Abstract
This paper describes implementation of an operational flow forecasting system for New Zealand that couples a high-resolution (1.5 km) weather forecast model to a hydrological model. The impact of the high-resolution weather forecast is tested for two large storms in a case study catchment where convective storms and orographic rains – only modelled by high-resolution models – play a crucial role. The high-resolution rainfall forecast is compared with that of a low-resolution (12 km) model and a 5 km gridded observed rainfall product, and the corresponding flow forecasts are assessed using deterministic forecasts and probabilistic lagged ensemble forecasts. The results show the benefit of the increased high-resolution convective-permitting lagged ensemble forecast over the large scale (lower resolution) model for flood forecasting. However, rainfall underestimation, from all the weather models and observed data, highlights the ongoing need to better account for precipitation and model uncertainties with bias correction, ensemble and flow data assimilation techniques to produce more accurate flood forecasts.

Keywords
Flow forecasting, hydrological model, numerical weather forecast, deterministic and ensemble forecasts, data assimilation, case study.

Introduction
Recent flooding and other water-related disasters around the world have led to increased interest in hydrologic forecasting systems. In 2013, flooding was responsible for nearly half of all natural hazards related losses, costing billions of dollars in damage (Pagano et al., 2014). In New Zealand, the Insurance Council1 reports costs related to flood losses of 442.3 million dollars between 1996 and June 2014. Storm- and flood-related insurance claims constituted 90% of all the insurance claims between January 2013 and June 2014 (Priest, 2014). Increasing trends in losses due to natural hazards, in terms of fatalities and economic damages, are reported in the literature (Barredo, 2009; Kunreuther et al., 2011; Bevere et al., 2012; Bevere et al., 2013; Kundzewicz et al., 2013). Increased urbanisation and the modification of the precipitation regime associated with climate change will likely increase our vulnerability to flooding in the coming decades (Palmer and Räisänen, 2002; Rosso and Rulli, 2002; Parry, 2007).

Flood forecasting can mitigate the impacts of these natural disasters by providing accurate
and timely warning of major flooding events. Operational flood forecasting systems are being used to make decisions on upcoming floods by hydro-meteorological agencies around the world (Pagano, 2013; Wetterhall et al., 2013; Pappenberger et al., 2015). Flood forecasts are often provided by coupling numerical weather prediction models with hydrological models, and hence are highly dependent on the quality of the weather forecast. In particular, precipitation is the most important meteorological component for driving hydrological models.

The latest international advances in Numerical Weather Prediction (NWP) models are being made by using very high-resolution models, which are able to more accurately represent the atmospheric processes that can lead to frontal, orographic and convective rainfall, as well as having a more detailed representation of surface features such as coastlines and topography, compared to low-resolution models. Frontal rain occurs when warm and cold air masses meet, with the warm air, being less dense, pushed up over the cold air mass where it cools and condenses. Orographic rain originates from clouds formed as air masses interact with the topography of the land, e.g., air rising over mountainous areas. Convective rain is produced by clouds formed in vertical motions resulting from atmospheric instabilities (e.g., heating from the sun). Convective cells are much smaller than the large-scale lifting of air occurring from fronts and give smaller areas of rain. Frontal rain is relatively well modelled by NWP models of 12 km horizontal grid resolution. However, such models have insufficient resolution to represent local orographic effects and convective storms on the scales required by hydrological models.

For an orographically-enhanced, stratiform rainfall event in the United States, the best hydrological model simulations were achieved when using the highest NWP resolution (4 km, compared to 12 km and 36 km models) (Westrick and Mass, 2001). The simulations with the best-resolved topography lead to a roughly 50% increase of maximum precipitation over mountains and foothills using the weather model WRF (Flesch and Reuter, 2012).

In several Mediterranean flash-flood case studies, the precipitation underestimation was significantly less marked for high-resolution, convection-permitting NWPs (2.5 km horizontal resolutions) (Anquetin et al., 2005; Chancibault et al., 2006; Younis et al., 2008; Vincendon et al., 2010). However, despite the improvement in meteorological performance scores, the models still contained uncertainties in rainfall location and did not systematically lead to improved flow forecasting when coupled to hydrological models. Similar results were obtained with a case study in the United Kingdom using the Met Office Unified Model (MetUM) at 12, 4 and 1 km resolutions coupled to a hydrological model: the unpredictability of convective systems required a more probabilistic approach (Roberts et al., 2009). In Liu et al. (2015), data assimilation in both the weather forecast model (WRF) and the hydrological model improved the flood forecast.

The state of the art operational flood forecasting systems are using ensembles of weather forecasts (probabilistic approach) rather than single deterministic forecasts to drive hydrological models (Cloke and Pappenberger, 2009; Wetterhall et al., 2013). Ensemble prediction systems are used to take account of the uncertainties in the stochastic and chaotic nature of the atmospheric physical processes modelled, and result in multiple weather forecasts at the same place and time (Buizza et al., 1999; Palmer and Buizza, 2007). Ensemble prediction systems have generally better skill than for both mean forecast performance and forecasting more extreme events (Wetterhall et al., 2013).
However, generating high-resolution weather forecast ensembles is computationally expensive and time-lagged ensembles, which do not require further computational resources to create a small ensemble, can be used as an alternative. A time-lagged ensemble is a set of consecutive overlapping forecasts produced at six hour intervals, for example. The time-lagged ensemble forecasting concept was introduced by Hoffman and Kalnay (1983) as a precursor to ensemble forecasting, and has been applied to medium range (6-10 days) (Dalcher et al., 1988; Van den Dool and Rukhovets, 1994) and short range forecasts (Lu et al., 2007; Mittermaier, 2007; Ushiyama et al., 2014). The initial conditions of the ensemble members are generated from different initializations of the atmospheric state (Hoffman and Kalnay, 1983; Lu et al., 2007; Mittermaier, 2007; Ushiyama et al., 2014).

