The potential of qualitative field observations to improve model structure adequacy

A case study of the Gulp

MSc HWM Thesis – Peter Jansson

Supervisors: Lieke Melsen & Ryan Teuling | Examiner: Tim van Emmerik

Abstract

Current modelling practices are built on predetermined model structures with calibrated parameters. A reliance on such paradigm-bound conventions may not lead to adequate models given the context of the catchment under study. This study proposes a modeler's methodology that involves collecting qualitative field observations from the studied catchment, theoretically leading to an improved model. The use of qualitative data, namely first-hand field observations and interviews with catchment inhabitants, was proposed as the methodological starting point, to inform a field-based perceptual model. The methodology was tested using a case study in the Gulp catchment, situated between the Netherlands and Belgium. Based on the developed perceptual model, eight new model structures were conceptualised. The performances of these models were compared to those of benchmark HBV models, which used either regionalised or calibrated parameters. Though the field-based perceptual model coincides with literature findings on the catchment geology, the resulting field-based models generally performed worse (NSE = -1.15 to 0.09) than the baseline models (NSE = 0.07 to 0.33) in recreating the observed catchment streamflow. Still, the field-based models represented actual, observed, catchment processes, which indeed signals the potential of the proposed methodology to be used by modellers, particularly for data-scarce catchments. Modellers could take inspiration from this study to develop models with improved adequacy using qualitative observations, provided changes in model structures take place in an integrated manner rather than through stepwise changes to a pre-existing model structure as was performed in this study.

1. Introduction

> *"Somebody from the EA, phoned me up personally, 'You are gonna flood tonight'. And I looked out the window and I said, 'No, I'm not'. He said, 'Oh, I am very sorry, you are, because our computers say this, that or the other'. And I said, 'My two sticks say, we're fine'. And we didn't. We know. We see it every day. We know […] And all they've got is computers."*
>
> *(female, 63 years) interview excerpt from McEwen et al.*(2012)

Hydrological models are used globally to predict floods and droughts, informing local stakeholders on effective water management. However, model uncertainties can affect the reliability of model output, potentially misleading stakeholders if these uncertainties are not mitigated, communicated or even identified (Melsen et al., 2018; Wagener & Gupta, 2005). Large model uncertainties can be linked to defining the model structure and its parameters (Abramowitz et al., 2006). These uncertainties can be eventually traced back to the subjective modelling decisions a modeller makes, especially during the initial stage of the modelling process, the development of the perceptual model (Beven, 2011; Melsen, 2022). Current modelling practices rely mainly on a reductionist approach of applying pre-determined model structures to a catchment, with little use of qualitative data from field visits, nor from engaging with local stakeholders (Addor & Melsen, 2019; Holländer et al., 2009; Seibert & McDonnell, 2002; Sivapalan, 2005). This study explores the benefits of including qualitative field observations to improve hydrological models, answering calls for hydrological modellers to work with more holistic perceptual models (Vogel et al., 2015).

Despite a lack of consensus among hydrologists regarding the best practices in catchment modelling (e.g. Blair & Buytaert, 2016; Shen et al., 2022), some core activities have been outlined theoretically and have become common practice (Beven, 2011). To reach a hydrological model that can simulate discharge, the modeller should first develop a perceptual model, which reflects the modeller's understanding of the catchment. The perceptual model can range from simply consisting of a set of qualitative impressions based on the modeller's experiences with different catchments (Beven, 2011) to a more complex collection of different processes, storage terms and fluxes across the biophysical continuum (Fenicia et al., 2008; Kavetski & Fenicia, 2011; Mcmillan et al., 2009). A mathematical model could then be developed, either 'from scratch' or from implementing an existing model structure, to best represent the perceptual model.

In practice, modellers choose predetermined model structures based on the legacy of its past insights and its established modelling infrastructure to model the catchment (Addor & Melsen, 2019). Stains in these past legacies are often overlooked. Applying the model structures in a different catchment could impose the assumptions made in their 'home' catchments to the catchment in question, which may not be valid in the new context (Melsen et al., 2018). Predefined model structures can hence contain uncertainties beyond what is currently quantified and can also be biased by the subjectivity of the modeller or by paradigm-bound conventions (Melsen et al., 2018). With high model performance made possible through parameter calibration, the choice of model structure tends to go unquestioned. Structural model uncertainty may then be underrepresented by common validation procedures. Resultant predictions are hence considered "false prophecies", which were made assuming that the model structure is correct in the first place (Beven, 1993).

But, once faced with a rigorous validation procedure, the false prophets are put to the sword. Holländer et al. (2009) demonstrated that, using predefined model structures to model an intensively-monitored, artificial catchment that modellers have never visited, large errors in discharge estimates are generated when compared to observational data from this catchment. Upon visiting the catchment, the modellers revised their models and estimates. Based on these results, the authors advised modellers to visit the field to gain insights, even qualitative ones, to aid their perceptual model development. To this end, this study further explores the benefits of qualitative field insights in developing perceptual and conceptual hydrological models.

Qualitative field insights that can aid a modeller's perceptual model development can be obtained from collecting qualitative field observations (QFO). Unlike quantitative data, qualitative observations can be cheaply done, and can be applied to any catchment that is reachable by the researcher, including data-scarce and ungauged basins. Such observations are highly dependent on the researcher's knowledge and experience, as well as the context of the catchment when visited. This context dependence encourages modelling issues to be treated in a case-by-case basis as opposed to finding a solution that fits for all catchments (Blöschl et al., 2013; Clark et al., 2011; van Emmerik et al., 2015). QFO can include primary observations collected first-hand, or secondary observations obtained from interviewing members of the local population. Primary observations are commonly used by field hydrologists, who look to observe catchment processes and characteristics that are relevant in developing their perceptual model (Seibert & McDonnell, 2002).

Further insights could be gained from secondary qualitative observational data, which can be provided by local stakeholders (Abu et al., 2020; Eden et al., 2016). Local sources of knowledge may provide insights on the historical behaviour of the catchment, from a timespan exceeding that of the historical data collected by the measuring gauges (Abu et al., 2020). Also, the locally-informed model may elaborate upon catchment dynamics that are more relevant to those who live there (Abu et al., 2020). Through interviews with local farmers, Massuel et al. (2018) found that their insights on hydrological processes were relevant and compatible with hydrodynamic models of the area, further enhancing the hydrologists' understanding of the area. Expert knowledge has shown to help constrain ranges in inputs and parameters, improving the realism of model output (Antonetti & Zappa, 2018). Van Emmerik et al. (2015) gained valuable insights from interviewing stakeholders on their observations and knowledge of catchment processes, revealing that their perception of catchment processes was tied to water levels at an irrigation reservoir. By including this reservoir in the perceptual model, local perceptions could be translated into adjustments in model structure and boundary conditions to improve model "realism", while also having a source of observations for model validation.

Though the use of qualitative field observations may not appear replicable nor objective, the naturalist line of inquiry would promote this approach for its flexibility in developing knowledge specific to the catchment under study (Lincoln & Guba, 1985). This paradigm is rarely considered in the field of natural sciences, including hydrology, where positivism prevails (Blöschl, 2006). Positivism assumes the existence of *one* truth, which can be found if enough data is collected and analysed the 'right' way. To this end, positivism considers objectivity as an important prerequisite for valid qualitative research (Kirk & Miller, 1986). Conversely, the naturalistic paradigm assumes the existence of multiple realities, meaning that inquiries (e.g. repeated observations) do not necessarily converge. Hence, the more qualitative field observations there are, subjective as they may be, the more versions of reality (of the catchment) will be explained, all of which can be interlinked in the perceptual model. Though a planned field campaign can attempt to improve objectivity, the researcher should prioritise having the freedom to enter the field with an open mind, being able to observe and incrementally revise their perception of the catchment, as proposed by grounded theorists (Glaser & Strauss, 1967). The researcher should then be viewed as a research instrument, whose observations should be considered as a truth, yet whose biases and subjectivity should be also considered when interpreting the results.

In this study, I aim to explore how primary and secondary qualitative field observations can lead to an evolved perceptual model, leading to differing model structures and discharge predictions. I will explore the Gulp – a free-flowing, gauged catchment – as a case study of a catchment where the integration of qualitative field observations (QFO) in catchment models may aid the model's performance. By testing the approach in a gauged basin, I will investigate how insightful this approach can be, and its potential to guide a modeller's approach to modelling gauged and ungauged basins. As a baseline, the HBV model using parameters from a globalised parameter set will be used (Beck et al., 2016). Tracking the evolution of the model structure and parameters, as I acquire more qualitative field observations will lead me to answer the following research questions:

How do models informed by qualitative field observations represent – and simulate discharge in – the Gulp, compared to an HBV model?

Sub RQ1: What catchment characteristics and hydrological processes make up the field-based perceptual model of the Gulp catchment?
Sub RQ2: How can the main catchment characteristics and processes represented in the field-based perceptual model be represented mathematically on SuperflexPy?
Sub RQ3: How do the field-based mathematical models perform in representing discharge of the Gulp, compared to the baseline models?

2. Approach

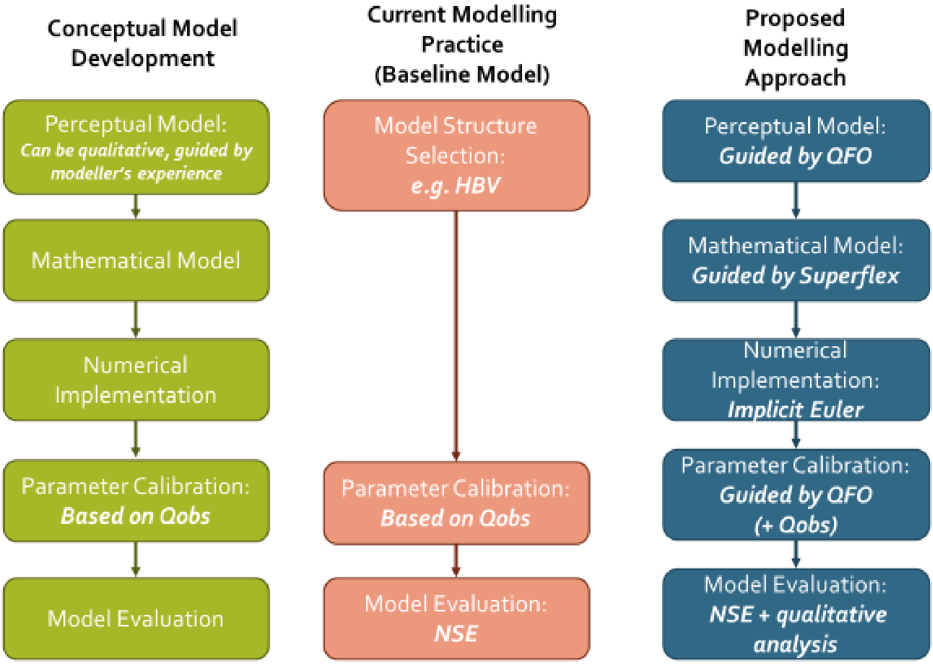

Figure 1 - Comparison of different approaches to hydrological model development. In green: stages of 'prudent' conceptual model development, as synthesised by Jansen et al. (2021). In red: common practice in model development (adapted from L. Melsen, lecture notes, April 2021). In blue: modelling approach taken in this study, incorporating QFO in the flexible modelling framework Superflex. Note that the definitions from each source may vary.

I used qualitative field observations – both primary and secondary – to develop multiple mathematical models of the Gulp. Figure 1 shows how the approach taken in this study contrasts with current modelling practices, which are generally based on choosing and calibrating a predefined model structure (Addor & Melsen, 2019). Indeed, the steps I proposed more closely align with the stages of model development agreed among hydrologists to be more prudent, since the development of the mathematical model structure (which generates discharge predictions) is based upon the perceptual model of the catchment (Beven, 2011). Using qualitative insights to better represent a certain catchment follows the naturalist paradigm (Guba, 1981), as the research and insights therefrom are assumed to be practitioner- context- and time-sensitive. To ensure that the conclusions of this case study are valid and potentially transferable to other (ungauged) catchments, the local context and scope needed to be well-defined, and my positionality as a researcher was reflected upon.

