cluster == i] )
 kannualene$cluster[clene$cluster == i] <- i
}
setwd("D:/PDMProg/BW_15rainfall/")
write.table(kannualene,"kannualene.csv",sep = ", ",col.names = FALSE, row.names = FALSE)
```

```
#-----------------------------------------------------------------------------------------------------------------------------------
# Elbow Method for finding the optimal number of clusters
#-----------------------------------------------------------------------------------------------------------------------------------
set.seed(123)
# Compute and plot wss for k = 2 to k = 15.
k.max <- 12
data <- kannual[,c(13)]
wss <- sapply(1:k.max,
          function(k){kmeans(data, k, nstart=50,iter.max = 12 )$tot.withinss})
wss
plot(1:k.max, wss,
    type="b", pch = 19, frame = FALSE,
    xlab="Number of clusters K",
    ylab="Total within-clusters sum of squares")

#-----------------------------------------------------------------------------------------------------------------------------------
#  kmeans clustering using monthly rainfall (12 months)
#-----------------------------------------------------------------------------------------------------------------------------------
del_stations <- c("80731","80730","80729", "80725","80709","80724","80703","80723","80726")
kMn_rain <- subset(Mn_rain[ !(Mn_rain$Stn_No_ %in% del_stations ), ])
set.seed(50)
clMn_rain <- kmeans(kMn_rain[,13:24],6,nstart = 50)
#create csv outout file adding cluster column
kMn_rain$cluster <- 0
for (i in 1:12){
  print(kMn_rain$Stn_No_[clMn_rain$cluster == i] )
  kMn_rain$cluster[clMn_rain$cluster == i] <- i
}
write.table(kMn_rain,"kMn_rain.csv",sep = ", ",col.names = FALSE, row.names = FALSE)

#-----------------------------------------------------------------------------------------------------------------------------------
#  function to plot the Kmeans clustering result
#-----------------------------------------------------------------------------------------------------------------------------------
kmeans_plot <- function(fn,clname,varname,f){
  # y <- colnames(fn)[colnum]
  ggplot(fn, aes(clname$cluster,varname, group = as.factor(clname$cluster), color =as.factor(
clname$cluster),label = Stn_No_)) + geom_text(size = 3)+
    labs(x = "Cluster", y = paste("Rainfall in",f)) + scale_colour_discrete(name  ="Clusters")
}

kmeans_plot(kannual,clrain,kannual$total_rain, "year 2010")
i = 12
kmeans_plot(kMn_rain,clMn_rain,kMn_rain[,12+i],month.abb[i])


#-----------------------------------------------------------------------------------------------------------------------------------
#    summary after fill gaping
#-----------------------------------------------------------------------------------------------------------------------------------
miss_per_year <- function(gdata,stn){
  list.year <- c("2010","2011","2012","2013","2014","2015","2016","2017")
  for (year in list.year){
    print(paste(year, nrow(gdata[[stn]][!complete.cases(gdata[[stn]]) & gdata[[stn]]$Year == year  ,])))
  }
}
miss_per_year(gMnlist.data,7)
miss_per_year(list.data,7)
```

```
#-------------------------------------------------------------------------------------------------------------------------------------------
#  calculate annual rainfall for a given year for each rainguges after gap filling
# gMNlist.data (list of clean rainfall files from OPW
# ------------------------------------------------------------------------------------------------------------------------------------------
nx <- nrow(gMnlist.data[[1]][gMnlist.data[[1]]$Year== "2010",])
ann_rain <- data.frame()
for (i in 1:32)
{
 for( j in 1:12){
   ann_rain[i,j] <- sum(gMnlist.data[[i]][gMnlist.data[[i]]$Year == 2010 & gMnlist.data[[i]]$mm == j ,]$Iprec)
   if (is.na(ann_rain[i,j]))
     ann_rain[i,j] <- 0
 }
 write.table(ann_rain,"ann_rain_10.csv",sep = ",",col.names = FALSE, row.names = FALSE)
}
ann_rain <- data.frame()
ann_rain <- list()
for (i in 1:32)
{
 ann_rain[i] <- sum(gMnlist.data[[i]][gMnlist.data[[i]]$Year == 2010 ,]$Iprec)
 if (is.na(ann_rain[i]))
   ann_rain[i] <- 0
 write.table(ann_rain,"ann_rain_10.csv",sep = "\n",col.names = FALSE, row.names = FALSE)
}

#-------------------------------------------------------------------------------------------------------------------------------------------
# test group fill gaping
#-------------------------------------------------------------------------------------------------------------------------------------------
cummulat_plot <- function(cl_n,year){
 p1 <- ggplot()
 nx <- length(gMnlist[[cl_n]])
 for (i in gMnlist[[cl_n]]){
   p1 <- p1 + geom_line(data = gMnlist.data[[i]][gMnlist.data[[i]]$Year== year,], aes(x=1:nx,
y=cumsum(Iprec)),size =1.3)
 }
 p1 <- p1 + xlab("Time -15 minutes") + ylab("Cumulative rainfall- mm") +
   ggtitle(paste("Cumulative rainfall- clean data reduce missing value", year))
 # + scale_colour_manual("Raingauge",
 #    values = c("1" = "blue", "2" = "green", "3" = "red", "4" = "black","5" = "yellow"))
 return(p1)
}