Time-lagged ensembles improved the high-resolution (4 km) model precipitation forecast skill of the MetUM, and created a similar spread to the one created from perturbed initial conditions in conventional ensembles (Mittermaier, 2007). The flood alert system improved with time-lagged ensemble forecasts (5 km) coupled to a hydrological model for a case study in Pakistan (Ushiyama et al., 2014).

Other sources of uncertainty in flow forecasting come from the hydrological model and include structural errors, model parameter errors, data errors (in rainfall and flow data used for model calibration) and model framework uncertainty (Beven and Freer, 2001; Montanari, 2005; McMillan et al., 2009; McMillan et al., 2012). Hydrological Ensemble Prediction Systems (HEPS) can be used to quantify the different types of uncertainties when used in operational systems (Cloke and Pappenberger, 2009).

In this paper, we discuss the implementation and impact of driving the hydrological model TopNet with weather forecasts from a new high-resolution (1.5 km) New Zealand Convective Scale Model (NZCSM). We investigate how the new NZCSM-TopNet coupled model compares with the current flow forecasting system in use at NIWA, which uses the New Zealand Local Area Model (NZLAM) weather forecast model. Both weather forecast models are local implementation of the MetUM for New Zealand, with a grid spacing of 1.5 km (NZCSM) compared to 12 km (NZLAM). We base our investigation on a case study in the Hutt River basin in the Wellington region, where convective storms and orographic rain, only modelled by high-resolution grids, play a crucial role.

Data and models
Weather models NZLAM and NZCSM
NZLAM is a regional NWP model that uses lateral boundary conditions from the global version of the MetUM. The model grid includes New Zealand and surrounding oceans (Fig. 1). NZLAM uses 3DVAR data assimilation of observations from land, sea and atmosphere (Lorenc et al., 2000). The model provides meteorological data on a 12 km spatial grid, with 70 vertical levels and an hourly temporal resolution. The model is run by NIWA every 6 hours, producing a 48 hour forecast.

The NZCSM has a smaller spatial extent than NZLAM and uses NZLAM forecasts to derive its boundary conditions to generate meteorological data at a spatial resolution of 1.5 km and a half-hourly temporal resolution. Its initial conditions are generated via a pseudo data assimilation scheme that optimally combines the large-scale features of the NZLAM forecast, especially away from the surface over complex topography, with the higher resolution NZCSM forecast from the previous forecast at the same forecast time. This finer grid allows a
more accurate representation of the New Zealand topography, which is especially beneficial in mountainous regions, but the higher resolution implies a significantly greater computational power requirement (information on runtime and number of parallel tasks required is described in the Results section). NZCSM is run four times a day and generates a 36 hour forecast.

**Hydrological model Topnet**

TopNet is a distributed hydrological model, based on runoff generation concepts from TOPMODEL (Beven et al., 1995), aimed at predicting flow in large watersheds using smaller sub-basins as model elements (Bandaragoda et al., 2004). TopNet includes a kinematic wave routing algorithm to route flow from multiple sub-basins along the channel network (Goring, 1994). For this application, we used sub-basins of approximately 1 km².

Within each subcatchment, TopNet includes a potential evapotranspiration component, a snow storage component, a canopy storage component and a soil zone component to allow for the modelling of infiltration-excess runoff and vertical drainage to the saturated zone. The saturated zone component uses the concept of the topographic index (Beven and Binley, 1992) and computes the local depth to water table to simulate baseflow and saturation-excess runoff. Figure 2 gives a schematic view of the physical processes represented by TopNet, and complete model equations are given by Clark et al. (2008). An initial estimate of TopNet model parameters is made using information on catchment topography,

---

**Figure 1** – Global model zoomed in over Australia and New Zealand showing the NZLAM (outermost) and NZCSM (innermost) domain boundaries.

**Figure 2** – Schematic representation of the water balance component of TopNet (adapted with permission from Bandaragoda et al., 2004).
physical and hydrological properties derived from the River Environment Classification (Snelder and Biggs, 2002), soil, land use and geology databases (Newsome et al., 2000). TopNet is used extensively in New Zealand for applications including flow forecasting and climate change impact assessment (Poyck et al., 2011; Gawith et al., 2012; McMillan et al., 2013).

**Case study catchment**

The case study catchment is the Hutt catchment in the Wellington region (Fig. 3). To test model predictions of rainfall and river flow, the Hutt catchment was divided into three sub-basins, each associated with a flow gauging station. The smallest sub-basin (86 km²) is the area draining to the Hutt River at Kaitoke flow station, and is the main focus of this study. The Hutt River at Taita Gorge flow station represents the largest sub-basin (558 km²).

The elevation range of the catchment is large, with mountainous areas (the Tararua and Rimutaka ranges) in the north-west and an extensive floodplain in the lower reaches (Fig. 3). Land cover in high elevation areas is dominated by indigenous forest, mid-elevation areas have mixed land cover of exotic forest, scrub and agricultural pasture, and low elevation areas are largely urban (Fig. 4). The soil type does not vary significantly over the catchment area, predominantly consisting of loam with different degrees of silt, sand or stone. The catchment lies on basement geology of Torlesse Greywacke. Below the Taita gorge (i.e., downstream of the catchments used in this study), the Hutt River is underlain by the Waiwhetu aquifer, which is 20-70 m thick and is artesian in the lower reaches. In the upper catchment, particularly in hilly and mountainous terrain, groundwater interaction with surface flows is thought to be less important.