Figure 2 provides a summary of the main activities and models that will be generated as part of the study. The circled numbers refer to the section number of the proposal that describes the respective stage.

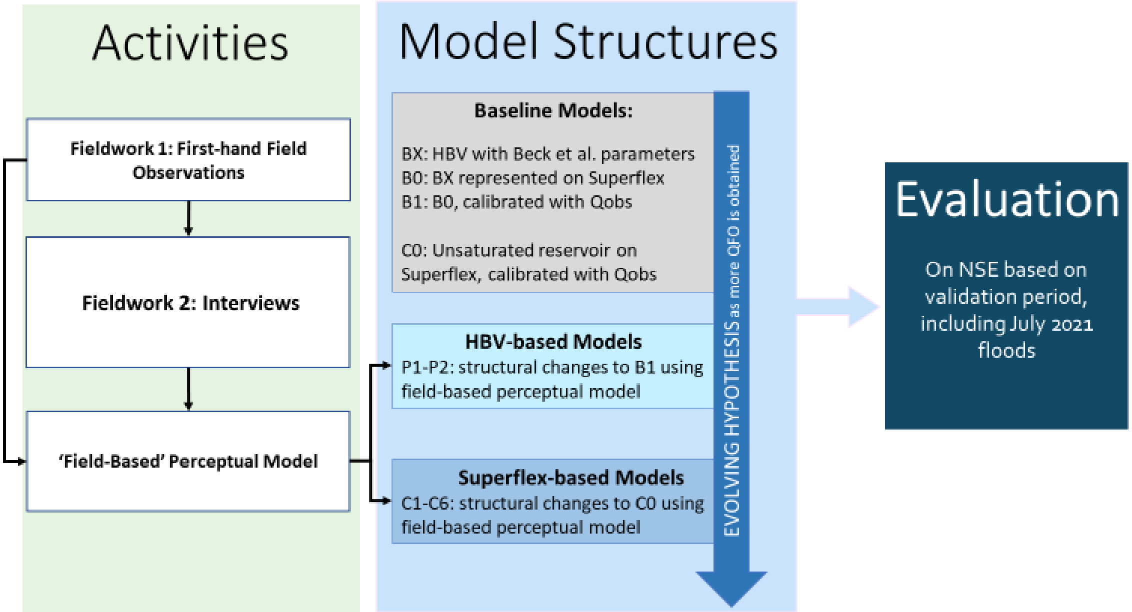

Figure 2 - Summary of activities, and model sets generated therefrom.

2.1 Case Study: The Gulp Catchment

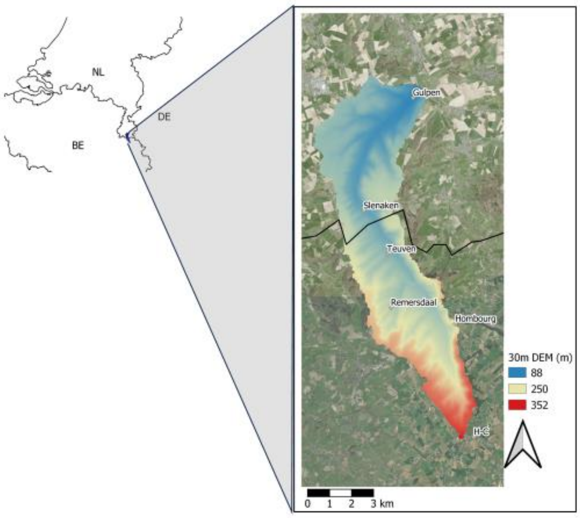

Figure 3 - Location of the Gulp catchment on the Dutch-Belgian frontier. Elevations generated by a DEM at 30m resolution.

The Gulp, a stream located between Netherlands and Belgium, is a monitored catchment that appears complex to conceptualise in hydrological modelling. Its geology has long been of interest among local water scientists due to its topography and dissected plateau landscape, which can affect groundwater and runoff processes (Nota & van de Weerd, 1988). With a size of roughly 46km^2, the catchment appeared suitably small, such that the perceptual model resulting from my brief fieldwork campaigns could be considered spatially representative of the entire catchment.

The catchment outlet has an active monitoring gauge, which has (>5 years) historical discharge data, sufficient for model calibration and validation. Other forcing data, such as precipitation and evapotranspiration, were obtained from the meteorological station in Maastricht, located 14km away from the catchment outlet. Input metadata is provided in table 1.

Table 1 - Metadata of input variables used for all models in this study. The study will use all input data at a daily timestep (mm/d)

Variable	**Source**	**Range**	**Highest Resolution**
Precipitation (0.1mmh^{-1}), Temperature (0.1°C),	Measurement Station *Maastricht (no. 380)*, from KNMI (Royal Dutch Meteorological Institute)	11/02/2018 until 09/02/2023	Hourly
Evaporation (mmh^{-1})	Hargreaves E_{Pot} based on $\text{Temperature}_{avg/min/max}$	11/02/2018 until 09/02/2023	Daily
Discharge (m^3s^{-1})	Measurement Station *Gulp Azijnfabriek*, managed by Waterschap Limburg (Water board Limburg)	11/02/2018 until 09/02/2023	Every 15 minutes

2.2 Model Development

Primary and secondary QFO were used to develop field-based model structures using the flexible modelling framework Superflex (Dal Molin et al., 2021). Using Superflex (as available on python as SuperflexPy), the models generated in this study did not have to conform to a pre-determined model structure, allowing for a high flexibility in generating mathematical models based on the various local perceptions of catchment processes (Fenicia et al., 2011).

Before implementing QFO, I first developed baseline models (figure 3). The baseline model structures acted as a starting point to which QFO-informed modifications can be made before their performances and structures can be compared. The first baseline model used (BX) was based on an HBV model structure which has been applied globally (Jansen et al., 2021; Seibert & Bergström, 2022), with parameters obtained from a regionalised global dataset developed by Beck et al. (2016). As seen in figure 6, the HBV structure used for BX consists of 4 reservoirs. The first reservoir, the snow routine, splits incoming precipitation into rainfall and snow fluxes. Water then enters the unsaturated reservoir (UR), where it either evaporates or infiltrates into the upper groundwater reservoir (FR). If storage exceeds a certain threshold (defined by parameter UZL), the excess overflow leaves the catchment as the flux Q0. Other fluxes out of FR are a baseflow flux Q1, and percolation to the lower groundwater reservoir (SR), which generates a slower baseflow flux (Q2).

This baseline model structure reflects a modeller's typical approach to modelling a data-scarce basin, using the popular technique of parameter regionalisation (Razavi & Coulibaly, 2013) on the HBV model (Bergström, 1992), a model which has been applied in catchments worldwide. To allow for a straightforward comparison of the baseline model with the QFO-informed models made on Superflex, the BX model structure was reproduced on Superflex based on elements developed by Jansen et al. (2021), yielding model B0. The parameters were then calibrated based on the Gulp discharge observations, to account for the effect of calibration on model performance, yielding model B1.

Another baseline model (C0) was developed, consisting of a single unsaturated reservoir calibrated on discharge observations, as used by Kavetski & Fenicia (2011). This baseline model represents a simple 'black box' and indicates how a model with minimal structural complexity simulates discharge when aided by parameter calibration. This baseline model would also allow for simpler field-based model structures to be developed, resulting in a more straightforward interpretation of the functions of the model's components. Model development was also simplified to only include structural elements that were readily available on the Superflex repository (version 1.3.1), which excluded the unsaturated, upper and lower groundwater reservoirs used in HBV.

A review of studies using Superflex (see appendix A) ascertained how different catchment processes have been conceptualised, which guided the modeller in conceptualising the field-based perceptual models into mathematical ones on Superflex. Most of the processes were conceptualised through including a reservoir to represent the respective storage/flux, while other conceptualisations involved adjusting the forcing data and implementing parameter boundaries. The resulting model structure consisted of a combination of these reservoirs, making the structure unique to the insights provided by the personal observations and/or interviews. The reservoirs were added incrementally, following a stepwise approach (Fenicia et al., 2008) to yield multiple "field-informed" models of different complexities (models P1,2 and C1-6).

In summary, I evaluated four different baseline models: an HBV model informed by globally regionalised parameters (BX), this same HBV model (and parameters) transferred to SuperflexPy (B0), B0 with parameters calibrated on discharge observations of the Gulp (B1), and an unsaturated reservoir with calibrated parameters (C0). Using QFO, I developed the field-based models P1 and P2 through making structural changes to B1, while the field-based models C1-C6 were based on structural changes made to C0.

2.3 Primary and Secondary QFO Collection and Analysis

I collected primary QFO on catchment characteristics over a 5-day field campaign (*Fieldwork 1* in figure 3). I covered pre-described transects of the catchment, also visiting various land cover types in the catchment. I iteratively developed field notes, resulting in a perceptual model of the catchment based on primary QFO. Following the lines of naturalistic inquiry and grounded theory, I gave myself freedom in being able to continuously refine the observation plan as findings were made and hypotheses developed during fieldwork. A field guide (appendix B) was developed to guide myself to observe as much of the catchment as possible within the limited timeframe, based on maps and satellite imagery. Observations, interpretations and hypotheses made in the field consisted of photos and field notes, which were reflected upon iteratively, synthesising the obtained primary QFO into a coherent narrative.

On the second round of fieldwork (*Fieldwork 2* in figure 3), I obtained secondary QFO through conducting semi-structured field interviews with local inhabitants. Since this study is exploratory in nature, semi-structured interviews were suitable, as they provided a balance between structured questions and improvised follow-up questions (Kallio et al., 2016). An interview guide (see appendix C) was developed with pre-defined probes, based upon hydrological processes which are known to be conceptualizable on Superflex (see appendix A). Due to their extensive experiential knowledge and their reliance on hydrological processes in their daily lives, farmers were approached for interviews (Massuel et al., 2018). Though other stakeholder groups, namely local water authorities, may have more relevant knowledge about the local hydrology, farmers form a stakeholder group that exists in catchments worldwide, and whose livelihoods are closely tied to floods and droughts, as well as from water management policies often informed by hydrological models. Focusing purely on farmers may hence yield a depth of insight which could be obtained in other catchments, increasing the transferability of this approach.

Extensive notes were taken based on the interview recordings. I analysed the interview data thematically, following Braun and Clarke (2006), yielding local inhabitants' perceptions of dominant catchment processes. Finally, as shown in the final activity in figure 3, all (primary and secondary QFO) insights were synthesised to form the field-based perceptual model (figure 3) that will aid in the development of all field-based model structures.

2.4 Model Performance and Analysis

Apart from baseline models BX and B0, both QFO-informed and baseline model structures were calibrated and validated in the same manner, as was carried out in another Superflex study by Wrede et al. (2015). 5 years of historical data was used to generate and evaluate estimates. This timespan accommodated for a 1-year 'warm-up' period (to correct the initial storage values in the model reservoirs), a 2-year calibration period and a 2-year validation period, as follows:

11/02/2018-11/02/2019: Warm-up period.

12/02/2019-12/02/2021: Calibration period.

13/02/2021-17/07/2023: Validation period.

Nash-Sutcliffe efficiency of streamflow predictions (NSE), a widely accepted measure of model performance within hydrology (Nash & Sutcliffe, 1970), was used to evaluate model performance during calibration and validation. Calibration involved using Latin Hypercube sampling (McKay et al., 1979). The number of iterations were restricted to 2000, beyond which increases in NSE were found to be minimal during preliminary runs. During calibration, the parameter space was constrained using values obtained in HBV and Superflex studies in geologically and climatologically similar catchments, as seen in appendix D (e.g. Beck et al., 2016; Kavetski & Fenicia, 2011; Seibert, 1997; Uhlenbrook et al., 1999; Wrede et al., 2015).