#-------------------------------------------------------------------------------------------------------------------------------------------
# find sequence of rainfall > 4
#-------------------------------------------------------------------------------------------------------------------------------------------
seq_rain <- function(fn,num){
 len <- nrow(fn) - num
 for (start in 1:len ){
   end <- start + num -1
   tests <- fn[start:end,]
   if (nrow(tests[!is.na(tests$Iprec)&  tests$Iprec > 5, ]) == num ){
     print (start)
   }
 }
}
seq_rain(ave1,4)
```

```
#-----------------------------------------------------------------------------------------------------------------------------------
# change zero rainfall between 2 high amount of rainfall
#-----------------------------------------------------------------------------------------------------------------------------------
midel_rain <- function(fn,tr){
 len <- nrow(fn)-1
 for(i in 2:len){
  b <- i -1
  a <- i+1
  # print(i)
  if (!is.na(fn$Iprec[b]) & !is.na(fn$Iprec[a]) & !is.na(fn$Iprec[i]) )
   if( fn$Iprec[b] >= tr & fn$Iprec[i] == 0 & fn$Iprec[a] >= tr ){
    fn$Iprec[i] <- (fn$Iprec[b] + fn$Iprec[a]) /2
    print (paste(i,fn$Year[i],fn$mm[i],fn$dd[i], fn$hh[i],fn$mn[i],fn$Iprec[b], fn$Iprec[i],fn$Iprec[a]))
   }
 }
 return(fn)
}


#-----------------------------------------------------------------------------------------------------------------------------------
#  Gap filing with average of 2 nearest record
#-----------------------------------------------------------------------------------------------------------------------------------
average_gap_filling <- function(fn){
 #tmp_fn  <- merge(my_data,fn,by = "V1" , all=TRUE)
 tmp_fn <- fn
 len <- nrow(tmp_fn)-1
 if (is.na(tmp_fn$Iprec[1]) & !is.na(tmp_fn$Iprec[2]))
  tmp_fn$Iprec[1] <- tmp_fn$Iprec[2]
 for(i in 2:len){
  b <- i -1
  a <- i+1
  # print(i)
  if (is.na(tmp_fn$Iprec[i]) & !is.na(tmp_fn$Iprec[b])) {
   if (!is.na(tmp_fn$Iprec[a]))
    tmp_fn$Iprec[i] <- (tmp_fn$Iprec[b] + tmp_fn$Iprec[a]) /2
   else
    tmp_fn$Iprec[i] <- tmp_fn$Iprec[b]
   #print (paste(i,tmp_fn$V1[i],tmp_fn$V2[b], tmp_fn$V2[i],tmp_fn$V2[a]))
  }
 }
 if (is.na(tmp_fn$Iprec[len+1])){
  tmp_fn$Iprec[len+1] <- tmp_fn$Iprec[len]
 }
 return(tmp_fn)
}
```

```
#-------------------------------------------------------------------------------------------------------------------------------------------
# Compare monthly rainfall from OPW and MERA (for raingauge 80711)
#-------------------------------------------------------------------------------------------------------------------------------------------
i = 11
rain_mon <- data_frame('month' = integer(24),'rain' = numeric(24)  ,'type' = character(24) )
for (m in 1:12){
  rain_mon$month[m] <- m
  rain_mon$type[m] = 'MERA'
  rain_mon$rain[m] <- Mn_rain1[i,m+12]
  rain_mon$month[m+12] <- m
  rain_mon$type[m+12] = 'OPW'
  rain_mon$rain[m +12] <- sum(gMnlist.data[[i]][gMnlist.data[[i]]$Year == 2010 & gMnlist.data[[i]]$mm ==
m,]$Iprec)
}

p1<- ggplot(data = rain_mon, aes(x=month ,y = rain , fill = type, width = 0.8)) +
  geom_bar(  stat="identity" , position= 'dodge' ) +
  xlab("Month") + ylab("Monthly rainfall- mm")  +
  scale_x_continuous (breaks = c(1:12),
                      labels= c('Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','Sep','Oct','Nov','Dec') )  +
  scale_fill_discrete("Raingauge") +
  theme(legend.position = "top")
p1

#-------------------------------------------------------------------------------------------------------------------------------------------
#plot times series of raingauge (e.g 80711) in spesific month
#-------------------------------------------------------------------------------------------------------------------------------------------
pl_rain <- function(rn,year, mn){
  date_l <-  2* c(1:15)
  break_l <- 2  * 24 * 4 * c(1:15)
  nx <- nrow(gMnlist.data[[rn]][gMnlist.data[[rn]]$Year== year  & gMnlist.data[[rn]]$mm == mn &
gMnlist.data[[rn]]$Iprec <= 15,])
  p1 <- ggplot() +
    geom_line(data = gMnlist.data[[rn]][gMnlist.data[[rn]]$Year== year & gMnlist.data[[rn]]$mm == mn &
gMnlist.data[[rn]]$Iprec <= 15 ,] , aes(x=1:nx, y = Iprec) ) +
    xlab(paste("15-minute time step Nov", year)) + ylab("Rainfall- mm")  +
    scale_x_continuous(breaks = break_l,
                       labels = date_l )
  return(p1)
}

pl_rain(11,2010,11)
```