The Hutt catchment was chosen for this study for several reasons. Firstly, approximately 150,000 people live within the Hutt basin floodplain, resulting in significant risks to human life, property and economy in the event of a major flood (GWRC, 2014; Priest, 2014); therefore improved forecasting methods would be extremely valuable for this area. Secondly, the topography of the catchment, including mountainous areas, leads to orographic and convective rainfall mechanisms (Sturman et al., 1999) that we expect to be more accurately simulated when using high-

![Figure 3 – Digital elevation of the Hutt catchment, showing mountainous areas where convective rain is expected to play an important role in the weather system.](image-url)
resolution numerical weather prediction methods compared to low-resolution models. Thirdly, smaller catchments demonstrate a larger uncertainty in the flood forecast as they lack the ‘smoothing effects’ that occur when modelling a larger catchment (Bálint et al., 2006). With a small area of 86 km$^2$, the Kaitoke subcatchment is an ideal candidate to test the uncertainty in the high-resolution numerical weather prediction method. Lastly, the Hutt has a network of rainfall and flow gauges sufficient for hydrological model calibration and evaluation.

**Storm events**

On 24 May 2014, a large cold front moved from the West Coast over the South Island and lower North Island, splitting into two larger storms before merging again. The storm produced intense rain in several parts of the country and a 20-year return period flow (415 m$^3$s$^{-1}$) was observed at Hutt River at Kaitoke (Fig. 5). Rainfall gauges in the Tararua Range recorded up to 228 mm of rainfall over a 24-hour period.

On 2-3 August 2014, a strong convergence line, originating from the central north Tasman Sea, remained in place over the lower North Island for more than 24 hours. A cold front approaching from the south-west eventually merged with this feature over the lower North Island prolonging the period of rain over this area. Convective cells, already in the convergence line, passed over the mountainous areas of the Wellington region producing heavy rain and generating a two-year return period flow event at the Hutt River at Kaitoke monitoring site (Fig. 5). Rainfall gauges in the Tararua Range recorded 100 mm of rainfall over a 24-hour period during the peak of the storm.

For both storm events the large meteorological system lasted several days. While neither of these two storms caused severe damage, they are representative of the type of storms that regularly occur in the Hutt catchment. Furthermore, these storms were the largest events to occur after the high-resolution weather forecast NZCSM became operational in April 2014.

**Rain gauges**

Rainfall data are available from 11 automatic rain gauges operated by either NIWA or Greater Wellington Regional Council (GWRC) in the Hutt and neighbouring catchments. The two high altitude rain gauges, Penn Creek at McIntosh (1286 m)
Figure 5 – Hydrograph of the Hutt River at Kaitoke during April to October 2014 with a close-up on the 20-year flood event on 24 May 2014 and the 2-year flood event on 3 August 2014.

and Tauherenikau at Bull Mound (1030 m), located in the mountainous area near the Kaitoke catchment, provide the most accurate measurements of the true mountain rainfall totals during intense storms. The south-east mountainous area of the Hutt catchment contains a mid-altitude rain gauge, Pakuratahi at Centre Ridge (510 m). The remaining rain gauges are mid-altitude or spread out in the valley along the Hutt River. On average, the rain gauges have 24 years of available data. All the rain gauges used in this study are tipping bucket rain gauges with rainfall recorded at 5-minute intervals.

**Gridded rainfall product VCSN**

The Virtual Climate Station Network (VCSN) interpolates observed meteorological values onto a grid covering New Zealand at a 5 km spatial and a daily temporal resolution (Tait *et al.*, 2006). The Climate Database (CLIDB) underlying the VCSN holds data from about 6500 climate stations, which have been operating for various periods since the earliest observations were made in 1850. The database continues to receive data from more than 600 stations that are currently operating. The starting date for most VCSN weather variables is 2 January 1972, except for the wind variable which starts on 1 January 1997.

The Akatarawa at Cemetery and Wallaceville rain gauges were used for the interpolation to generate the VCSN data between 80% and 100% of the time. However, the rain gauge network used to create the VCSN is relatively sparse over the mountainous area and northern part of the Wellington region in general, because only publicly available, telemetered rainfall data is included.

**Flow data**

Three flow stations were used in this study, Hutt River at Kaitoke, Hutt River at Birchville and Hutt River at Taita Gorge. The catchments upstream of each of these gauges are shown in Figure 4. Flow gauge and catchment characteristics are summarised in Table 1.
Methods

Coupling of weather models and hydrological model

In order to produce current operational river flow forecasts, NIWA maintains a software system that couples the NZLAM weather model with the TopNet hydrological model (McMillan et al., 2013). Model data is automatically passed between models, and model outputs are displayed using the software platform and Web application EcoConnect (Uddstrom et al., 2006; Lane et al., 2009; Moore et al., 2012). For the current project, the system was rewritten to allow coupling between the 1.5 km NZCSM weather model and TopNet. To update the flow forecasting system and allow for the much larger quantities of data generated by the higher resolution weather model, the coupled model system was redesigned to use the Rose and Cylc software (FCMteam; Oliver, 2015) for scheduling dependent tasks and data management. Rose is a framework for managing and running forecasting suites and Cylc is the suite engine. The input and output routines of the hydrological model were re-implemented to work with the new self-contained Rose-based Cylc suites. Other changes that were required to accommodate the NZCSM weather input were to aggregate weather forecasts from a half-hourly to an hourly timestep and to extract a subset domain to reduce file archiving space with an added benefit of speeding-up the hydrological model input and output routines. The operational flow forecasting system uses flow data assimilation with the recursive ensemble Kalman filter, and a space and time statistical perturbation of the precipitation provides a flow forecast ensemble of 50 members (Clark et al., 2008; McMillan et al., 2013).

For model testing, we used an off-line version of the operational system on a case study event. This system used archived weather forecasts produced by the NZLAM and NZCSM models. We were also able to speed up the test process by using composite forecasts rather than multiple overlapping forecasts (see below).