As the main point of comparison, the model structures were compared over the entire validation period (see *Evaluation* in figure 3), to reflect the model's performance in predicting the general discharge behaviour. The validation period includes the July 2021 floods, which were centred near the Gulp catchment, and led to severe damage in the surrounding region (Koks et al., 2021). The inclusion of this time period allows model comparisons to be made in predicting that specific flood. Comparisons can also be made during the fieldwork period (13-18/06/2023), to indicate whether the P-QFO collected during that period led to a model that was better adapted to predicting discharge in those conditions.

3. Qualitative Field Observations and Model Development

This section covers results obtained during fieldwork, data analysis, modelling and the analysis of model performances. Firstly, the results of the fieldwork rounds that collected primary and secondary QFO are presented, leading to a field-based perceptual model that summarises the catchment characteristics observed in the field which could be used to adjust the model structure. Then, the development of the baseline models based on the HBV model structure (BX, B0, B1) are described. The translation of QFO insights into adjustments in the HBV model structure are presented, leading to models P1 and P2, whose model performance compared to the baseline models were presented in terms of NSE and optimal parameter ranges. Finally, the alternative baseline model, based on the elements readily available on Superflex (C0) was developed, and changes were implemented based on QFO insights, leading to models C1-C5. Finally, the performances of these models were evaluated.

3.1 Primary Observations

I conducted the first round of fieldwork on the 13-18th June 2023. As shown in table 2, I identified 4 catchment characteristics by combining various observations made in the field, and hypothesised how these observations translate to catchment discharge dynamics. Finally, guided by the literature represented in appendix A, each hypothesis was represented in terms of changes to the model structure.

Table 2 - Summary of the interpretive steps the modeler made from collecting P-QFO (based on the narrative in appendix E) to forming hypotheses on catchment dynamics, and how these hypotheses could be conceptualised on SuperflexPy.

	Primary QFO	Perceived Hypotheses	Model Implementation
Soil & Interflow	Top 60cm of soil profile had little to no moisture.	Interflow is negligible during dry period. If heavy rainfall were to occur, the soil's low infiltration capacity can lead to high surface runoff	Separate precipitation into surface runoff and infiltration. Ensure infiltration capacity is dependent on precedent moisture or rainfall.
	Soil is fine-grained, yellow.	Catchment mainly covered in loess soil, which can have high matrix flow in wet conditions.	
Side Streams	Discharge in the side streams visible in BE, most originate from springs. Discharge of side streams similar magnitude as Gulp discharge at most upstream stretches.	Springs provide significant contribution to Gulp discharge in BE portion of Gulp during dry periods. In NL, groundwater flow is the dominant source of baseflow.	Split catchment in to 2 'sub-catchments/HRUs' (NL and BE). Ensure that springflow discharge in BE corresponds with observations during period encompassing fieldwork. Add baseflow component, ensuring that minimal discharge >0 during entire modelling period.
	Side streams in NL were completely dry. Within the Dutch portion, Gulp discharge increases as one moves downstream.		
Hollewegen	Paths/roads in valleys (locally known as 'hollewegen') observed in slopes, with deep erosion marks.	Hollewegen act as conduits for surface runoff. Erosion marks indicate that this flux is rapid and powerful, possibly making it a dominant runoff-generating flux during heavy rainfall.	Explicitly represent runoff as a rapid flux, activated during high precipitation intensity.
Topography	BE had generally steeper, more undulating topography. NL had a more clearly defined topography of floodplain-slope-plateau.	Surface runoff is a more significant contributor of peakflow in BE than in NL.	Ensure surface runoff fluxes in BE exceed that of NL.

3.2 Elaborating P-QFO with Secondary Observations

I returned to the field to conduct the second round of fieldwork on the 21-22 August 2023. In line with the methodology, I initially approached only farmers. However, since many farmers were not available to be interviewed, other inhabitants of the Gulp were also approached. In total, 9 semi-structured interviews were conducted, all with inhabitants of the Gulp, 3 of whom being farmers. All interviewees have resided in the Gulp catchment for more than 10 years and have witnessed various 'flood' and 'dry' events over that timespan, implying extensive experiential knowledge about the catchment. All interviews were recorded with permission. Recordings from the interviews, along with field notes, were inductively coded, to generate a narrative based on secondary QFO (appendix F).

The interview and field notes, combined with previous knowledge obtained from the earlier fieldwork round, led me to refine the hypotheses formed from the first round of fieldwork (table 2). Some hypotheses, such as the spatial differences between the NL and BE portions of the catchment, were confirmed by the interviewees. Interviewee #1 noted that there are “many” springs in Teuven (BE) but virtually none across the border in Slenaken (NL). Interviewee #7 also agreed that there were more springs in BE, and hypothesised that it was due to the soils in BE being generally more clayey. Interviewee #7 also went on to note that the topography in BE had more abrupt elevation differences, hypothesising that it leads to there being more surface runoff in BE compared to NL. Interviewee #9 also mentioned a relationship between topography, land use and surface runoff: NL has more defined plateaus, which are suitable for crops. Cropland, in his experience, allow for more interception and infiltration, leading to less runoff. Conversely, the steeper topography in BE leads to a lack of arable land. Hence, BE is dominated by grassland, which is more conducive to surface runoff. These observations confirm the earlier hypothesis that BE is more conducive to surface runoff, which is attributed to a combination of topographical, lithological and land-use differences between the 2 portions of the catchment.

When discussing floods and runoff processes, the interviewees minimised the importance of the hollewegen in generating or transporting water. When describing a flood event near his home on the slope of the catchment, interviewee #4 described water coming along the slopes “from all directions […] it came as a sort of film layer of the hill, over the pastures”. This insight, along with those from interviewee #9 mentioned previously, indicated that surface runoff does not necessarily require a conduit such as hollewegen to induce flow, but instead can occur over various surfaces, especially over grassland. Interviewee #9 also mentioned that higher runoff often occurs when preceding soil moisture is low. As a result, the model elements that would represent runoff should not try to emulate the hydrological properties (i.e. hydraulic conductivity) of specific land-use types, but instead should try to encapsulate runoff as a spatially-averaged process, in line with similar conceptualisations in lumped models, and link the parameters to precipitation intensity and/or antecedent soil moisture.

Interviewees also indicated the temporal range in which surface runoff could be observed. Interviewee #4 also observed that, during floods, “the water disappeared as quickly as it came”, which usually occurs in the span of a few minutes to a few hours. It could hence be argued that the shape of the runoff peak could be modelled to be symmetrical within the span of a day, though other interviewees mention that runoff can persist for multiple days before subsiding.

Lastly, springflow was perceived by 5 interviewees to be closely linked to the water level in a nearby ‘shallow’ (2-10m deep) well. Interviewee #7 described how there are layers of ‘loamy soil’ a few metres below the surface, which prevents further infiltration. Water moves instead along the loam layer, and discharges in springs where the loam layer reaches the surface. This can occur in various locations, including through the wall of the interviewee’s living room[1]. Interviewees #2 and #7 state that most springs have relatively constant discharge, with changes in discharge only noticeable after weeks of rain or dry weather.

[1] Interviewee #7’s house was built into the slope of the catchment. According to him, the living room wall comes into contact with the loam layer. During heavy rainfall, water regularly leaks through the wall, such that a portion of the floor underneath had to be tiled to channel the runoff away, to keep the rest of the living room dry. Such a particular insight was not explicitly included in the field-based perceptual model, though the story indicates how varied runoff processes can be in a catchment.

In summary, the catchment inhabitants whom I interviewed confirmed, adjusted and elaborated upon my previous catchment narrative. The hypothesised differences in runoff between BE and NL were confirmed to be linked to topographical differences, but also to land-use and soil type. When generated, surface runoff can occur over various land cover types – particularly grasslands – and not only over hollewegen. Springflow discharge in BE is fairly stable and is linked to the water level in nearby shallow wells. These observations were hypothesised to be part of a groundwater table that is perched upon a loam layer, which is less prevalent in NL.

3.3 The Field-Informed Perceptual Model

Figure 4 represents the 'field-informed' perceptual model, where I include elements that capture the catchment characteristics described by primary observations (P QFO) and interviews (S QFO). The Belgian portion of the catchment (BE) is characterised by a more undulating topography (P QFO), with more grasslands, leading to more surface runoff being generated during high-precipitation events (P+S QFO). A 'loam layer' is present in BE, at 2-10m depth, which directs infiltrating water towards the surface (S QFO), leading to springs (P QFO). This process provides the multiple active springs in BE with a significant magnitude of discharge, which persists during dry periods, leading me to hypothesise that springflow forms a significant pathway for baseflow of the Gulp (P+S QFO). Meanwhile, the topography in Dutch portion of the catchment (NL) is characterised by well-defined plateaus (P QFO), dominated by arable farming, leading to higher amounts of infiltration compared to BE (P+S QFO). Geological and topographical differences between the two regions is evident through a lack of active springs in NL, as baseflow in the Gulp in NL is generated by groundwater recharge, having percolated into the deeper aquifers (P+S QFO).

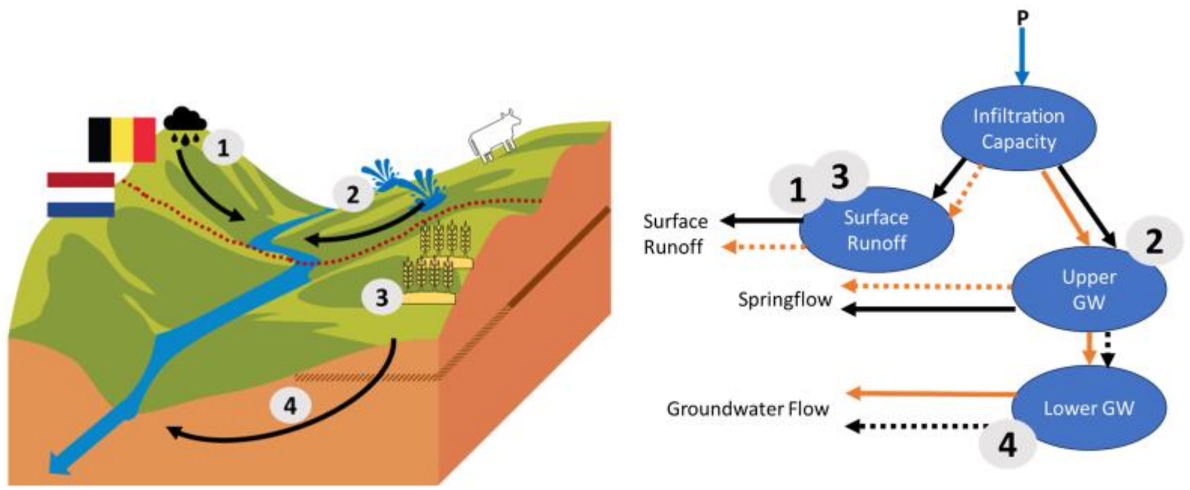

Figure 4 - Perceptual model based on primary and secondary QFO, represented pictorially (left). The main findings are highlighted: 1) steeper topography in BE leading to more surface runoff; 2) multiple springs observed in BE, with significant amount of springflow that persists during dry periods, with groundwater flow being redirected by the 'loam layer'; 3) plateaued landscape with arable land in NL, leading to less surface runoff and more infiltration compared to BE; 4) groundwater flow, a dominant contributor to baseflow during dry periods in NL. Right: schematic representation of the perceptual model with fluxes in NL (orange) and BE (black). Fluxes that are more dominant in one portion of the catchment is represented as a solid arrow, minor fluxes with a dashed arrow.

4 Model Development and Performances

This section describes the series of mathematical models I developed to conceptualise the various catchment processes identified by the field-informed perceptual model. As seen in figure 5, 2 'families' of models were developed – one based on the HBV model structure and parameters developed by Beck et al (2016), and one based on a single unsaturated reservoir (C0), with implementations being limited to using existing elements on Superflex. Of the 12 models used in this study, 4 models acted as baseline models and 3 models included field-based insights in a lumped manner – most notably including a runoff reservoir to simulate fast runoff processes. The remaining 5 models (blue) were spatially distributed to account for the hypothesised differences in hydrological dynamics between the Dutch (NL) and Belgian (BE) portions of the catchment. This chapter will explain, per model family, how each model was developed through implementing insights from the field-based perceptual model. Finally, the performances of each model will be compared to ascertain the impact of QFO on developing better-performing hydrological models.

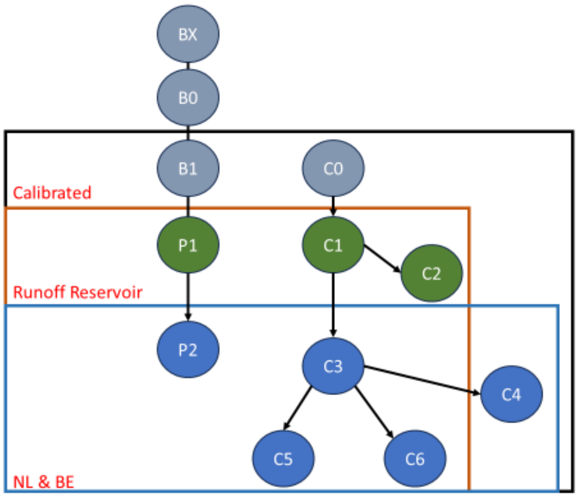

Figure 5 - Summary of the models developed during this study, separated into the baseline models (grey), lumped models (green) and semi-distributed models (blue).

4.1 HBV-Based Baseline Models: Development and Calibration

The baseline (BX,B0,B1) models (figure 6) were developed based on the HBV model used by Beck et al. (2016). The authors developed 10 global parameter datasets, which were found to have achieved equifinality in simulating discharge at the global scale. Each of the 10 global parameter sets are spatially subdivided into grid cells of 0.1x0.1 degrees. The parameter sets from the grid cell that best aligns with the catchment boundary (50.7-50.8N, 5.8-5.9E) were selected. Of the 10 parameter sets, set 00 had the best model performance (in terms of $NSE_{sim,obs}$) during the calibration period ($NSE_{BX,Qobs}$ = -0.55). As a result, BX contains 11 parameters extracted from set 00, which were incorporated into 4 HBV elements – the snow routine, unsaturated zone, and the upper and lower groundwater routines.

Janssen et al. (2021) noted differences in model output in different representations of HBV – including that of HBV represented on SuperflexPy – attributing the differences to discrepancies in the mathematical and numerical implementations of HBV between representations. To ensure comparability between the HBV model structure and the field-based models developed on Superflex, model BX was represented on SuperflexPy using the mathematical and numerical elements presented by Janssen et al. (2021), resulting in model B0. The snow routine was omitted to reduce model uncertainties related to overparameterisation (Schoups et al., 2008), as the snow-related fluxes simulated in model BX contributed minimally to catchment discharge over the entire observation period (<0.5%). PCORR was also omitted to reduce overparameterisation, as no rainfall correction occurred in BX (PCORR = 1). The condition $E_p = 0$ *for* $S > S_{tol}$ was removed from B0 to reduce systematic discrepancies in simulated discharge compared to BX. As a result, B0 contains 3 reservoirs, and is governed by 9 parameters obtained from the global dataset. In the calibration period, B0 was found to replicate BX with $NSE_{B0,BX}$ = 0.978. Despite the high correlation in the outputs of the two models, their performances in simulating observed discharge during the calibration period varied greatly ($NSE_{BX,Qobs}$ = -0.88, $NSE_{B0,Qobs}$ = -0.55).

B1 was generated by replicating the model structure of B0 using calibrated parameters. B1 was found to greatly outperform the previous uncalibrated models in the calibration period, with $NSE_{B1,Qobs}$ = 0.730, though the model still replicates the outputs of the previous baseline models to a large extent ($NSE_{B1,B0}$ = 0.666, $NSE_{B1,BX}$ = 0.620).

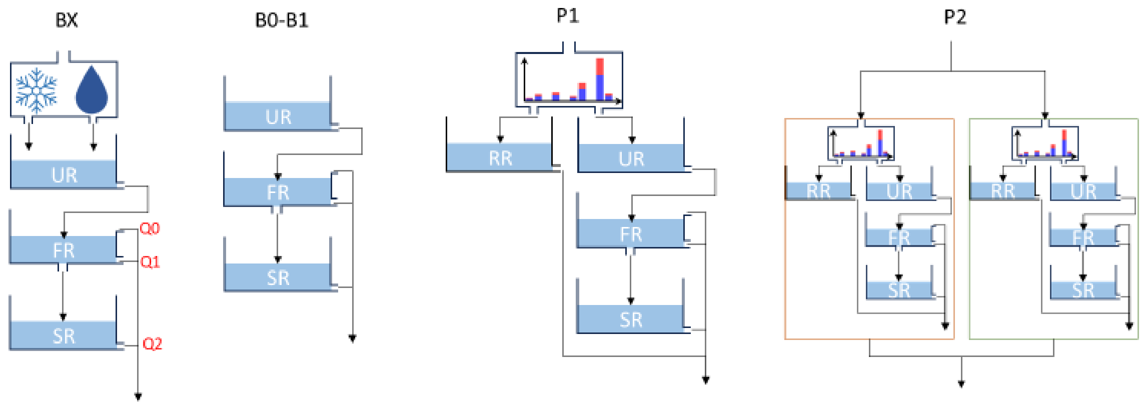

Figure 6 - Schematic representation of each HBV-based model developed in this study. The components of the HBV model output (Q0,Q1,Q2) are indicated in red.

4.2 Changes to HBV Structure based on QFO

This section relates to the choices I made in developing the HBV-based model structures (figure 6) and the Superflex-based model structures (figure 7). The perceptual model developed from primary and secondary QFO indicated a fast surface runoff process that may play a dominant role in generating peakflow during heavy rainfall. The magnitude of this runoff process was linked to precipitation intensity and antecedent soil moisture (by interviewee #9), with higher runoff occurring during an intense rainfall event following a dry period. I represented this process in model P1 through including a linear reservoir (RR) with a high recession coefficient (kRR). This reservoir was included parallel to the HBV elements as this flow path negates the saturated zone, while a dynamic splitter was implemented to divide incoming precipitation into infiltrating and runoff fluxes. A field study (Langhans et al., 2011) that was also carried out in the Belgian portion of the Western and Central European Loess Belt (Bertran et al., 2021) revealed an empirical relationship between precipitation intensity and infiltration capacity at an hourly timescale. The sensitivity of this relationship to local (e.g. lithological, topological) conditions hinders its translatability when 'lumped' in a model at the catchment scale and at the daily timescale. Instead, I implemented a log-log relationship between precipitation intensity (p) and infiltration capacity (f) (equation 1), where α and β are parameters that can be calibrated.

$$Equation\ 1: ln(f) = \alpha\, ln(p) + \beta$$

During calibration, the parameter ranges of α and β were defined based on the range of values obtained by Langhans et al. (2011). As interview results indicate that such runoff events can occur within the span of hours of a rain event, the upper limit of kRR was defined such that 95% of the reservoir is discharged within 1 hour of the rain event, corresponding to kRR ≈ $17d^{-1}$.

My perceptual model shows a distinct difference in the catchment processes between the Dutch (NL) and Belgian (BE) portions of the catchment. This was attributed to the differences in the topography and geology of the two areas, leading to runoff and springflow processes being a more substantial contributor to catchment discharge in BE compared to NL. I represented these insights in model P2 through dividing the catchment into 2 hydrological response units (HRU, Flügel (1995)), which were then combined, weighted by the relative area of each HRU in the catchment ($weight_{BE}$ = 0.532, $weight_{NL}$ = 0.468). As seen in figure 6, the HRUs are structurally identical but differ through parameter calibration, as the parameter ranges obeyed additional conditions outlined in equation 2, to represent BE having a higher tendency for peakflow than NL.

$$Equation\ 2: \quad kRR^{BE} > kRR^{NL}; \quad \alpha^{NL}, \beta^{NL} > \alpha^{BE}, \beta^{BE}$$

4.3 Superflex-Based Model Development

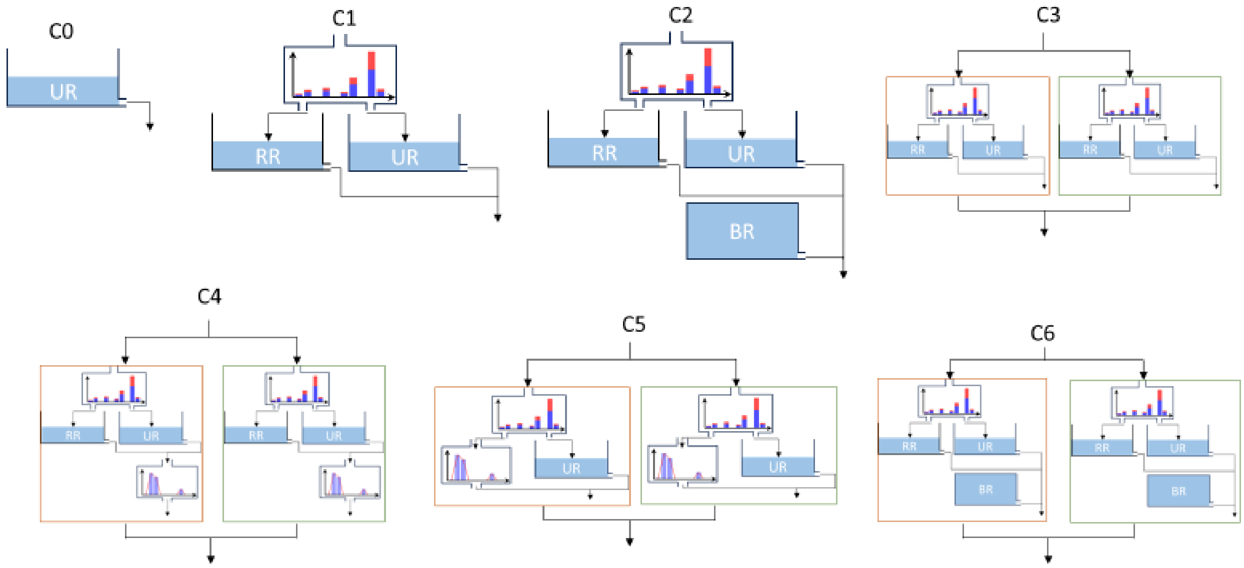

Figure 7 - Schematic representation of each Superflex-based model structure developed in this study.

As seen in figure 7, the baseline model (C0) simply consists of a single unsaturated reservoir, as was available in the Superflex repository. This element, as with most others in the Superflex repository, differs mathematically from the HBV elements. For instance, the Superflex unsaturated reservoir calculated actual evapotranspiration (AET) from potential evapotranspiration (PET) as shown in equation 3, introducing the parameters C_e and m. When used in this study, these parameters were kept constant at 1.0 and 0.01 respectively, as was done in previous Superflex studies (Dal Molin et al., 2020; Kavetski & Fenicia, 2011).

$$Equation\ 3: \ AET = C_e PET(\frac{\bar{S}(1+m)}{\bar{S}+m})$$

As with model P1, model C1 was developed by including a linear runoff reservoir in parallel, with precipitation being split using a parameterised log-log relationship with precipitation intensity. Model C2 emulated C1, but also includes a constant baseflow component, equal to the average observed discharge of the Gulp during the days when primary observations were collected (0.618mm/d). Considering that the catchment was in an exceptionally dry period (31 days preceding fieldwork were without rain), the assumption was made that the observed discharge was equal to a minimal baseflow component that always persists in the Gulp. This assumption was partially confirmed by interviewer #6, who has lived next to the Gulp for 49 years and has never observed the Gulp running dry near his house.