<table>
<thead>
<tr>
<th>Name</th>
<th>Kaitoke</th>
<th>Birchville</th>
<th>Taita Gorge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area (km²)</td>
<td>86</td>
<td>427</td>
<td>558</td>
</tr>
<tr>
<td>Elevation range (m)</td>
<td>194-1363</td>
<td>64-1363</td>
<td>14-1363</td>
</tr>
<tr>
<td>Land cover %</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forest</td>
<td>93</td>
<td>80</td>
<td>78.5</td>
</tr>
<tr>
<td>Cultivated</td>
<td>2.4</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>Urban</td>
<td>0</td>
<td>0.9</td>
<td>4.2</td>
</tr>
<tr>
<td>Shrub</td>
<td>2.4</td>
<td>6.8</td>
<td>6</td>
</tr>
<tr>
<td>Other</td>
<td>2.2</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Mean annual precipitation (mm/yr)</td>
<td>3141.6</td>
<td>2255.1</td>
<td>2063</td>
</tr>
<tr>
<td>Mean annual runoff (mm/yr)</td>
<td>2821.3</td>
<td>1644.7</td>
<td>1354.3</td>
</tr>
<tr>
<td>Flow series length</td>
<td>48 years</td>
<td>45 years</td>
<td>36 years</td>
</tr>
<tr>
<td>Flow series source</td>
<td>NIWA</td>
<td>NIWA</td>
<td>GWRC</td>
</tr>
</tbody>
</table>

Table 1 – Flow gauges and case study catchment characteristics.
**Creation of rainfall products**

*VCSN and interpolated station data*

When the daily, gridded VCSN-interpolated rainfall data is used to drive the hydrological model TopNet, it must be disaggregated to the hourly model time resolution. When hourly rainfall depths are available from one or more nearby rain gauges (as is the case here), the proportion of the 24-hour rainfall occurring in each hour is interpolated from the gauges onto the catchment centroid, and this temporal pattern is used to make the disaggregation. TopNet can also be forced directly with rainfall measurements from a set of rain gauges. For any rainfall input (rain gauge, gridded VCSN or gridded forecasts), spatial interpolation of the rainfall input onto the hydrological catchment is done within the hydrological model, using inverse distance interpolation onto the catchment centroid.

**Deterministic composite forecast data**

The NWP rainfall time series (NZLAM and NZCSM) used in this study are composite forecasts generated by concatenating time slices from multiple successive forecasts (Roberts *et al.*, 2009). A 6-hour period is extracted from each 6-hourly forecast (from \(T+(IA/2)\) to \(T+(IA/2)+6\)hr, where \(IA\) is the forecast incremental analysis and \(T\) is the data assimilation time period). These slices are concatenated to generate forecast series of any length. The incremental analysis period is used for data assimilation in the operational system. For NZLAM the incremental analysis occurs over the period \(T-3\) to \(T+3\) of the analysis time (\(IA = 6\) hours) whereas for NZCSM it occurs in the period \(T-1\) to \(T+1\) of the analysis time (\(IA = 2\) hours). The NZCSM rainfall forecast is provided at half-hourly intervals and these were aggregated to hourly before use.

**Lagged ensemble**

A lagged ensemble is created from a series of successive forecasts and can be interpreted as forecasts from a set of perturbed initial conditions. Only deterministic forecasts within a six hour cycle were considered for the ensemble member pool. The maximum size of the time-lagged ensemble within a six hour cycle was six members. The main difference between the ensemble members is that each comes from a different forecast initialized at a different time.

A flow ensemble was generated by driving the hydrological model with a time-lagged ensemble for both weather forecast models (high and low resolution). The spread of the flow ensemble gives a more accurate representation of the uncertainty in the rainfall in time and space than the alternative method of statistical perturbation.

**Comparison of rainfall products**

*Storm events*

For both storm events the rainfall products were examined in two ways. First, maps of the cumulative storm rainfall over the catchment were created, to show differences between rainfall products in terms of rainfall intensity and storm locations. We expect the higher resolution rainfall products with the best orographic representation, in particular, the NWP convective-permitting model, to give a more accurate rainfall distribution and intensity over the mountainous area north-east of the Kaitoke sub-catchment. Secondly, rainfall depths from the four weather input data series (interpolated station data, VCSN, NZLAM and NZCSM) were compared with data from nearby rain stations. The interpolated station data is derived from the rain gauge totals using an inverse square distance interpolation. All the rainfall products were processed to calculate the average rainfall in the Kaitoke sub-catchment. As the daily VCSN rainfall product is disaggregated to hourly data based on nearby rain stations, the timing of the VCSN rainfall is expected to be close to the true observed rainfall and interpolated station data model; however, we expect...
some differences in rainfall volumes for this product because it uses a different set of rain gauges (see Discussion).

**Long-term comparison**

To investigate the long-term effect of the convective rain and orographic resolution in the Hutt catchment, two-year composite time series of available NZLAM forecasts and VCSN product time series were created, and maps of the cumulative rainfall totals for the two years were compared. This will allow us to test whether long-term rainfall totals have a similar intensity and spatial pattern to that of the two storm events (and therefore whether underestimation of precipitation over the hills by the low-resolution forecast is likely to be a common occurrence).

**Hydrological model calibration**

Before using the hydrological model to compare the impacts of different rainfall products, the values of seven model parameters were calibrated against recorded flow data at the Kaitoke gauge. In the absence of available long-term rainfall forecast data, historical observed rainfall data (VCSN) was used to drive the model. The VCSN precipitation data were bias-corrected using a long-term water balance approach correction grid (Woods *et al.*, 2006). The VCSN rainfall data were provided at daily time steps and were disaggregated to hourly values, based on the temporal pattern of observed rainfall data, before use in the model. In order to reduce the dimensionality of the optimisation problem, the spatial pattern of the parameters derived from catchment characteristics was retained, and the optimisation was implemented by calibrating a spatially-constant set of parameter multiplier values. We used the ROPE (RObust Parameter Estimation) calibration method (Bardossy and Singh, 2008) over a five-year period to produce an ensemble of potential parameter sets. A refined parameter set ensemble was then used to run the model for a further five years, and one preferred set was selected based on Nash Sutcliffe efficiency scores (linear and log space) and visual inspection by a hydrologist. An additional two ‘runner-up’ parameter sets and the original uncalibrated set were retained for comparison.

**Comparison of flow forecasts**

The impact of the different rainfall products on the flow forecasts was investigated by running the hydrological model for the two selected storm events. Initial soil moisture conditions can be a source of hydrological modelling uncertainties (Zehe *et al.*, 2005; Le Lay and Saulnier, 2007). Thus, to obtain accurate initial conditions of the model state variables (e.g., soil moisture, depth to groundwater), the model was run for a warm-up period of one year, using VCSN rainfall data. The warm-up period was run for each parameter set separately; this was found to be necessary as each parameter set resulted in different equilibrium model states.

Rainfall depths derived from the interpolated station data, VCSN, NZLAM and NZCSM forecasts were compared to the predicted and measured cumulative runoff depth. By comparing observed runoff totals with forecast rainfall and runoff totals, we can test the data in volumetric terms; for example, if input rainfall depth is less than output runoff depth, the model will not be able to close the water balance, and this data would be considered disinformative (Beven *et al.*, 2011). The shape and timing of the predicted hydrographs and observed hydrographs were also compared. We tested the predicted flow forecast performance in each of the three nested study catchments to investigate the effect of catchment size on forecast quality. The initial hypothesis was that larger catchments would show better performance, due to the smoothing effect as rainfall quantities and catchment characteristics are averaged over a larger area.
The hydrological model performance was assessed using the selected optimum parameter set, the default parameter sets and the two ‘runner-up’ parameter sets. Our intention was to investigate the sensitivity of the model to parameter set choice, and to compare the parametric uncertainty with the uncertainty due to rainfall product choice.

Results

**Coupling of weather models and hydrological model**

The high-resolution weather forecast model was successfully coupled to the hydrological model in an operational flow forecasting system. The NZCSM weather model takes just under three hours to run on 810 parallel tasks (on IBM Power 6 compute nodes), and the TopNet model takes 20 minutes to run in serial for a 42 hour forecast (recently extended from 36 hour forecast) for the largest catchment (6305 km$^2$) and five minutes to run for a smaller catchment of 2000 km$^2$. The runtime scaling is slightly nonlinear due to increasing inputs and outputs with catchment size. The data ingestion to the EcoConnect web database, visualization, and archiving processes add another 20 minute overhead for the Cylc suite. Since all the catchments in the suite are run in parallel, the maximum runtime for the hydrological model holds for the largest catchment. Therefore, the hydrological forecast is produced within four hours of the high-resolution weather forecast model start time. In comparison, the lower resolution weather forecast NZLAM takes under 10 minutes with 256 parallel tasks (on IBM Power 6 nodes). Despite the increase in the number of parallel tasks, the runtime difference between the weather models illustrates the high computational costs incurred by the 1.5 km resolution weather forecast over the 12 km model. With NZLAM, the TopNet model takes 15 minutes to run in serial for a 48 hour forecast for the largest catchment of 6305 km$^2$, with another 20 minute overhead for the Cylc suite. There is a slight speed-up of running the hydrological model when using NZLAM as the inverse distance interpolation method is quicker with fewer grid cells. Consequently, the hydrological forecast is produced within one hour of the low-resolution forecast model start time. These results highlight the negligible runtime increase (i.e., 20 minutes compared to 15 minutes) of the hydrological model coupled with the higher resolution weather forecast.

**Comparison of rainfall products**

**Storm events**

The cumulative rainfall totals and spatial distribution for all the rainfall products over periods of six and five days for the May and August storm events, respectively, highlight different weather patterns between the models (Figs. 6 and 7). The large differences between the station data and the interpolated VCSN data clearly show the difficulty in ascertaining the true rainfall distribution. Although rainfall totals recorded by gauges can be expected to be within 10% of the true point rainfall, the typical situation where rain gauges are assumed to be representative on scales of 1-10 km has been shown to lead to standard errors in catchment average rainfall of 33-65% (McMillan et al., 2012). The storm pattern forecast by NWP models, especially the high-resolution NZCSM, provides distributed information but which is uncertain due to model limitations. The difference between the observed rainfall product VCSN – which includes both automatic 5-minute and manual daily gauges, but is restricted to publicly-funded telemetered gauges – and the station data shown – which includes only 5-minute automatic gauges, but is not restricted to publicly-funded gauges – shows the very large differences that can occur due to use of a subset of gauges.
than station data. The lowest resolution composite forecast, NZLAM, did not forecast enough rainfall with no increase in rainfall amounts with altitude, highlighting the inadequate orography with this model resolution.

For the May event, the VCSN produced the highest rain accumulation over the Kaitoke mountainous area, of the products compared, and the closest rainfall totals to those observed by rain gauges, although the highest amounts were somewhat too low.

For both storm events, the highest cumulative rainfall totals were in the mountainous areas (north-east part of the catchment). As expected, the higher resolution models (VCSN and NZCSM), with a more accurate orography representation, gave a more accurate spatial distribution of rainfall. However, the modelled storm centre location was outside of the Kaitoke catchment for the high-resolution weather forecast NZCSM, which may explain why NZCSM rainfall totals within the catchment were still lower than station data. The lowest resolution composite forecast, NZLAM, did not forecast enough rainfall with no increase in rainfall amounts with altitude, highlighting the inadequate orography with this model resolution.