Model C3 represents C1 in a semi-distributed manner, distinguishing between the NL and BE HRUs as with model P2. Model C4 builds upon C3 by including a symmetrical time-lag element (Unit hydrograph 2 of the model GR4J) (Perrin et al., 2003) on the catchment discharge. Through this implementation, I aimed to represent observations that peak flows increased and subsided symmetrically. This element is governed by 1 parameter - the time-lag – which could be calibrated to take values ≥ 1 day. Model C5 implements the unit hydrograph in place of the runoff reservoir, ensuring that only peak flows undergo the symmetrical transformation. Lastly, model C6 is based on C3, but includes a baseflow component that is treated as an external flux. Unlike model C2, the magnitude of the flux was not fixed, but was left to calibration.

4.4 Model Performances

CAL	Simulated Param											
Observed Param	BX (00)	B0	B1	P1	P2	C0	C1	C2	C3	C4	C5	C6
Qobs	-0.88	-0.55	0.73	-0.25	-0.25	-1.36	-0.36	-3.05	-0.64	-0.89	-14.28	0.39
BX (00)	X	0.97	0.62	0.43	0.47	0.35	0.43	-0.81	0.53	0.58	-2.94	0.39
B0	X	X	0.67	0.47	0.51	0.26	0.46	-0.96	0.49	0.51	-3.79	0.40
B1	X	X	X	0.19	0.20	-1.04	0.09	-3.07	-0.25	-0.60	-14.20	0.58
C0	X	X	X	-0.90	-0.99	X	-0.96	-5.82	0.61	0.30	-8.44	-0.02

VAL	Simulated Param											
Observed Param	BX (00)	B0	B1	P1	P2	C0	C1	C2	C3	C4	C5	C6
Qobs	0.07	0.08	0.33	-0.31	-0.31	-0.57	-0.38	-1.12	-0.25	-0.46	-1.15	0.09
BX (00)	X	0.94	0.62	0.38	0.40	0.01	0.35	-1.09	0.45	0.41	-0.36	0.40
B0	X	X	0.70	0.51	0.52	-0.09	0.46	-1.21	0.47	0.40	-0.40	0.50
B1	X	X	X	-0.02	-0.02	-0.89	-0.14	-1.81	-0.16	-0.44	-1.42	0.54
C0	X	X	X	-0.47	-0.52	X	-0.50	-2.93	0.69	0.69	0.79	0.21

Figure 8 - Summary of NSEs of all field-based models compared to observed discharge (Qobs) and discharge simulated by baseline models (BX, B0, B1, C0). Separate comparisons were made for the calibration and validation periods.

Figure 8 represents the performance of all baseline and field-based models over the calibration and validation periods. The best-performing model is the calibrated HBV baseline model, model B1, with NSE_{Qobs} of 0.33 in the validation period and 0.73 in the calibration period. Baseline models B0 and BX perform worse, with NSE_{Qobs} of 0.08 and 0.07 respectively in the validation period. Discharge dynamics generated by BX and B0 appear almost identical to each other, with BX generating slightly higher peakflows.

It is worth noting that multiple calibrated models (C0, C2, C3, C4, C5) performed better during the validation period compared to the calibration period. This discrepancy could be due to slight differences in precipitation dynamics between the calibration and validation periods. For instance, 48% of the days in the calibration period had precipitation, compared to 45% during the validation period. Indeed, the non-calibrated models BX and B0 performed markedly worse during the calibration period (NSE_{Qobs} = -0.88 & -0.55) compared to the validation period (NSE_{Qobs} = 0.07 & 0.08). As seen in figure 11, model C0 simulates no baseflow and severely overestimates peakflow. Severe errors being systematically generated across precipitation events may hence imply a high sensitivity of model NSE to the frequency and intensity of precipitation events.

The field-based models (P1,P2, C1-C6) generally performed worse than the baseline models, with only one model (C6, NSE = 0.09) outperforming the time-averaged mean observation (NSE = 0). C6 is also the only field-based model that outperforms baseline models (BX, NSE = 0.07 and B0, NSE = 0.08). However, it would be difficult to view this result as a success for field-based modelling, as BX and B0 both used regionalised parameters, while C6 used parameters that were calibrated based on the observed catchment discharge.

Instead, when only comparing the performance of each field-based model with the baseline model from its own family, it is evident that the HBV-based models perform worse (NSEs drop from 0.33 for B1 to -0.31 for P1 and P2) when incorporating field-based insights in its structure. Meanwhile, the Superflex-based models C1, C3, C4 and C6 all outperform their corresponding baseline model C0. Models C2 and C5 performed substantially worse than C0, scoring NSEs of -1.12 and -1.15 respectively. The poor performance of these models could be attributed to the implementation of the baseflow component in C2, leading to systematic overestimations of observed discharge, and to the representation of the runoff reservoir as a symmetrical unit hydrograph in model C5.

Qualitative inspection of the outputs from the HBV-based models (figure 9) reveals that the best-performing baseline model (B1) systematically overestimates peakflows while underestimating baseflow, leading to baseline NSE_{Qobs} ranging between 0.07 (for BX) and 0.31 (for B1) during the validation period. P1 and P2 have a near-identical signal and have a baseflow component that is well-defined and of a similar magnitude as B0. P1 and P2 also systematically overestimate peakflow, which reflects in their low model performance. The similarity between the outputs of P1 and P2 indicate that the best-performing parameter sets did little to distinguish between the two HRUs, suggesting that the conditions I applied to distinguish the two HRUs had a negligible effect on improving model performance.

All models failed to replicate the July 2021 flood, with most models instead overestimating a smaller peak event 15 days before the flood. However, due to the lack of precipitation in the days leading up to the flood, it could be reasonably assumed that the inability for any model to predict the flood was due to the precipitation data being underestimated. Reanalysis of precipitation data from that event was shown to be spatially concentrated in the catchment area, with the area surrounding the weather station (Maastricht) experiencing minimal precipitation (Asselman et al., 2022).

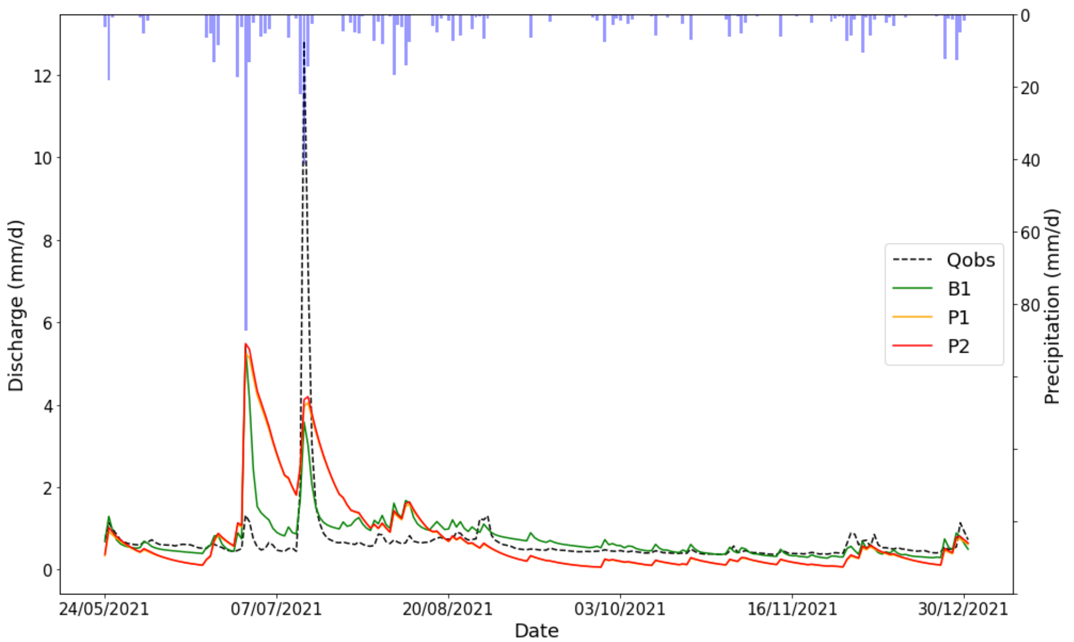

Figure 9 - Validation-period comparisons of observed (grey) and simulated discharges from the baseline model B1 (green) along with the HBV-based models (yellow and red).

Along with the results shown in figures 8 and 9, a sensitivity analysis of model P2 reveals further weaknesses in the performance of the field-informed models. Model performance is insensitive to changes in most parameter combinations, though high sensitivity was observed with the parameters kBE and kNL. Figure 8 plots the better-performing parameter sets generated during the calibration procedure (NSE>-2), revealing that P2 model performance is optimised when kBE and kNL are minimised. Calibrated optimal values for kBE and kNL were 0.112 and 0.103 respectively; these values imply that the reservoir would be 95% empty 30 days after a rainfall event, implying that the best-performing parameter set does not represent a fast runoff process in the runoff reservoir.

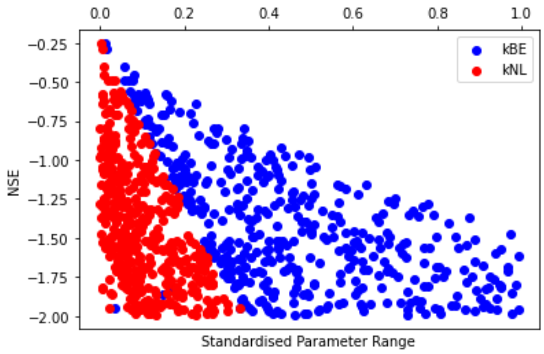

Figure 10 - Parameter sensitivities of kBE and kNL in model P2. The x-axis represents the parameter range standardised by its calibration bounds (0-16).

Figure 11 shows the simulated discharge from the 2 best-performing Superflex-based models – C3 and C6 – in terms of NSE, alongside the baseline model C0 and Qobs. Compared to the HBV-based models, models C0 and C3 severely underestimate baseflow. The inclusion of the runoff reservoir in C3 appears to have 'smoothened out' the peakflows compared to model C0, further demonstrating the tendency for the runoff reservoir to represent 'slower' runoff process than intended as a result of calibration. Meanwhile, the best-performing field-based model (C6) is characterised by a flat hydrograph, with only some small individual peaks visible. It is highly likely that this model's high performance can be attributed mainly to its calibrated baseflow parameter (low_flow), which closely resembles the mean discharge observed (Qobs = 0.62mm/d during the fieldwork period, = 0.63mm/d during the validation period). During model development, low_flow was intended to emulate a minimal baseflow component that is constant over time. It was hence expected that low_flow would take a value close to the minimal discharge observed in the time series (0.19mm/d). Conversely, the better-performing parameter sets tended to assign higher values for low_flow, representing this component as a mean rather than a minimal baseflow component. Indeed, figure 12 illustrates that models with higher values of low_flow simulated discharge with a higher NSE. If one considers alignment with the perceptual model to be another objective function, the frontier shown in the figure may resemble a pareto front, where assigning lower values for low_flow will worsen model performance in terms of NSE, despite providing a closer representation of the field-based perceptual model.