For the May event, the VCSN produced the highest rain accumulation over the Kaitoke mountainous area, of the products compared, and the closest rainfall totals to those observed by rain gauges, although the highest amounts were somewhat too low.
The NZCSM composite forecast produced less rainfall accumulation than the VCSN product over the hills.

For the August event, the high-resolution composite forecast NZCSM predicted the most accurate spatial distribution of rainfall, with the highest cumulative amounts over the Kaitoke catchment mountainous area. This model forecasted rainfall totals closer to those observed by the meteorological stations than the VCSN rainfall product. However, even the high-resolution model underestimated the cumulative precipitation over the storm period. For the VCSN model, although the rainfall quantities were higher over the hills, the rainfall amounts did not accurately simulate the intensity of the storm event.

For each of the rainfall products, the cumulative rainfall depths over a 5- and 6-day period for the May and August storm events, respectively, highlighted the spatial variation in intensity and timing of the catchment average rain (Figs. 8 and 9). For both storm events, the highest cumulative rainfall depths were observed by the high altitude rain gauges, outside of the Kaitoke catchment. Of all the rainfall products, the interpolated station data model gave the highest rain volumes and a temporal pattern closest to the individual station records. However, the catchment-average precipitation totals calculated as input into the hydrological model were much lower than those recorded by the high altitude stations. This is because the model uses an inverse-distance interpolation between all nearby rain gauges, and therefore also includes several lower altitude gauges where smaller amounts were observed. The lowest resolution model NZLAM performed the worst, missing the storm intensity in both cases. Overall, the intensity of the rainfall was substantially under-estimated by all the models; however, the timing of the rainfall was correct or late only by a couple of hours compared to the timing of the rainfall observed by nearby meteorological stations.

For the May storm event, the rainfall intensity and volume was predicted best by the VCSN model. However, for the August storm event the high-resolution weather forecast NZCSM model predicted the closest rainfall volumes to those observed by the

**Figure 8** – May event at the Hutt at Kaitoke gauge: runoff and rainfall depths (left) and hydrograph (right) for the observed flow and modelled flow using NZCSM, VCSN, NZLAM and interpolated station data.
high elevation station and the interpolated station data model.

**Long-term comparison of rainfall**

A comparison of the cumulative rainfall over two years between the low-resolution composite forecast NZLAM and the VCSN product showed a similar rain intensity and pattern to the two storm events (Fig. 10). The high elevation parts of the catchment had higher rainfall totals with the 5 km resolution model, whereas NZLAM is not resolved enough (with reduced peak heights and terrain variability in the mountainous areas) to generate orographic and convective rain.

---

**Figure 9** – August event at the Hutt at Kaitoke gauge: runoff and rainfall depths (left) and hydrograph (right) for the observed flow and modelled flow using NZCSM, VCSN, NZLAM and interpolated station data.

**Figure 10** – Cumulative precipitation over two years (8/12-8/14) for NZLAM and VCSN. The interpolated observed data set shows that convective rainfall is especially important for higher elevation. Additional convective rainfall included by NZCSM is a significant addition.
Hydrological model calibration

The TopNet model was calibrated over the period 1 December 1998 to 31 April 1994 including a 2.5 year warm-up period to obtain accurate initial conditions on the model state variables for calibration. This period was chosen to include data for hourly-disaggregation from a high elevation rain station Hutt at Philips, which was located in the Kaitoke catchment but was closed in 1998. To compare the best parameter sets (A, B and C) and the default uncalibrated parameter set (N) in terms of Nash-Sutcliff performance scores, the model was run over a verification period of 1 February 2010 to 31 January 2014.

In spite of the sparsity of the meteorological station network data used to generate the VCSN product in the Kaitoke catchment area, the model was calibrated with the VCSN data as the best available long-term observation product. Over a long-term period (two years), we compared rainfall and runoff depths for several large storms and found many events with non-closing water balance budgets, i.e., with greater runoff than rainfall depths, indicating a serious under-estimation of rainfall in the Kaitoke catchment area for large storm events.

Despite the difficulty in calibrating the model in such a small catchment (86 km²), where the uncertainty in the precipitation is the highest (Bálint et al., 2006), the Nash-Sutcliffe performance scores (Table 2) showed a good improvement of the calibrated model over the uncalibrated set for both the calibration and verification periods. In particular, the Nash-Sutcliffe (NS), indicative of flood, gave similar or better results for the optimal parameter sets for the larger Hutt catchment (558 km²) during the verification period. In the small Kaitoke subcatchment, all the models had a tendency to underestimate the observed discharge largely due to the tendency to underestimate the precipitation in this mountainous area.

Table 2 – Nash-Sutcliffe (NS) and logarithmic Nash-Sutcliffe (Log NS) performance scores for the best calibrated parameter sets (A, B and C) and the original uncalibrated parameter set (N).

<table>
<thead>
<tr>
<th>Parameter Sets</th>
<th>Calibration Period (01/12/1988-31/04/1994)</th>
<th>Verification Period (01/02/10 - 31/01/14)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kaitoke catchment (86 km²)</td>
<td>Kaitoke catchment (86 km²)</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>Log NS</td>
</tr>
<tr>
<td>A</td>
<td>0.53</td>
<td>0.57</td>
</tr>
<tr>
<td>B</td>
<td>0.48</td>
<td>0.58</td>
</tr>
<tr>
<td>C</td>
<td>0.43</td>
<td>0.58</td>
</tr>
<tr>
<td>N</td>
<td>0.24</td>
<td>0.42</td>
</tr>
</tbody>
</table>
Comparison of flow forecasts

Runoff totals

For all of the rainfall products, the catchment-average rainfall depth and corresponding runoff depth were much lower (or just equal in one case) than the observed for both storm events (Figs. 8 and 9). This clearly indicates that none of the rainfall products, including the observed data from nearby stations, produced enough rain volume to close the water balance and account for the observed runoff. In contrast, the high elevation station rainfall totals (outside of the Kaitoke catchment), Penn Ck at McIntosh and Tauherenikau at Bull Mound, were much higher than the observed runoff by up to an additional 89 mm in May (the runoff volume matched only 56% of the observed rainfall volume) and 112 mm in August (the runoff volume matched only 66% of the observed rainfall volume). The highest modelled runoff was simulated when using observed rainfall from nearby station data, including the high elevation stations, but was still under-estimated by 50 mm (producing only 56% of the observed runoff volume) in May and by 111 mm (producing only 49% of the observed runoff) in August. Uncertainty in the observed flow, which is typically highest during flood events (McMillan et al., 2012), may explain part of the discrepancy between rainfall and runoff totals. However, we also infer that rainfall totals in the ungauged parts of the catchment were higher than interpolation between nearby gauges would suggest.

For the May event, the VCSN-driven model performed better than the forecast-driven models, but still underestimated the observed runoff by 72 mm (producing only 37% of the observed runoff). The water volume difference between the cumulative rainfall and runoff of the VCSN driven model was 69 mm, thus 38% of the modelled rainfall produced flow. With a similar conversion percentage volume applied to the observed flow, the actual catchment average rain could have reached totals of 300 mm, which is well in excess of the cumulative rainfall amounts from the mountainous stations (Penn Ck at McIntosh 122 mm, Tauherenikau at Bull Mound 203 mm). However, this total is likely to be an over estimate, because in larger storm events the initial soil water deficit becomes less important and so a higher percentage of rainfall becomes flow.

For the August storm event, the best modelled runoff (other than with the interpolated station data) was obtained when using the high-resolution weather forecast NZCSM, with 82 mm of runoff (producing 38% of the observed flow volume) ahead of the VCSN and NZLAM with 54 mm and 30 mm of runoff (producing 25% and 14% of the observed runoff), respectively. For this storm event, the water volume difference between the cumulative rainfall and runoff of the NZCSM-driven model was 70 mm, thus 54% of the modelled rainfall produced flow. Applying this volume correction to the observed flow, the actual catchment average rain could have reached totals of 400 mm, which is in excess of the quantities observed by the high elevation rain stations (Penn Ck at McIntosh 328 mm, Tauherenikau at Bull Mound 290 mm.) Again the total is likely to be an over estimate, because in larger storm events the initial soil water deficit becomes less important and so a higher percentage of rainfall becomes flow.

Hydrograph shapes

The modelled flow peaks in the hydrographs from the deterministic rainfall products and observed nearby station data did not match the observed flow peaks for either of the storm events (Figs. 8 and 9).

For the May storm event, the observed discharge had one extreme peak and two smaller but distinct peaks either side, corresponding to three rain events (Fig. 8).
The most responsive model was driven by the VCSN with a maximum hourly peak discharge of 88 m\(^3\)s\(^{-1}\), ahead of the interpolated station data with 61 m\(^3\)s\(^{-1}\), but still greatly under-estimated the observed maximum peak discharge of 327.5 m\(^3\)s\(^{-1}\). Both forecast-driven models produced a much smaller peak, 9% (NZCSM) and 4.5% (NZLAM) of that observed, with inaccurate (late) timing.

For the August storm event, the modelled flows produced using observation-based rainfall data (VCSN and observed station data) two peaks corresponding to two rain events with the same timing as the observed flow peaks (Fig. 9). In contrast, the weather forecast-driven hydrographs only produced one peak, matching the timing of the single rain event forecasted by the NWP forecasts. The maximum modelled peaks were produced by using the interpolated station data as input but were under-estimated by 52% for the first peak and 60% for the second peak. The high-resolution composite weather forecast NZCSM produced the second largest peak when used as input to the hydrological model. However, the modelled discharge represented only one peak instead of two and only 33% of the peak height observed.

**Lagged ensemble forecasts**

For both storm events, the flow ensemble spread was much wider when generated with the 1.5 km resolution time-lagged ensemble model than for the lower 12 km resolution ensemble. This is likely due to the higher rainfall volumes and ability to model convective storms in the high-resolution model forecasts rather than a better stochastic representation of the rain uncertainty between consecutive forecasts. The NZCSM-driven lagged flow ensemble improves the users’ understanding of the uncertainty in the forecast, highlighting the uncertainty in timing, location and intensity of the storms, and the highest ensemble member was closer to the observed flow (Fig. 11).

For both storm events and both NWP resolutions, the earlier forecasts produced the highest peak discharge. The closer the forecasts were to the storm event peak intensity the less rain was produced, leading to much smaller peak discharges. This is an unexpected behaviour; typically the highest forecast accuracy is expected in the first 12 hours of a forecast. A potential explanation resides in the nesting of the weather forecast resolutions for determining the initial and lateral boundary conditions. The 1.5 km resolution model NZCSM is linked to the

---

**Figure 11** – Comparative discharge for the August storm event between the observed flow and lagged modelled flow ensemble using NZCSM for the small Hutt catchment (Hutt at Kaitoke, left) and the large Hutt catchment (Hutt at Taita Gorge, right).
12 km grid NZLAM via its pseudo-analysis scheme where large-scale air flow features from the NZLAM forecast are inserted in the NZCSM analysis. The convective cells and other characteristic forecast behaviour of the NZCSM model may require some time to develop after initialisation by the low-resolution large-scale air flow.