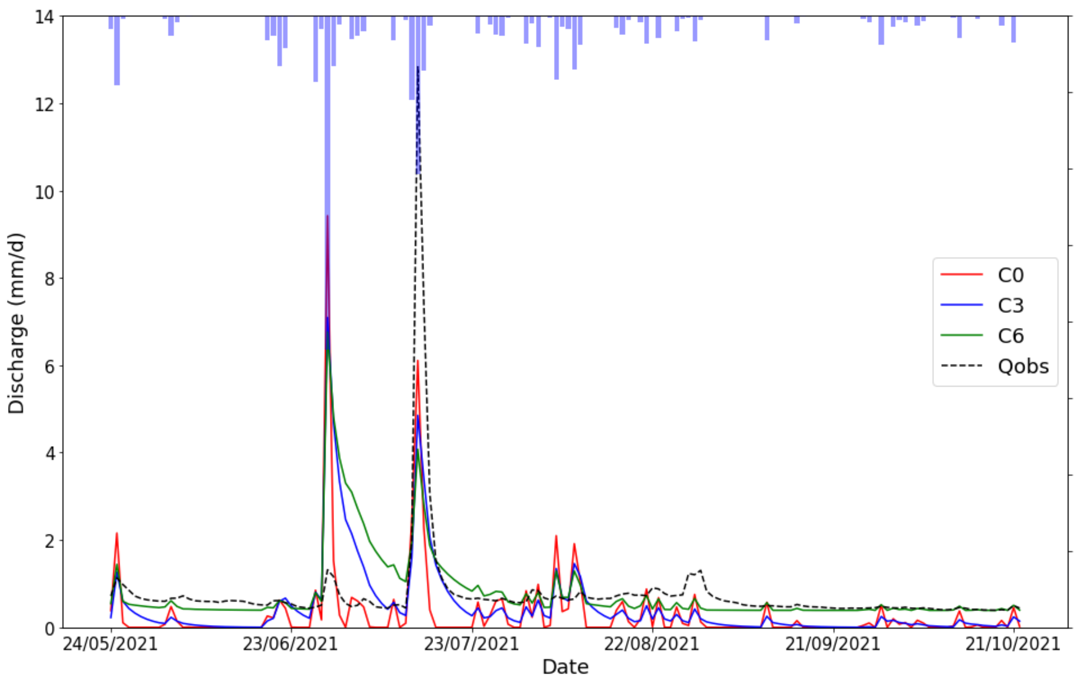

Figure 11 - Observed and simulated discharged generated by the Superflex-based models C0, C3 and C6, during the validation period.

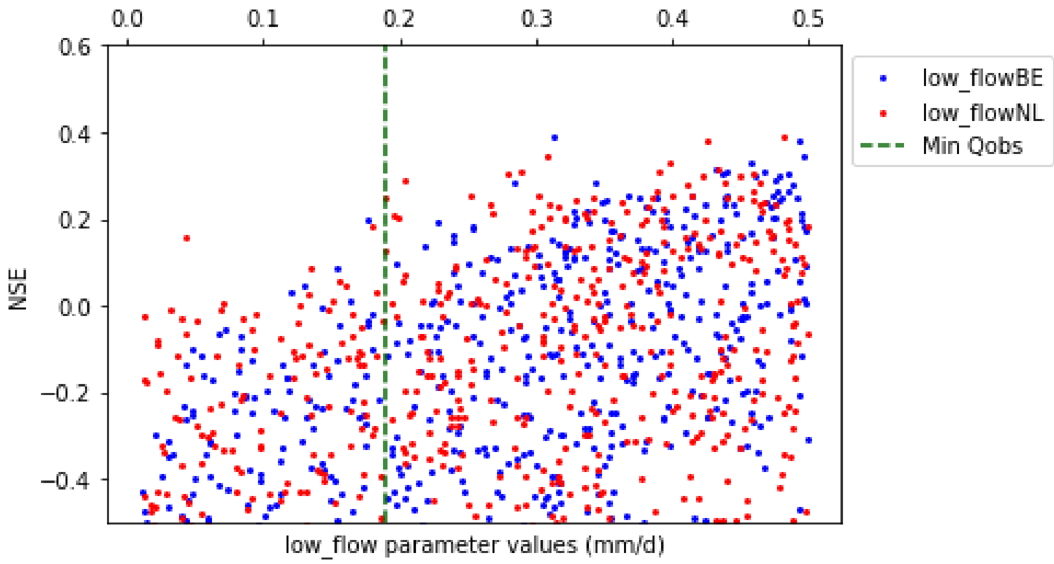

Figure 12 - Sensitivity analysis of the parameters low_flowBE and low_flowNL in model C6. The dotted line indicates the minimal discharge of the Gulp during the entire 5-year time period (0.

Another point of comparison between the models developed is the variation in optimal parameter values each model takes. Figure 13 reveals that there is little agreement between the model structures on the optimal parameter values. A clear example of this disagreement is with parameter k2 (left panel), where models B0 and BX took values of the maximum parameter bound (k2 = 0.15) while the better-performing model B1 took the minimal value (k2 = 0.02). The main exception is kRR, with optimal parameter sets in both HBV-based and Superflex-based model families taking minimum-bound values of kRR. Though it could be expected that optimal parameter values differ (in both value and meaning) across different model structures, the convergence in optimal values for kRR towards the minimum bound indicates that all model structures perform better if the reservoir RR simulates a slow process rather than a fast one.

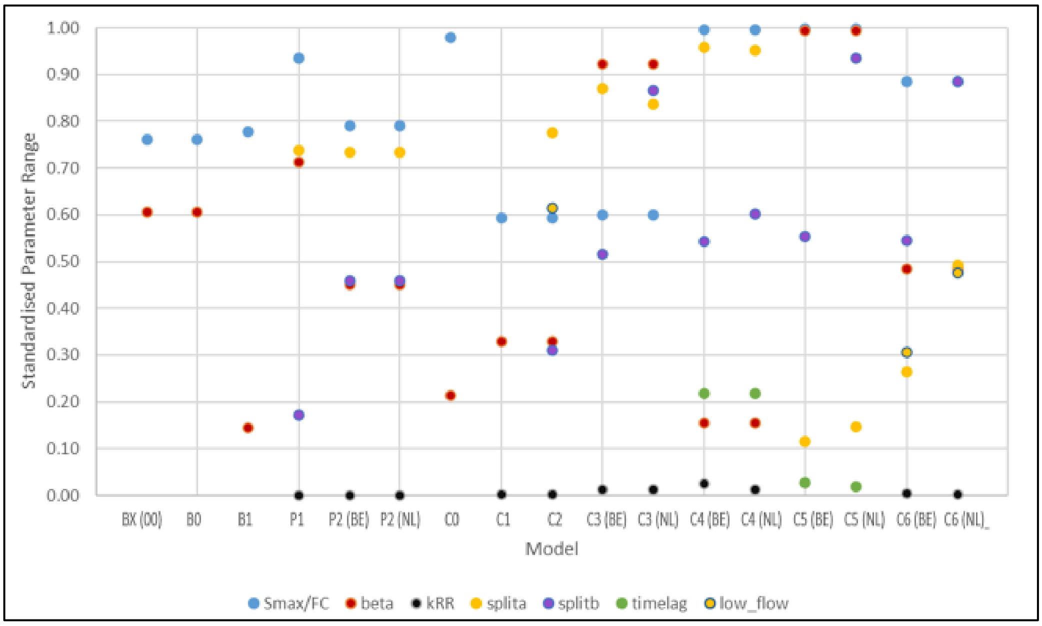

Figure 13 - Plot of the values taken by each optimal parameter set in each model, standardized by its parameter bounds. Note that some HBV parameters were omitted to aid readability.

4. Discussion

In this study, I sought to find out whether I could develop an evolved hydrological model structure based on qualitative field observations (QFO) made in the field. Insights from personal observations and interviews led to developing a field-based perceptual model, with 4 main hypotheses on catchment characteristics. These insights were used to adapt the structure of 2 baseline models, leading to 8 field-based models to be developed. Once calibrated, the field-base models generally failed to outperform the baseline models in simulating discharge. The results appear discouraging for the case of developing catchment-specific model structures, instead indicating the versatility of the HBV model in outperforming field-based models in a geologically complex catchment. But are there any lessons to be learnt from this exercise, that still support the case of field-based modelling?

Various hydrological modelling exercises have used field-based knowledge to improve their modelling outcomes. Van Emmerik et al. (2015) used data gathered from field interviews and observations to restrict the model's parameter space and output uncertainty. Despite the ungauged basin providing limited validation data, the model was used by local stakeholders to improve the local irrigation system. These results indicate that, in practice, models can be deemed useable by stakeholders, given alignment with field-based insights, despite no extensive quantitative validation being possible. Seibert & McDonnell (2015) found that, starting without any flow observations, observations of a single high-flow event can be as informative for catchment calibration as 3 months of streamflow observations. During this study however, fieldwork took place during a low-flow period. Low-flow observations were incorporated into model structures but yielded little improvement in model performance. Finally, Antonetti & Zappa (2018) have shown that the manner of incorporating expert insights will impact model performance. Though transferability between the results of each study is made difficult by the different catchment and research contexts, their insights indicate that the low NSE performances of my field-based models could be attributed to the choices I made in developing the mathematical model based on field-based insights, as opposed to using the latter at all.

Since the field-based perceptual model was developed based on observations and interviews made in the field, the model itself is an inherently adequate source of knowledge about a catchment. However, no quantitative methods (geophysical methods, tracers or otherwise) common to catchment hydrology were used (Seibert & McDonnell, 2002; Wrede et al., 2015), making it difficult to ascertain the uncertainty or realism of the perceptions developed. Cross-sections (see appendix G) taken of hydrogeological models BRO REGIS II v2.2.1 (TNO Geologische Dienst Nederland, 2023) and G3Dv3.1 (DOV, 2023) illustrate the existence of multiple geological formations in the catchment, which also differ spatially between the two sections. As one moves north, the valley incision that delineates the Gulp catchment encompasses fewer geological formations. In BE (transect C), the Gembloux, Vaals and Gulpen formations are exposed, with the interface to the Aachen formation situated 2-3m below the valley floor. Moving northwards into NL (transect A), the valley incision remains within the Gulpen formation. The lack of interfaces between different formations in NL promotes continuous groundwater flow from the surface to the stream, while the interfaces in BE can bring infiltrating water back to the surface through springflow.

The insights from the field-based perceptual model appear to coincide with hydrogeological maps of the area. This agreement could be viewed as a 'confirmation' (Oreskes et al., 1994) of the perceptual model, as empirical observations (i.e. the hydrogeological maps) align with phenomena hypothesised by the perceptual model. Indeed, the perceptual model itself was formed through an evolving hypothesis, based on common ground between my own observations with those of the interviewees. Meanwhile, the resulting mathematical models performed worse than the baseline models, indicating that these models failed to be confirmed by the observed discharge. So, how should the field-based insights and models be assessed in light of these conflicting results?

The field-based models were an attempt to improve model structural adequacy by making structural changes based on qualitative field observations. An adequate model (Addor & Melsen, 2019) indicates that the choice of model structure is 'fit for purpose'. The components, and benefits, of model adequacy could be further explored through the assessment of information generated by hydrological models (Brauman et al. (2022), adapted from Cash et al. (2003)), which is centred around the concepts of salience, credibility, and legitimacy. Of the three concepts, model performance (high NSE) only reflects the model's credibility, which assesses how well the model output matches observations. However, Brauman et al. (2022) note that an increase in credibility does not necessarily affect decision-makers if the hydrological information lacks salience and legitimacy. A salient model is one that is relevant to the needs of the decision-makers, while legitimacy is gained if the model's users perceive the model process – data collections, assumptions and assessment – to be trustworthy, which can be developed through stakeholder engagement. Through engaging with the catchment inhabitants, the field-based models have the potential to represent processes that were more important to its inhabitants, improving structural adequacy. Hence, a worse-performing model may be of more use to the catchment inhabitants, given the explicit consideration to runoff and spring processes that may be of more interest to them than catchment discharge.

For example, runoff processes during peak events were identified by the interviewees to be a salient phenomenon. Observations were collected and numerically conceptualised in the form of the runoff reservoir. However, without qualitative observations to clearly parameterise and calibrate the reservoir fluxes, the reservoir was not able to capture this process effectively without sacrificing the model's performance, which is linked to its ability to recreate the catchment outflow. This trade-off is illustrated by the 'pareto' front observed in the sensitivity analysis in figure 12. A model that had conceptualised this process well would have been more adequate and salient in the eyes of the inhabitants. Conversely, the model would be seen by hydrologists as a less credible one given the lack of verification data beyond catchment discharge. Indeed, it could then also be argued that a model of the Gulp that is optimised to best reproduce the catchment discharge may not be the most salient one in the eyes of the catchment's inhabitants.

Another constructivist viewpoint on hydrological model evaluation is the concept of *fidelius* (Gharari et al., 2021). A model with high fidelity accurately represents the modeller's understanding of the natural system ('realism', in the eyes of the perceptual model) while also simulating outputs which agree with the observed data within an acceptable bound of certainty (in short, performance). The results from the Gulp indicate that these two concepts do not necessarily go hand in hand – the field-based models better represent the modeller's understanding of the natural system yet perform worse than the baseline models. As reflected by the case of the runoff reservoir, a model with a slower surface runoff would perform better in replicating catchment discharge but would lose its 'realism' in representing peakflow runoff processes. Meanwhile, the HBV model structure appears to align with the main fluxes described in the perceptual model, including surface runoff, a faster baseflow process, and a near-constant baseflow. However, the internal fluxes of the best-performing HBV model (B1) do not align with the observed magnitude of the main fluxes in the perceptual model – that is surface runoff, springflow and baseflow. This mismatch may reflect that the HBV model may conform less to the modeller's understanding of the natural system, despite performing better in representing catchment discharge. As a result, the HBV model may not necessarily have higher fidelity than the field-based models developed here.

Operating within a naturalist perspective, I recognise that the results of this study are dependent on my observations, interview probes, interpretations, and modelling choices. Hence, speaking from my perspective as a research instrument, it is important to recognise that my positionality may have affected the research outcomes. Firstly, the Gulp catchment was chosen due to its convenient size, geological complexity and for its practical feasibility for fieldwork. However, despite being in the relative wake of the 2021 floods, the interviewees did not voice a need for improved flood forecasting or other measures that would be served by hydrological models. Instead, I was regarded merely as a curious student, who was interested in floods. Interviewee #1 mentioned "if you want to hear something interesting [flood observations] you need to go to the Berwine", while a farmer who lived next to the Gulp (not interviewed) mentioned that he did not have much to tell me regarding floods. Hence, it was possible that the insights gained from the interviews I did conduct were skewed and embellished, as interviewees elaborated on flood-related processes despite not being severely affected by them. The models formed may hence reflect processes that were most observable rather than what was most important in generating runoff. Furthermore, it is possible that my previous knowledge of the regional hydrogeology led me to predefine my hypothesis on catchment function. I was aware of the multiple geological formations lining the slopes in (the Dutch) South Limburg, whose interface can restrict groundwater flow and redirect flow to springs. It is hence possible that the perceptual model overstates the importance of springflow to catchment discharge. Though this was not a comprehensive reflection of my positionality in this study, the points discussed convincingly indicate that, if this case study was carried out in another catchment and/or by another researcher, the resulting field-based models may be structured, and perform, differently.

Despite the context-dependence of the results of this case study, recommendations can be made on how to apply this methodology could be applied in the future, to maximise model adequacy and fidelity. The field-based perceptual model did unite personal observations with insights gained from the inhabitants of the catchment and is coherent with geological maps of the region. Instead, changes should be made with how the insights from the perceptual model are integrated into mathematical models. The results indicated that model performance reduced when additional elements are added to the HBV model structure, but increased when elements were added to a single reservoir. This could indicate structural redundancy: if the HBV model can conceptualise surface runoff through flux Q0, adding an additional reservoir to represent surface may be superfluous. To avoid such redundancies, insights from the perceptual model should be applied in an integrated manner. A field-based model 'made from scratch' would resemble the C family of models, who outperformed the P-models possibly due to the lack of redundant model elements. For instance, the best-performing field-based model C6 relied on the baseflow component to generate baseflow, and the runoff and unsaturated reservoirs to govern peakflow generation. It is hence recommended that field-based models are 'made from scratch' with parameter ranges informed by (if possible) the perceptual model, with the aim of each element serving a clear function with a unique link to the perceptual model.

If, for any reason, a modeller chooses to rely on a pre-existing model structure, the low performance of the P-models indicates that insights from the field-based perceptual model should not be translated to changes to the model structure, to avoid structural redundancy. Proponents of the HBV model would argue that this its limitations to represent physical, observable processes is one inherent to conceptual models, whose parameters are typically generated through calibration instead of being physically defined (Devia et al., 2015). Instead, the adequacy of pre-defined model structures could be refined using multi-objective criteria, which contain elements from the perceptual model, to guide parameter calibration and to assess model performance (Efstratiadis & Koutsoyiannis, 2010).

To summarise, the underperformance of the field-based numerical models does not necessarily imply that the field-based approach to modelling should be abandoned. The field-based perceptual model manages to harmonise personal observations with those of the catchment inhabitants and aligned well with geological literature of the region. When using holistic measures to judge results from hydrological models, model performances indicate a trade-off between model performance and realism, implying that field-based models developed in this study may not necessarily be of lower fidelity than the baseline models. Instead, insights from the field allowed me to position the modelling process to align with the observations and experiences communicated by the catchment inhabitants, improving the realism, salience, and legitimacy of the field-based models. Modellers who seek to improve the adequacy of their models could refer to this methodology and refine the approach with changes to the conceptualisation stage. Insights from the perceptual model should be implemented in an integrated manner, leading to a field-based model 'made from scratch'. When working with a pre-existing model structure, the perceptual model could instead be used to develop multi-objective criteria to assess the adequacy of the model structure.

5 Conclusions

I conducted this study to explore how primary and secondary qualitative field observations can lead to an evolved perceptual model, leading to differing numerical model structures and discharge predictions.

The field-based perceptual model, based on first-hand observations and interviews with catchment inhabitants, revealed differences in catchment characteristics between the Belgian and Dutch portions of the catchment. Topographical and land-cover differences led to the Belgian portion being more conducive to surface runoff. Additionally, geological differences are reflected in the dominant baseflow-generating processes in each portion – springflow in the Belgian portion, and groundwater fluxes in the Dutch portion. I generated 8 mathematical models to represent the hypotheses stated in the perceptual model in different ways. For instance, all models explicitly represented surface runoff, of which 5 were semi-distributed to distinguish the Belgian and Dutch portions of the catchment. The parameter space governing the partitioning and generation of runoff were restricted to ensure a higher runoff response in the Belgian portion. Various approaches were also employed to represent baseflow and flood dynamics.

The results indicated that the mathematical field-based models had limited success in outperforming the baseline models. Changes made to the HBV model worsened model performance compared to a pure HBV structure, while some changes made to the unsaturated reservoir led to an improvement in model performance compared to a single unsaturated reservoir model. Sensitivity analyses indicated multiple trade-offs between model performance and alignment with the perceptual model. As a result, best-performing parameter sets often did not reflect values that best align with field-based insights.

When judging a model using more holistic concepts, such as model fidelity, saliency and legitimacy, it is clear that the most adequate model is not necessarily one that performs best in representing catchment discharge. In the case of this study, a worse-performing field-based model may be of more relevance to catchment inhabitants, as its structure better aligns with a field-based perceptual model developed with the help of their insights. Additionally, since the field-based model structures better align with the modeller's understanding of the catchment dynamics, these models are not necessarily of lower fidelity than a pre-defined model structure.

Insights from the field-based perceptual model could be confirmed by geological maps of the catchment. This suggests that the low performances of the mathematical models were a result of the choices made in conceptualising the field-based insights, as opposed to the field-based insights being misleading. Moving forward, modellers should refrain from making stepwise changes to pre-existing model structures, instead opting for creating a field-informed model structure 'from scratch' to minimise structural redundancy. If a pre-existing model structure is used, the field-based perceptual model can be used to develop additional criteria that the model parameters and outputs should fulfil to better represent field-based insights. Based on these recommendations, I encourage modellers to enter the field to develop a more holistic perception of how the catchment functions and how its hydrology is experienced by its inhabitants. From such perceptions, a more adequate, salient and legitimate model can be formed.

Acknowledgements

It has been said that it takes a village to run a model. Over the course of this study, I have been surrounded by a village of brilliant, helpful people who have helped make my idea of ‘bringing the modeller to the field’ into a reality.

My first thanks undoubtedly go to the ‘modelling village chief’, my main supervisor, Lieke Melsen. I came to you a year ago with a mix of ideas on ontologies and qualitative research and wanted to make a thesis out of it in a Quantitative Hydrology chair group. Between all those thoughts, you saw a way forward, and so an out-of-the-paradigm idea was born. Our high-speed discussions helped developed my idea into a full study, and your insights, feedback and guidance made our meetings something I always looked forward to. Your belief in this idea sometimes exceeded my own and gave me confidence at the times I needed it. If I one day find myself in a supervisory role, I would look to your example.

Ryan Teuling has co-supervised me and has been a great help in various stages of the thesis, especially in numerical modelling and interpreting the results. Many thanks also go to Rozemarijn ter Horst, who helped guide my proposal writing. Thank you for being there for me, even from Palestine. Thanks also go to Janneke Remmers and Roel Dijksma for our interesting discussions, and to Koen Jansen for sharing me your HBV-on-Superflex code.

The people in the Dutch and Belgian Limburg(s) were just as essential to my field-based journey. Thank you to all 9 interviewees for sharing your insights on the hydrology, as well as on topics ranging from railway maintenance to snail farming, and how, in life, perspectives may all come together yet may differ between person to person. After my two trips to the Gulp, I can surely claim that there is more to a catchment than just numbers. On the quantitative side, thanks also go to Arvid Voncken from Waterschap Limburg for providing me with discharge observations.

Of my friends and family, there is almost too many to thank for their help and support over the past months. Awad, Sebastian, Lenka – thanks for all the feedback, encouragement and the valuable hours spent at the ping-pong table. Similar credit also goes to Tycho, who also helped conduct the interviews and helped translate from Dutch and French. Also, many thanks to you and Janis for accompanying me for a lovely end to fieldwork in Belgium.

A special shoutout goes to my friends from my Milan days. Thank you, Rohan, for spending your free afternoon to help me with my programming issues. Your wise words *“for every problem, there is more than one solution”* may have been intended for programming, but I’ll take it as a life lesson as well. And thank you Michael for encouraging me to connect with experts. You are great at that.

So many thanks also to Nathalie for introducing me to frameworks in social science and teaching me a bit on how social scientists think. Also, thank you for being there for all the highs and lows, for dealing with my constant indecision, and for always believing that things will work out. It did!

Finally, some final thanks (in advance) to Tim van Emmerik for your positive response to my idea and for being the examiner of my thesis. I hope you enjoyed the read.

References

Abramowitz, G., Gupta, H., Pitman, A., Wang, Y., Leuning, R., Cleugh, H., & Hsu, K.-L. (2006). *Neural Error Regression Diagnosis (NERD): A Tool for Model Bias Identification and Prognostic Data Assimilation*.