**Effects of catchment size**

The deterministic modelled flows did not improve for either storm when the modelled catchment area was increased (not shown) to the wider Hutt catchment. In contrast, when a time-lagged ensemble fed the hydrological model, a much wider spread in the modelled flow was observed for the larger catchment (Fig. 11). At the Hutt at Taita Gorge station, with an upstream catchment area of 558 km², the earlier NZCSM-driven flow forecasts produced maximum peak discharges very close (within 16%) to the observed peak. These good results are somewhat expected behaviour, as larger catchments are typically less susceptible to the spatial uncertainty on the rain, producing smoother results than their smaller catchment counterparts.

**Model sensitivity to choice of parameter set**

All the deterministic and ensemble flow models showed less sensitivity to the choice of parameter set than the input rainfall product (Fig. 12). For the two storm events the uncertainty in the rainfall volumes completely dominated the uncertainty in the model calibration.

**Summary and discussion**

This research showed that a high-resolution weather forecast model could be coupled to a hydrological model for flow forecasting in New Zealand. The requirements for high volumes of data transfer between the models were achieved by running the models under workflow engine and metadata scheduler software (Rose and Cylc) specialising in cyclic workflows (FCMteam; Oliver, 2015). Since the Rose-based Cylc suites are self-contained, the new operational flow forecasting system has one critical advantage over the old forecasting system: we can more easily set-up and run various experiments with different calibration parameters, new catchments, or hydrological model versions to test new features.

![Figure 12](image-url)  
**Figure 12** – Model sensitivity to the choice of parameter set for the August storm event. The hydrographs were all produced by the high-resolution NZCSM-driven model.
In the small mountainous catchment considered in the case study, use of the 1.5 km resolution NZCSM weather forecast model produced higher rainfall totals and rainfall patterns closer to that measured by recording rain gauges in the area, compared to the lower resolution NZLAM model. The convective-permitting nature of the NZCSM model and its better orographic representation are key to representing localised storms more accurately over high ground. However, none of the rainfall products tested produced sufficient rainfall depth to accurately simulate river flows during the flood events studied. For a flash-flood case study, Chancibault et al. (2006) also observed that the high-resolution (2.4 km) convective-permitting NWP model allowed for higher rainfall amounts than the large scale weather forecast model, but still underestimated the observed rainfall amount. Furthermore, the centre of the convective storms was over the mountain crests instead of over the foothills as observed.

These results suggest that further research and adaptation of the model forecasts is required before confident use of the flow forecasts is possible. Our work resulted in the important conclusion that forecast sensitivity to input rainfall product is far greater than to hydrological model calibration. Therefore, future efforts should be concentrated on improvement of rainfall prediction and its associated uncertainty. The ability to accurately predict forecast uncertainty is essential if flow data assimilation procedures are used to update hydrological model states based on observed flow data, as was demonstrated by McMillan et al. (2012). The ability of the 1.5 km NZCSM model to accurately locate and represent the shape of storm events means that bias correction techniques to correct systematic underestimation of rainfall totals in mountainous areas are more likely to be successful than in previous low-resolution models. Our future work in this area will firstly seek to create a static bias correction method, and then to build on this method to allow for rainfall bias that changes with rainfall quantity, wind direction and wind speed. Accounting for wind is important, because wind ultimately provides the energy that drives moist air up and over mountains, producing rain as it does so.

The reasons for rainfall under-prediction as seen in this study may be varied. Even the NZCSM topography resolved at 1.5 km represents a substantial smoothing when compared to the steep slopes common in New Zealand mountains. Therefore, orographic rainfall is likely to be underestimated. NWP models are also well known to underestimate heavy rainfall, and overestimate light rainfall or drizzle (Shroder et al., 2007; Helmis and Nastos, 2012; Blacutt et al., 2015), even when annual totals are approximately correct. From a meteorological perspective, a spatial error in the location of a weather front or band of convective cells of the order of a 50 km is considered quite small, but can lead the coupled hydrological model to completely miss the flood event if the heaviest rain is wrongly forecast as falling outside the catchment (Vincendon et al., 2011). For the May event used as a case study in this paper, the storm centre predicted by NZCSM-driven model was outside the small 86 km² Kaitoke catchment, which greatly impacted the modelled river flow peak. Even at 1.5 km resolution the stochastic nature of the weather models are better represented with time-lagged ensemble members to reflect the intensity and spatial variability of storms. Initial soil condition uncertainties could also be a source of flow under-prediction. The hydrological model was initialised using a warm-up period of one year with the VCSN rainfall data, which contained serious under-estimation of observed rainfall for several large intensity storms in the particular area studied.
Further verification with a larger ensemble size and flow data assimilation is still needed to fully evaluate the new operational flow forecasting system over New Zealand. In particular, other case study catchments in high elevation areas, where snow plays a crucial role in winter, will provide a test bed for model temperature sensitivity and performance.

Conclusions
In this paper, we described the implementation of a new flood forecasting system for New Zealand. We showed the feasibility of operationally running a high-resolution (1.5 km) weather model coupled to a hydrological model. The performance of the coupling and the impact of the high resolution was tested using a case study in the Wellington region, by comparing rainfall and flow forecasts during two storm events for the low (12 km) and high (1.5 km) resolution weather models and for a gridded observed rainfall product (5 km). The lagged ensemble flow forecast, which accounts for some of the uncertainty in the precipitation, produced peak discharges closest to the observed flows – within 16% – for the largest catchment studied. Despite the benefits and improvement of the convective-permitting model over the low-resolution model during extreme rain events, rainfall was underestimated during storm events, and therefore rainfall bias correction, and ensemble and flow data assimilation techniques, are required to produce more accurate and timely flood forecasts.

Acknowledgements
The authors wish to acknowledge the contribution of the New Zealand eScience Infrastructure (NeSI) to the results of this research. New Zealand's national compute and analytics services are supported by NeSI and funded jointly by NeSI's collaborator institutions and through the Ministry of Business, Innovation and Employment. URL http://www.nesi.org.nz

References


Manuscript received 9 October 2015; accepted for publication 28 April 2016