Abu, R., Reed, M. G., & Jardine, T. D. (2020). Using two-eyed seeing to bridge Western science and Indigenous knowledge systems and understand long-term change in the Saskatchewan River Delta, Canada. *International Journal of Water Resources Development*, *36*(5), 757–776. https://doi.org/10.1080/07900627.2018.1558050

Addor, N., & Melsen, L. A. (2019). Legacy, Rather Than Adequacy, Drives the Selection of Hydrological Models. *Water Resources Research*, *55*(1), 378–390. https://doi.org/10.1029/2018WR022958

Antonetti, M., & Zappa, M. (2018). How can expert knowledge increase the realism of conceptual hydrological models? A case study based on the concept of dominant runoff process in the Swiss Pre-Alps. *Hydrology and Earth System Sciences*, *22*(8), 4425–4447. https://doi.org/10.5194/hess-22-4425-2018

Asselman, N., van Heeringen, K.-J., de Jong, J., & Geertsema, T. (2022). *Juli 2021 overstroming en wateroverlast in Zuid-Limburg*.

Beck, H. E., van Dijk, A. I. J. M., De Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., & Bruijnzeel, L. A. (2016). Global-scale regionalization of hydrologic model parameters. *Water Resources Research*, *52*(5), 3599–3622.

Bergström, S. (1992). *The HBV model–its structure and applications*. SMHI.

Bertran, P., Bosq, M., Borderie, Q., Coussot, C., Coutard, S., Deschodt, L., Franc, O., Gardère, P., Liard, M., & Wuscher, P. (2021). Revised map of European aeolian deposits derived from soil texture data. *Quaternary Science Reviews*, *266*, 107085.

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

Beven, K. J. (2011). *Rainfall-runoff modelling: the primer*. John Wiley & Sons.

Blair, P., & Buytaert, W. (2016). Socio-hydrological modelling: a review asking ``why,~what~and~how?''. *Hydrology and Earth System Sciences*, *20*(1), 443–478. https://doi.org/10.5194/hess-20-443-2016

Blöschl, G. (2006). Hydrologic synthesis: Across processes, places, and scales. *Water Resources Research*, *42*(3).

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., & Savenije, H. (2013). *Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales*. Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9781139235761

Brauman, K. A., Bremer, L. L., Hamel, P., Ochoa-Tocachi, B. F., Roman-Dañobeytia, F., Bonnesoeur, V., Arapa, E., & Gammie, G. (2022). Producing valuable information from hydrologic models of nature-based solutions for water. *Integrated Environmental Assessment and Management*, *18*(1), 135–147.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Cash, D. W., Clark, W. C., Alcock, F., Dickson, N. M., Eckley, N., Guston, D. H., Jäger, J., & Mitchell, R. B. (2003). Knowledge systems for sustainable development. *Proceedings of the National Academy of Sciences*, *100*(14), 8086–8091.

Clark, M. P., Kavetski, D., & Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research*, *47*(9). https://doi.org/10.1029/2010WR009827

Dal Molin, M., Kavetski, D., & Fenicia, F. (2021). SuperflexPy 1.3.0: An open-source Python framework for building, testing, and improving conceptual hydrological models. *Geoscientific Model Development*, *14*(11), 7047–7072. https://doi.org/10.5194/gmd-14-7047-2021

Dal Molin, M., Schirmer, M., Zappa, M., & Fenicia, F. (2020). Understanding dominant controls on streamflow spatial variability to set up a semi-distributed hydrological model: the case study of the Thur catchment. *Hydrology and Earth System Sciences*, *24*(3), 1319–1345.

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

DOV. (2023, September). *Databank Ondergrond Vlaanderen*. https://www.dov.vlaanderen.be/portaal/?module=verkenner&bm=b9fb93ab-83ea-404f-ac8c-e0fd741673b6

Eden, S., Megdal, S. B., Shamir, E., Chief, K., & Lacroix, K. M. (2016). Opening the black box: Using a hydrological model to link stakeholder engagement with groundwater management. *Water (Switzerland)*, *8*(5). https://doi.org/10.3390/w8050216

Efstratiadis, A., & Koutsoyiannis, D. (2010). One decade of multi-objective calibration approaches in hydrological modelling: a review. *Hydrological Sciences Journal–Journal Des Sciences Hydrologiques*, *55*(1), 58–78.

Fenicia, F., Kavetski, D., & Savenije, H. H. G. (2011). Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development. *Water Resources Research*, *47*(11). https://doi.org/10.1029/2010WR010174

Fenicia, F., Savenije, H. H. G., Matgen, P., & Pfister, L. (2008). Understanding catchment behavior through stepwise model concept improvement. *Water Resources Research*, *44*(1). https://doi.org/10.1029/2006WR005563

Flügel, W. (1995). Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the River Bröl, Germany. *Hydrological Processes*, *9*(3-4), 423–436.

Gharari, S., Gupta, H. V, Clark, M. P., Hrachowitz, M., Fenicia, F., Matgen, P., & Savenije, H. H. G. (2021). Understanding the information content in the hierarchy of model development decisions: Learning from data. *Water Resources Research*, *57*(6), e2020WR027948.

Glaser, B., & Strauss, A. (1967). Grounded theory: The discovery of grounded theory. *Sociology the Journal of the British Sociological Association*, *12*(1), 27–49.

Guba, E. G. (1981). *ERIC/ECTJ Annual Review Paper: Criteria for Assessing the Trustworthiness of Naturalistic Inquiries* (Vol. 29, Issue 2). https://about.jstor.org/terms

Holländer, H. M., Blume, T., Bormann, H., Buytaert, W., Chirico, G. B., Exbrayat, J.-F., Gustafsson, D., Hölzel, H., Kraft, P., Stamm, C., Stoll, S., Blöschl, G., & Fï Uhler, H. (2009). Comparative predictions of discharge from an artificial catchment (Chicken Creek) using sparse data. In *Hydrol. Earth Syst. Sci* (Vol. 13). www.hydrol-earth-syst-sci.net/13/2069/2009/

Jansen, K. F., Teuling, A. J., Craig, J. R., Dal Molin, M., Knoben, W. J. M., Parajka, J., Vis, M., & Melsen, L. A. (2021). Mimicry of a Conceptual Hydrological Model (HBV): What's in a Name? *Water Resources Research*, *57*(5), e2020WR029143.

Kallio, H., Pietilä, A.-M., Johnson, M., & Kangasniemi, M. (2016). Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing*, *72*(12), 2954–2965. https://doi.org/10.1111/jan.13031

Kavetski, D., & Fenicia, F. (2011). Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights. *Water Resources Research*, *47*(11). https://doi.org/10.1029/2011WR010748

Kirk, J., & Miller, M. L. (1986). *Reliability and validity in qualitative research*. Sage.

Koks, E., Van Ginkel, K., Van Marle, M., & Lemnitzer, A. (2021). Brief communication: Critical infrastructure impacts of the 2021 mid-July western European flood event. *Natural Hazards and Earth System Sciences Discussions*, *2021*, 1–11.

Langhans, C., Govers, G., Diels, J., Leys, A., Clymans, W., Van den Putte, A., & Valckx, J. (2011). Experimental rainfall–runoff data: Reconsidering the concept of infiltration capacity. *Journal of Hydrology*, *399*(3–4), 255–262.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. sage.

Massuel, S., Riaux, J., Molle, F., Kuper, M., Ogilvie, A., Collard, A. L., Leduc, C., & Barreteau, O. (2018). Inspiring a Broader Socio-Hydrological Negotiation Approach With Interdisciplinary Field-Based Experience. In *Water Resources Research* (Vol. 54, Issue 4, pp. 2510–2522). Blackwell Publishing Ltd. https://doi.org/10.1002/2017WR021691

McEwen, L., Krause, F., Hanson, J., & Jones, O. (2012). Flood histories, flood memories and informal flood knowledge in the development of community resilience to future flood risk. *Hydrology for a Changing World*, 01–04. https://doi.org/10.7558/bhs.2012.ns34

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, *21*(2), 239–245.

Mcmillan, H., Woods, R. A., & Duncan, M. (2009). *Improving perceptual and conceptual hydrological models using data from small basins Hydrologic change in rainfall-streamflow characteristics View project*. https://www.researchgate.net/publication/228644977

Melsen, L. A. (2022). It Takes a Village to Run a Model—The Social Practices of Hydrological Modeling. *Water Resources Research*, *58*(2). https://doi.org/10.1029/2021WR030600

Melsen, L. A., Vos, J., & Boelens, R. (2018). What is the role of the model in socio-hydrology? Discussion of "Prediction in a socio-hydrological world"*. In *Hydrological Sciences Journal* (Vol.

63, Issue 9, pp. 1435–1443). Taylor and Francis Ltd. https://doi.org/10.1080/02626667.2018.1499025

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Nota, D. J. G., & van de Weerd, B. (1988). A hydrogeological study in the basin of the Gulp Creek-A reconnaissance in a small catchment area. *Agricultural University Wageningen Papers*, *87*, 7.

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. In *New Series* (Vol. 263, Issue 5147).

Perrin, C., Michel, C., & Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, *279*(1–4), 275–289.

Razavi, T., & Coulibaly, P. (2013). Streamflow prediction in ungauged basins: review of regionalization methods. *Journal of Hydrologic Engineering*, *18*(8), 958–975.

Schoups, G., Van de Giesen, N. C., & Savenije, H. H. G. (2008). Model complexity control for hydrologic prediction. *Water Resources Research*, *44*(12).

Seibert, J. (1997). Estimation of parameter uncertainty in the HBV model: Paper presented at the Nordic Hydrological Conference (Akureyri, Iceland-August 1996). *Hydrology Research*, *28*(4–5), 247–262.

Seibert, J., & Bergström, S. (2022). A retrospective on hydrological catchment modelling based on half a century with the HBV model. *Hydrology and Earth System Sciences*, *26*(5), 1371–1388.

Seibert, J., & McDonnell, J. J. (2002). On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. *Water Resources Research*, *38*(11), 23-1-23–14. https://doi.org/10.1029/2001wr000978

Seibert, J., & McDonnell, J. J. (2015). Gauging the ungauged basin: relative value of soft and hard data. *Journal of Hydrologic Engineering*, *20*(1), A4014004.

Shen, H., Tolson, B. A., & Mai, J. (2022). Time to update the split-sample approach in hydrological model calibration. *Water Resources Research*, *58*(3), e2021WR031523.

Sivapalan, M. (2005). Pattern, Process and Function: Elements of a Unified Theory of Hydrology at the Catchment Scale. In *Encyclopedia of Hydrological Sciences*. John Wiley & Sons, Ltd. https://doi.org/10.1002/0470848944.hsa012

TNO Geologische Dienst Nederland. (2023, September). *DINOloket Data en Informatie van de Nederlandse Ondergrond*. https://www.dinoloket.nl/ondergrondmodellen/kaart

Uhlenbrook, S., Seibert, J. A. N., Leibundgut, C., & Rodhe, A. (1999). Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structure. *Hydrological Sciences Journal*, *44*(5), 779–797.

van Emmerik, T., Mulder, G., Eilander, D., Piet, M., & Savenije, H. (2015). Predicting the ungauged basin: Model validation and realism assessment. *Frontiers in Earth Science*, *3*. https://doi.org/10.3389/feart.2015.00062

Vogel, R. M., Lall, U., Cai, X., Rajagopalan, B., Weiskel, P. K., Hooper, R. P., & Matalas, N. C. (2015). Hydrology: The interdisciplinary science of water. *Water Resources Research*, *51*(6), 4409–4430. https://doi.org/10.1002/2015WR017049

Wagener, T., & Gupta, H. V. (2005). Model identification for hydrological forecasting under uncertainty. *Stochastic Environmental Research and Risk Assessment*, *19*(6), 378–387. https://doi.org/10.1007/s00477-005-0006-5

Wrede, S., Fenicia, F., Martínez-Carreras, N., Juilleret, J., Hissler, C., Krein, A., Savenije, H. H. G., Uhlenbrook, S., Kavetski, D., & Pfister, L. (2015). Towards more systematic perceptual model development: A case study using 3 Luxembourgish catchments. *Hydrological Processes*, *29*(12), 2731–2750. https://doi.org/10.1002/hyp.10393